

HABILITATION À DIRIGER DES RECHERCHES

Présentée à

L'Université Joseph Fourier de Grenoble

Spécialité

Informatique

Soutenue par

Hervé BLANCHON

Le 20 décembre 2004

Comment définir, mesurer et améliorer la qualité, l'utilisabilité et l'utilité des systèmes de TAO de l'écrit et de l'oral

Une bataille contre le bruit, l'ambiguïté, et le manque de contexte

Composition du Jury

Président Yves CHIARAMELLA

Rapporteurs Yves LEPAGE
Jean-Marie PIERREL
Éric WEHRLI

Examineurs Christian BOITET
Jean CAELEN
Patrice POGNAN
Cécile ROISIN

Travaux préparés au sein de l'équipe GETA du laboratoire CLIPS
(Communication Langagière et Interaction Personne-Système)
de l'Institut IMAG

Table des matières

INTRODUCTION	1
---------------------	----------

Première Partie Qualité en communication interpersonnelle médiatisée par la machine

INTRODUCTION	11
---------------------	-----------

CHAPITRE I COMMUNICATION INTERPERSONNELLE ET TRADUCTION AUTOMATIQUE

I.1. Caractérisation de la communication interpersonnelle	13
I.1.1. Modalités de communication _____	13
I.1.2. Émetteurs et récepteurs de la communication _____	14
I.1.3. Unités de temps et de lieu de la communication _____	14
I.1.4. Domaine et planification de l'information à communiquer _____	15
I.1.5. Médiation traduisante dans la communication _____	18
I.2. Caractérisation des systèmes de traduction	19
I.2.1. Chronologie des types de TAO _____	19
I.2.2. Automatisation et types de TAO _____	21
I.2.3. Architectures linguicielles _____	21
I.2.4. Moments de l'interaction avec l'utilisateur _____	25

CHAPITRE II CARACTERISATION ET MESURE DE LA QUALITE DES SYSTEMES DE TRADUCTION

II.1. Mise en place de l'évaluation	27
II.1.1. ALPAC, the (in-)famous report _____	28
II.1.2. Travaux du JEIDA _____	31
II.1.3. Campagne ARPA (1992-1994) _____	32
II.2. Efforts de formalisation	34
II.2.1. EAGLES _____	34
II.2.2. ISLE & FEMTI _____	35
II.2.3. TIDES & NIST _____	37

II.3. Point sur les méthodes d'évaluation proposées	37
II.3.1. Méthodes subjectives _____	38
II.3.2. Méthodes objectives _____	43
II.3.3. Évaluation de l'évaluation _____	53
II.4. Premiers commentaires	55
II.4.1. Méthodes subjectives _____	55
II.4.2. Méthodes objectives _____	55
II.4.3. Que retenir du point de vue de la qualité ? _____	55
CHAPITRE III	OBSTACLES A LA QUALITE EN COMMUNICATION
	INTERPERSONNELLE MEDIATISEE PAR LA MACHINE
III.1. Variabilité de la parole	57
III.1.1. Variabilité personnelle _____	57
III.1.2. Variabilité inter-personnelle _____	58
III.1.3. Variabilité environnementale _____	58
III.2. Spontanéité	59
III.2.1. Phénomènes syntaxiques particuliers _____	59
III.2.2. Extragrammaticalité _____	60
III.2.3. Implications pour l'analyse automatique _____	60
III.3. Ambiguïté	60
III.3.1. Point de vue linguistique _____	60
III.3.2. Implications pour l'analyse automatique _____	64
III.3.3. Spécification formelle de l'ambiguïté _____	65
CHAPITRE IV	REVUE DES DIFFERENTES SITUATIONS DE
	COMMUNICATION INTERPERSONNELLE
IV.1. Communication monolingue non médiatisée	67
IV.1.1. Communication écrite _____	67
IV.1.2. Communication orale d'un orateur _____	68
IV.1.3. Communication orale entre interlocuteurs _____	69
IV.2. Communication multilingue médiatisée par l'homme	69
IV.2.1. Communication écrite traduite via un traducteur humain _____	69
IV.2.2. Communication orale d'un orateur traduite via un interprète humain _____	70
IV.2.3. Communication orale entre interlocuteurs via un interprète humain _____	71
IV.3. Communication multilingue médiatisée par la machine	71
IV.3.1. Communication écrite traduite via un système de TA _____	72
IV.3.2. Communication orale d'un orateur traduite via un système de TA _____	72
IV.3.3. Communication orale entre interlocuteurs via un système de TA _____	73
CONCLUSION	75

Deuxième Partie Communication interpersonnelle écrite réfléchie tout terrain

INTRODUCTION **79**

CHAPITRE V AMELIORATION DE LA QUALITE EN TA

V.1. TA sans interaction	81
V.1.1. Heuristiques psycholinguistiques _____	81
V.1.2. Heuristiques sémantiques _____	82
V.1.3. Heuristiques adaptées à un corpus donné _____	82
V.2. TA interactive : nouveaux développements	82
V.2.1. Bref rappel historique _____	82
V.2.2. KANT _____	85
V.2.3. TRANSTYPE _____	87
V.3. Contribution pour aller plus loin : LIDIA	88
V.3.1. Motivations _____	88
V.3.2. Critères de choix de l'approche par dialogue _____	91
V.3.3. Situations traductionnelles adaptées à la TAFD _____	91
V.3.4. Situation dans la classification proposée _____	92

CHAPITRE VI LIDIA-1 (89-95)

VI.1. Architecture linguistique et voies intermédiaires nouvelles	93
VI.1.1. Transfert multiniveau à acceptions, propriétés et relations interlingues _____	94
VI.1.2. Encodage d'informations linguistiques pouvant provenir d'une prédiction _____	94
VI.1.3. Étapes du processus de traduction _____	96
VI.2. Architectures informatiques	97
VI.2.1. Aspects ergonomiques _____	97
VI.2.2. Différentes implémentations _____	98
VI.3. Démonstration	100

CHAPITRE VII PROGRES EN DESAMBIGUISATION INTERACTIVE, LIDIA-2 (94-03) ET LIDIA-3 (04-)

VII.1. Étude sur les ambiguïtés en situations de dialogue oral finalisé	104
VII.1.1. Corpus d'étude _____	105
VII.1.2. Corpus d'ambiguïtés _____	106
VII.1.3. Observations _____	108
VII.1.4. Conclusions _____	109
VII.2. Nouvelle architecture pour la DI	112
VII.2.1. Proposition _____	113
VII.2.2. Illustration _____	121
VII.2.3. Évaluation en utilisabilité, utilité sur l'anglais _____	121
VII.3. LIDIA-2 et LIDIA-3	133
VII.3.1. Nouvelles propositions _____	134
VII.3.2. Première mise en œuvre (LIDIA-2) _____	136

VII.3.3. Vers une meilleure implémentation et de nouveaux scénarios d'usages (LIDIA-3) _____	138
VII.4. Perspectives de recherche	138
VII.4.1. Permettre une désambiguïsation interactive incomplète _____	138
VII.4.2. Certifier le sens _____	139
VII.4.3. Désambiguïsation par analogie _____	139
CHAPITRE VIII QUAND LA QUALITE NE SUFFIT PAS : LES DOCUMENTS AUTO-EXPLICATIFS	
<hr/>	
VIII.1. Proposition	143
VIII.2. Première mise en œuvre dans le cadre de LIDIA-2	144
VIII.3. Perspectives de recherche	145
VIII.3.1. Présenter le support de l'ambiguïté _____	145
VIII.3.2. Créer des DAE en langue cible _____	146
CONCLUSION	149
Troisième Partie Communication interpersonnelle orale spontanée spécialisée	
INTRODUCTION	153
CHAPITRE IX HISTORIQUE ET ANALYSE DU DOMAINE	
<hr/>	
IX.1. Communication orale entre personnes et systèmes d'information	155
IX.1.1. Stratégies de dialogue et domaines d'application _____	156
IX.1.2. Architectures et composants _____	156
IX.1.3. Galaxie GALAXY _____	157
IX.2. Traduction de parole	158
IX.2.1. Architecture et composants _____	158
IX.2.2. Spoken Language Translator _____	159
IX.2.3. AT&T _____	160
IX.3. Traduction de dialogue	161
IX.3.1. États-Unis _____	161
IX.3.2. Europe _____	164
IX.3.3. Asie _____	165
IX.3.4. Approches particulières _____	169
CHAPITRE X GRANDS PROJETS EN TRADUCTION DE DIALOGUES ORAUX FINALISES	
<hr/>	
X.1. C-STAR	171
X.1.1. C-STAR I (1986-1993) _____	171
X.1.2. C-STAR II (1993-1999) _____	172

X.2. Verbmobil	174
X.2.1. Présentation	174
X.2.2. Architectures linguistiques	174
X.2.3. Évaluation	175
X.3. NESPOLE!	176
X.3.1. Présentation	176
X.3.2. Architecture informatique	177
X.3.3. Interface utilisateur	178
X.3.4. Démonstrateurs et évaluation	179

CHAPITRE XI CONTRIBUTIONS A C-STAR II ET NESPOLE!

XI.1. C-STAR II : première approche en groupe	181
XI.1.1. Langage pivot de C-STAR	181
XI.1.2. Composants développés	182
XI.1.3. Démonstrateur du CLIPS++	184
XI.1.4. Démonstrations du CLIPS++	187
XI.1.5. Bilan	190
XI.2. NESPOLE! : modules implémentés	190
XI.2.1. Langage pivot de NESPOLE!	190
XI.2.2. Analyse français vers IF	194
XI.2.3. Génération	199
XI.2.4. Bilan sur l'IF	203
XI.3. NESPOLE! : évaluation de la qualité des traductions	204
XI.3.1. Principes	204
XI.3.2. Premier démonstrateur en tourisme réduit	204
XI.3.3. Second démonstrateur en tourisme étendu	206
XI.3.4. Mesure des progrès accomplis	210

CHAPITRE XII CONTRIBUTION A C-STAR III ET REFLEXIONS SUR LE DOMAINE

XII.1. Évaluation dans C-STAR III	213
XII.1.1. Première évaluation interne	213
XII.1.2. Contribution à la première campagne d'évaluation compétitive ouverte (IWSLT 2004)	214
XII.1.3. Conclusion	220
XII.2. Problèmes	220
XII.2.1. Mesure de la qualité linguistique	220
XII.2.2. Mesure de l'utilisabilité	221
XII.2.3. Comparaison avec les systèmes commerciaux	222
XII.2.4. Évaluation d'autres aspects	223
XII.3. Perspectives de recherche	223
XII.3.1. En évaluation	223
XII.3.2. En architecture des systèmes	224
XII.3.3. Premières idées concrètes et réalisations autour de l'architecture des systèmes	229

CONCLUSION 241

CONCLUSION		243
-------------------	--	------------

BIBLIOGRAPHIE		247
----------------------	--	------------

BIBLIOGRAPHIE PERSONNELLE		277
----------------------------------	--	------------

ANNEXES		289
----------------	--	------------

Annexe I	Ambiguïtés dans les corpus de ATR	291
Annexe II	Données pour l'évaluation pilote en compréhension des questions	297
Annexe III	Données pour la seconde évaluation en compréhension des questions	301
Annexe IV	Ambiguïtés prises en compte dans le module de désambiguïsation de l'anglais	307
Annexe V	Ambiguïtés prises en compte dans le module de désambiguïsation du français	311
Annexe VI	Extraits de la spécification de l'IF de C-STAR II	319
Annexe VII	Extraits de la spécification de l'IF de NESPOLE!	321
Annexe VIII	Annotation en IF de tours de parole d'un client	331
Annexe IX	Extraits du code du premier analyseur NESPOLE!	335
Annexe X	Extraits de code des modules du second démonstrateur NESPOLE!	339
Annexe XI	Fichier complété par un juge lors de la seconde évaluation NESPOLE!	343
Annexe XII	Traductions produites pour la paire japonais-anglais (IWSLT 2004)	349

Listes des figures

Figure I-1 :	Composants de la communication proposés par Biber _____	16
Figure I-2 :	Différences situationnelles de la communication proposés par Biber _____	17
Figure I-3 :	Vision chronologique des différents types de TAO _____	20
Figure I-4 :	Triangle de Vauquois _____	22
Figure I-5 :	Traduction avec pré-édition _____	26
Figure I-6 :	Traduction avec interaction médiane _____	26
Figure I-7 :	Traduction avec post-édition _____	26
Figure II-1 :	Principe des évaluations ARPA/DARPA _____	33
Figure II-2 :	Version compacte de la hiérarchie FEMTI _____	36
Figure II-3 :	Version développée de la section 2.1.1 de la hiérarchie FEMTI _____	36
Figure II-4 :	Interface web d'évaluation de la bonne formation linguistique _____	41
Figure II-5 :	Interface web d'évaluation de l'adéquation _____	42
Figure II-6 :	Définition d'une opération d'édition élémentaire _____	43
Figure II-7 :	Définition des opérations d'édition classiques _____	43
Figure II-8 :	Définition de la distance d'édition sur les chaînes _____	43
Figure II-9 :	Comparaison de la valeur des facteurs de pénalité de BLEU et NIST _____	49
Figure II-10 :	Plan de l'évaluation pour les applications du TALN _____	54
Figure III-1 :	Schéma général d'une extragrammaticalité _____	60
Figure III-2 :	Définition formelle d'une ambiguïté _____	66
Figure III-3 :	Représentation graphique d'une ambiguïté _____	66
Figure IV-1 :	Communication interpersonnelle monolingue écrite réfléchie _____	68
Figure IV-2 :	Communication orale monolingue d'un orateur _____	68
Figure IV-3 :	Communication orale monolingue entre interlocuteurs _____	69
Figure IV-4 :	Communication écrite traduite via un traducteur humain _____	70
Figure IV-5 :	Communication orale d'un orateur traduite via un interprète humain _____	70
Figure IV-6 :	Communication orale entre interlocuteurs via un interprète humain _____	71
Figure IV-7 :	Communication écrite traduite via un système de TA _____	72
Figure IV-8 :	Communication orale d'un orateur traduite via un système de TA (encore utopique) _____	72
Figure IV-9 :	Communication orale entre interlocuteurs via un système de TA _____	73
Figure V-1 :	KANTOO et le module de réécriture _____	86
Figure V-2 :	Réécriture de la ponctuation _____	86
Figure V-3 :	Explicitation de la fonction syntaxique d'une proposition _____	86
Figure V-4 :	Ajout d'un article manquant _____	86
Figure V-5 :	Explicitation de la subordination _____	86
Figure V-6 :	Réécriture des gérondifs _____	86
Figure V-7 :	Résolution d'une ambiguïté de polysémie _____	86
Figure V-8 :	Capture d'écran de TransType en action _____	87
Figure V-9 :	Étapes de traduction interactive, pour un traducteur non professionnel, proposée par NEC _____	88
Figure V-10 :	Caractéristiques de la TAFD _____	92

Figure VI-1 :	Exemple de structure mmc _____	96
Figure VI-2 :	Architecture linguicielle de LIDIA _____	97
Figure VI-3 :	Architecture informatique LIDIA-1 (3270) _____	99
Figure VI-4 :	Architecture informatique de LIDIA-1 (Mail) _____	100
Figure VI-5 :	Carte d'histoires concernant « Le capitaine à rapporté un vase de Chine. » _____	100
Figure VI-6 :	Sélection d'un champ textuel à traduire sur une carte de phrase _____	101
Figure VI-7 :	Demande de l'état d'avancement des traitements _____	101
Figure VI-8 :	Fenêtre flottante de suivi de l'état des traitements _____	101
Figure VI-9 :	Signalement de questions de désambiguïsation en suspens _____	101
Figure VI-10 :	Désambiguïsation structurale avec LIDIA-1 _____	102
Figure VI-11 :	Désambiguïsation de polysémie avec LIDIA-1 _____	102
Figure VI-12 :	Forme annotée de l'analyse désambiguïsée _____	102
Figure VI-13 :	Rétrotraduction permettant de vérifier la traduction _____	102
Figure VI-14 :	Traduction en allemand d'une carte d'histoires _____	103
Figure VII-1 :	Situations expérimentales de collectes de corpus avec EMMI (interactions) _____	106
Figure VII-2 :	Ambiguïtés pour 100 mots dans les trois situations d'expérimentation EMMI _____	108
Figure VII-3 :	Ambiguïtés de chaque type pour 100 mots dans les trois situations d'expérimentation EMMI _____	108
Figure VII-4 :	Ambiguïtés de subordination corrigées pour 100 mots dans les trois situations d'expérimentation EMMI _____	109
Figure VII-5 :	Répartition des effets de la répétition sur l'ambiguïté _____	110
Figure VII-6 :	Moteur et linguiciels pour des modules de désambiguïsation interactive _____	112
Figure VII-7 :	Instance d'un module de désambiguïsation particulier _____	112
Figure VII-8 :	Analyses produites pour la phrase «Le capitaine a rapporté un vase de Chine. » _	113
Figure VII-9 :	Vision abstraite par forêt des analyses de la phrase « Le capitaine a rapporté un vase de Chine » _____	114
Figure VII-10 :	Faisceau de patrons définissant le type d'ambiguïté présent dans la phrase « le capitaine a rapporté un vase de Chine. » _____	114
Figure VII-11 :	Patron 1 de définition d'une ambiguïté de structure argumentaire du verbe de type 3 _____	115
Figure VII-12 :	Patron 2 de définition d'une ambiguïté de structure argumentaire du verbe de type 3 _____	115
Figure VII-13 :	Définition de la reconnaissance d'une ambiguïté _____	115
Figure VII-14 :	Définition de la distance entre les ensembles de variables de forêt de deux patrons _____	116
Figure VII-15 :	Synchronisation de l'appariement des patrons avec les solutions _____	117
Figure VII-16 :	Méthode de rephrasage associée au patron sav-t3-1 (Figure VII-11) _____	118
Figure VII-17 :	Méthodes de rephrasage pouvant être associées au patron sav-t3-1 (Figure VII-12) _____	118
Figure VII-18 :	Construction d'un arbre de questions (vue externe) _____	120
Figure VII-19 :	Construction d'un arbre de questions (vue interne) _____	120
Figure VII-20 :	Écran du sujet avec présentation d'une question de désambiguïsation _____	123
Figure VII-21 :	Nombre de réponses correctes pour les interprétations faciles et difficiles pour la première expérience en compréhensibilité des questions de désambiguïsation __	124
Figure VII-22 :	Nombre de réponses correctes pour les questions humaines et machines pour la première expérience en compréhensibilité des questions de désambiguïsation __	124
Figure VII-23 :	Nombre de réponses incorrectes pour les questions humaines et machines pour la première expérience en compréhensibilité des questions de désambiguïsation __	124
Figure VII-24 :	Nombre de questions restées sans réponse pour les questions humaines et machines pour la première expérience en compréhensibilité des questions de désambiguïsation _____	125
Figure VII-25 :	Interface de lecture et de désambiguïsation sous forme de boîte de dialogue de la seconde expérience en compréhensibilité des questions de désambiguïsation __	126

Figure VII-26 :	Analyse des données de la seconde évaluation en compréhensibilité des questions de désambiguïsation interactive _____	130
Figure VII-27 :	Architecture informatique de LIDIA-2 _____	134
Figure VII-28 :	Descripteur d'un document LIDIA-2 _____	135
Figure VII-29 :	Extrait d'un fichier LIDIA-2 pour une phrase _____	135
Figure VII-30 :	Fenêtre de document LIDIA-2 (après analyse) _____	136
Figure VII-31 :	Première question de désambiguïsation de la phrase "I want the symposium on interpreting telecommunications at the international conference center" _____	137
Figure VII-32 :	Seconde question de désambiguïsation de la phrase "I want the symposium on interpreting telecommunications at the international conference center" _____	137
Figure VII-33 :	Fin de la session de désambiguïsation de la phrase "I want the symposium on interpreting telecommunications at the international conference center" _____	137
Figure VII-34 :	Affichage d'un bitexte _____	137
Figure VII-35 :	Homomorphisme entre structures analogiques _____	140
Figure VII-36 :	Traduction par analogie _____	140
Figure VII-37 :	Désambiguïsation par analogie - vue abstraite _____	141
Figure VII-38 :	Désambiguïsation par analogie - vue concrète _____	141
Figure VIII-1 :	Architecture linguicielle de LIDIA avec intégration de la construction de DAE ____	144
Figure VIII-2 :	Extrait d'un document auto-explicatif LIDIA-2 _____	145
Figure VIII-3 :	Interface du visualiseur de DAE _____	145
Figure VIII-4 :	Affichage de la lecture à retenir pour la phrase sélectionnée _____	145
Figure VIII-5 :	Présentation des supports d'ambiguïté d'un DAE _____	146
Figure IX-1 :	Architecture et composants des systèmes de COP-SI _____	156
Figure IX-2 :	Schéma de l'architecture GALAXY-II _____	157
Figure IX-3 :	Représentation sémantique en GALAXY-II _____	158
Figure IX-4 :	Architecture d'un système de traduction de parole _____	158
Figure IX-5 :	Architecture linguicielle de SLT _____	159
Figure IX-6 :	Résultat de l'évaluation SLT _____	160
Figure IX-7 :	Speechlator, un système de traduction de dialogue anglais-arabe de CMU ____	163
Figure IX-8 :	Exemple de traduction anglais-chinois avec MASTOR _____	164
Figure IX-9 :	Interfaces du système MASTOR de IBM sur PDA _____	164
Figure IX-10 :	Modèle de traduction de NEC _____	167
Figure IX-11 :	Modèle de traduction proposé par Spoken Translation Inc. _____	169
Figure IX-12 :	Modèle de traduction de ERIM-Interprète _____	170
Figure X-1 :	Partenaires de C-STAR II et de C-STAR III _____	172
Figure X-2 :	C-STAR II dans la classification à cinq branches proposée _____	172
Figure X-3 :	Architecture informatique de C-STAR II _____	173
Figure X-4 :	Avantages liés à l'utilisation d'un pivot pour la traduction _____	173
Figure X-5 :	Architecture linguicielle de Verbmobil _____	174
Figure X-6 :	NESPOLE! dans la classification à cinq branches proposée _____	177
Figure X-7 :	Architecture informatique de NESPOLE! _____	178
Figure X-8 :	Interface Utilisateur du démonstrateur NESOPLE! _____	179
Figure XI-1 :	Intégration des composants du démonstrateur C-STAR II du groupe CLIPS++ _	184
Figure XI-2 :	Écran de visioconférence multisite du démonstrateur CSTAR II du groupe CLIPS++ _____	185
Figure XI-3 :	Interface utilisateur du démonstrateur C-STAR II du groupe CLIPS++ _____	185
Figure XI-4 :	Interface de contrôle du démonstrateur C-STAR II du groupe CLIPS++ _____	186
Figure XI-5 :	Page web d'activité touristique proposée par l'agent coréen à un client français _	187
Figure XI-6 :	Architecture de l'analyseur français vers IF du premier démonstrateur NESPOLE! _____	195
Figure XI-7 :	Architecture du module d'analyse français vers IF du second démonstrateur NESPOLE! _____	197

Figure XI-8 :	Architecture du module de génération IF vers français du premier démonstrateur NESPOLE! _____	200
Figure XI-9 :	Architecture du module de génération IF vers français du second démonstrateur NESPOLE! _____	202
Figure XII-1 :	Interaction entre la reconnaissance de la parole et le module d'analyse _____	225
Figure XII-2 :	Interaction des modules d'analyse et de génération avec le module de reconnaissance de la parole _____	225
Figure XII-3 :	Interface d'évaluation de la traduction de la première hypothèse et de choix d'une meilleure hypothèse _____	230
Figure XII-4 :	Étapes de calcul du modèle de langage sémantique _____	234
Figure XII-5 :	Architecture de gestion du dialogue _____	236

Liste des Tableaux

Table I-1 :	Émetteur et récepteur en fonction de la modalité de communication _____	14
Table I-2 :	Unités de temps et de lieu en communication interpersonnelle orale _____	15
Table I-3 :	Différentes planifications de l'écrit et de l'oral _____	18
Table II-1 :	Échelle d'intelligibilité ALPAC _____	29
Table II-2 :	Échelle de fidélité ALPAC _____	29
Table II-3 :	Définition des notes de bonne formation linguistique proposées par le NIST _____	40
Table II-4 :	Définition des notes de l'adéquation proposées par le NIST _____	42
Table II-5 :	Calcul de précision modifiée de BLEU – exemple 1 _____	46
Table II-6 :	Calcul de précision modifiée de BLEU – exemple 2 _____	46
Table II-7 :	Précision BLEU sur un exemple _____	50
Table II-8 :	Précision NIST sur un exemple _____	50
Table V-1 :	Comparaison des taux de succès en désambiguïsation automatique et en désambiguïsation interactive avec KANT _____	85
Table VII-1 :	Situations expérimentales de collecte de corpus avec EMMI (données factuelles)	106
Table VII-2 :	Classes d'ambiguïtés (base de travail) et exemples tirés de l'expérience pilote _____	123
Table VII-3 :	Classes d'ambiguïtés (base de travail) et exemples tirés de la seconde expérience	127
Table VII-4 :	Facilité des questions _____	131
Table VII-5 :	Recommandations des sujets _____	132
Table VII-6 :	Stratégie adoptée pour répondre aux questions _____	132
Table VII-7 :	Prédictibilité des questions _____	132
Table IX-1 :	Résultats d'évaluation subjective du système de NEC en 2002 _____	167
Table X-1 :	Résultats de l'évaluation de masse dans Verbmobil en qualité de la traduction ____	175
Table XI-1 :	Comparaison des IF C-STAR II et NESPOLE! sur quelques exemples _____	194
Table XI-2 :	Résultats de la première évaluation dans NESPOLE! _____	205
Table XI-3 :	Accord entre juges – It ou Fr vers Fr _____	207
Table XI-4 :	Accord entre juges – Fr vers It _____	208
Table XI-5 :	Fidélité de chacun des juges au cours du temps – It ou Fr vers Fr _____	208
Table XI-6 :	Fidélité de chacun des juges au cours du temps – Fr vers It _____	208
Table XI-7 :	Coefficients Kappa des juges pour l'évaluation du second démonstrateur NESPOLE! _____	208
Table XI-8 :	Résultats de la seconde évaluation du projet NESPOLE! _____	209
Table XII-1 :	Résultats de l'évaluation pilote interne du consortium C-STAR III : _première condition _____	214
Table XII-2 :	Résultats de l'évaluation pilote interne du consortium C-STAR III : _seconde condition _____	214
Table XII-3 :	Résultats de l'évaluation objective des systèmes Systran pour la paire chinois-anglais _____	216
Table XII-4 :	Résultats de l'évaluation objective des systèmes Systran pour la paire japonais-anglais _____	216
Table XII-5 :	Résultats de l'évaluation objective du système J_3 Systran japonais-anglais révisé	216

Table XII-6 :	Résultats de l'évaluation subjective des systèmes japonais-anglais illimités _____	217
Table XII-7 :	Résultats de l'évaluation objective des systèmes chinois-anglais illimités _____	217
Table XII-8 :	Résultats de l'évaluation subjective des systèmes japonais-anglais illimités _____	217
Table XII-9 :	Résultats de l'évaluation objective des systèmes japonais-anglais illimités _____	217
Table XII-10 :	Résultats de l'évaluation objective des systèmes japonais-anglais illimités avec J_4 _____	218
Table XII-11 :	Coefficients Gamma et Kappa pour l'accord inter-juges de l'évaluation subjective IWSLT 2002 pour Systran _____	219
Table XII-12 :	Liste des vingt hypothèses de reconnaissance sur les énoncés : « oui je le vois » et « d accord très bien» _____	229
Table XII-13 :	Évaluation de la meilleure hypothèse et des vingt hypothèses _____	231
Table XII-14 :	Distribution des meilleures traductions selon le rang de l'hypothèse _____	231
Table XII-15 :	Évolution des notes lorsque la qualité de la traduction s'améliore _____	232
Table XII-16 :	Exemples de classes IF obtenues automatiquement _____	233
Table XII-17 :	Exemples d'hypothèses de reconnaissance obtenues avec le modèle de langage sémantique _____	234

Liste des équations

Équation I-1 :	Décodage en traduction statistique _____	25
Équation II-1 :	Calcul de la précision avec BLEU _____	47
Équation II-2 :	Calcul de la pénalité de brièveté de BLEU _____	47
Équation II-3 :	Calcul de BLEU _____	47
Équation II-4 :	Mesure d'informativité d'un n -gramme proposée par NIST _____	48
Équation II-5 :	Calcul de la pénalité de brièveté de NIST _____	48
Équation II-6 :	Calcul de NIST _____	49
Équation II-7 :	Calcul de ROUGE-L _____	50
Équation II-8 :	Calcul explicite de la F-mesure en fonction de n , m , et $LCS(X, Y)$ _____	51
Équation II-9 :	Calcul de ROUGE-L _{multi} _____	51
Équation II-10 :	Calcul de ROUGE-W _____	52
Équation II-11 :	Calcul de ROUGE-S _____	52
Équation VII-1 :	Propriétés des termes participant à une équation analogique _____	139
Équation XII-1 :	Efficacité relative d'un système de traduction de dialogue _____	224

Introduction

J'ai commencé à faire de la recherche en 1989 au sein du Laboratoire GETA (Groupe d'Étude pour la Traduction Automatique), en travaillant sur une nouvelle approche de la TA Interactive, dans le but de démocratiser un jour la TA de qualité. Après un DEA et une thèse soutenue en janvier 1994 suivis d'un stage post-doctoral de 15 mois (1994-1995) à ATR-ITL au Japon, je suis devenu Maître de Conférences en septembre 1995 au département informatique de l'IUT2 à l'Université Pierre Mendès France.

Mes travaux de recherches ont donc d'abord concerné la **traduction de l'écrit** dans le cadre du **projet LIDIA** (Large Internationalisation des Documents par Interaction avec l'Auteur) de Traduction Automatisée Fondée sur le Dialogue (TAFD) pour auteur monolingue. Ce concept avait été proposé en 1989 par Christian Boitet pour essayer d'obtenir, sur des textes non contraints une qualité équivalente à celle obtenue sur des sous-langages restreints (comme des manuels techniques).

L'idée est de permettre à un auteur monolingue rédigeant un document de le faire traduire à partir de son environnement de rédaction. Dans ce cadre, l'auteur aide le système à traduire vers une ou plusieurs langues cibles, via une standardisation (normalisation) et une clarification (désambiguïsation) interactives effectuées en langue source, et en une seule fois, quel que soit l'ensemble des langues cibles visées.

En TAFD, le processus de **désambiguïsation interactive** joue un rôle crucial. Il s'agit d'abord de détecter les **ambiguïtés** que le module d'analyse du système de TA n'a pas été capable de résoudre automatiquement. Comme nous ne voulons exiger aucune compétence particulière de l'auteur ni en linguistique, ni dans les modèles et formalismes de représentation utilisés par le système de TA, les questions de désambiguïsation doivent être libellées de la manière la plus simple possible.

Pour ce faire, et afin de ne pas introduire de nouvelles ambiguïtés dans les questions, nous avons choisi d'utiliser un certain nombre d'opérations simples (parenthésage, utilisation de ponctuation, déplacement des mots de la phrase) sur le texte afin de révéler les différentes interprétations plutôt qu'un vrai module de génération.

Nous avons donc proposé et implémenté une architecture logicielle pour la désambiguïsation interactive. Un désambiguïseur interactif est construit autour d'un moteur qui est indépendant des langues et des ambiguïtés à résoudre. Ce moteur propose: – un langage de description des ambiguïtés auquel est associé un module de reconnaissance des ambiguïtés décrites, – un ensemble d'opérateurs qui permettent de produire des items de questions à partir des informations collectées lors de l'étape de reconnaissance d'une

ambiguïté, et – un mécanisme de présentation des questions à l'utilisateur afin que celui-ci puisse y répondre.

Pour instancier un module de désambiguïsation le moteur est complété par un logiciel, dépendant de la langue, qui décrit des ambiguïtés particulières à résoudre ainsi que la manière de construire les questions qui seront proposées à l'auteur. En suivant cette approche, j'ai implémenté des modules de désambiguïsation interactive du français et de l'anglais. Afin de valider l'**utilisabilité** de cette approche, j'ai conduit deux évaluations, en collaboration avec Laurel Fais (ATR-ITL), afin de savoir si les questions posées étaient ou non compréhensibles. Nous avons heureusement obtenu des résultats satisfaisants.

Mon retour en France a coïncidé avec la création du laboratoire CLIPS (Communication Langagière et Interaction Personne-Système) intégrant les équipes GETA et GEOD (Groupe d'Étude de l'Oral et du Dialogue). Ce rapprochement d'équipes spécialisées dans le traitement de l'écrit et le traitement de l'oral a été le catalyseur d'une nouvelle aventure, la traduction de parole.

En rejoignant le consortium C-STAR II (Consortium for Speech Translation Advanced Research) en 1996, nous avons pu coopérer avec des groupes pionniers au Japon, en Corée du Sud, en Allemagne, en Italie et aux États-Unis. L'objectif du consortium était alors de faire des démonstrations communes en 1999 afin de faire connaître la traduction de dialogue auprès du plus grand nombre.

En **traduction de dialogues** oraux, il faut reconnaître ce que dit celui qui parle, traduire ce qui a été dit dans la langue de celui qui écoute, et synthétiser le résultat obtenu afin que celui qui écoute puisse entendre la traduction. On cumule donc les problèmes de la reconnaissance de la parole, de la traduction d'une production linguistique orale qui n'est pas aussi propre qu'une production écrite, et de la synthèse naturelle de la parole.

Dans le cadre du **projet C-STAR II**, j'ai été chargé d'animer, puis de diriger le groupe CLIPS++, composé des équipes GEOD et GETA du CLIPS, ainsi que du LATL (Genève, Suisse) et du LAIP (Lausanne, Suisse) jusqu'aux démonstrations du 22 juillet 1999. Il eût été bien trop ambitieux que le CLIPS s'occupe seul du français. Personnellement, j'ai participé à l'intégration des différents modules et réalisé un module d'analyse (français \rightarrow IF [un langage pivot]) très efficace (approche multimoteur) fondé sur la reconnaissance de patrons de phrases.

Forts de leur expérience et du succès des démonstrations, les partenaires européens et américain de **C-STAR II** ont proposé conjointement le **projet NESPOLE!** (NEgotiating through SPOken Language in E-commerce) devant la Communauté Européenne et la National Science Foundation. Ce projet avait pour but de réaliser deux démonstrateurs (en domaine restreint et en domaine élargi) de traduction de dialogue (en 4 langues) sur Internet en utilisant une architecture par pivot qui avait démontré son intérêt dans le cadre de **C-STAR II**.

Dans le cadre de **NESPOLE!**, j'ai assuré la direction de l'équipe du CLIPS et implémenté les modules d'analyse (français \rightarrow IF) et de génération (IF \rightarrow français) pour chacun des deux démonstrateurs. Pour l'analyse, j'ai choisi une approche par patron qui permet de lutter contre le **bruit** et d'instancier des fragments d'IF. Pour la génération, nous avons expérimenté deux approches : une approche visant une couverture complète de la spécification de l'IF avec ARIANE-G5 (le générateur de systèmes de TA développé au GETA), et une approche fondée sur la concaténation de segments textuels produits à partir de certains éléments de l'IF.

Dans le cadre du projet **NESPOLE!**, une importance particulière a été accordée à l'évaluation de la **qualité** des traductions produites avec les deux démonstrateurs. Ces évaluations ont été réalisées en utilisant des protocoles d'évaluation subjective (réalisés par des juges humains). Elles ont permis de vérifier que les systèmes implémentés ont des performances comparables aux autres systèmes état de l'art. Nous avons aussi cherché à mesurer l'évolution des performances entre le démonstrateur en domaine restreint et le démonstrateur en domaine élargi.

À la suite du projet C-STAR II, le **projet C-STAR III**, dont j'ai pris la présidence en mars 2004, est officiellement né en octobre 2000 en accueillant un nouveau partenaire en Chine. Ce consortium poursuit deux objectifs. Le premier consiste en la construction d'un corpus multilingue (japonais, coréen, chinois, allemand, italien, français, anglais) aligné d'énoncés dans le domaine du tourisme. Ces énoncés sont du type de ceux que l'on trouve dans un livre de phrases pour le tourisme. Le consortium a aussi pour objectif d'organiser des campagnes d'évaluation compétitive de systèmes de traduction de parole sur le corpus que nous construisons.

Depuis la fin du projet NESPOLE!, j'ai cherché à faire converger les deux thèmes sur lesquels j'ai successivement travaillé, tout en approfondissant chacun d'eux. Je suis revenu sur la désambiguïsation interactive de document en améliorant l'architecture d'un système de TAFD, et j'ai commencé à prototyper le concept de **Document Auto-Explicatif** (DAE) que nous avons inventé avec Christian Boitet en 1994, comme une sorte de « retombée » monolingue des travaux en TAFD.

J'ai aussi commencé à regarder comment une dose raisonnable de désambiguïsation interactive pourrait améliorer la TA de dialogue en liaison avec les efforts de recherche et développement de Mark Seligman (ancien de ATR lui aussi) qui travaille sur un tel produit pour sa société Spoken Translation Inc. qu'il a lancé en 2001.

J'en suis venu à réaliser qu'il s'agissait finalement de résoudre des aspects différents d'un problème très générique, celui de la communication multilingue médiatisée par la machine. Certes, les contextes, et donc les contraintes varient beaucoup, du textuel au multimodal en passant par le vocal seul, ou de la « monographie » asynchrone au dialogue en face à face en passant par le dialogue écrit (tchat) et la téléconférence multimodale.

En définitive, il s'agit toujours de combattre le bruit et l'ambiguïté, de construire des systèmes génériques facilement spécialisables à des situations particulières, et d'évaluer soit des systèmes, soit des architectures.

En écrivant ce mémoire, j'ai voulu préciser cette vue unificatrice, et m'appuyer sur elle pour revisiter mon activité de recherche sous un nouveau jour. Puisqu'il s'agit finalement de construire de « bons » ou de « meilleurs » systèmes, je mettrai en un certain sens l'évaluation au centre de ma présentation. Le thème fédérateur en sera alors l'évaluation de la qualité, de l'utilisabilité et de l'utilité des systèmes de TAO de l'écrit et de l'oral.

Dans la mesure du possible, chaque approche ou solution présentée sera confrontée à une évaluation de ses conséquences et de ses perspectives par rapport à ces trois propriétés.

Ce document comporte trois parties, constituée chacune de quatre chapitres.

La première partie, intitulée « **Qualité en communication interpersonnelle assistée par la machine** », présente notre domaine de recherche en introduisant le vocabulaire et les concepts essentiels à la suite de l'exposé.

Le premier chapitre (chapitre I) propose des critères de comparaison permettant de construire des taxonomies. Nous essayons d'abord de caractériser la communication interpersonnelle selon plusieurs dimensions, en nous limitant à celles qui sont pertinentes pour nos travaux. Il s'agit de la ou des modalités mises en œuvre, de la qualité des participants, de l'unité de temps et de lieu, de la planification du contenu à communiquer, et des assistants possibles à la communication.

Comme nous nous intéressons en particulier à la communication interpersonnelle multilingue assistée par la machine, nous présentons aussi une caractérisation des systèmes de traduction. Nous nous attachons en particulier à décrire les différentes caractérisations et mesures de la qualité des systèmes (chapitre II). Le domaine de l'évaluation de la TA est ancien, et connaît un regain d'intérêt certain depuis 1999-2000. Nous détaillons d'abord quelques grands travaux qui constituent les fondations des méthodes actuelles. Nous dressons ensuite un état des lieux des différentes méthodes d'évaluation pratiquées, qu'elles soient subjectives (mesures produites par des humains) ou objectives (mesures produites automatiquement par des programmes informatiques). Nous insistons, en particulier, sur la base de publications très récentes, à l'heure de la rédaction de ce mémoire, sur la prolifération de travaux en évaluation objective. Nous retenons finalement une terminologie et les éléments les plus importants.

Produire des résultats de qualité est, en soi, un objectif ultime et louable de tous les concepteurs de systèmes de traduction, mais c'est très difficile. Nous nous intéressons donc dans le troisième chapitre (chapitre III) aux différents obstacles à la qualité en traduction. Nous identifions trois classes d'obstacles à la qualité. La première classe d'obstacles est liée à la *modalité orale* et concerne tous les phénomènes que les systèmes automatiques vont considérer en quelque sorte comme des parasites qui nuisent à leur performance. La seconde ressortit à la *taille du domaine* à couvrir par le système. La troisième est, bien sûr, liée à *l'ambiguïté de la langue*.

En réfléchissant à la notion d'ambiguïté, nous sommes conduit à proposer une définition formelle de l'ambiguïté que nous utiliserons largement dans la seconde partie du mémoire. Cette première partie se conclut (chapitre IV) par une description de différentes situations. Nous insistons, en particulier, sur les deux situations de communication interpersonnelle multilingue assistée par la machine qui sont au cœur des deux parties suivantes : la **communication interpersonnelle écrite réfléchie tout terrain** et la **communication interpersonnelle orale spontanée**.

La seconde partie présente notre contribution à l'amélioration de la qualité en « **communication interpersonnelle écrite réfléchie tout terrain** ».

Le premier chapitre (chapitre V) présente assez rapidement différentes familles de techniques mises en œuvre en traduction automatique afin de tendre vers une meilleure qualité. Nous sommes assez bref et renvoyons le lecteur vers des références qui traitent du sujet plus en détail. En effet, un ouvrage de la taille de ce mémoire n'y suffirait pas. Nous

actualisons ensuite l'état de l'art en traduction interactive que nous avons proposé en 1994.

Nous concluons ce chapitre par la proposition d'une approche permettant d'aller plus loin en TA de qualité, la **Traduction Automatisée Fondée sur le Dialogue** (TAFD). C'est une idée que nous avons mise en œuvre dans le cadre du projet LIDIA à partir de 1990. En TAFD, bien que les sources de connaissances linguistiques soient encore cruciales, et que des connaissances extra-linguistiques puissent être utilisées si elles sont disponibles, l'accent est mis sur la pré-édition indirecte au moyen d'un dialogue de désambiguïsation avec l'auteur, permettant d'obtenir des traductions de haute qualité sans révision.

Le chapitre suivant (chapitre VI) donne un aperçu de la première phase de nos travaux (1989-95) au travers des différentes versions de la maquette LIDIA-1. Nous décrivons l'architecture linguicielle proposée, architecture qui n'a jamais été révisée depuis notre proposition originale. Nous donnons ensuite un aperçu de l'évolution vers une architecture informatique qui s'affranchit d'un mode d'accès au système de traduction trop câblé, au sens propre. Une petite démonstration conclut ce chapitre consacré à la version historique de la maquette LIDIA-1.

Le chapitre suivant (chapitre VII) fait le point sur l'évolution de nos travaux sur la **désambiguïsation interactive** et l'**architecture informatique « idéale » pour la TAFD**. Nos travaux sur la désambiguïsation interactive se sont déroulés dans deux directions. La première concerne le recensement et l'annotation de corpus. Cette activité s'est développée dans le cadre du projet MIDDIM. Ce projet, d'une durée de trois ans entre le GETA et ATR (Kyoto, Japon), s'est déroulé sous l'égide du CNRS. Le deuxième axe de nos travaux sur la désambiguïsation interactive concerne l'implémentation informatique d'un module de désambiguïsation interactive. Une instance d'un module de désambiguïsation interactive utilise un moteur de désambiguïsation et un linguiciel de désambiguïsation. Le moteur offre les services de base utilisés par tout module de désambiguïsation. Le linguiciel décrit d'une part les ambiguïtés propres à l'application au sein de laquelle le module de désambiguïsation doit être intégré, et d'autre part la manière de préparer les questions de désambiguïsation destinées à les résoudre.

Nous rendons aussi compte de deux expériences que nous avons conduites afin de vérifier que les dialogues de désambiguïsation interactive produits avec la méthode que nous proposons sont compréhensibles. Sur le thème d'une meilleure architecture informatique pour la TADF, les évolutions ultérieures, LIDIA-2 et LIDIA-3, sont liées à plusieurs facteurs : lutte contre l'obsolescence, lutte contre le support sur une plate-forme matérielle unique, et nécessité de tendre vers une architecture flexible qui rende la TAFD disponible en tout lieu et sur toute plate-forme matérielle.

Le dernier chapitre de cette partie (chapitre VIII) introduit et développe au moyen d'une première implémentation le concept de **Document Auto-Explicatif** (DAE). Au cours du développement de la maquette LIDIA-1, nous avons naturellement été conduits à l'idée que les informations obtenues par le système lors de la phase de désambiguïsation interactive peuvent être conservées afin d'enrichir le document avec le sens qu'il véhicule. Un tel document enrichi est un DAE. Un visualiseur de DAE, dont nous présentons une première réalisation, peut montrer au lecteur où se trouvent les ambiguïtés et, à sa demande, préciser le sens choisi par l'auteur.

Finalement, la troisième partie aborde la « **communication interpersonnelle orale spontanée spécialisée** ».

Dans le premier chapitre (chapitre IX) nous présentons un historique et une analyse du domaine. Nous nous intéressons d'abord aux systèmes de communication entre une personne et un système informationnel. Nous pensons en effet que ce domaine est extrêmement proche du domaine de la traduction de parole par les verrous scientifiques et technologiques auxquels il est confronté. Il nous semble donc essentiel de présenter ce domaine dont, à notre connaissance, les travaux en traduction de parole ne s'inspirent pas assez. Nous montrerons ensuite comment s'opère la transition de la traduction de parole vers la traduction de dialogue.

Le chapitre suivant (chapitre X) est consacré aux grands projets achevés en traduction de dialogues oraux finalisés. Il s'agit des projets C-STAR II (1996-99), Verbmobil (1993-2000) et NESPOLE! (2000-02). Nous présentons principalement les architectures informatique et linguistique. Pour le projet Verbmobil, nous commentons aussi les résultats des évaluations conduites dans le cadre du projet.

Nous présentons alors en détail (chapitre XI) notre contribution aux projets C-STAR II et NESPOLE!. Dans ces deux projets, l'architecture linguistique que nous avons mise en œuvre est une architecture à pivot sémantique. Notre contribution concerne, entre autres, le développement de modules d'analyse du français vers le pivot "IF" et de génération du pivot vers le français. Nous montrons comment nous avons essayé de maintenir une certaine qualité malgré le « **bruit** » présent dans l'entrée du module d'analyse (produit par un module de reconnaissance de la parole), et malgré les problèmes de spécification des formules IF constituant l'entrée du module de génération. Nous nous arrêtons aussi sur l'évolution du pivot IF utilisé dans le cadre de ces deux projets et en proposons une analyse critique. Nous présentons finalement en les commentant les résultats de deux campagnes d'évaluation conduites dans le cadre du projet NESPOLE!.

Le dernier chapitre (chapitre XII) propose une revue critique des travaux en cours en **communication interpersonnelle orale spontanée spécialisée**. Nous décrivons d'abord les travaux du consortium C-STAR III autour d'une évaluation compétitive de systèmes de traduction de parole. Nous présentons quelques verrous importants qui semblent encore à résoudre dans le domaine. Nous proposons enfin quelques pistes pour les résoudre que nous comptons bien sûr suivre dans la suite de notre recherche.

Première Partie

Qualité en communication interpersonnelle médiatisée par la machine

Introduction

Les systèmes de traduction automatique de l'écrit et de l'oral peuvent être vus comme des médiateurs qui permettent la **communication interpersonnelle¹ multilingue** entre plusieurs individus. Selon la qualité des traductions, la conception des interfaces, les domaines couverts, et les temps de réponse, l'**utilisabilité** et l'**utilité** des systèmes, une fois déployés, pourront varier considérablement. L'efficacité de cette forme de communication en dépendra crucialement. Pour évaluer *a priori* cette efficacité, avant de mettre en service à grands frais des systèmes opérationnels, on cherche à faire des projections (inférences) crédibles à partir d'évaluations faites sur des prototypes, ou même sur certains composants seulement. Cela n'est pas du tout facile, car l'expérience a prouvé que des systèmes produisant de mauvaises traductions du point de vue des traducteurs professionnels peuvent être jugés comme très utiles², tandis que d'autres, donnant des traductions de grande qualité sous certaines conditions, sont restés au niveau préopérationnel³.

L'évaluation de la traduction automatique est un sujet très important depuis les années 60. La littérature sur le sujet est peut-être même plus importante que la littérature sur les techniques de traduction automatique elles-mêmes.

L'évaluation de la traduction automatique peut avoir plusieurs objectifs :

- aider les utilisateurs potentiels à choisir le système qui correspond le mieux à leurs besoins,
- aider les pourvoyeurs de fonds à choisir les technologies sur lesquelles investir,
- aider les développeurs à mesurer différents aspects de leurs systèmes et à mesurer leurs progrès.

Pour ce dernier objectif, le coût important des méthodes d'évaluation subjective (conduites avec des juges humains) a motivé le passage vers des méthodes d'évaluation objectives (ce sont des programmes informatiques qui évaluent). Le nombre des méthodes proposées en évaluation objective a explosé depuis l'année 2002. On assiste maintenant, en 2004, à une véritable inflation dans ce domaine.

¹ Que cette communication par traduction soit *impersonnelle* comme en diffusion de documents écrits (manuels, pages web) ou de messages oraux (informations, avertissements...), ou *personnelle* comme en dialogue écrit ou oral, elle est bien *interpersonnelle*, par opposition à la communication entre personnes et systèmes.

² Par exemple, une ancienne version de Systran russe-anglais utilisée par Euratom à Ispra en 1972.

³ Par exemple, le système russe-français du GETA, pourtant testé avec succès sur de grands volumes entre 1981 et 1987 par le CEDOCAR, ne tournait que sous un OS particulier, et n'a pu de ce fait être installé et utilisé qu'en interne.

La traduction automatique est une tâche extrêmement difficile. La langue naturelle n'est pas un langage formel créé de toutes pièces au moyen d'une syntaxe et d'une sémantique formelles, et pour lequel on pourrait construire un analyseur exact. Bien sûr, on peut observer et formaliser des régularités et des propriétés des énoncés acceptables en langue naturelle. Si cela n'était pas possible, comment serions-nous capables d'apprendre et de parler une langue ? Mais on doit se contenter de formaliser une syntaxe et une sémantique approchées.

D'autre part, toute langue naturelle (écrite ou orale) est intrinsèquement ambiguë pour les êtres humains. La situation est donc bien pire pour les analyseurs automatiques qui n'ont à leur disposition qu'un « petit » sous-ensemble des connaissances humaines. En analyse de la langue orale, on est confronté au relâchement de la « syntaxe standard » de l'écrit, la formalisation est donc plus difficile. Enfin, en reconnaissance automatique de la parole, on est confronté à d'autres types de problèmes. Tous ces éléments constituent des obstacles à la qualité.

Dans cette première partie, nous proposons :

- des éléments de caractérisation des diverses situations de communication interpersonnelle, et des systèmes de traduction automatique,
- un panorama des méthodes d'évaluation subjective et objective proposées à ce jour, et
- un recensement des différents obstacles à la qualité en communication interpersonnelle médiatisée par un système de traduction automatique.

Nous présentons enfin une classification des différents types de communication interpersonnelle, organisée autour de cinq axes. Elle nous servira par la suite à proposer une vue unificatrice de la communication interpersonnelle multilingue médiatisée par la machine.

Chapitre I

Communication interpersonnelle et traduction automatique

Pour définir et analyser la communication interpersonnelle et la communication personne-système, D. Vernant [Vernant, D. 2003] puise son discours dans la philosophie du langage et de la communication. Il définit d'abord la « personne » comme co-agent d'un procès de transactions et d'interactions, puis analyse la communication interpersonnelle dans ses dimensions transactionnelles et interactionnelles. Il aborde ensuite la communication personne-système et mesure l'écart qui sépare ces deux types d'interaction. Nous reviendrons sur l'interaction personne-système dans la section IX.1.

Nous adoptons ici un point de vue résolument différent. Nous nous intéressons uniquement aux traces d'activités langagières qui sont produites lors de la communication. Notre domaine d'étude concerne plus particulièrement la communication entre personnes ne partageant pas la même langue. Avant d'être transmis, un contenu informationnel, en langue source, doit donc être traduit vers une langue cible. La fonction traduisante peut être assurée par un humain ou un système de traduction automatique (ou automatisé).

I.1. Caractérisation de la communication interpersonnelle

Nous proposons d'abord des critères de caractérisation de la communication interpersonnelle qui nous permettront de classer les situations communicationnelles.

I.1.1. Modalités de communication

La première dimension de caractérisation de la communication interpersonnelle qui est utile est la modalité de communication.

Le support de la communication peut être écrit, nous parlerons alors de communication écrite.

Le support de la communication peut être oral, nous parlerons alors de communication orale.

La communication peut aussi être multimodale. Dans ce cas, l'émetteur échange de l'information avec le récepteur en utilisant plusieurs supports. En communication interpersonnelle, il s'agit principalement de co-utiliser l'oral et l'écrit.

I.1.2. Émetteurs et récepteurs de la communication

La modalité employée pour la communication interpersonnelle permet de nommer les participants.

En communication interpersonnelle écrite, l'émetteur est un **auteur** et le récepteur un **lecteur**, sauf dans le cas du tchat, mais nous y reviendrons.

En communication interpersonnelle orale, deux familles de situations peuvent être envisagées. La première famille est constituée des situations lors desquelles les participants changent de rôle au cours du temps à des intervalles assez brefs (ce qui est le cas lors d'une conversation, d'un dialogue). Les deux participants sont alors tous deux actifs du point de vue langagier. La seconde famille est constituée des situations lors desquelles le rôle des participants ne change pas au cours du temps à des intervalles assez brefs (ce qui est le cas lors d'une conférence, d'un cours, d'une diffusion audio).

Pour le dialogue, [Morel, M.-A. 2004] choisit d'appeler « parleur patenté » l'énonciateur de plein droit à un moment particulier de l'échange et « écouteur » son interlocuteur. Afin de bien distinguer les deux types de situation que nous avons identifiés, nous proposons une autre terminologie.

Pour la première famille de situations, nous dirons qu'à un instant t , l'émetteur est le **locuteur** et le récepteur est l'**interlocuteur**. Pour la seconde famille de situations, le statut de l'émetteur n'est pas remis en question, il est encore considéré comme locuteur. Par contre, le statut du récepteur change, puisque celui-ci demeure passif du point de vue langagier, il est donc **auditeur**.

Type de communication	Émetteur	Récepteur
Écrite	auteur	lecteur
Orale	Bidirectionnelle	interlocuteur
	Monodirectionnelle	auditeur

Table I-1 : Émetteur et récepteur en fonction de la modalité de communication

I.1.3. Unités de temps et de lieu de la communication

La communication interpersonnelle peut aussi être caractérisée par le caractère instantané ou différé de la réception de l'information transmise par l'émetteur au récepteur. Il s'agit alors de savoir si la transmission vérifie le principe d'unité de temps. Si l'intervalle de temps écoulé entre l'émission et la réception est nul, la transmission est considérée comme **instantanée**. Lorsque l'intervalle de temps écoulé entre l'émission et la réception est égal à la durée de production de l'information auquel s'ajoute un bref intervalle, fonction de cette durée, la transmission est dite **consécutives**. Enfin, si on ne peut pas quantifier l'intervalle de temps écoulé entre l'émission et la réception, la transmission est **différée**.

La seconde dimension de caractérisation de la transmission est l'unité de lieu. Il s'agit de savoir si émetteur et récepteur partagent un espace physique commun. En d'autres termes, il s'agit de savoir si les participants (**émetteur** et **récepteur**) sont en situation de face à face ou non. Dans le dialogue interpersonnel en face à face, l'unité de lieu est vérifiée. Dans le

dialogue interpersonnel au moyen d'un système de visioconférence, on peut dire que l'unité de lieu est rétablie dans une certaine mesure. S'ils ne partagent pas un espace physique commun, les participants peuvent se voir. Par contre, lors d'une conversation téléphonique ou lors de l'écoute d'une émission de radio ou de télévision, l'unité de lieu n'est pas respectée.

La communication interpersonnelle écrite ne vérifie pas l'unité de temps, sauf en tchat. En effet, le récepteur n'est pas identifié *a priori* et le moment de la production de l'information et le moment de sa réception sont disjoints. Ils peuvent même être très éloignés dans le temps.

Pour la communication interpersonnelle orale, l'éventail des situations est en revanche plus grand. Le tableau suivant résume ce que nous venons de dire.

Situation	Unité de temps	Unité de lieu
Communication en face à face	oui	oui
Communication en visioconférence	oui	partiellement rétablie
Communication téléphonique	oui	non
Audio diffusion en direct	oui	non
Audio diffusion en différé	non	non

Table I-2 : Unités de temps et de lieu en communication interpersonnelle orale

I.1.4. Domaine et planification de l'information à communiquer

Notre propos n'est pas ici de proposer une nouvelle classification des formes de contenu. [Habert, B. 2000] oppose deux sortes de classifications :

- Une classification *a priori* qui repose sur les conditions de production, sur les buts (pris ici au sens de fonctions) visés, sur l'inscription dans des « genres », sur l'emploi ou non de certaines marques linguistiques, ou bien sur une combinaison de ces critères.
- Une classification *a posteriori* pour laquelle les types obtenus reposent sur des propensions d'un groupe de contenus à recourir à un ensemble de traits linguistiques et à en éviter d'autres. Ces types émergent grâce à un traitement statistique des textes [Habert, B. 2000, section 4].

[Biber, D. 1989, pp. 5-6] définit la notion de « genre » comme une typologie de sens commun : « les genres sont les catégories de textes distinguées spontanément par les locuteurs confirmés (*“matures”*) d'une langue ; par exemple, les genres de l'anglais incluent les romans, les articles de journaux, les éditoriaux, les articles de recherche (*“academic articles”*), les discours en public, les nouvelles radiophoniques, et la conversation de tous les jours »⁴.

Pour la communication [Biber, D. 1988] distingue huit composantes principales raffinées comme suit :

1. Le rôle et les caractéristiques des participants
 - 1.1. Rôle communicatif des participants**
 - 1.1.a. Émetteur
 - 1.1.b. Destinataire
 - 1.1.c. Audience

⁴ Traduction de Benoît Habert (Ibid.)

- 1.2. Caractéristiques personnelles
 - 1.2.a. Stables : personnalité, intérêts, croyances, ...
 - 1.2.b. Temporaires : humeur, émotions, ...
- 1.3. Caractéristiques de groupe
 - 1.3.a. Classe sociale, groupe ethnique, genre, âge, métier, éducation, ...
- 2. Les relations entre participants
 - 2.1. Relation de rôle social : pouvoir social relatif, statut, ...
 - 2.2. Relation personnelle : affinité, respect, ...
 - 2.3. Étendue des connaissances partagées
 - 2.3.a. Connaissances culturelles du monde
 - 2.3.b. Connaissances spécifiques personnelles
 - 2.4. « Pluralité » des participants
- 3. Le cadre**
 - 3.1. Contexte physique
 - 3.2. Contexte temporel
 - 3.3. Type d'activité supérieure
 - 3.4. Manière dont le temps et l'espace sont partagés entre les participants
- 4. Le sujet**
- 5. Le but**
 - 5.1. Objectifs conversationnels
 - 5.2. Objectifs personnels
- 6. L'évaluation sociale
 - 6.1. Évaluation des événements communicatifs
 - 6.1.a. Valeurs partagées au sein d'une culture
 - 6.1.b. Valeurs d'une sous-culture ou personnelles
 - 6.2. Attitude des participants à propos du texte (contenu communiqué)
 - 6.2.a. Sentiments, jugements, posture
 - 6.2.b. Ton, manière de parler
 - 6.2.c. Degré d'implication envers le contenu, posture épistémologique
- 7. La relation des participants au texte
- 8. Le canal**
 - 8.1. Canal primaire : parole, texte, signes, ...
 - 8.2. Autres canaux disponibles

Figure I-1 : Composants de la communication proposés par Biber

Biber distingue aussi des fonctions linguistiques, qui identifient des types d'information et sont marquées dans le discours ; elles ne nous intéressent pas ici. Finalement, Biber propose des différences situationnelles et fonctionnelles entre l'écrit et l'oral « typique ». Il propose six composants situationnels qu'il définit comme suit :

- 1. Canal physique**
 - 1.1. Écrit ou parole (l'écrit peut être produit pour être parlé — théâtre, discours ; la parole peut être prononcée pour être écrite — dictée d'une lettre)
 - 1.2. Disponibilité ou non des sous-canaux prosodique et paralinguistique
- 2. Usage culturel
 - 2.1. Acquis à la maison ou à l'école
 - 2.2. Évaluation sociale (l'écrit est par exemple souvent mieux considéré)
 - 2.3. Maintien du statut social (refus d'utiliser des simplifications phonologiques et syntaxiques observé chez les classes socio-économiques supérieures)
- 3. Relation de communication entre les participants**
 - 3.1. Degré d'interaction (celui qui écoute peut interagir — poser des questions ; ... — avec celui qui parle ; les lecteurs ne peuvent pas avoir ce type d'interaction avec un texte)
 - 3.2. Degré des connaissances partagées
 - 3.3. Degré de négociation des objectifs
 - 3.4. Efforts consentis pour maintenir la relation
 - 3.5. Étendue de la connaissance culturelle du monde partagée

4. Relation des participants avec le contexte externe

- 4.1. Temps à partager (celui qui parle et celui qui écoute partagent un contexte temporel ; ce n'est pas le cas entre un rédacteur et un lecteur)
- 4.2. Espace partagé (celui qui parle et celui qui écoute partagent normalement un espace commun ; ce n'est pas le cas entre le rédacteur et le lecteur)

5. Relation des participants avec le texte (contenu communiqué)

- 5.1. Degré de permanence du texte ; opportunité d'une interaction avec le texte lors de la production (planification, révision) et en compréhension
- 5.2. Rapidité de production
- 5.3. Rapidité de compréhension
6. Objectif premier de la communication
 - 6.1. Transmettre des idées, transmettre pour des raisons personnelles, interpersonnelles, contextuelles

Figure I-2 : Différences situationnelles de la communication proposés par Biber

Plus récemment, [Biber, D. 1995] a introduit le terme « registre » dans une conception élargie des genres dont l'inventaire est à géométrie variable [Habert, B. 2000, section 3.4.2]. À chaque registre de langue peuvent correspondre une syntaxe, un temps grammatical⁵, une prononciation, ou encore un vocabulaire spécifique, comme le montrent les exemples ci-dessous :

- familier : « J'crèche dans c'te baraque pourrie... »
- courant : « J'habite dans cette vieille maison »
- soutenu : « Je réside dans cette vétuste demeure »

La norme, traditionnellement dénommée « le bon usage », sert de repère à la détermination des registres de la langue et comprend aussi bien la norme prescriptive que la norme d'usage.

Pour utile qu'elle soit en traitement automatique des langues naturelles, la distinction en genre ou en registre ne nous est pas utile ici.

Nous préférons utiliser certains composants de la communication et certaines différences situationnelles ; nous les avons mis en relief dans la Figure I-1, et la Figure I-2.

Nous dirons qu'un domaine du discours (à l'écrit ou à l'oral) — point 4. Figure I-1 — est un **domaine fermé** (ou clos) lorsque les objets du discours sont prédictibles, circonscrits et figés. En ce sens, les bulletins météorologiques, les recettes de cuisine, la réservation d'hôtel, etc. constituent des domaines fermés. À l'opposé, lorsque les objets du discours ne sont pas prédictibles, non circonscrits et mouvants, nous parlerons de **domaine ouvert**. Les articles d'un journal, les conversations de tous les jours, etc. constituent des domaines ouverts, même si localement il est possible de délimiter des domaines fermés.

La planification du discours, du contenu — point 5.1. Figure I-2 — ressortit au caractère réfléchi ou spontané de celui-ci et à ses répercussions sur le matériau langagier produit. Le caractère réfléchi du discours n'est pas l'apanage de l'écrit, comme le caractère non réfléchi n'est pas celui de l'oral.

⁵ Moins le registre est élevé, moins sont employés le passé simple, le passé antérieur, le subjonctif imparfait et le plus-que-parfait.

Voici ici quelques exemples :

Communication	Planification	Écrit	Oral
Réfléchie vers	++	Document diplomatique Traité	Discours diplomatique Journal télévisé
	+	Article scientifique Article de presse Bulletin météorologique	Exposé scientifique Cours, conférence Radio-guidage
	+/-	Listes de diffusion Courrier personnel	Conversation de travail Conversation téléphonique
	-	Forum de discussion	Remue méninges Improvisation
Spontanée	--	Tchat	Conversation impromptue Dialogue entre « jeunes »

Table I-3 : Différentes planifications de l'écrit et de l'oral

Nous nous intéresserons finalement à l'opportunité d'une interaction en compréhension — point 5.1. Figure I-2.

I.1.5. Médiation traduisante dans la communication

Pour les critères de caractérisation que nous avons présentés jusqu'à présent, nous ne nous sommes pas préoccupés de la langue utilisée pour « transporter » le message. Nous nous intéressons maintenant aux situations particulières pour lesquelles la langue de réception du message est différente de la langue d'émission. Dans ce cas, la communication est **médiatisée** par une fonction traduisante.

Cette fonction traduisante peut être assurée par une personne ou par une machine.

Lorsqu'elle est assurée par une personne, s'il s'agit de communication écrite, on parle de **traduction**. S'il s'agit de communication orale, on parle d'**interprétation** et on en distingue deux types : l'interprétation de conférence (simultanée ou consécutive) et l'interprétation de liaison.

En **interprétation simultanée**, l'interprète traduit le discours à mesure que l'orateur le prononce. Le discours et sa traduction sont (presque) simultanés et l'auditeur a, en quelque sorte, l'impression de suivre le discours ou l'intervention comme si l'orateur s'exprimait dans sa propre langue. Entre le discours et sa traduction, il y a juste l'incompressible décalage nécessaire à la compréhension et surtout au fait que l'interprète doit savoir où va la phrase de l'orateur (comment elle va finir) avant de « plonger ».

En **interprétation consécutive**, l'interprète traduit des portions successives du discours. L'orateur parle pendant un temps raisonnable - ni trop longtemps (sinon l'interprète perdrait le fil) ni pas assez (sinon les auditeurs auraient une impression de discours haché et incohérent) - s'interrompt le temps que l'interprète traduise ce qu'il vient de dire, reprend son discours, s'interrompt à nouveau, et ainsi de suite. Les portions du discours et leurs traductions sont consécutives. Le travail en consécutive convient lorsqu'un orateur s'adresse à une assemblée. L'orateur divise son exposé en plusieurs parties, généralement en unités d'information. L'interprète prend des notes puis, quand l'orateur s'interrompt, l'interprète traduit ce qui a été dit. L'interprète se trouve à côté de l'orateur et s'adresse à l'ensemble de l'auditoire.

L'**interprétation de liaison** est une forme d'interprétation « de terrain » où il s'agit de traduire dans les deux sens les propos de deux individus ou de deux groupes présents sur un même site et généralement engagés dans une tâche ou une transaction particulière. L'interprétation se fait à voix haute (parfois même en criant pour couvrir le bruit ambiant) et doit se concentrer exclusivement sur l'efficacité du message, sachant que l'interprète peut se permettre tous les aménagements, toutes les répétitions, toutes les explications, et tous les retours qu'il juge nécessaires.

En **traduction**, sur la base d'un document papier ou d'un fichier électronique rédigé en langue source (la langue maternelle de l'auteur), le traducteur produit une version du même document dans une langue cible.

Lorsqu'elle est assurée par la machine, la fonction traduisante est prise en charge par un système de traduction.

I.2. Caractérisation des systèmes de traduction

L'acronyme TA peut être compris comme Traduction Automatique ou comme Traduction Automatisée. Par Traduction Automatique, on entend une traduction réalisée par un système même s'il y a intervention humaine, pourvu que cette intervention ne nécessite que des compétences monolingues. Par Traduction Automatisée, on entend que la traduction est réalisée par un système assisté par un humain bilingue éventuellement spécialiste, le système pouvant proposer des traductions (par Traduction Automatique ou MT).

Cela correspond à deux activités professionnelles différentes :

- celle du réviseur qui part du texte en langue cible et ne regarde le texte en langue source qu'en cas de problème manifeste (contresens supposé, possible erreur grammaticale).
- celle du traducteur qui lit le texte en langue source d'abord, le comprend, puis le traduit en langue cible.

Avant de poursuivre, nous voudrions aussi rappeler que la TA n'est pas une science ni une collection de recettes, mais une technologie scientifique, soit un ensemble de méthodes qui progressent par intégration d'idées nouvelles et amélioration incrémentale des savoir-faire.

I.2.1. Chronologie des types de TAO⁶

Vers 1949, les USA, puis l'URSS, ont lancé des programmes motivés par le besoin de renseignements. C'est la **TAO du veilleur**, toujours actuelle. Il s'agit de traduire automatiquement, pour obtenir des traductions « grossières », rapidement, en grand volume et à bas coût. La qualité de ces traductions n'est pas essentielle, car elles servent à filtrer des documents, dont les plus intéressants peuvent, si nécessaire, être traduits ou communiqués à des spécialistes bilingues. Préédition et postédition doivent être absentes ou très limitées (ex : séparer les phrases, les formules, les figures...)⁷. Ce besoin est toujours actuel. Cependant, il s'agit maintenant plus de veille scientifique, technique, économique et financière que de renseignement militaire⁸.

⁶ Cette section est extraite de [Boitet, C. 1990a], nous avons jugé inutile de la reformuler.

⁷ Les systèmes Systran sont essentiellement de ce type (par exemple, le système russe-anglais installé depuis 20 ans à la Wright-Patterson Air Force Base traduit, d'après nos informations, environ 18 millions de mots par an, avec une qualité tout à fait satisfaisante pour l'usage visé).

⁸ À titre d'exemple, on peut citer l'accès en anglais à des bases de données japonaises depuis la Suède [Sigurdson, R.

Une quinzaine d'années plus tard, les chercheurs se sont orientés vers la **TAO du réviseur**, dans laquelle on veut produire des traductions « brutes », destinées à être révisées. Dans cette optique, la machine doit remplacer le traducteur, promu réviseur⁹. Typiquement, en moyenne industrielle, une page technique de deux cent cinquante mots est traduite en une heure et révisée en vingt minutes. Avec quatre personnes, on passerait donc de trois pages par heure à douze pages/heure, et on multiplierait donc la productivité par quatre ? Il s'agit d'une limite, le chiffre le plus vraisemblable étant plutôt de huit pages par heure, en comptant une révision plus lourde, de trente minutes par page.

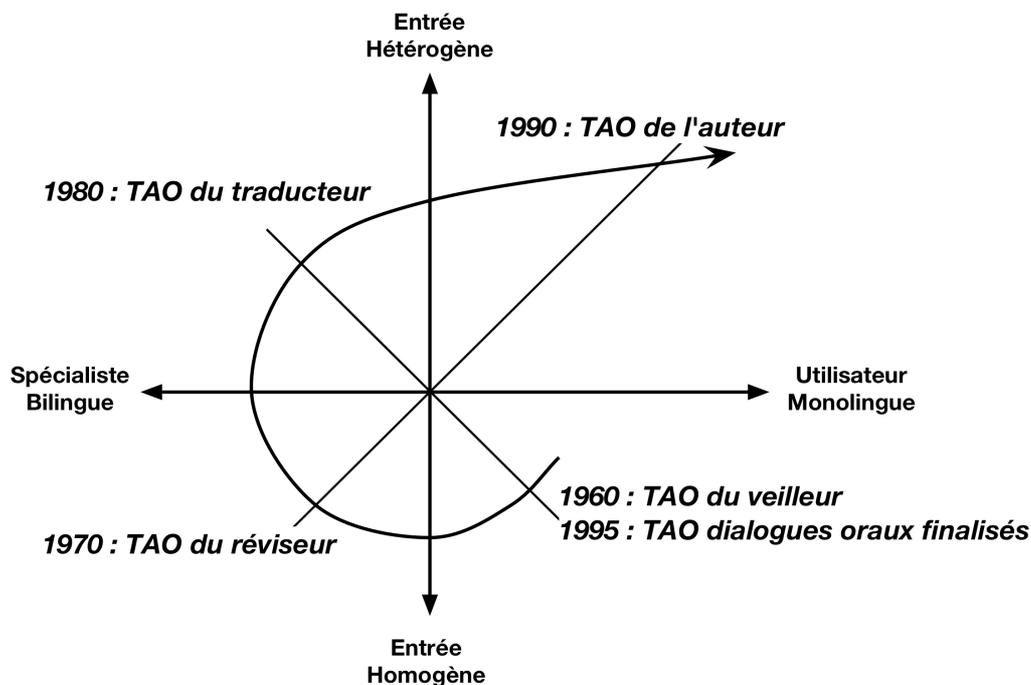


Figure I-3 : Vision chronologique des différents types de TAO

Que faire pour la plus grande partie des textes dont on veut obtenir de bonnes traductions ? La bureautique apporte maintenant des solutions variées, sous forme d'outils de **TAO du traducteur**. Il s'agit de traduction humaine assistée par la machine, ou THAM, et pas de traduction automatique. C'est l'utilisateur qui traduit, en s'aidant de dictionnaires bilingues, de bases terminologiques, de thésaurus de « bitextes » (textes + traductions) etc., accessibles depuis un traitement de texte, le tout formant un « poste de travail pour le traducteur », réalisé sur microordinateur ou station de travail¹⁰. Pour la majorité des besoins, et en particulier pour la traduction de manuels d'enseignement dans des pays où la langue nationale ne s'est que récemment affirmée comme support de l'enseignement secondaire et universitaire, la THAM est actuellement la seule voie réaliste.

Enfin, on commence à voir apparaître des outils de **TAO de l'auteur**, destinés à des personnes ignorant la langue cible. Cela correspond à des besoins croissants. Pour l'instant,

and Greatex, R. 1987], ou depuis la CEE/UE (service Japinfo utilisant des traductions « grossières » d'ATLAS-II de Fujitsu, à peine « arrangées »).

⁹ La plupart des systèmes commerciaux récents visent ce créneau. On peut citer des systèmes japonais (AS-Transac de Toshiba, ATLAS-II de Fujitsu, PIVOT de NEC, HICAT de Hitachi, DUET de Sharp, SHALT d'IBM-Japon, PENSÉE de OKI, Majestic de l'Université de Kyoto et du JICST,...), ou américains (LOGOS, METAL), ou français (Ariane/aéro/F-E de SITE-B'VITAL, fondé sur les outils informatiques et les méthodes linguistiques du GETA). Il est aussi possible d'adapter des systèmes de conception plus ancienne à cet usage, si on les spécialise à un langage fortement contrôlé, comme cela a été fait chez Xerox avec Systran pour la traduction de notices d'anglais en 4 langues.

¹⁰ Il s'agit d'accessoires de bureau (MTX™ sur PC, WinTool™ sur Macintosh), ou d'outils intégrés et plus riches (Trados d'Ink, Translation Manager d'IBM, et Similis de Lingua&Machina).

il s'agit en fait de textes multilingues préenregistrés, personnalisables grâce à des parties variables. Par exemple, Ambassador™, disponible en anglais-japonais, anglais-français, anglais-espagnol et français-japonais, offre environ deux cents "formats" de lettres et formulaires, et 450 « moules » de texte (phrase ou paragraphe)¹¹. Nous nous intéressons plutôt à la traduction de textes libres. Dans ce contexte, il n'y a pas encore de produits commerciaux, bien que le système JETS d'IBM-Japon [Maruyama, H., *et al.* 1990, Tsutsumi, T., *et al.* 1993] ait été proche d'être commercialisé vers 1992-94 ; mais le marketing en décida autrement. En tout état de cause, il ne pourra s'agir que de systèmes fondés sur le dialogue (TAFD), comme nous le verrons plus loin.

1.2.2. Automatisation et types de TAO

Nous distinguons trois types de TAO ;

La TAO 100% automatique correspond à la **TAO du veilleur** ou la **TAO de dialogues oraux**. Elle produit des traductions de basse ou très basse qualité, en satisfaisant ses utilisateurs. Les situations pour lesquelles elle est utile sont la veille, la navigation sur Internet, la traduction dépiage et l'interprétation.

La TAO de bonne qualité et 75-90% automatique correspond à la **TAO du réviseur** et à la **TAO de l'auteur**. Elle produit des traductions « brutes », pour des sous-langages ou des langages contrôlés, destinées à être révisées, ou alors des traductions ne nécessitant pas de révision en faisant appel à l'auteur au moyen de désambiguïsation interactive en langue source.

La TAO de bonne qualité et 15-50% automatique correspond à la Traduction Humaine Assistée par la Machine (THAM) ou **TAO du traducteur**. Elle est constituée d'outils destinés à un utilisateur humain bilingue. Ces outils peuvent être des éditeurs de bitextes, des dictionnaires en ligne ou des mémoires de traductions.

On peut situer historiquement les différents types de TAO selon deux dimensions : le type d'utilisateur et le type de document à traiter (Figure I-3). Les deux types extrêmes d'utilisateurs sont, d'une part, les utilisateurs monolingues et, d'autre part, les utilisateurs spécialistes bilingues. En ce qui concerne les documents à traiter, il est pertinent de savoir s'il s'agit de documents homogènes du point de vue du contenu, ou bien s'il s'agit de documents hétérogènes. Du point de vue de ces deux dimensions, la TAO de dialogues oraux finalisés se situe sur le même point que la TAO du veilleur.

1.2.3. Architectures linguicielles

Le triangle de Vauquois Figure I-4 permet de classer les systèmes de TAO en fonction de l'architecture linguicielle mise en œuvre pour leur implémentation. Les niveaux de représentation linguistique sont indiqués sur la gauche, les types de structures utilisées pour représenter le texte à traduire sont à droite, et le nom de l'architecture est indiqué à l'intérieur du triangle.

¹¹ Disponible sur Macintosh et PC, produit par Language Engineering Corp., Belmont, MA 02178.

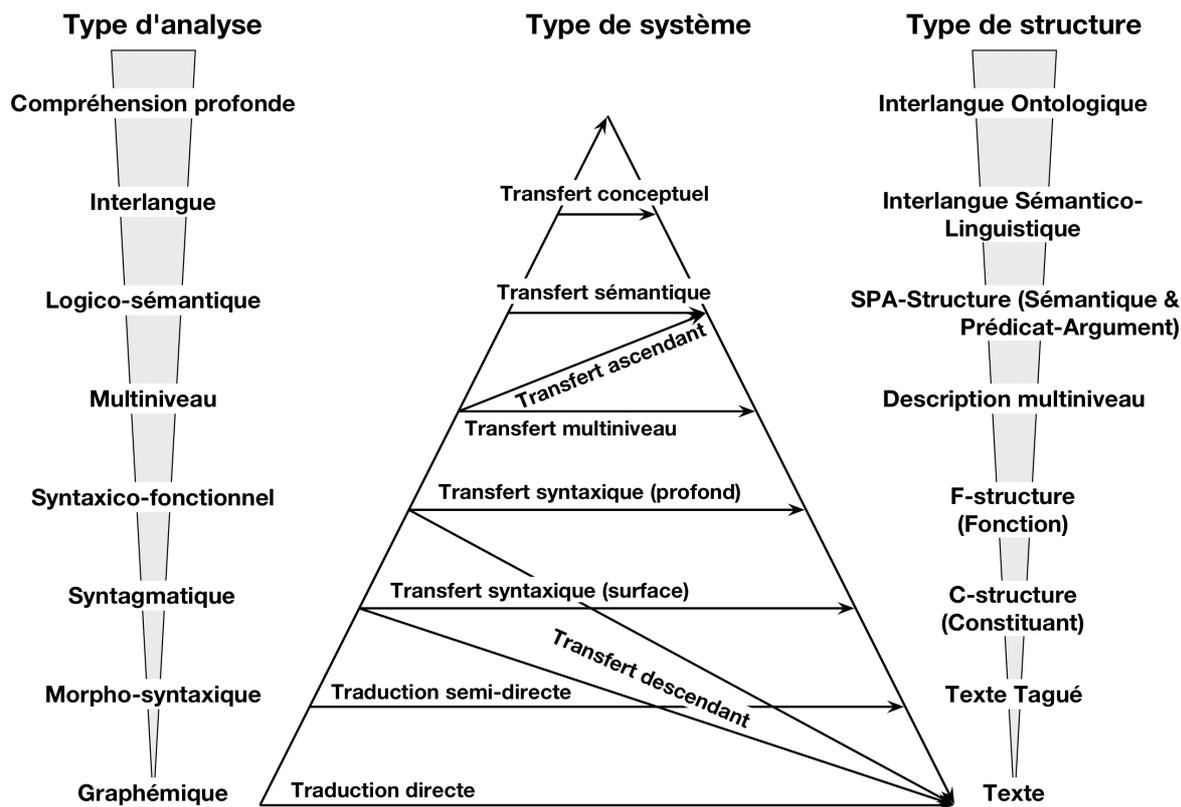


Figure I-4 : Triangle de Vauquois

I.2.3.1. Traduction directe ou semi-directe

Historiquement, le premier paradigme d'implémentation d'un système de TA était fondé sur la similarité supposée entre traduction et déchiffrement¹². Avec cette approche, un texte cible est vu comme le résultat du codage d'un texte en langue cible utilisant des remplacements de mots et des permutations de groupes de mots.

La traduction du texte source vers un texte cible essaie donc de retrouver les meilleurs équivalents en langue cible des mots ou groupes de mots du texte en langue source en regardant leur contexte et en faisant quelques modifications de surface du texte ainsi obtenu (permutations) afin de construire un texte grammaticalement correct en langue cible.

Cette approche ne marche pas très bien, mais les mauvais choix lexicaux et les fautes grammaticales peuvent souvent être corrigés par un lecteur humain cherchant simplement à acquérir des informations. Cette approche, encore utilisée dans la plupart des systèmes commerciaux, est celle des systèmes dits de première génération.

L'approche purement statistique [Brown, P. F., *et al.* 1990], au niveau des mots, est la version actuelle de l'approche directe. L'idée est d'apprendre des correspondances entre mots ou groupes de mots de deux langues en utilisant des corpus de phrases alignées et d'utiliser les correspondances ainsi apprises pour traduire de nouvelles phrases.

¹² Dans une note rédigée à l'intention de la Fondation Rockefeller en 1949, Warren Weaver écrit : "I have a text in front of me which is written in Russian but I am going to pretend that it is really written in English and that it has been coded in some strange symbols. All I need to do is to **strip off the code** in order to retrieve the information contained in the text".

L'expérience a montré que cette technique ne produit pas de meilleurs résultats que la première approche directe. Mais ses défenseurs appliquent maintenant la même idée à des niveaux d'interprétation linguistique plus élevés — utilisant par exemple des grammaires stochastiques aux niveaux syntagmatique ou syntaxique, et plus récemment des grammaires sémantiques — qui rentrent dans les catégories des approches par transfert ou par pivot.

1.2.3.2. Approches par transfert

L'approche par transfert au sens strict se divise en trois étapes : une analyse strictement monolingue qui produit un certain niveau **L** de représentation linguistique, un transfert bilingue (normalement décomposé en phases lexicale et structurale) vers une structure de la langue cible de niveau **L**, et une génération monolingue.

On peut distinguer plusieurs types de transfert : de surface¹³, de syntaxe profonde¹⁴, sémantique¹⁵, et multiniveau (si la structure produite par l'analyseur contient des informations syntaxiques et sémantiques)¹⁶. Quelle que soit la variante utilisée, une architecture linguistique à transfert permet de réutiliser l'analyse et la génération pour plusieurs couples de langues (anglais-français, anglais-allemand, ... ; allemand-anglais, français-anglais, ...).

Par manque d'outils appropriés, certains développeurs n'ont pas été capables de développer la partie structurale de l'étape de transfert. Les systèmes produits¹⁷ mélangent alors le transfert et la génération syntaxique en une seule étape implémentée habituellement par un parcours récursif descendant de l'arbre d'analyse. On obtient en quelque sorte un « transfert descendant ».

Par contraste, B. Vauquois introduisit vers 1978 la technique du transfert ascendant. L'analyse produit une structure multiniveau qui contient des informations syntaxiques et sémantiques, soit : d'une part, à un niveau neutre du point de vue de la langue, des informations logico-sémantiques (composées de structures prédicat-arguments et de relations sémantiques) et, d'autre part des informations de surface comme les fonctions syntaxiques (groupe nominal, groupe verbal, etc.) et les catégories syntagmatiques (sujet, objet direct, etc.). Durant le transfert, des informations de surface peuvent être transférées de la représentation en langue source vers la représentation en langue cible ; mais, lors de la génération, toutes ces informations de surface sont recalculées à partir des informations sémantiques neutres du point de vue de la langue. À ce point, les informations de surface proposées par le module de transfert sont utilisées uniquement comme guide stylistique ou filet de sécurité dans le cas où des informations sémantiques nécessaires sont absentes de la structure de transfert. L'objectif général est de minimiser le transfert des informations de structure de surface pour s'appuyer, autant que possible, sur le transfert d'informations sémantiques. Cela permet d'exprimer cette information sémantique sous forme idiomatique sans ancrage excessif à la structure de surface de la langue source.

¹³ Exemple : TDMT (ATR) [Iida, H., *et al.* 1996, Takezawa, T., *et al.* 1998]

¹⁴ Exemple : LMT (IBM) [McCord, M. 1989a, McCord, M. 1989b, Rimon, M., *et al.* 1991]

¹⁵ Exemple : MU (Université de Kyoto) [Nagao, M. and Tsujii, J.-I. 1986, Tsujii, J.-I. 1986] et son successeur MAJESTIC (JICST)

¹⁶ Exemple : système russe-français du GETA [Boitet, C. and Nédobekine, N. 1981]

¹⁷ Exemples : METAL (LRC) [Slocum, J. 1987, Slocum, J., *et al.* 1985], Shalt-1 (IBM) [Tsutsumi, T. 1986]

I.2.3.3. Approches pivot

De façon formelle, l'approche pivot consiste en l'utilisation d'un format intermédiaire standardisé (le pivot) pour représenter l'information à transmettre durant la traduction depuis une langue source vers une langue cible. La nature de ce pivot peut varier. Ce peut être une langue naturelle¹⁸ ou une représentation du sens dans un formalisme artificiel.

Utiliser une langue naturelle comme pivot est envisageable seulement si la sortie du système d'analyse est de haute qualité et linguistiquement bien formée. Sinon, le système de génération devra essayer de produire une traduction en langue cible à partir d'une entrée pivot avec des erreurs qui peut contenir des mots en langue source ou des constructions qui n'ont pas été correctement reconnues ou traduites. Cela ne peut conduire qu'à des traductions de mauvaise qualité.

On peut aussi utiliser une représentation structurale d'une langue naturelle L_p comme pivot (plutôt qu'un texte dans la langue L_p). Pour traduire entre L_p et une autre langue L_c , un seul module de transfert est nécessaire. Dans le cas précédent, deux transferts sont nécessaires entre la langue source et la langue pivot, puis entre la langue pivot et la langue cible. L'approche par pivot structural permet de construire des systèmes à transfert normal et ascendant, mais pas à transfert descendant.

L'intérêt de passer par un pivot structural réside dans le fait qu'aucune ambiguïté n'est ajoutée lors de la « descente » vers la langue L_p qui devra ensuite être analysée. Cette technique a été initialement proposée et implémentée par le CETA à Grenoble [Vauquois, B. 1975] dans les années soixante sous le nom de « pivot hybride ». Elle a été utilisée avec succès dans le cadre du projet DLT [Schubert, K. 1988, Witkam, A. P. M. 1983, Witkam, T. 1988]. Le pivot alors utilisé était l'espéranto augmenté de balises structurales et grammaticales. La même technique est actuellement utilisée par l'équipe IPPI à Moscou [Boguslavsky, I., *et al.* 2000] pour traduire depuis le pivot UNL (Universal Networking Language) vers le russe via des structures de dépendances profondes de l'anglais.

Avec la technique du pivot structural, le niveau de description lexicale utilise des « unités lexicales » qui sont des familles dérivationnelles de lemmes (forme du dictionnaire des mots) liées par des dérivations (« fonctions lexico-sémantiques ») telles que potentialité passive (**observer_verbe** → **observable_adjectif** avec la paraphrase « qui peut être observé »). Les dérivations possibles sont très nombreuses, mais seules celles qui sont productives et utiles pour générer des paraphrases en traduction sont conservées afin de diminuer le coût de création des dictionnaires. Afin de caractériser le sens des unités lexicales, il faut leur adjoindre un identificateur de sens tel que #1 pour **observer/observation** et #2 pour **observer/observance**.

Le pas suivant est de créer un vocabulaire autonome de symboles qui représentent les « acceptions interlingues » (ou des ensembles d'acceptions qui sont des sens de mots). Si la représentation des textes n'utilise pas d'informations linguistiques de surface, alors le résultat est un « interlingua linguistique ». Le meilleur exemple en est le langage UNL [Tsai, W.-J. 2004].

Il est aussi possible de développer un formalisme de représentation pivot purement sémantique et orienté vers une tâche donnée pour un domaine limité tel que la réservation hôtelière, la prise de rendez-vous, etc. Cette approche a été suivie dans plusieurs projets de

¹⁸ Dans ce cas, les modules d'analyse et de génération sont des systèmes de traduction : l'analyse, de la langue source vers la langue naturelle pivot ; la génération, de la langue naturelle pivot vers la langue cible.

traduction de parole tels que **C-STAR** et **Nespole!** qui ont utilisé un pivot appelé IF (Interchange Format). Nous y reviendrons dans le Chapitre XI.

I.2.3.4. Approches fondées sur les données

On distingue deux types d'approches fondées sur les données, les systèmes à mémoire de traduction et les systèmes statistiques.

Le cœur d'un système à mémoire de traduction est une base de données dans laquelle chaque segment déjà traduit est représenté par sa représentation en langue source et sa représentation en langue cible. La représentation peut être plus ou moins complexe allant d'une simple chaîne de caractères à une représentation utilisant plusieurs niveaux de représentation linguistique [Planas, É. 1994]. Traduire un nouveau segment **S** signifie chercher une combinaison de segments ou sous-segments de la base de données qui se rapproche le plus de **S**. Les traductions des segments ou sous-segments choisis sont combinées pour produire une traduction de **S** en langue cible.

Avec un système de traduction statistique proposé pour la première fois par [Brown, P. F., *et al.* 1990], on considère la traduction comme un problème de décodage [Ney, H. 2004]. Le modèle de base utilise deux sources de connaissances statistiques définies par des distributions de probabilités calculables séparément :

- un modèle de traduction **Pr(s|c)** qui nécessite de définir une correspondance entre les mots de la phrase cible **c** et les mots de la phrase source **s**.
- un modèle de langage de la langue cible **Pr(c)** qui définit la bonne formation d'une suite de mots en langue cible **c**.

L'étape de traduction proprement dite est une étape de décodage (de recherche) qui cherche à maximiser le produit **Pr(c).Pr(s|c)** pour l'ensemble des phrases cibles. Soit :

$$\hat{c} = \underset{c}{\operatorname{argmax}} Pr(c)Pr(s|c)$$

Équation I-1 : Décodage en traduction statistique

I.2.4. Moments de l'interaction avec l'utilisateur

Pour les systèmes interactifs, quelle que soit leur architecture linguistique, on distingue trois moments de l'interaction : avant la traduction, pendant la traduction, et après la traduction.

I.2.4.1. Pré-édition

La pré-édition intervient avant la phase de traduction proprement dite. Cette étape vise à produire le document à traduire ou à en transformer le contenu dans une forme telle que le système puisse produire automatiquement une traduction (un document cible) de qualité.

Le document source doit être conforme à une norme. Il s'agit de vérifier que le lexique est bien conforme à la norme attendue et que la syntaxe correspond à la syntaxe admise par le système.

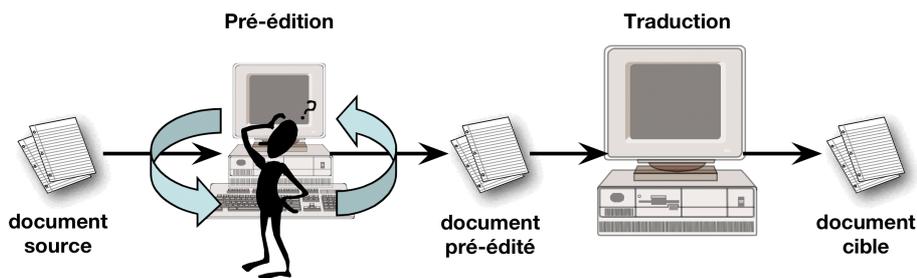


Figure I-5 : Traduction avec pré-édition

I.2.4.2. Interaction médiane

Les systèmes à interaction médiane font appel à l'utilisateur durant les étapes du processus de traduction proprement dit. L'utilisateur doit guider le système dans ses choix lorsqu'il ne peut choisir automatiquement. On parle dans ce cas de système de traduction interactif. L'interaction avec l'utilisateur doit permettre de produire des traductions de qualité.

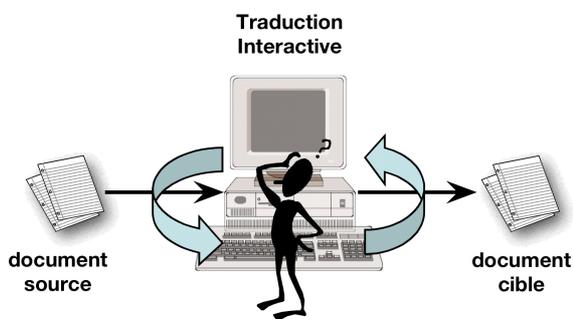


Figure I-6 : Traduction avec interaction médiane

I.2.4.3. Post-édition

Lorsque le système produit une traduction brute qui n'est pas d'excellente qualité, il faut la réviser afin de produire un document cible diffusable. C'est la TA du réviseur que nous avons présentée précédemment.

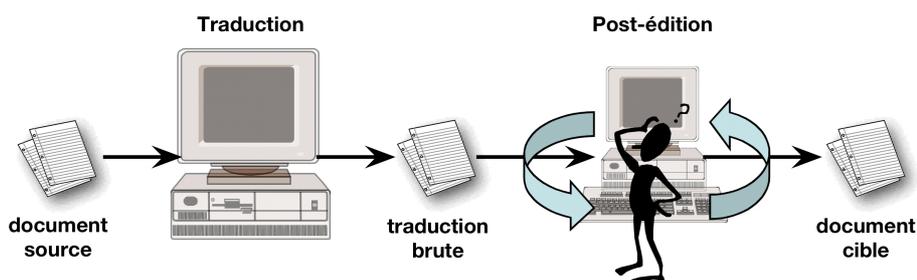


Figure I-7 : Traduction avec post-édition

Chapitre II

Caractérisation et mesure de la qualité des systèmes de traduction

L'évaluation des systèmes de traduction est importante pour les chercheurs, les développeurs de systèmes commerciaux, et les utilisateurs. Ce n'est donc pas une préoccupation récente. Le premier effort, resté célèbre, est reproduit dans le rapport ALPAC [ALPAC 1966]. On le verra, la méthode d'évaluation proposée alors a jeté les bases des méthodes d'évaluation subjective (réalisées par des juges humains) encore utilisées aujourd'hui.

L'impact du rapport ALPAC est indéniable. Par exemple, lors de COLING-84 (18 ans après), les participants d'une table ronde [King, M. 1984] s'interrogeaient sur la date et les motivations d'un nouveau rapport de type ALPAC. Plus tard, en 1989 (23 ans après !), les Japonais [JEIDA 1989] étudient la situation dans leur pays à la lumière des considérations et recommandations dudit rapport.

Bien sûr, l'évaluation n'est pas une tâche anodine et ce domaine n'échappe pas à un besoin de formalisation¹⁹. Différents projets ont été conduits dans cette direction (EAGLES, ISLE). À l'instar de la reconnaissance de la parole, la communauté de la TA s'est aussi dotée de méthodes d'évaluation objective (réalisées par des programmes informatiques).

Dans ce chapitre, nous nous intéressons à la notion de qualité d'un énoncé et d'une traduction. Nous présentons les deux familles de méthodes d'évaluation (objectives, subjectives) des systèmes de traduction. Nous essayons d'en tirer quelques enseignements.

II.1. Mise en place de l'évaluation

Le premier effort d'évaluation ayant connu une large publicité [ALPAC 1966] est resté célèbre. [Hutchins, J. 2003] rappelle et commente son contenu. Le rapport ALPAC est bref, 34 pages, et complété par 20 annexes qui totalisent 90 pages. En particulier, l'annexe 10 [ALPAC 1966, pp. 67-75] rend compte de l'évaluation conduite. Nous en rappelons ici certains éléments pour montrer qu'il a jeté des bases importantes pour le domaine.

¹⁹ [Van Slype, G. 1979], [King, M. 1996], [EAGLES-EWG 1999, section3.3] et [Hovy, E., *et al.* 2002] en dressent un panorama.

II.1.1. ALPAC, the (in-)famous report²⁰

Deux critères indépendants étaient déjà proposés pour évaluer une traduction : son **intelligibilité**, et sa **fidélité** au sens véhiculé par le texte original. En effet, une traduction peut être intelligible et manquer de fidélité ou d'exactitude (précision). Inversement, une traduction peut être fidèle et manquer d'intelligibilité, bien que cela ne puisse se produire que dans des cas pour lesquels la source est déjà peu intelligible.

II.1.1.1. Données et protocole

L'intelligibilité de la traduction était évaluée sans référence à la phrase source. La fidélité était mesurée de façon indirecte : le juge devait assimiler tous les éléments de sens présents dans la traduction puis évaluer la phrase originale du point de vue de son informativité par rapport à ce qu'il avait compris de la traduction. Ainsi, la phrase source est très informative par rapport à la traduction lorsque que cette dernière n'est pas très fidèle puisque la phrase source contient plus d'information que la traduction.

Deux groupes de juges ont été entraînés pour la tâche. Le premier groupe monolingue était constitué de dix-huit personnes anglophones ne connaissant pas la langue source (le russe en l'occurrence). Afin d'évaluer l'informativité, ils avaient à leur disposition une traduction en langue anglaise soigneusement préparée des phrases d'origine. Ils comparaient donc deux phrases anglaises (le résultat de traduction à évaluer, la traduction soignée en anglais de la phrase originale). Les dix-huit membres du second groupe bilingue possédaient une solide connaissance de la langue russe scientifique. Afin d'évaluer l'informativité, ils avaient à leur disposition la phrase originale en russe et le résultat de traduction.

Les données d'évaluation étaient constituées de 144 phrases sélectionnées au hasard dans quatre extraits d'un livre russe. Pour chacune des phrases, six traductions différentes ont été produites : trois traductions humaines, et trois traductions de systèmes de TA différents. Les traductions ont alors été mélangées au hasard pour former six lots (de 144 phrases) sous la contrainte qu'une traduction des phrases originales n'apparaisse qu'une seule fois dans chacun des lots. Chaque lot a été proposé à six juges, trois du groupe monolingue et trois du groupe bilingue ayant eu chacun une heure d'entraînement sur un lot de 30 phrases extraites des mêmes passages du livre russe.

Pour chaque phrase, le juge devait d'abord mesurer son **intelligibilité** puis sa **fidélité**. Ce faisant, il devait aussi chronométrer le temps qu'il mettait pour lire et mesurer l'intelligibilité de chaque traduction.

L'évaluation en intelligibilité était réalisée sur une échelle de neuf valeurs (Table II-1), chaque valeur étant définie verbalement.

L'évaluation en fidélité était aussi réalisée sur une échelle de neuf valeurs (Table II-2).

²⁰ Titre de l'article de John Hutchins [Hutchins, J. 2003] qui fait une analyse point par point du rapport.

Valeur	Définition
9	Parfaitement clair et intelligible. Se lit comme un texte ordinaire ; exempt de maladresse.
8	Parfaitement ou presque clair et intelligible, mais contenant quelques maladrotes grammaticales ou stylistiques mineures, et/ou des mots utilisés dans un usage inhabituel qui pourraient être aisément corrigés.
7	Généralement clair et intelligible, mais le style et les choix lexicaux et/ou les constructions syntaxiques sont un peu plus pauvres que dans la catégorie 8.
6	L'idée générale est presque immédiatement intelligible, mais une pleine compréhension est obérée par un style pauvre, de mauvais choix lexicaux, des expressions alternatives ; des expressions non traduites, de mauvaises constructions syntaxiques. Post-éditer peut permettre d'obtenir une forme presque acceptable.
5	L'idée générale n'est compréhensible qu'après une étude considérable, mais après cette étude, le lecteur est assez confiant quant à sa compréhension. Mauvais choix lexicaux, syntaxe grotesque, mots non traduits sont présents, mais ils constituent principalement du « bruit » derrière lequel l'idée principale est encore perceptible.
4	Le texte semble intelligible, mais en réalité, il est plus inintelligible qu'intelligible. Néanmoins, l'idée peut vaguement être appréhendée. Les choix lexicaux, les constructions syntaxiques, et/ou les expressions alternatives sont en général bizarres, et il peut y avoir des mots essentiels qui ne sont pas traduits.
3	Généralement inintelligible ; le texte semble un non-sens, mais avec un effort considérable de réflexion et d'étude, le lecteur peut au moins faire une hypothèse sur l'idée du contenu.
2	Presque irrémédiablement inintelligible même après réflexion et étude. Néanmoins, le texte ne semble pas dénué de sens.
1	Désespérément inintelligible. Il semble qu'aucune étude et réflexion ne puisse permettre de révéler le sens du texte.

Table II-1 : Échelle d'intelligibilité ALPAC

Valeur	Définition
9	Extrêmement informatif. L'original change complètement ou inverse le sens communiqué par la traduction.
8	Très informatif. L'original contribue énormément à la clarification du sens communiqué par la traduction.
7	Entre 8 et 7.
6	Clairement informatif. L'original ajoute une information considérable sur la structure de la traduction et sur les choix lexicaux, en mettant le lecteur sur la bonne voie du sens voulu.
5	Entre 6 et 4.
4	Par contraste avec 3, l'original ajoute une certaine quantité d'information sur la structure de la traduction, et sur les relations syntaxiques ; elle peut aussi corriger des incompréhensions mineures du sens général de la traduction ou sur le sens de certains mots.
3	En corrigeant un ou deux sens critiques, principalement au niveau des choix lexicaux, l'original donne un tour assez différent au sens communiqué par la traduction. Elle n'ajoute cependant pas d'information nouvelle sur la structure de la traduction.
2	Aucune nouvelle information sur le sens n'est apportée par l'original, que ce soit au niveau du lexique ou de la grammaire, mais le lecteur est en quelque sorte plus sûr qu'il appréhende le sens voulu.
1	Pas du tout informatif ; aucun sens nouveau n'est ajouté par l'original, la confiance du lecteur vis à vis de sa compréhension n'est ni diminuée ni augmentée.
0	L'original contient moins d'information que la traduction. Le traducteur a ajouté du sens, apparemment pour rendre l'original plus compréhensible.

Table II-2 : Échelle de fidélité ALPAC

Les six jeux de traductions soumis à l'évaluation, donnés ici dans leur ordre de classement, étaient les suivants :

- T1 Traductions Humaines travaillées
- T4 Traductions Humaines rapidement exécutées
- T2 Traductions Humaines rapidement exécutées
- T7 Traductions Machines Système B
- T5 Traductions Machines Système A
- T9 Traductions Machines Système C

II.1.1.2. Résultats

Les traductions humaines sont clairement supérieures aux traductions produites par les systèmes. Au sein de chaque classe de traduction, on distingue cependant deux groupes au niveau de la qualité ((T1, T4), T2) et ((T7, T5), T9). Il faut noter aussi que les traductions produites par la machine sont bien plus longues à lire et à évaluer (deux fois plus de temps). La performance moyenne des systèmes de traduction se situe au milieu de l'échelle de notation. Pour l'époque, avec le recul, on peut dire que ces résultats ne sont pas mauvais.

II.1.1.3. Réactions

Pour [King, M. 1996], « il est clair que les mesures utilisées sont entièrement qualitatives, et que la définition perverse de la fidélité jette des doutes sur leur validité. »

Ross Macdonald²¹, dans [Pankowicz, Z. 1966], que nous citons ici *in extenso*, rapporte ceci à propos de l'évaluation : « Dans la description, en annexe 10, d'une tentative d'évaluation de la TA, il apparaît que des traductions de textes ont été évaluées individuellement phrase par phrase. Des traductions de différentes parties du texte produites par différents systèmes de traduction étaient mélangées au hasard. Dans ces circonstances, même des phrases en anglais écrites par des locuteurs natifs peuvent ne pas avoir semblé particulièrement intelligibles. À moins qu'une sélection assez grande du texte ne soit présente, afin que la personne qui souhaite l'utiliser ait certains repères, le test ne peut être valide. La perte de linéarité et de repère dont un lecteur parcourant un matériau continu bénéficie peut très bien faire une différence frappante dans la compréhensibilité du texte. De plus, si les phrases utilisées étaient assez attentivement sélectionnées, presque n'importe quel résultat pouvait être obtenu au moyen d'un tel test. Le seul but de la TA, à ce moment de l'histoire, n'est pas de produire des phrases complètement claires et grammaticales, mais de transmettre le contenu et l'intention d'un document complet.

Si les traductions candidates doivent être évaluées, elles doivent certainement être évaluées sur des unités non segmentées [pas sur les phrases uniquement], et elles doivent être évaluées par ceux qui vont vraisemblablement les utiliser ; pas par des étudiants en licence à Harvard, leurs qualités oratoires, leur formation en sciences en général fussent-elles excellentes, ni par des traducteurs professionnels qui doivent, quoi qu'ils disent à ce propos, être suspectés d'avoir des préjugés sur ce sujet, mais par des scientifiques qui veulent des traductions de documents scientifiques et qui ont un intérêt profond à en extraire les informations qu'ils contiennent. »

²¹ Membre de la Georgetown University, Washington D.C., dont le système a été utilisé pour l'évaluation.

II.1.2. Travaux du JEIDA

Le JEIDA (Japanese Electronic Industry Development Association) a publié deux rapports en 1989 [JEIDA 1989] et en 1992 [JEIDA 1992].

II.1.2.1. JEIDA 1

Ces premiers travaux du JEIDA devaient permettre de répondre à trois questions :

- 1) Quels sont les changements technologiques et sociaux du marché depuis le rapport ALPAC ?
- 2) À la lumière de ces changements, les conclusions du rapport ALPAC sont-elles encore valides aujourd'hui ?
- 3) Sinon, comment doit-on évaluer l'état actuel et le futur de la traduction automatique ?

Pour faire cette étude, le comité a étudié les systèmes alors en usage dans différents pays et conduit des enquêtes sur la demande en traduction au Japon, l'état de l'activité du domaine de la traduction et l'utilisation des systèmes de traduction. Cependant, à notre avis, aucune réponse claire aux questions initiales n'est vraiment proposée.

II.1.2.2. JEIDA 2

En 1992, cela faisait dix ans que le premier système de traduction commercial avait été proposé sur le marché japonais. Lors de la rédaction du rapport, plus de dix systèmes étaient commercialisés. L'objectif était de développer une méthodologie d'évaluation suivant différents points de vue : l'évaluation des facteurs économiques par les utilisateurs ; l'évaluation technique des systèmes par les utilisateurs ; et l'évaluation technique par les développeurs.

II.1.2.2.1. Évaluation technique par les utilisateurs

Les critères d'évaluation technique par les utilisateurs doivent permettre de mesurer le degré de satisfaction des utilisateurs d'un système de traduction automatique en fonction de leurs besoins plutôt que d'évaluer le système de traduction lui-même. C'est une approche de type système expert qui doit permettre de vérifier si le système rend les services attendus.

Sachant que ces systèmes ne peuvent produire des traductions de haute qualité, les utilisateurs potentiels doivent avoir les moyens de choisir le système qui correspond le mieux à leurs besoins. Il s'agit en quelque sorte de produire un outil d'aide à la décision. Au cours de cette étape, 5 formulaires sont renseignés : (A) analyse des besoins de l'utilisateur, (B) spécifications du système, (C) évaluation du système, (D) configuration du système requise par l'utilisateur, et (E) une fiche de test.

Les besoins de l'utilisateur sont d'abord précisés (formulaire A) en termes de ce que le système doit faire (couples de langues, qualité de la traduction brute, qualité de la traduction révisée, ...) et comment il doit le faire (gestion des entrées et des sorties, ..., budget, plate-forme matérielle et logicielle). Les réponses à ce formulaire permettent de contraindre les réponses aux questions du formulaire D afin d'identifier les configurations système et matérielle requises. Un document représentatif des besoins de traduction est traduit par le système. Le résultat obtenu permet de remplir le formulaire E. Le formulaire B est rempli par le fournisseur du système en tenant compte des réponses fournies dans les formulaires D (configuration matérielle) et E (rapidité de la tâche de traduction). Lorsque

ces quatre formulaires sont remplis, l'utilisateur peut vérifier, à l'aide du formulaire C, si le système couvre ses besoins.

II.1.2.2.2. Évaluation technique par les développeurs

Cet ensemble de critères doit aider les chercheurs et les développeurs de systèmes de traduction, ainsi que les chefs de projet de recherche et de développement lorsqu'ils souhaitent évaluer le système qu'ils ont développé. Cet ensemble de critères est organisé selon six axes :

- Général/spécifique à un domaine,
- Couverture fonctionnelle (le système offre-t-il ou non les fonctions de base ?),
- Degré de précision (le système considère-t-il des détails ?),
- Originalité (les techniques de base mises en œuvre sont-elles originales ou non ?),
- Ouverture du système (le système peut-il être ou non personnalisé, enrichi ?),
- Facilité d'utilisation.

Selon ces six axes, on examine les techniques employées pour les composants : dictionnaires (types, nombre, ...), analyse (morphologique, syntaxique, sémantique, contextuelle), structures intermédiaires et traitement (méthode de traduction), génération (choix parmi des équivalents, grammaticalité, génération des ellipses, conjonctions, aspect, temps), traitement des formes spéciales (titres, listes, formules, dates, ...), personnalisation et fonctions d'apprentissage (typage et/ou reconnaissance des types de fragments, préférences grammaticales, dictionnaires, grammaires), environnement et mise en œuvre (pré-édition, édition de structures intermédiaires, post-édition, environnement de développement).

On obtient finalement une mesure quantitative, sur chacun des six axes, qui permet de comparer les systèmes.

II.1.2.2.3. Évaluation de la qualité des traductions

Le rapport propose enfin une méthode d'évaluation de la qualité de la traduction dont l'objectif principal est de détecter les phénomènes linguistiques que le système ne sait pas gérer. Les résultats de l'évaluation fournissent des résultats objectifs.

La méthode proposée est fondée sur l'utilisation d'un jeu de test²². Ce jeu de test doit prendre en compte les capacités des systèmes de traduction. Pour construire ce corpus, le rapport propose d'examiner des sorties de système de traduction, de détecter les problèmes, et produire des données qui présentent les problèmes rencontrés. Il doit être complet (couvrir tous les problèmes détectés) et ne pas contenir des phénomènes exclus de l'évaluation.

Nous verrons que ce genre d'évaluation est proposé dans d'autres travaux.

II.1.3. Campagne ARPA (1992-1994)

Dans la série de campagnes d'évaluation lancées par l'ARPA (Advanced Research Projects Agency) [White, J. S., *et al.* 1994], l'objectif correspondait à un besoin d'évaluation comparative de systèmes utilisant différentes architectures linguistiques depuis différentes langues sources vers l'anglais (unique langue cible considérée). Il s'agissait de faire participer des systèmes académiques et des systèmes

²² Ce sont les "test suites" proposées par Maghi King.

commerciaux répondant à plusieurs types d'usage. Ainsi, on trouvait des systèmes totalement automatiques et des systèmes d'aide à la traduction. Dans ce contexte, deux critères de fonctionnalité ont été retenus, la compréhensibilité et la qualité de la traduction.

Pour la première évaluation, un ensemble d'articles de presse traitant de fusions et d'acquisitions financières a été rassemblé. Ces articles ont été traduits par des traducteurs professionnels vers les différentes langues sources des systèmes considérés. Les systèmes devaient retraduire les articles vers l'anglais. Les traductions ainsi obtenues étaient alors évaluées. La mesure associée à la qualité était basée sur une mesure utilisée par le gouvernement américain pour évaluer la compétence de traducteurs humains. Elle s'est avérée inopérante à cause de la nature et de la quantité des erreurs présentes dans les traductions et n'a pas été retenue pour les évaluations suivantes.

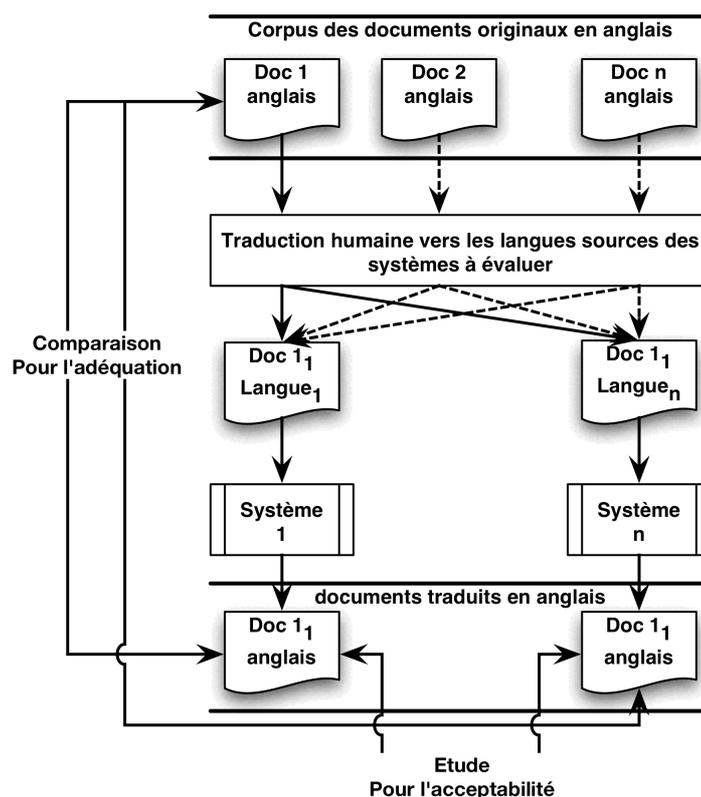


Figure II-1 : Principe des évaluations ARPA/DARPA

Le critère de qualité a été remplacé par deux critères : **adéquation** (*adequacy*) et **acceptabilité** (*fluency*). La mesure associée à l'adéquation permet à un juge anglais de déterminer le degré auquel les unités d'information présentes dans une traduction professionnelle peuvent être retrouvées dans la sortie d'un système de traduction automatique. Une unité d'information est un constituant syntaxique qui contient assez d'informations pour permettre une comparaison. Pour mesurer l'acceptabilité, on demande aux mêmes juges de déterminer si la traduction est rédigée dans un anglais convenable. Cette opération est réalisée sans faire référence à l'original, ainsi l'adéquation ne rentre-t-elle pas en ligne de compte.

Cependant, cette pratique a soulevé plusieurs questions sur la validité des résultats obtenus. La mesure de l'adéquation est fondée sur la comparaison d'une traduction professionnelle avec une traduction produite par un système. Dans quelle mesure la compétence du traducteur humain responsable de la traduction interfère-t-elle avec la validité de la mesure ? D'autre part, la mesure d'acceptabilité repose entièrement sur un jugement

subjectif : on sait bien que le jugement humain sur le caractère « bien formé » d'un énoncé peut varier. Nous le montrerons plus tard (sections XI.3.3.2 et XII.1.2.3.3).

Les auteurs du rapport [EAGLES-EWG 1996 ; p. 53] se demandent s'il ne serait pas plus profitable de mesurer si une traduction est assez bonne pour un besoin spécifique, plutôt que d'essayer de définir une notion, somme toute assez abstraite, de la qualité d'une traduction.

II.2. Efforts de formalisation

Nous situons l'époque de la maturité du domaine vers les années 1995, en particulier avec le projet EAGLES dont les efforts pour formaliser et organiser l'évaluation des systèmes de Traitement Automatique des Langues Naturelles (TALN) seront poursuivis dans le cadre du projet ISLE.

II.2.1. EAGLES

Le projet européen EAGLES (Expert Advisory Group on Language Engineering Standards) est une tentative de création de standards dans le domaine des industries de la langue. Les premiers domaines concernés dans la première phase du projet (1993-95) [EAGLES-EWG 1996] furent les corpus, les lexiques, les formalismes grammaticaux et les méthodes d'évaluation. Il a été reconnu qu'une stratégie unique d'évaluation ne pouvait pas être développée, même pour un domaine d'application spécifique. En effet, la qualité d'un système est fortement dépendante de son usage et de ses utilisateurs potentiels. Cependant, il semblait possible et nécessaire de proposer un cadre général pour la conception d'évaluations qui permettrait de concevoir des évaluations particulières adaptées à une application dans un contexte bien identifié.

Influencé par les travaux précédents, et en particulier par le standard ISO/IEC 9126 [King, M. 2003], le groupe de travail sur l'évaluation a proposé un modèle de qualité pour les systèmes de TALN en général. Ce modèle est organisé en une hiérarchie de propriétés et d'attributs dont les feuilles sont des attributs mesurables auxquels sont associées des méthodes de mesure.

Ce premier effort pour proposer un cadre théorique a été validé sur différentes familles de systèmes, par exemple : les correcteurs grammaticaux [EAGLES-EWG 1996, Annexe D, p. 116] et les aides au traducteur humain [id. Annexe E p. 136].

Dans la seconde phase du projet (1995-1996) [EAGLES-EWG 1999], le travail sur l'évaluation s'est essentiellement limité à la consolidation des résultats (correcteurs grammaticaux [id. Section 5.1, p. 60], correcteurs orthographiques [id. Section 5.2, p. 76], mémoires de traduction [id. section 5.4, p. 106]) et à la dissémination de la proposition.

II.2.1.1. Concrètement, une évaluation doit être réalisée en sept étapes²³ :

1. Pourquoi faire une évaluation ?

- Quel est le but de l'évaluation ? Toutes les personnes impliquées ont-elles la même compréhension du but de l'évaluation ?
- Que va-t-on évaluer exactement ? Un système dans son entier ou un module de celui-ci ? Un système pris isolément ou un système pris dans un contexte spécifique ? Quelles sont les limites du système ?

²³ Texte extrait de <http://www.osil.ch/sept-etapes.html> (Véronique Sauron, 2000)

2. Élaborer un modèle des tâches

- Identifier les rôles et les agents pertinents.
- Comment le système va-t-il être utilisé ?
- Par qui ? Pour quelle utilisation ? Qui sont les utilisateurs ?

3. Définir des caractéristiques de qualité

- Quelles caractéristiques du système doivent être évaluées ? Ont-elles toutes la même importance ?

4. Spécifier des critères détaillés pour le système évalué sur la base des étapes 2 et 3

- Pour chaque critère identifié, peut-on trouver un moyen valable et cohérent de mesurer les performances du système en fonction de ces critères ?
- Dans le cas contraire, ces derniers devront être décomposés en critères secondaires mesurables de manière à les rendre valides.
- Ce processus doit être répété jusqu'à ce que les critères soient mesurables.

5. Définir les mesures à appliquer au système pour les critères déterminés au point 4

- Tant la mesure que la méthode pour obtenir cette mesure doivent être définies pour chaque critère.
- Pour chaque critère mesurable, en fonction du modèle des tâches (2), que va-t-on considérer comme un bon résultat, un résultat satisfaisant, un résultat insatisfaisant ? Où se situe le seuil d'acceptabilité ?
- Habituellement, un critère est décomposé en plusieurs critères secondaires. Comment les valeurs des différents critères secondaires se combinent-elles à la valeur définie pour le critère de référence afin de refléter leur importance relative (là encore en fonction du modèle des tâches) ?

6. Préparer l'exécution de l'évaluation

- Développer le matériel de test qui permettra d'évaluer le système.
- Qui sera effectivement chargé de procéder aux différentes mesures ? Quand ? Dans quelles circonstances ? Sous quelle forme le résultat final sera-t-il présenté ?

7. Exécuter l'évaluation

- Procéder aux mesures.
- Comparer les résultats à l'échelle des degrés de satisfaction préalablement déterminés.
- Résumer les résultats dans un rapport d'évaluation, cf. point 1.

II.2.2. ISLE & FEMTI

Les activités du groupe de travail sur l'évaluation du projet EAGLES se sont poursuivies dans le cadre du projet ISLE²⁴ (International Standards for Language Engineering) (1999-2002) financé par la Communauté Européenne et la National Science Foundation (NSF, USA). Dans le cadre du projet, les efforts ont porté sur l'évaluation des systèmes de traduction automatique et ont donné lieu à une série d'ateliers²⁵ et à la proposition FEMTI (Framework for the Evaluation of MT in ISLE) [Hovy, E., *et al.* 2002].

²⁴ Site web du projet : <http://www.issco.unige.ch/projects/isle/femti/>

²⁵ Consulter <http://www.issco.unige.ch/projects/isle/ewg.html> pour la liste complète.

FEMTI propose :

- une classification des caractéristiques principales qui définissent un contexte d’usage (première partie), soit : le type d’utilisateur, le type de tâches et la nature des données d’entrée.
- une classification des caractéristiques de qualité des systèmes (seconde partie) dans une hiérarchie organisée en sous-catégories dotées, au niveau des feuilles, d’attributs internes et/ou externes (des métriques). Les niveaux supérieurs de cette hiérarchie coïncident avec les caractéristiques définies dans le standard ISO/IEC 9126 [ISO/IEC 2001].
- une mise en correspondance entre la première classification et la seconde qui définit ou suggère les caractéristiques de qualité, sous-caractéristiques et attributs/métriques qui sont appropriés à chaque contexte d’usage.

La hiérarchie est bien trop grande pour que nous puissions la proposer complètement ici. Nous proposons la version compacte (Figure II-2) et le développement de la section 2.2.1 (Figure II-3).

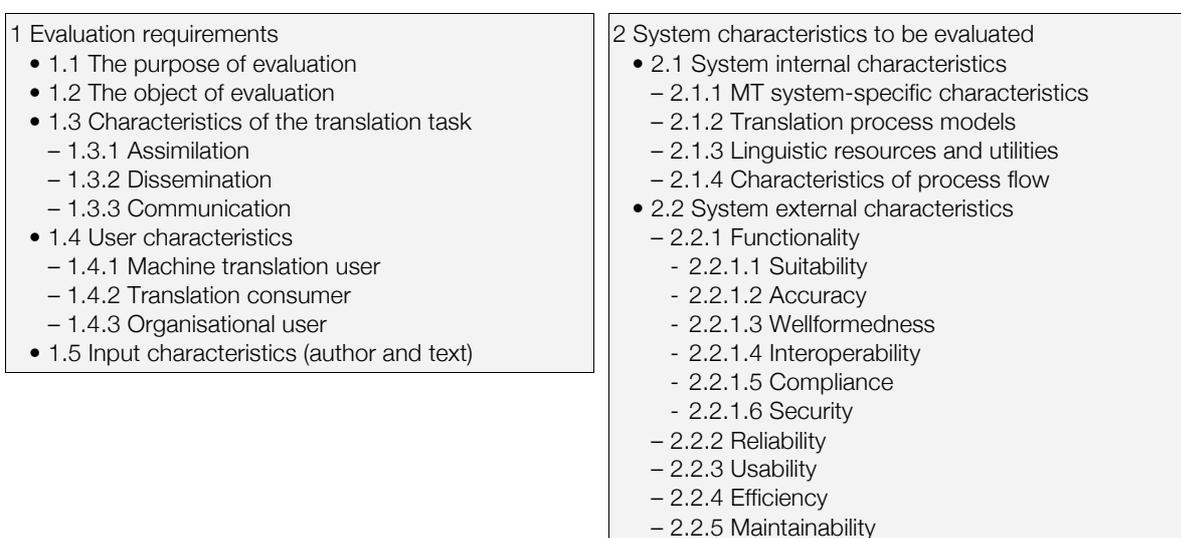


Figure II-2 : Version compacte de la hiérarchie FEMTI

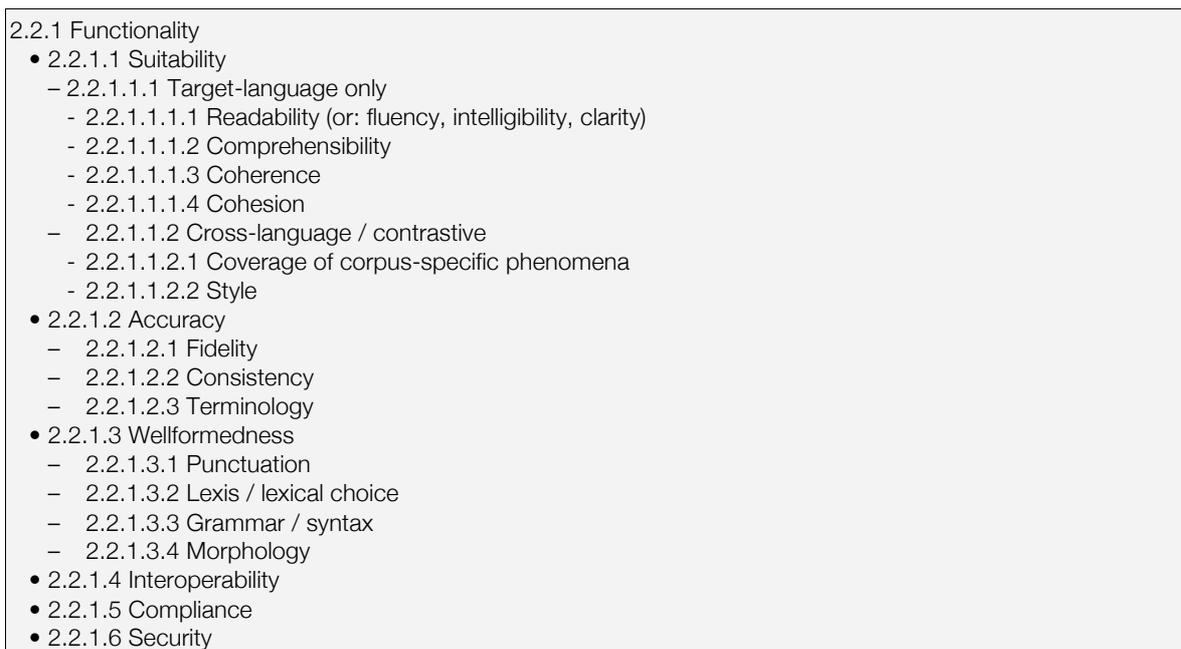


Figure II-3 : Version développée de la section 2.1.1 de la hiérarchie FEMTI

FEMTI devrait aider :

- les utilisateurs potentiels de la TA. Ils peuvent définir les caractéristiques qui sont les plus importantes pour eux et donc choisir le système le mieux adapté à leurs besoins.
- les personnes (utilisateurs potentiels et développeurs) désireuses de comparer plusieurs systèmes de TA. Elles peuvent choisir les caractéristiques les plus proches de leur situation, et ainsi trouver les mesures d'évaluation et les tests pertinents.
- les développeurs de systèmes de TA qui peuvent trouver des informations sur les besoins des utilisateurs et trouver des niches pour leurs applications.

II.2.3. TIDES & NIST

Le programme TIDES (Translingual Information Detection, Extraction and Summarization) vise le développement d'outils de traitement des langues naturelles pour permettre à des locuteurs anglophones de trouver et d'interpréter des informations importantes dans de multiples langues sans connaissance *a priori* de celles-ci. La partie évaluation est supervisée par le NIST²⁶ (National Industry Standards and Technology) qui fournit des outils de calcul.

Nous reviendrons sur les évaluations organisées par le NIST dans la section II.3.1.2.

II.3. Point sur les méthodes d'évaluation proposées

Il existe deux classes de méthodes d'évaluation. La première classe regroupe les méthodes pour lesquelles l'évaluation est réalisée par des juges humains qui doivent donner des notes (scores) définies en langue naturelle. Le résultat de l'évaluation donne alors une évaluation concrète directement interprétable du genre « bon » ou « mauvais ». On parle d'**évaluation subjective** car on ne peut contrôler finement la manière dont les juges remplissent leur tâche et choisissent une note plutôt qu'une autre.

Ainsi, deux juges auxquels on donne le même jeu de données à évaluer peuvent fournir des notes différentes ; on peut et on devrait alors évaluer l'accord entre juges. De même, au cours du temps, le jugement d'un juge sur un même ensemble de données peut varier ; on peut et on devrait alors évaluer la fidélité²⁷ de la mesure d'un juge au cours du temps. Nous reviendrons sur les problèmes de cohérence en faisant des mesures dans les sections XI.3.3.2 et XII.1.2.3.3.

La seconde classe de méthodes regroupe les méthodes pour lesquelles l'évaluation, la décision sur les notes, est réalisée par un programme informatique déterministe qui réalise un calcul sur les données qui lui sont fournies en entrée et fournit un résultat sous forme numérique. On parle alors d'**évaluation objective** car, pour un ensemble de données fournies au programme, celui-ci fournit toujours le même résultat unique ; il n'y a donc plus de problème de fidélité de la mesure.

Dans le reste de cette section, nous examinons les méthodes d'évaluation subjectives et objectives proposées dans la littérature.

²⁶ Accessible via : <http://www.nist.gov/speech/tests/mt/>

²⁷ Il s'agit bien ici de vérifier que la mesure (dans les faits, la note) proposée par un juge humain est bien fidèle, c'est à dire que la même note est obtenue lorsque le juge répète la mesure sur un même objet.

II.3.1. Méthodes subjectives

La première famille de méthodes d'évaluation subjectives essaie de comparer les performances des systèmes de traduction à des performances humaines. La deuxième famille de méthodes vise à faire évaluer les performances des systèmes par des humains en fournissant des résultats directement interprétables, quantifiés par une note définie en langue naturelle.

II.3.1.1. Comparaison à des mesures de performance humaine

II.3.1.1.1. TOEFL

Afin d'évaluer des systèmes de traduction automatique de l'anglais vers le japonais [Tomita, M., *et al.* 1993] ont adopté un protocole visant non pas à faire une évaluation au niveau des phrases, mais plutôt au niveau d'un segment de document.

Des tests de compréhension d'un manuel d'entraînement au TOEFL (Test Of English as a Foreign Language) ont été sélectionnés. Un test de compréhension est constitué de deux ou trois passages, chacun de ces passages est suivi de plusieurs questions. Ces passages ont été traduits manuellement et avec les trois systèmes à évaluer.

Les traductions de ces passages ont été mélangées afin d'obtenir des tests de compréhension du japonais constitués de traductions de systèmes différents. Les questions ont été traduites manuellement. Les tests de compréhension ainsi obtenus peuvent être considérés comme des tests de compréhension du japonais.

Deux expériences ont été conduites, il s'agit pour des juges de répondre aux questions de compréhension. Les résultats permettent donc de donner une note au test TOEFL à chaque système évalué.

Lors de la première expérience, les tests ont été passés par soixante Japonais, ce qui fait vingt juges par système. Les tests ont alors été notés et la performance de chaque système a été calculée en pourcentage de bonnes réponses. Les concepteurs ont noté que les juges n'ont pas fait de leur mieux pour répondre aux questions. Il est parfois pénible de lire les textes traduits par la machine. De plus, il y avait des erreurs dans certaines traductions manuelles des questions.

Lors de la seconde expérience, la qualité de la traduction des questions a été consciencieusement vérifiée et les candidats étaient motivés par un classement doté de prix. Les résultats ont été bien meilleurs grâce à la forte mobilisation des candidats à bien faire.

II.3.1.1.2. TOEIC

S'inspirant des travaux précédents, mais au niveau de la phrase, [Sugaya, F., *et al.* 1999, Sugaya, F., *et al.* 2000, Sugaya, F., *et al.* 2001] de ATR ont proposé une méthode d'évaluation fondée sur le TOEIC (Test Of English for International Communication) pour un système de traduction de parole.

Données et protocole

Les données de test sont constituées de 330 tours de parole en japonais extraits de vingt-trois conversations. Les transcriptions de ces tours de parole serviront de référence.

Il s'agit de comparer des traductions humaines (japonais vers anglais) produites par des Japonais avec les traductions produites par le système. On connaît les notes au test TOEIC des traducteurs. On souhaite donc mesurer la note TOEIC du système.

Les traducteurs humains écoutent le tour de parole à traduire et écrivent une traduction. Des fiches d'évaluation sont produites avec l'énoncé source, en japonais, et les traductions humaines et les traductions du système présentées dans un ordre aléatoire. Pour chacune des fiches, des juges américains capables de comprendre le japonais, doivent évaluer la qualité des deux traductions en utilisant une échelle de quatre valeurs : A (parfait), B (correct), C (acceptable), D (incompréhensible). Ils doivent aussi sélectionner la meilleure traduction.

Résultats

Les résultats de l'évaluation montrent que le système est meilleur que les traducteurs humains ayant une note TOEIC comprise entre 300 et 400 et moins performant que ceux qui ont une note supérieure ou égale à 800. La zone d'équivalence se situe autour de 700, ce qui est confirmé par une analyse de régression (707,6). La même expérience a été conduite à partir des hypothèses de reconnaissance. Dans ce cas, la note TOEIC du système est 548,0 soit une chute de 150 points. Cette baisse est bien sûr due aux erreurs du module de reconnaissance de la parole.

Avec le même système, une évaluation automatique (objective) de similarité a été conduite avec un ensemble de paraphrases en anglais de chacun des énoncés japonais (14,42 paraphrases par énoncé). Les notes en TOEIC calculées à partir de cette mesure de similarité sont de 682,9 sur les références et 547,3 sur les hypothèses.

Remarques

ATR présente ici une méthode d'évaluation subjective pour justifier ensuite une évaluation objective à moindre coût. Cette mesure n'est cependant pas directement interprétable à la lecture des résultats. Le seul point d'accroche avec la réalité est que la note TOEIC de 568 points correspond au niveau moyen d'anglais des étudiants universitaires japonais.

II.3.1.2. Proposition FEMTI et idées proches

Les autres types d'évaluation subjective permettent de mesurer la qualité des traductions candidates produites par un système de TA à l'aide de critères dont les valeurs sont interprétables directement car définies en langue naturelle. On retrouve ici l'idée de l'évaluation ALPAC.

Dans FEMTI, on distingue trois critères de qualité des traductions : évaluation de la forme (bonne formation linguistique), évaluation du fond (aptitude), et évaluation de la justesse. L'évaluation de la forme et du fond n'a pas besoin d'être contrastive et doit donc se faire uniquement sur les traductions produites par le système. L'évaluation de la justesse est forcément contrastive ; elle ne peut se faire que par rapport à une référence. Nous verrons que les expériences d'évaluation rapportées dans la littérature ne retiennent que deux critères : la **bonne formation linguistique**, et selon l'un des axes de la justesse, la **fidélité**.

Il nous semble cependant qu'en pratique, dans les protocoles d'évaluation que nous avons consultés, la fidélité recouvre aussi des critères qui sont classés dans la catégorie aptitude.

II.3.1.2.1. Bonne formation linguistique²⁸

La mesure de bonne formation linguistique (2.2.1.3) [Loffler-Laurian, A.-M. 1983] dépend de quatre critères :

- la ponctuation (2.2.1.3.1). Respecte-elle les règles de la langue ?
- les choix lexicaux (2.2.1.3.2). Ces erreurs concernent les mots, ou groupes de mots qui sont inappropriés (dans leur collocation, leur connotation, leur registre, leur idiomaticité).
- la grammaire/syntaxe (2.2.1.3.3) qui permet d'évaluer le respect des règles grammaticales de la langue.
- la morphologie (2.2.1.3.4) qui permet d'évaluer le respect des règles flexionnelles de la langue.

Dans la pratique cependant, la bonne formation linguistique est évaluée de manière globale sous le vocable de “**fluency**”.

Par exemple, les consignes d'évaluation de la bonne formation linguistique dans les campagnes d'évaluation supervisées par le NIST se lisent comme suit²⁹ : « [ce critère] se réfère au degré de bonne formation de l'énoncé du point de vue des règles de l'anglais écrit standard. Un segment est bien formé si :

1. il respecte les règles de bonne formation linguistique (grammaire),
2. il contient des mots correctement orthographiés (morphologie),
3. il respecte l'usage commun du vocabulaire, des titres et noms,
4. il est intuitivement acceptable,
5. il peut être interprété de manière raisonnable par un locuteur natif de l'anglais.

Un jugement de bonne formation linguistique peut avoir les valeurs suivantes :

Note	Définition
5	Anglais sans défaut (“Flawless English”)
4	Bon anglais (“Good English”)
3	Anglais non natif (“Non-native English”)
2	Anglais non standard (“Disfluent English”)
1	Incompréhensible (“Incomprehensible”)

Table II-3 : Définition des notes de bonne formation linguistique proposées par le NIST

Lorsque la traduction en anglais contient des caractères ou des mots de la langue source, les juges doivent donner une note comprise entre 1 et 3 en fonction du degré auquel les caractères non traduits, parmi les autres facteurs, affectent le caractère de bonne formation de la traduction. »

La Figure II-4 montre une interface d'évaluation de la bonne formation linguistique.

²⁸ Dans cette section et dans la suivante, les numéros renvoient à l'indice des critères mentionnés dans la classification FEMTI (www.issco.unige.ch/projects/isle/femti/fullclassification.html) au 4 septembre 2004.

²⁹ Les instructions d'évaluation proposées par le NIST données ici en français sont tirées d'un document accessible à l'adresse suivante : <http://www ldc.upenn.edu/Projects/TIDES/Translation/TransAssess02.pdf>.

Figure II-4 : Interface web d'évaluation de la bonne formation linguistique

II.3.1.2.2. Aptitude

L'aptitude du système (**suitability**) est mesurée suivant deux axes. Le premier axe mesure les aptitudes du système à évaluer, du point de vue de la langue cible uniquement :

- la lisibilité (2.2.1.1.1), ou encore le naturel, l'intelligibilité ou la clarté d'une **phrase**. Elle concerne la facilité de compréhension, sa clarté pour le lecteur.
- la compréhensibilité (2.2.1.1.2) permet d'évaluer dans quelle mesure un **texte** considéré dans son ensemble est facile à comprendre et donc dans quelle mesure le lecteur peut y retrouver les informations et faire des inférences valides à partir du texte considéré.
- la cohérence (2.2.1.1.3) est le degré auquel le lecteur peut assigner un rôle à chaque phrase (ou groupe de phrase) par rapport au texte dans son entier.
- la cohésion (2.2.1.1.4) se réfère aux chaînes lexicales et autres éléments — anaphores, ellipses — qui relie les différentes unités au cours d'une phrase.

Le second axe permet de mesurer la qualité de la traduction sur la base de la langue source et du résultat produit par le système dans la langue cible. Les critères retenus sont :

- la couverture de phénomènes spécifiques au corpus (2.2.1.1.2.1).
- le style (2.2.1.1.2.2).

La description de ces critères est assez floue.

On peut remarquer que le point 5 des instructions d'évaluation de la formation linguistique recouvre au moins la lisibilité qui est listée dans la catégorie **aptitude** de FEMTI.

II.3.1.2.3. Justesse

La justesse (**accuracy**) concerne la capacité du système à produire les bons résultats ou effets avec le degré de précision requis. La justesse est mesurée par rapport au texte en langue source. Elle est évaluée avec trois critères :

- la fidélité (2.2.1.2.1) mesure la quantité d'information présente dans le texte original qui est reproduite dans le texte cible sans distorsion. La fidélité a un score égal ou inférieur à la compréhensibilité. Les variations entre ces deux scores sont dus à la perte d'information (silence), les interférences (bruit), ou à des distorsions provoquées par la combinaison des deux phénomènes précédents. La fidélité est très difficile à évaluer, car chaque phrase ne transmet pas une unité d'information complète et unique, mais plutôt une portion de message ou une collection de messages dont l'importance dans la phrase n'est pas facile à évaluer.
- la cohérence³⁰ (2.2.1.2.2) mesure la capacité du système à produire pour une entrée donnée à un point donné dans le temps, la même sortie. Ce critère est essentiel pour

³⁰ Il s'agit de ce que l'on nomme la « fidélité » dans la terminologie des appareils de mesure.

les développeurs car il permet de détecter des effets de bord non prévus. Il est aussi important pour la traduction de documents techniques.

- la terminologie (2.2.1.2.3) permet de vérifier que les termes spécifiques au domaine du document sont correctement traduits.

Dans les campagnes d'évaluation subjective qui respectent les instructions du NIST, la mesure d'évaluation contrastive est l'adéquation (**adequacy**). Cette mesure est définie comme suit²⁹ : « Après avoir fait l'évaluation en bonne formation linguistique, le juge peut voir l'une des traductions de référence. En comparant la traduction candidate en langue cible avec la référence, il détermine si la traduction est adéquate. L'adéquation concerne le degré auquel les informations présentes dans l'original³¹ sont aussi communiquées par la traduction. Ainsi pour l'adéquation, la traduction de référence représente (tient le rôle de) l'originale.

Les juges doivent répondre à la question qui leur est posée (Quelle quantité du sens exprimé dans la référence est aussi présente dans la traduction ?) avec les notes suivantes :

Note	Réponse
5	Tout
4	Presque tout
3	Beaucoup
2	Peu
1	Aucun

Table II-4 : Définition des notes de l'adéquation proposées par le NIST

Lorsque la traduction en anglais contient des caractères ou des mots de la langue source, les juges doivent donner une note comprise entre 1 et 3 en fonction du degré auquel les caractères non traduits, parmi les autres facteurs, affectent le caractère de bonne formation de la traduction. »

La Figure II-5 montre une interface d'évaluation de l'adéquation.

Figure II-5 : Interface web d'évaluation de l'adéquation

L'adéquation du protocole du NIST est comparable à la fidélité de FEMTI. Il faut cependant bien rappeler que, pour les promoteurs de FEMTI l'évaluation est contrastive entre la langue source et la langue cible, alors qu'avec le protocole du NIST l'évaluation contrastive a lieu entre deux phrases en langue cible. Cela est dû à des nécessités purement pratiques. Il est plus facile de trouver des locuteurs monolingues que des locuteurs

³¹ Il y a ici abus de langage puisque ce n'est pas la phrase originale en langue source *stricto sensu* qui est présentée au juge, mais une traduction humaine de référence en langue cible de l'originale. La validité de la démarche repose donc sur la présupposition que la traduction de référence est déjà elle-même adéquate !

bilingues, surtout pour des couples de langue rares. On peut cependant regretter que la validité de cette approche n'ait jamais été ni discutée ni vérifiée !

Pour ce faire, il faudrait mesurer la fidélité, au sens de FEMTI, des traductions humaines de référence par rapport aux segments en langue source.

II.3.2. Méthodes objectives

La première méthode d'évaluation objective proposée fut la distance d'édition sur les mots. Cette méthode est empruntée au domaine de la reconnaissance de la parole et plus connue sous le nom de taux d'erreur en mots (Word Error Rate). Plus récemment des méthodes d'évaluation objective fondées sur la co-occurrence de n-grammes ont été proposées. Les mesures BLEU et NIST, sont maintenant devenues des standards. Mais ce domaine est en constante évolution. Nous verrons qu'une multitude de nouvelles méthodes vient d'être proposée.

II.3.2.1. Distance d'édition

Nous rappelons d'abord brièvement ce qu'est une opération d'édition élémentaire.

Soit V un vocabulaire donné (un ensemble de symboles). Soit $V^* = V^+ \cup \{\varepsilon\}$ l'ensemble des chaînes finies de caractères où ε représente la chaîne vide.

Une opération d'édition élémentaire concerne un couple $(a, b) \in V \cup \{\varepsilon\} \times V \cup \{\varepsilon\}$ et elle est notée $a \rightarrow b$

Alors la chaîne B est le résultat de l'application de l'opération d'édition $a \rightarrow b$ sur la chaîne A s'il existe des chaînes σ et τ de V^* telles que $A = \sigma a \tau$ et $B = \sigma b \tau$. On note ceci $A \Rightarrow B$.

Figure II-6 : Définition d'une opération d'édition élémentaire

On considère généralement trois opérations élémentaires :

$a \rightarrow b$ est une opération de **substitution** si $a \neq \varepsilon$ et $b \neq \varepsilon$

$a \rightarrow b$ est une opération d'**insertion** si $a = \varepsilon$ et $b \neq \varepsilon$

$a \rightarrow b$ est une opération d'**effacement** si $a \neq \varepsilon$ et $b = \varepsilon$

Figure II-7 : Définition des opérations d'édition classiques

Pour définir une distance d'édition, on affecte un coût à chacune de ces opérations. La distance d'édition entre deux chaînes S_1 et S_2 est alors la somme minimale des coûts des opérations élémentaires permettant le passage de S_1 à S_2 . Une suite de ces opérations élémentaires s'appelle une S -dérivation.

Soit une fonction de coût positive γ associée à chacune des opérations d'édition élémentaires qui vérifie : $\gamma(a \rightarrow a) = 0$, $\gamma(a \rightarrow b) = \gamma(b \rightarrow a)$, $\gamma(a \rightarrow b) + \gamma(b \rightarrow c) \geq \gamma(a \rightarrow c)$.

Soit l'extension de cette fonction de coût à une dérivation S telle que

$$\gamma(S) = \sum_{i=1}^m \gamma(S_i) \quad \text{si } m=0 \text{ alors } \gamma(S)=0$$

On définit alors la distance d'édition entre deux chaînes comme :

$d(A, B) = \min(\gamma(S))$, S étant une S -dérivation de A vers B .

Figure II-8 : Définition de la distance d'édition sur les chaînes

II.3.2.1.1. Distance de Levenshtein et de Damerau-Levenshtein

La distance de Levenshtein [Levenshtein, V. I. 1966] ou distance d'édition est couramment utilisée dans de nombreuses applications où il faut mesurer la similarité entre deux séquences. Le coût de la substitution stricte vaut 1, celui de la substitution identique vaut 0. Le coût de la suppression et de l'insertion vaut 1.

La distance dite de Damerau-Levenshtein [Damerau, F. J. 1964], [Levenshtein, V. I. 1966] (souvent notée DLM dans la littérature pour Damerau-Levenshtein Metric) ne considère plus simplement chaque opération d'édition comme possédant un coût unitaire mais affecte des poids à chaque opération.

Wagner et Fischer [Wagner, C.-K. and Fisher, M.-J. 1974] ont montré que la minimisation de la DLM peut s'exprimer sous la forme d'une récurrence et ont proposé un algorithme de résolution par programmation dynamique. Il faut signaler enfin que [Fisher, W. M. and Fiscus, J. G. 1993] ont proposé des procédures d'alignement plus performantes pour limiter le cumul des erreurs.

II.3.2.1.2. Variations

[Leusch, G., *et al.* 2003] étendent la distance de Levenshtein à la notion de bloc de mots, ou plutôt dans ce cas à la notion de syntagme au sein d'un arbre d'analyse. On cherche ici à ne pas pénaliser la mesure au niveau des mots, mais au niveau des syntagmes, ce qui diminue la pénalité. En effet, on pénalise la permutation de syntagmes (une opération) et non celle des mots composant le syntagme (plusieurs opérations de suppression et d'insertion).

[Akiba, Y., *et al.* 2001] proposent seize distances d'édition (ED1, ..., ED16) ! Elles utilisent les mêmes trois opérations que [Wagner, C.-K. and Fisher, M.-J. 1974] (substitution, effacement et insertion), auxquelles est adjointe une opération additionnelle d'édition : l'échange, comme l'ont proposé [Lowrence, R. and Wagner, C.-K. 1975, Su, K.-Y., *et al.* 1992]. Les distances d'édition ED1 à ED8 utilisent les trois opérations standard, les distances D9 à D16 utilisent en plus l'échange. Dans chacune de ces deux familles de distances, on utilise trois classes de vocabulaire (restriction aux mots pleins, référence aux codes sémantiques, restriction aux mots-clés), qui sont combinées.

Le lecteur peut consulter [Planas, É. 1994] qui utilise des distances d'édition sur les différents niveaux de la structure de données dont il se sert pour représenter un segment dans une mémoire de traduction.

II.3.2.2. Taux d'erreur sur les mots (WER, PER)

La mesure du taux d'erreur sur les mots³² (Word Error Rate en anglais), comme son nom l'indique mesure le nombre de mots qui diffèrent entre un candidat et une référence. Cette mesure est utilisée pour évaluer les performances des modules de reconnaissance de la parole. Dans ce cas, on compare l'hypothèse de reconnaissance (candidat) produite pour un signal donné et une transcription manuelle de ce signal.

Le WER est mesuré relativement à un alignement de deux séquences de mots, le candidat et la référence. Lors de l'alignement, les opérations d'édition élémentaires appliquées sur les mots constituent des erreurs. L'alignement optimal est trouvé par une procédure de calcul dynamique qui minimise la distance de Levenshtein des deux séquences de mots.

³² La mesure prend en compte chacune des occurrences de chaque mot.

Le WER est le nombre d'erreurs divisé par le nombre de mots de la référence. Ce taux peut en pratique excéder 100% lorsque l'on observe un nombre trop important d'insertions.

Comme l'ont proposé [Nießen, S., *et al.* 2000], on peut aussi calculer le taux d'erreur sur les mots d'un candidat par rapport à un ensemble de références (multi reference WER, mWER). On calcule le WER du candidat par rapport à chaque référence et on prend le minimum.

[Tomás, J., *et al.* 2003] proposent une mesure qui est un cas particulier du mWER, il s'agit du aWER (all references WER). Cette mesure part du principe que l'on disposerait de toutes les références possibles lors de l'évaluation. Comme ce n'est jamais le cas, pour contourner ce problème, les juges peuvent proposer eux-mêmes de nouvelles références plus adaptées aux candidats.

[Leusch, G., *et al.* 2003] proposent une distance de Levenshtein à laquelle ils ajoutent une opération de permutation de groupes de mots qui a un coût constant quelle que soit la taille des blocs de mots permutés. Sans cette opération, si l'on veut permuter les blocs **B** et **C** dont l'intersection en terme de mots est vide dans la suite de blocs **ABCD**, il faut effacer tous les mots de **B** puis les insérer après **C** (ou vice versa). Le coût de cette transformation est $2 \cdot \min(|B|, |C|)$.

[Och, F. J. and Ney, H. 2002] proposent encore une distance de Levenshtein indépendante de la position des mots (PER – Position-independent Word Error Rate) ; une phrase est un ensemble de mots et la distance entre une phrase et n'importe laquelle de ses permutations est égale à zéro³³. Par exemple la distance PER entre « je voudrais une chambre simple » et « je une simple voudrais chambre » est nulle.

II.3.2.3. Mesures fondées sur la co-occurrence de n-grammes

II.3.2.3.1. Bleu

La méthode BLEU (BiLingual Evaluation Understudy), introduite par [Papineni, K., *et al.* 2002] est la première méthode d'évaluation objective fondée sur une mesure de co-occurrence de n-grammes. Il s'agit de mesurer la similarité entre une traduction candidate et une ou plusieurs traductions de référence. Cette comparaison s'appuie sur une mesure de précision modifiée.

a. Calcul de BLEU

Au niveau des monogrammes (les mots) le calcul de la précision modifiée d'une traduction candidate est effectué comme suit :

1. pour chaque mot d'une traduction candidate
 - trouver le nombre maximum d'occurrences dans les traductions de références (*max_ref_count*)
 - minorer ce maximum par le nombre d'occurrences du mot dans la traduction candidate (*count*). On obtient $Count_{clip} = \min(count, max_ref_count)$, et
2. diviser la somme des $Count_{clip}$ par la somme des *count*.

³³ Cette mesure n'est donc pas une distance dans son acception mathématique.

Pour l'exemple suivant [id.] :

- Candidat 1 : It is a guide to action which ensures that the military always obeys the commands of the party.
- Candidat 2 : It is to ensure the troops forever hearing the activity guidebook that party direct.
- Référence 1 : It is a guide to action that ensures that the military will forever heed Party command.
- Référence 2 : It is the guiding principle which guarantees the military forces always being under the command of the Party.
- Référence 3 : It is the practical guide for the army always to heed the direction of the Party.

La précision modifiée du Candidat 1 par rapport aux trois références est 17/18 (Table II-5), celle du Candidat 2 par rapport aux trois références est 8/14 (Table II-6).

mots de Candidat 1	count	max_ref_count	Count _{clip}
it	1	1	1
is	1	1	1
a	1	1	1
guide	1	1	1
to	1	1	1
action	1	1	1
which	1	1	1
ensure	1	1	1
that	1	2	1
military	1	1	1
always	1	1	1
obeys	1	0	0
the	3	4	3
commands	1	1	1
of	1	1	1
party	1	1	1
Somme	18		17

Table II-5 : Calcul de précision modifiée de BLEU – exemple 1

mots de Candidat 2	count	max_ref_count	Count _{clip}
it	1	1	1
is	1	1	1
to	1	1	1
insure	1	0	0
the	2	4	2
troops	1	0	1
forever	1	1	1
hearing	1	0	0
activity	1	0	0
guidebook	1	0	0
that	1	2	1
party	1	1	0
direct	1	0	0
Somme	14		8

Table II-6 : Calcul de précision modifiée de BLEU – exemple 2

La formule de calcul de précision de BLEU pour un corpus de traductions candidates chacune associée à des traductions de références s'énonce comme suit :

$$p_n = \frac{\sum_{C \in \{Candidates\}} \sum_{n-gram \in C} Count_{clip}(n-gram)}{\sum_{C \in \{Candidates\}} \sum_{n-gram \in C} Count(n-gram)} \quad (B 1)$$

Équation II-1 : Calcul de la précision avec BLEU

où

$Count_{clip}(n-gram)$ est le nombre maximum de n-grammes qui apparaissent simultanément dans une traduction candidate et une traduction de référence, et

$Count(n-gram)$ est le nombre de n-grammes d'une traduction candidate.

La co-occurrence de monogrammes tend à satisfaire l'adéquation. Les co-occurrences de longs n-grammes tendent à satisfaire la fluidité.

p_n est calculé sur tout le corpus au niveau des n-grammes, plutôt que phrase par phrase.

Une traduction candidate ne doit être ni trop longue, ni trop courte. Les traductions candidates plus longues que les traductions de référence sont déjà pénalisées par la mesure de précision. Il faut alors pénaliser les traductions trop courtes qui maximisent la précision. Par exemple dans la situation suivante :

Candidat :	of the
Référence 1 :	It is a guide to action that ensures that the military will forever heed Party command.
Référence 2 :	It is the guiding principle which guarantees the military forces always being under the command of the Party.
Référence 3 :	It is the practical guide for the army always to heed the direction of the Party.

La précision modifiée sur les monogrammes vaut 2/2, et la précision modifiée sur les bigrammes vaut 1/1, alors que la traduction n'est manifestement pas bonne !

Afin de ne pas donner une note trop importante aux traductions courtes qui maximisent la précision, BLEU utilise une pénalité pour brièveté calculée comme suit pour tout le corpus :

$$BP_{BLEU} = \left\{ \begin{array}{ll} 1 & \text{if } |c| > |r| \\ e^{(1-|r|/|c|)} & \text{if } |c| \leq |r| \end{array} \right\} \quad (B 2)$$

Équation II-2 : Calcul de la pénalité de brièveté de BLEU

où $|c|$ est la longueur totale du corpus de traductions candidates et $|r|$ est la somme des longueurs des traductions de référence qui s'apparient au mieux avec leur traduction candidate.

Finalement le score BLEU d'un corpus de traductions candidates est le produit de sa précision et de sa pénalité de brièveté :

$$BLEU = BP_{BLEU} \cdot \exp \left(\sum_{n=1}^N w_n \log p_n \right) \quad (B 3)$$

Équation II-3 : Calcul de BLEU

Le facteur de pondération w_n est une constante égale à $1/N$.

La mesure fournie par BLEU est comprise entre 0 et 1 au sens large ($0 \leq \text{score Bleu} \leq 1$).

b. *Remarques soulevées par le calcul de BLEU*

[Doddington, G. 2002] soulève deux problèmes liés à la méthode de calcul de BLEU :

- BLEU fait une moyenne géométrique (w_n est une constante) de la participation des co-occurrences de n-grammes. Ainsi, la participation des longs n-grammes (moins nombreux) est surévaluée par rapport aux n-grammes courts (plus nombreux). Pour éviter ce problème, on pourrait utiliser une moyenne arithmétique (utilisant le nombre d'appariements sur chaque classe de n-grammes).
- Avec BLEU, tous les n-grammes ont le même poids, or il peut être préférable d'assigner un poids plus important aux n-grammes les plus informatifs. On souhaite donc donner un poids plus important aux n-grammes qui apparaissent moins fréquemment en fonction de leur informativité.

Dans le même esprit, [Babych, B. and Hartley, A. 2004a] font remarquer que les mots d'un texte ont des poids informationnels différents et par conséquent leur importance pour la traduction varie. Ainsi pour évaluer l'adéquation d'une traduction candidate, le choix des équivalents traductionnels de ces mots importants doit avoir plus de poids que le choix des mots utilisés à finalité structurale et qui n'ont pas d'équivalent traductionnel précis dans le texte source. [Babych, B. and Hartley, A. 2004a] proposent de mesurer la participation des n-grammes dans le calcul de la précision en utilisant la saillance ("saliency") d'un mot via la mesure $tf.idf$ et un S-score défini dans [Babych, B. and Hartley, A. 2004b].

II.3.2.3.2. NIST

La mesure d'informativité proposée par [Doddington, G. 2002], calculée sur l'ensemble des traductions de référence, est la suivante :

$$Info(w_1, \dots, w_n) = \log_2 \left(\frac{\text{the \# of occurrences of } w_1, \dots, w_{n-1}}{\text{the \# of occurrences of } w_1, \dots, w_n} \right) \quad (N 1)$$

Équation II-4 : Mesure d'informativité d'un n-gramme proposée par NIST

a. *Calcul de NIST*

Une nouvelle pénalité de brièveté est proposée afin de réduire l'impact de petites variations dans la longueur de la traduction candidate par rapport aux traductions de référence. La nouvelle formule s'établit comme suit :

$$BP_{NIST} = \exp \left\{ \beta \log^2 \left[\min \left(\frac{L_{sys}}{\bar{L}_{ref}}, 1 \right) \right] \right\} \quad (N 2)$$

Équation II-5 : Calcul de la pénalité de brièveté de NIST

où

$$L_{sys} = \left[\begin{array}{l} \text{the number of words in the} \\ \text{translation being scored} \end{array} \right]$$

$$\bar{L}_{ref} = \left[\begin{array}{l} \text{the average number of words in a} \\ \text{reference translation averaged over} \\ \text{all reference translations} \end{array} \right]$$

$$\beta = -4.22 \text{ so that } BP = 0.5 \text{ when } L_{sys} = 2/3 \cdot \bar{L}_{ref}$$

La Figure II-9 montre la différence entre les facteurs de pénalité de BLEU et NIST.

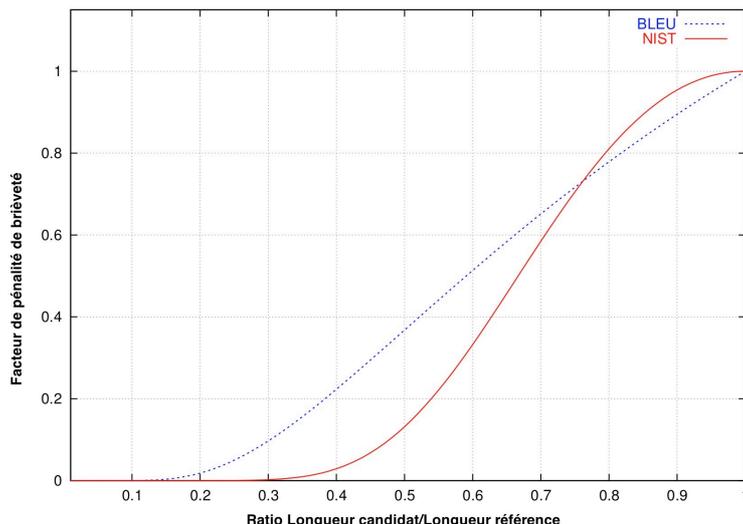


Figure II-9 : Comparaison de la valeur des facteurs de pénalité de BLEU et NIST

Finalement la mesure NIST se calcule comme suit :

$$NIST = BP_{NIST} \cdot \sum_{n=1}^N \left\{ \frac{\sum_{\substack{\text{all } w_1, \dots, w_n \\ \text{that co-occur}}} \text{Info}(w_1, \dots, w_n)}{\sum_{\substack{\text{all } w_1, \dots, w_n \\ \text{in sys output}}} (1)} \right\} \quad (N\ 3)$$

Équation II-6 : Calcul de NIST

La mesure fournie par NIST n’est pas théoriquement bornée. Elle est en pratique comprise entre 0 et 20 au sens large (0 ≤ score NIST ≤ 20).

b. Comparaison des mesures de précision de BLEU et NIST

Afin d’illustrer la tentative d’amélioration de la mesure de précision de BLEU par rapport à NIST, nous proposons l’exemple suivant évaluant la précision de deux hypothèses différentes avec une unique référence.

Référence : a b c d e f g h i j k l m n o p q r s
 Candidat 1 : a b c d e f g i h j l k m o n p r s q
 Candidat 2 : a b c d e f g x x x x x x x x x x x x

Du strict point de vue de l’observation, on peut dire que le Candidat 1 contient tous les mots de la Référence, seul l’ordre est bousculé à partir du rang 8 ; le Candidat 2 contient, dans le bon ordre, les 7 premiers mots de la Référence, ensuite, les mots ne sont pas dans la Référence. On souhaite donc intuitivement que le Candidat 1 ait une meilleure précision que le Candidat 2. Ce n’est pas le cas avec BLEU (Table II-7) . Par contre, le Candidat 1 a une meilleure précision que le Candidat 2 avec NIST (Table II-8).

	Candidat 1	Candidat 2
1-gramme	1,0000	0,3684
2-gramme	0,1666	0,3333
3-gramme	0,1176	0,2778
4-gramme	0,0588	0,2500
Précision totale	0,1871	0,3083

□ Table II-7 : Précision BLEU sur un exemple

	Candidat 1	Candidat 2
Précision totale	4,2479	1,5650

Table II-8 : Précision NIST sur un exemple

Les promoteurs de NIST ont donc bien atteint leur objectif.

II.3.2.3.3. Rouge

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) est une famille de méthodes proposée d'abord dans le cadre de l'évaluation automatique de résumés [Lin, C.-Y. 2004a, Lin, C.-Y. 2004b], puis mise en œuvre dans le contexte de l'évaluation de systèmes de traduction [Lin, C.-Y. and Och, F. J. 2004a]. Les méthodes sont basées sur la plus longue sous-séquence commune (pondérée ou non) entre une traduction candidate et un ensemble de traductions de référence.

ROUGE-L : plus longue (Longest) sous-séquence commune

Une séquence $Z=[z_1, z_2, \dots, z_n]$ est une sous-séquence d'une séquence $X=[x_1, x_2, \dots, x_m]$, si il existe une suite l d'indices de X strictement croissante $[i_1, i_2, \dots, i_k]$ telle que pour tout $j=1, 2, \dots, k$ on a $x_{i_j}=z_j$. La plus longue sous-séquence commune (LCS) est celle qui maximise la valeur de k .

Lorsqu'une traduction est vue comme une séquence de mots, l'intuition commande que plus la sous-séquence commune entre deux traductions est longue, plus ces traductions sont similaires. La mesure de similarité entre une traduction de référence X de longueur m et une traduction candidate Y de longueur n est une F-mesure³⁴ utilisant les plus longues sous-séquences communes calculée comme suit :

$$R_{lcs} = \frac{LCS(X, Y)}{m} \quad (\text{R-L 1})$$

$$P_{lcs} = \frac{LCS(X, Y)}{n} \quad (\text{R-L 2})$$

$$F_{lcs} = \frac{(1 + \beta^2)R_{lcs}P_{lcs}}{R_{lcs} + \beta^2P_{lcs}} \quad (\text{R-L 3})$$

Équation II-7 : Calcul de ROUGE-L

³⁴ La formule générale de calcul de la F-mesure avec R le rappel et P la précision est :

$$F = \frac{(1 + \beta^2)RP}{R + \beta^2P}$$

C'est le calcul d'une moyenne harmonique pondérée. Lorsque $\beta=1$, la F-mesure est la moyenne harmonique classique du rappel et de la précision

$$F = \frac{2RP}{R + P}$$

où $LCS(X,Y)$ est la longueur de la plus longue sous-séquence commune de X et Y, et $\beta=R_{lcs}/P_{lcs}$ ³⁵ lorsque $\partial F_{lcs}/\partial R_{lcs} = \partial F_{lcs}/\partial P_{lcs}$. ROUGE-L est la F-mesure (équation R-L 3).

L'équation de la F-mesure peut se ré-écrire comme suit :

$$F = \frac{n^2 + m^2}{nm(n + m)} LCS(X, Y)$$

Équation II-8 : Calcul explicite de la F-mesure en fonction de n, m, et LCS(X, Y)

En récompensant les appariements d'unigrammes en séquence, ROUGE-L prend aussi en compte la structure de la phrase. Si l'on considère l'exemple suivant :

Référence : police killed the gunman

Candidat 1 : police kill the gunman

Candidat 2 : the gunman kill police

BLEU donne le même score aux deux candidats (trois unigrammes et un bigramme). Par contre le Candidat 1 a un score Rouge-L de 3/4=0,75 et le Candidat 2 a un score de 2/4=0,5 avec $\beta=1$. Ainsi, le Candidat 1 est meilleur que le Candidat 2. On voit donc que ROUGE-L prend mieux en compte la « structure » de la phrase.

Lorsque plusieurs références X_j de longueur m_j sont disponibles, la formule de calcul de ROUGE-L_{multi} devient :

$$R_{lcs-multi} = \max_{j=1}^u \left(\frac{LCS(X_j, Y)}{m_j} \right) \quad (\text{R-Lm 1})$$

$$P_{lcs-multi} = \max_{j=1}^u \left(\frac{LCS(X_j, Y)}{n} \right) \quad (\text{R-Lm 2})$$

$$F_{lcs-multi} = \frac{(1 + \beta^2)R_{lcs-multi}P_{lcs-multi}}{R_{lcs-multi} + \beta^2 P_{lcs-multi}} \quad (\text{R-Lm 3})$$

Équation II-9 : Calcul de ROUGE-L_{multi}

où $\beta=R_{lcs-multi}/P_{lcs-multi}$ lorsque $\partial F_{lcs-multi}/\partial R_{lcs-multi} = \partial F_{lcs-multi}/\partial P_{lcs-multi}$. ROUGE-L_{multi} est la F-mesure (équation R-Lm 3).

ROUGE-W : plus longue sous-séquence commune pondérée (Weighted)

Si ROUGE-L a de bonnes propriétés, il ne prend pas en compte la distribution (connexité) de la plus longue sous séquence commune. Dans l'exemple suivant :

Référence : A B C D E F G

Candidat 1 : A B C D H I K

Candidat 2 : A H B K C I D

Candidat 1 et Candidat 2 ont le même score ROUGE-L. Cependant Candidat 1 devrait être un meilleur candidat. On associe donc la longueur k de la plus longue sous-séquence connexe commune à chacune des plus longues sous-séquences communes (WLCS). Les mesures de rappel, précision et F-mesure peuvent alors être paramétrées par une fonction de pondération f qui doit avoir la propriété suivante : $\forall \mathbf{x}, \mathbf{y} \in \mathbf{N}, f(\mathbf{x}+\mathbf{y}) > f(\mathbf{x})+f(\mathbf{y})$, afin de donner un poids plus important aux appariements consécutifs.

³⁵ Dans tous les articles de C.-Y. Lin que nous citons, tous les β sont définis comme le rapport P/R. Ce qui est faux et se montre facilement en résolvant l'égalité des deux dérivées partielles de F.

Les calculs pour une seule référence s'établissent alors comme suit :

$$R_{wlcS} = f^{-1} \left(\frac{WLC S(X, Y)}{f(m)} \right) \quad (\text{R-W } 1)$$

$$P_{wlcS} = f^{-1} \left(\frac{WLC S(X, Y)}{f(n)} \right) \quad (\text{R-W } 2)$$

$$F_{wlcS} = \frac{(1 + \beta^2)R_{wlcS}P_{wlcS}}{R_{wlcS} + \beta^2P_{wlcS}} \quad (\text{R-W } 3)$$

Équation II-10 : Calcul de ROUGE-W

ROUGE-W est la F-mesure (équation R-W 3).

En utilisant $\mathbf{f(k)}=\mathbf{k}^2$, avec l'exemple précédent, on obtient un score de 0,571 pour Candidat 1 et 0,286 pour Candidat 2.

ROUGE-S : statistique sur la co-occurrence de bigrammes non connexes (Skip-bigram)

Pour tirer avantage de l'idée de ROUGE-L, [Lin, C.-Y. and Och, F. J. 2004a] proposent aussi une mesure fondée sur la co-occurrence de bigrammes non connexes pour se rapprocher du calcul de BLEU de manière plus intuitive.

Un bigramme non connexe est une paire de mots dans leur ordre d'apparition dans la phrase en autorisant une distance quelconque entre eux. Une phrase de longueur n possède $C(n,2)$ bigrammes. Soient les données suivantes :

Référence :	police killed the gunman
Candidat 1 :	police kill the gunman
Candidat 2 :	the gunman kill police
Candidat 3 :	the gunman police killed

Chaque phrase a six bigrammes non connexes. Les bigrammes non connexes de la phrase Référence sont : "police killed", "police the", "police gunman", "killed the", "killed gunman", et "the gunman".

Les calculs de rappel, précision et F-mesure, pour une seule référence s'établissent alors comme suit :

$$R_{skip2} = \frac{SKIP2(X, Y)}{C(m, 2)} \quad (\text{R-S } 1)$$

$$P_{skip2} = \frac{SKIP2(X, Y)}{C(n, 2)} \quad (\text{R-S } 2)$$

$$F_{skip2} = \frac{(1 + \beta^2)R_{skip2}P_{skip2}}{R_{skip2} + \beta^2P_{skip2}} \quad (\text{R-S } 3)$$

Équation II-11 : Calcul de ROUGE-S

où $SKIP2(X, Y)$ est le nombre de co-occurrences de bigrammes non connexes entre X et Y, et $\beta=R_{skip2}/P_{skip2}$ quand $\partial F_{skip2}/\partial R_{skip2} = \partial F_{skip2}/\partial P_{skip2}$. ROUGE-S est la F-mesure (équation R-S 3).

Candidat 1, Candidat 2, et Candidat 3 ont respectivement trois, un, et deux bigrammes non connexes en commun avec Référence. En prenant $\beta=1$, les scores ROUGE-S de Candidat 1,

Candidat 2, et Candidat 3 sont respectivement de 0,5, 0,167, et 0,333. Ainsi, Candidat 1 est meilleur que Candidat 2, et Candidat 3 ; et Candidat 3 est meilleur que Candidat 2.

II.3.3. Évaluation de l'évaluation

Il est bien sûr important de valider les résultats d'une campagne d'évaluation. On observe cependant que, si des efforts importants sont consentis dans ce sens pour les méthodes objectives, nous n'avons pas trouvé dans la littérature d'effort d'évaluation de résultats d'évaluations subjectives. Notre explication de cet état de fait est que les résultats d'une évaluation subjective sont directement interprétables, alors que les résultats d'une évaluation objective ne le sont pas directement.

II.3.3.1. Corrélation avec les méthodes subjectives

La validation des méthodes d'évaluation objectives est en général faite via une mesure de corrélation [Coughlin, D. 2003], au niveau du corpus complet, des résultats obtenus avec les résultats d'une évaluation subjective au moyen de la corrélation produit-moment de Pearson ou de la corrélation des rangs de Spearman.

II.3.3.2. Limites observées des méthodes fondées sur la co-occurrence de n-grammes

[Culy, C. and Riehemann, S. Z. 2003] proposent une étude très intéressante qui compare, avec BLEU et NIST, différentes traductions allemandes produites par des traducteurs professionnels et des systèmes de traduction automatique de la Bible (les quatre premiers chapitres de Luc) et de Tom Sawyer (les deux premiers chapitres).

Les auteurs ont obtenu des résultats parfois surprenants. En voici quelques-uns :

- Pour le même document, une très mauvaise traduction automatique peut obtenir une meilleure note qu'une traduction humaine.
- En évaluant les traductions de Luc et Tom Sawyer avec Luc comme référence, on trouve bien sûr que les traductions de Luc ont de meilleures notes. Cependant, il existe une traduction humaine de Luc dont la note est plus proche d'une traduction de Tom Sawyer que d'une autre traduction de Luc !

II.3.3.3. ORANGE

Selon [Lin, C.-Y. and Och, F. J. 2004b], la comparaison des méthodes objectives avec les méthodes subjectives repose sur des jugements humains qui sont, malheureusement, souvent incohérents³⁶. Les auteurs montrent aussi que la corrélation n'est pas stable en fonction de la méthode d'évaluation choisie.

Afin d'évaluer objectivement le nombre de plus en plus important de méthodes d'évaluation objectives, [Lin, C.-Y. and Och, F. J. 2004b] proposent une nouvelle méthode³⁷ utilisable avec les systèmes de traduction statistique³⁸. Nous le rappelons encore une fois, l'hypothèse de base des méthodes d'évaluation objective est que les traductions de référence sont de bonnes traductions et que plus une traduction candidate est similaire à

³⁶ Les auteurs de l'article ne précisent pas ce qui leur fait porter ce jugement. On peut seulement constater qu'aucun des articles et documents que nous avons cités sur l'évaluation ne quantifie précisément la concordance entre juges en tenant compte de la chance ; concordance que l'on peut calculer, par exemple, avec le coefficient Kappa. Nous montrerons de tels résultats pour les évaluations que nous avons réalisées nous-mêmes.

³⁷ Il s'agit donc ici de méta-évaluation.

³⁸ Il faudra en effet utiliser plusieurs traductions candidates.

ses références, meilleure est cette traduction³⁹. [Lin, C.-Y. and Och, F. J. 2004b] font aussi l'hypothèse que les traductions candidates sont en général moins bonnes que leurs traductions de référence. Alors, les traductions de référence doivent, en moyenne, avoir un meilleur score que les traductions candidates si une bonne méthode d'évaluation est utilisée.

Étant donnée une phrase source, ses traductions candidates et ses traductions de référence, on calcule le rang moyen des traductions de référence parmi l'union des traductions candidates et des traductions de référence. Pour ce faire, on calcule d'abord le score des n meilleures traductions candidates et de chaque traduction de référence au moyen des méthodes d'évaluation objectives à évaluer. On classe ensuite ces traductions. On calcule alors le rang moyen des références parmi les n meilleures traductions candidates. On calcule finalement le rapport entre le rang moyen des références et la taille de la liste des n meilleures traductions candidates. Plus ce rapport, appelé ORANGE (Oracle⁴⁰ RANKing for Gisting Evaluation), est petit, meilleure est la méthode d'évaluation.

II.3.3.4. Quelle méthode pour quel critère à évaluer

[Soricut, R. and Brill, E. 2004] remarquent aussi qu'une méthode d'évaluation objective est intéressante si elle a une bonne corrélation avec une évaluation subjective. Ils font remarquer que les juges obéissent à un certain nombre d'instructions lors de l'évaluation. Les résultats sont donc largement influencés par ces instructions. L'évaluation de la corrélation entre une méthode d'évaluation objective et une méthode d'évaluation subjective ne serait donc intéressante que si les deux méthodes cherchent à évaluer le même critère.

Un cadre général d'organisation des méthodes d'évaluation selon deux axes est proposé : l'axe des applications, et l'axe du critère évalué. L'échelle de l'axe des applications est le rapport entre la fidélité et la compacité⁴¹. L'échelle de l'axe du critère évalué est le rapport entre la précision et le rappel.

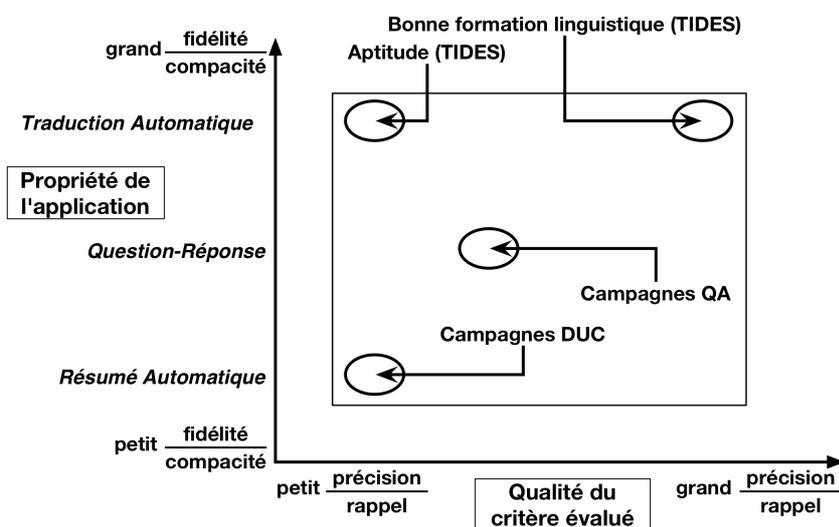


Figure II-10 : Plan de l'évaluation pour les applications du TALN

[Soricut, R. and Brill, E. 2004] proposent deux familles de mesures paramétrées : des mesures qui privilégient la précision, dont BLEU est une instance particulière ; des mesures

³⁹ Nous démontrons section XII.1.2.3.2 que dans le cas contraire, on ne peut rien conclure.

⁴⁰ L'Oracle est l'ensemble des traductions de référence utilisées pour l'évaluation.

⁴¹ Par exemple, un résumé doit être le plus compact (bref) possible.

qui privilégient le rappel, dont ROUGE est une instance particulière. Pour réaliser une évaluation avec des objectifs bien définis en fonction du critère à évaluer, les concepteurs sont donc en mesure de choisir la méthode la mieux adaptée à leur besoin.

II.4. Premiers commentaires

II.4.1. Méthodes subjectives

La comparaison monolingue pour l'évaluation de l'adéquation est peut-être biaisée, car dans les faits, on se sait pas si les traductions de référence sont de bonnes références. [Tomás, J., *et al.* 2003] proposent d'ailleurs que le juge puisse ajouter des références qui soient plus proches de la traduction candidate si celle-ci est satisfaisante.

II.4.2. Méthodes objectives

On ne le dit pas assez, mais les méthodes objectives calculent une **mesure de similarité** entre documents. Elles permettent donc de savoir si le système est capable de bien mimer les références. Si le score d'un système évalué objectivement est excellent, alors le système produit des traductions très similaires aux références. On peut alors dire que le système est bon. Par contre, si le score d'un système n'est pas bon, alors il est impossible de conclure quoi que ce soit sur la qualité des traductions produites. Nous avons fait cette expérience dans le cadre de l'évaluation IWSLT (cf. section XII.1.2.3.2).

Avec de telles méthodes, on peut objectivement classer les performances d'un système sur différents jeux de données, ainsi que les performances de plusieurs systèmes sur un même jeu de données. Par contre, aucune définition en langue naturelle ne peut être associée au résultat produit par le programme d'évaluation, sauf aux bornes si la méthode produit un résultat borné. Par exemple, si le résultat produit est compris entre 0 et 1 (0 étant la note minimale et 1 étant la note maximale), lorsque le résultat est 1 on peut dire que le système est parfait et lorsque le résultat est 0 on peut dire que le système ne réalise pas du tout la tâche (il ne vaut rien).

Pour résoudre ce problème, la méthode habituellement utilisée est une mesure de corrélation entre les résultats d'évaluations subjectives et d'évaluations objectives.

II.4.3. Que retenir du point de vue de la qualité ?

La qualité d'une traduction, c'est d'abord son **acceptabilité linguistique**. C'est aussi sa **fidélité** au message véhiculé par l'énoncé – oral ou écrit – source. Il nous semble aussi intéressant de s'intéresser à la **perception du récepteur** : la traduction proposée est-elle ambiguë ou non ? En effet, un énoncé source, à traduire, est rarement isolé. Il est produit dans un contexte (d'un document écrit, d'un discours oral, d'un dialogue oral). Le contexte est normalement désambiguïsant du côté de la source. Par contre, du côté de la cible rien ne garantit que le contexte, une fois traduit, restera désambiguïsant. Nous proposerons une solution à ce problème dans le Chapitre VIII.

Chapitre III

Obstacles à la qualité en communication interpersonnelle médiatisée par la machine

Une langue naturelle n'est pas un objet formel construit pour vérifier certaines bonnes propriétés. Ses mécanismes de production, de réception et de compréhension par l'homme sont extrêmement complexes et encore mal compris. Faire des traitements automatiques afin de reconnaître la parole ou faire de la traduction automatique nécessite de passer par une étape de modélisation et d'algorithmisation, ce qui implique des compromis et des approximations, qui eux-mêmes empêchent de résoudre un certain nombre de problèmes auxquels nous allons nous intéresser dans ce chapitre.

III.1. Variabilité de la parole

En traduction interpersonnelle orale médiatisée par la machine, la variabilité du signal à analyser est le premier obstacle auquel sont confrontés les modules de reconnaissance de la parole. [Junqua, J.-C. and Haton, J.-P. 1996 ; Chapitre 4] proposent de diviser la variabilité en cinq grandes catégories :

- Variabilité personnelle,
- Variabilité inter-personnelle,
- Variabilité environnementale,
- Variabilité linguistique,
- Variation contextuelle.

La variabilité linguistique recouvre les notions d'accent⁴² et de dialecte⁴³. La variation contextuelle concerne, entre autres, l'environnement phonétique local (la coarticulation : modifications acoustiques importantes aux frontières des mots adjacents).

III.1.1. Variabilité personnelle

Les sources principales de variation sont directement liées à l'état physiologique et psychologique de locuteur, et à l'objectif de la communication. En fonction de certains

⁴² Compris ici comme l'ensemble des caractères phonétiques distinctifs d'une communauté linguistique considérés comme différents de la norme.

⁴³ Les dialectes différents par leur syntaxe, vocabulaire et morphologie qui ne sont pas spécifiques à la parole et ne nous intéressent pas dans cette section.

facteurs, un locuteur peut changer la qualité de sa voix, sa vitesse d'élocution, sa fréquence fondamentale, et même ses patrons articulatoires.

Des changements environnementaux peuvent aussi induire des variations. Le bruit de fond (réflexe de Lombard) et des conditions stressantes conduisent à des efforts plus importants de la part du locuteur pour s'exprimer. Cela induit des variations acoustico-phonétiques. Enfin selon l'état de santé du locuteur, des caractéristiques telles que la nasalisation peuvent être affectées.

III.1.2. Variabilité inter-personnelle

Les différences physiologiques (longueur et forme du conduit vocal, physiologie des cordes vocales, forme du conduit nasal) sont d'importantes sources de variation entre les locuteurs. Ces différences induisent une grande variabilité acoustique.

Les habitudes articulatoires contribuent aussi à la variabilité. Elles sont généralement fonction de la personnalité du locuteur [Campbell, N. 2004]. Le sexe joue aussi un rôle important, ainsi que l'âge. La parole prononcée par un adulte est acoustiquement très différente de celle prononcée par un enfant.

III.1.3. Variabilité environnementale

Les conditions environnementales dans lesquelles la parole est produite jouent aussi un rôle important et affectent sa production. Les éléments qui induisent ces variations se rangent dans deux catégories :

- statiques (pour la durée d'une conversation) : acoustique du lieu et réverbération, matériel de capture (tels que micros et téléphones), et bien sûr, la personnalité du locuteur et ses caractéristiques physiques.
- dynamiques : bruit de fond, émotions [Id.], stress [Id.], charge cognitive et localisation du matériel de capture.

Les changements d'environnement statique et dynamique sont très difficiles à simuler en laboratoire à cause de leur très grande variabilité. Cela explique pourquoi on observe des différences de performance importantes pour un même système de reconnaissance de la parole entre les conditions de laboratoire et les conditions d'usage réel.

Pour avoir été locuteur privilégié du système C-STAR II du français (cf. Chapitre IX), je me rappelle très bien certains phénomènes. Par exemple, je devais contrôler ma manière de dire « bonjour ». En prononçant ce mot naturellement, neuf fois sur dix le système de reconnaissance de la parole rendait « bonsoir ». Ma prononciation de la dernière syllabe était naturellement trop longue. Pour être certain que le module de reconnaissance produise « bonjour », il fallait que j'accélère au moment de prononcer la dernière syllabe pour dire quelque chose du genre « bonjur ». J'ai toujours observé cependant qu'il fallait que je parle lentement en faisant des efforts d'articulation, en particulier pour les nombres dont les termes devaient être bien séparés et très articulés (« trente deux » et non pas « trent'deux »). Pour ce qui est du stress, je dois dire que le jour de la démonstration publique, les performances du module de reconnaissance de la parole étaient strictement équivalentes et peut-être même meilleures que lors des entraînements. En réécoutant les bandes, on remarque une diction très claire.

III.2. Spontanéité

Certaines spécificités du langage oral⁴⁴ que nous n'avons pas évoquées dans la section précédente constituent aussi des obstacles à sa reconnaissance et à son analyse automatique. Afin d'étudier la langue parlée, le premier obstacle que l'on rencontre est celui de sa transcription [Blanche-Benveniste, C. 2000, Chapitre 1]. Quand on dispose d'une transcription adaptée, on peut observer différentes classes de phénomènes linguistiques remarquables de la langue spontanée.

III.2.1. Phénomènes syntaxiques particuliers⁴⁵

La seconde classe de problème concerne des aspects syntaxiques particuliers que l'on ne rencontre pas à l'écrit, sauf dans les situations de tchat [Pierozak, I. 2003] et [Falaise, A. 2004]. On peut citer par exemple, sans prétendre être exhaustif, des aspects syntaxiques de base et des constructions syntaxiques complexes.

Au rang des aspects syntaxiques de base, on observe souvent⁴⁶ le non-respect de l'ordre Sujet Verbe Objet qui induit des constructions clivées comme dans les exemples suivants :

- Mon cahier je l'ai oublié à la maison.
- À deux cents mètres vous trouverez une pharmacie.
- Moi mon père je l'aime beaucoup.

On observe aussi des désaccords en genre et en nombre comme dans les exemples suivants [Sauvageot, A. 1972] tirés de corpus oraux :

- Une voiture émetteur
- Les dispositions que nous avons pris
- C'est mes amis

Les constructions syntaxiques complexes dont la réalisation est susceptible d'être particulière à l'oral sont les constructions interrogatives et les constructions relatives. La construction des phrases interrogatives peut :

- respecter les règles du français écrit : – inversion du sujet et du verbe du français écrit (la chambre est-elle libre [?]), – position en tête de proposition des pronoms, adjectifs, et adverbes interrogatifs (laquelle des deux est moins chère [?])
- ne pas respecter les règles du français écrit : – non inversion du sujet et du verbe ; c'est alors la prosodie qui marque l'interrogation (Elle est confortable [?]), – position en fin de proposition des pronoms, adjectifs, et adverbes interrogatifs (Elle est très chère n'est-ce pas [?], C'est très loin d'ici non [?], Son prix c'est combien [?])

Pour les relatives, la forme peut être standard (le professeur dont je parle), populaire (le prof que le parle de sa matière), pléonastique (le prof dont je parle de lui), ou, plus rarement défective (le mot (du directeur) que je parle) [Gadet, F. 1989].

⁴⁴ Nous employons langage oral pour désigner des productions linguistiques typiques observées en situation de dialogue oral ou écrit (tchat) spontanées.

⁴⁵ Nous reprenons ici la classification de [Kurdi, M. Z. 2003] qui différencie les aspects syntaxiques particuliers de l'exagrammaticalité. Les exemples que nous citons sont tirés de [Blanche-Benveniste, C. 2000] et de [Kurdi, M. Z. 2003].

⁴⁶ [Antoine, J.-Y. and Goulian, J. 1999, Antoine, J.-Y. and Goulian, J. 2001] ont montré que dans des dialogues oraux finalisés ce n'était pas tellement le cas.

III.2.2. Extragrammaticalité

[Kurdi, M. Z. 2003] propose d'utiliser le terme extragrammaticalité pour désigner les différents phénomènes spontanés de l'oral qui ne dépendent pas directement de la langue. La littérature divise une extragrammaticalité en trois zones (Figure III-1).

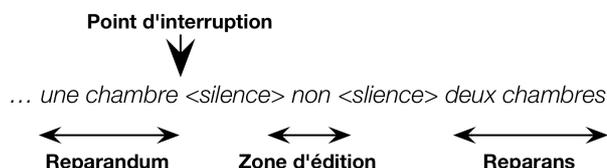


Figure III-1 : Schéma général d'une extragrammaticalité

Ce schéma ne permet cependant pas de rendre compte de tous les phénomènes d'extragrammaticalité, en particulier les extragrammaticalités lexicales (pauses simples, pauses remplies — hésitations —, mots incomplets, mots oraux — tels que ouais —, et amalgames — tels que ch'ui pour je suis —) et supralexicales (répétitions — un vol vers paris plus un vol un vol intérieur —, autocorrections — oui j'ai la j'ai les pages web... —, faux-départs — oui c'est à ça se prend au deuxième étage —, et inachèvements — et ils offrent les forf —).

[Goulian, J. 2002, Section 1.2] propose aussi une liste des phénomènes particuliers de l'oral.

III.2.3. Implications pour l'analyse automatique

Les phénomènes que nous venons d'évoquer sont malheureusement toujours inattendus et délicats à repérer ; ils sont donc très difficiles à traiter, il faut alors essayer de construire des analyseurs robustes. Pour un état de l'art du domaine, le lecteur peut consulter [Gala Pavia, N. 2003]. Disons que l'on va chercher à relaxer les contraintes syntaxiques et/ou faire de l'analyse par îlots.

III.3. Ambiguïté

La dernière classe de problèmes auxquels un analyseur est confronté concerne les ambiguïtés inhérentes à la langue naturelle. Nous en donnons ici un aperçu du point de vue linguistique avant d'en donner une définition formelle.

III.3.1. Point de vue linguistique

III.3.1.1. Définition

Pour Catherine Fuchs [Fuchs, C. 1996, chapitre 1], « pour qu'il y ait **ambiguïté linguistique**, il faut que les différentes significations en jeu soient prédictibles en langue, c'est-à-dire que l'analyse linguistique puisse en rendre compte : tous les **dictionnaires** du français doivent consigner le fait que la suite graphique *bière* correspond à deux unités lexicales, désignant respectivement la « boisson » et le « cercueil » ; toute **grammaire** du français doit prendre en compte le fait que la séquence « N1 faire V-infinitif N2 à N3 »⁴⁷ recouvre deux types de relations sous-jacentes correspondant respectivement à « N1 faire que X V N2 à N3 »⁴⁸ et à « N1 faire que N3 V N2 »⁴⁹. »

⁴⁷ Exemple : j'ai fait porter des fleurs à Lucie.

⁴⁸ Soit : j'ai fait que quelqu'un porte des fleurs à Lucie, pour notre exemple

L'ambiguïté peut être **virtuelle** ou **effective** : elle est virtuelle lorsque le contexte linguistique sélectionne l'une des significations ; elle est effective lorsque le contexte linguistique autorise plusieurs interprétations. Dans les exemples suivants, l'ambiguïté de bière reste virtuelle : [exemple 1] *la soif les poussa à commander deux bières au bar* (bière=boisson) et [exemple 2] *les croque-morts descendirent la bière dans la fosse* (bière=cercueil.) Par contre, dans la phrase *comme il faisait très chaud le jour de l'enterrement, on sortit la bière* (bière= ?), l'ambiguïté est effective.

Catherine Fuchs précise aussi que du point de vue linguistique, ni la sous-détermination⁵⁰ du sens, ni la sur-détermination⁵¹ du sens ne sont l'ambiguïté.

Une expression sous-déterminée est univoque, bien que sa signification reste ouverte. La sous-détermination ressortit au non-dit, à la généralité de sens, au sens flou et au sens en usage approximatif. Une expression sur-déterminée est plurivoque, mais ses différentes significations se cumulent. La sur-détermination ressortit à la surimposition d'un sens implicite et au cumul de sens. Une expression ambiguë est plurivoque, mais dans ce cas, les différentes significations sont mutuellement exclusives.

Ainsi, « sous-détermination et sur-détermination relèvent du discours, c'est-à-dire de l'usage de la langue par les personnes. Au contraire, le phénomène de l'ambiguïté relève du système de la langue elle-même : plus précisément des mécanismes par lesquels, dans la langue se trouvent associées des significations aux signes et aux combinatoires sur la chaîne. »

III.3.1.2. Classification

[Fuchs, C. 1996] propose une classification exhaustive des ambiguïtés du français⁵². Les ambiguïtés sont organisées en cinq classes dont nous passons rapidement en revue les trois premières^{53,54,55}.

III.3.1.2.1. Ambiguïtés morphologiques et lexicales

Au niveau de l'analyse morphologique et lexicale, il s'agit d'identifier les mots qui composent le texte, puis les unités lexicales. Pour ce faire, il faut être capable de :

a. Délimiter les frontières de mots

C'est le problème de la segmentation d'une chaîne phonique en mots. Il concerne la reconnaissance de la parole.

⁴⁹ Soit : *j'ai fait que Lucie porte des fleurs*, pour notre exemple

⁵⁰ Plusieurs phénomènes participent à la sous-détermination, par exemple : le sens indéterminable (« le silence vertébral indispose la voile licite »), le non dit (« à cette époque-là **ma femme** me gâtait » : la femme qui était mienne à l'époque et qui ne l'est plus, la femme qui était mienne à l'époque et qui l'est encore, ou encore celle qui ne l'était pas à l'époque et qui l'est maintenant ?).

⁵¹ Plusieurs phénomènes participent à la sur-détermination, par exemple : la surimposition d'un sens implicite (« la poubelle est pleine ! » qui signifie aussi certainement qu'il faut la vider), le jeu de mots (« à l'agence de voyage X, tout le monde s'y retrouve » physiquement et/ou financièrement ?)

⁵² Pour l'anglais, le lecteur peut consulter [Hirst, G. 1987] et [Sadler, V. 1989], par exemple.

⁵³ Nous n'examinerons pas ici, faute de place, les ambiguïtés sémantiques et les ambiguïtés pragmatiques.

⁵⁴ Nous nous contentons ici de donner un nom à chaque type d'ambiguïté et de proposer un exemple commenté. Pour une définition exhaustive, le lecteur peut consulter [Fuchs, C. 1996].

⁵⁵ Les exemples que nous présentons ici sont des « exemples de linguistes ». De notre point de vue, ils ont pour seule vocation d'illustrer le plus immédiatement possible les problèmes qui peuvent survenir. Nous aurions pu choisir nos exemples dans un exemplaire du journal *Le Monde*, la difficulté aurait alors été de repérer des exemples de chaque catégorie. Malheureusement, si avec l'habitude, certains types d'ambiguïté sont très faciles à détecter, d'autres sont plus durs ; le processeur humain met en œuvre des processus de filtrage très efficaces. Nous montrerons, Chapitre VII, que l'on peut très bien généraliser la description des ambiguïtés indépendamment des occurrences de mots apparaissant dans un segment ambigu.

Exemple a.1 passe moi la fiche/l'affiche

b. Reconnaître la catégorie morpho-syntaxique du mot

Exemple b.1 le pilote ferme la porte
le pilote barre l'ouverture/le pilote solide la tient dans ses bras

Exemple b.2 il s'est penché sur les dossiers sinistres
les dossiers qui sont sinistres/les dossiers qui concernent les sinistres

c. Reconnaître la nature des flexions

Ce problème peut survenir en reconnaissance vocale en cas d'homophonie. Il peut survenir à l'écrit en cas d'homographie.

Exemple c.1 il marche/ils marchent

Exemple c.2 à cette vue, il pâlit
indicatif présent/indicatif passé simple

d. Délimiter les unités lexicales

Exemple d.1 il convient avec le gardien de l'ordre du jour
convenir de l'ordre du jour avec le gardien/convenir du jour avec le gardien
de l'ordre

e. Reconnaître l'unité lexicale en cas d'homonymie

Ce problème concerne les homophones non homographes (à l'oral), les homographes non homophones (à l'écrit), et les homographes homophones (à l'oral et à l'écrit).

Exemple e.1 il faut monter ce sot/seau/sceau/saut

Exemple e.2 elle a perdu ses fils
ses enfants/ses objets fins et cylindriques

Exemple e.3 je suis un imbécile
je m'identifie à un imbécile/je marche derrière un imbécile

f. Reconnaître le sens de l'unité lexicale en cas de polysémie

Ce problème concerne les lexèmes polysémiques et les graphèmes polysémiques.

Exemple f.1 dans le milieu des conservateurs, cette mesure a été rejetée
le milieu des gardiens/le milieu des traditionnalistes

Exemple f.2 je vote pour ma mère
à la place/en faveur de ma mère

Exemple f.3 j'ai fait cela pour rien
gratuitement/sans but/en vain

III.3.1.2.2. Ambiguïtés syntaxiques

Au niveau de l'analyse syntaxique, il s'agit de délimiter les phrases et les syntagmes afin de les hiérarchiser. Il faut aussi attacher une fonction à chaque syntagme.

a. Délimiter les phrases et les propositions

À l'oral, le récepteur doit décomposer une chaîne phonique. Il y parvient grâce aux pauses et aux variations d'intonation de l'émetteur. Par contre, à l'issue d'une étape de reconnaissance vocale, ces informations ne sont, en général, plus disponibles.

Exemple a.1 l'instituteur dit : « le maire est un âne » / « l'instituteur », dit le maire, « est un
âne »

À l'écrit, les textes sont en général prédécoupés : des blancs séparent les mots, et des signes de ponctuation séparent les phrases, les propositions, voire les groupes. Cependant, ces marques ne sont pas fiables à 100%.

Exemple a.2 Quant à la réforme fiscale, on se demande qui en veut vraiment : « Les élus en parlent tant [=tellement,] qu' [=aussi longtemps qu'] ils n'ont pas à la voter » a dit le ministre. (alternative)

b. Délimiter les syntagmes pour les hiérarchiser

On retrouve ici les problèmes de rattachement des groupes prépositionnels, des groupes adjectivaux, et des groupes nominaux. [Fuchs, C. 1996] propose des patrons, tels que ceux que nous avons proposés aussi [Blanchon, H. 1992a, Blanchon, H. 1992b, Blanchon, H. 1994b, Blanchon, H. 1995a, Blanchon, H. 1995c, Blanchon, H. 1996, Blanchon, H. 1997], pour définir des problèmes d'attachement. Nous donnons ici seulement quelques exemples choisis.

Exemple b.1 Le capitaine a rapporté un vase de Chine.
rapporter de Chine un vase/rapporter un vase chinois

Exemple b.2 Il a dit qu'il donnerait son avis par fax.
dire par fax/donner par fax

Exemple b.3 Les voisins et les parents de Paul sont venus.
les parents de Paul et les voisins/les parents de Paul et les voisins de Paul

Exemple b.4 Ce produit vous aide à conserver votre teint frais.
conserver frais un teint/conserver un teint qui est déjà frais

Exemple b.5 C'est un professeur de droit anglais.
professeur anglais/droit anglais

Exemple b.6 Elle prend les cahiers et les classeurs noirs.
les classeurs noirs et les cahiers/les classeurs noirs et les cahiers noirs

c. Reconnaître la nature des propositions

Il existe aussi des problèmes de caractérisation syntaxique : relative/interrogative indirecte, circonstancielle/relative indirecte.

Exemple c.1 J'ai demandé au monsieur qui habitait là
le monsieur habitait là/qui habitait là ?

Exemple c.2 Écrivez-moi si vous comptez venir me voir
si vous comptez venir, écrivez-moi/écrivez-moi pour me dire si vous...

d. Reconnaître la fonction des syntagmes

On cherche ici à reconnaître la fonction syntaxique. Une même préposition peut introduire un complément régi par le verbe ou un circonstant, la place d'un syntagme peut aussi être fonctionnellement ambiguë.

Exemple d.1 La circulation a été déviée par la gendarmerie.
la gendarmerie dévie la circulation/la circulation est déviée et passe par le bâtiment de la gendarmerie

Exemple d.2 Quel auteur cite ce conférencier ?
un auteur est cité par un conférencier/un conférencier est cité par un auteur

Exemple d.3 Chaque jour grandissait sa renommée.
sa renommée est plus grande chaque jour/chaque jour fait grandir sa renommée

III.3.1.2.3. Ambiguïtés prédicatives

[Fuchs, C. 1996] classe dans cette catégorie plusieurs situations, nous nous intéressons ici aux cas pour lesquels il faut aussi restituer la construction des prédicats et restituer les prédications (en cas d'ellipse ou de construction réduite).

a. Restituer les arguments

- Exemple a.1 Les poulets sont prêts à manger.
 les poulets vont manger/les poulets vont être mangés
- Exemple b.2 Pierre fait porter des chocolats à Lucie.
 Lucie porte les chocolats/Lucie reçoit les chocolats
- Exemple b.3 Le choix du médecin peut se discuter.
 Le choix fait par le médecin/le médecin choisi

b. Restituer les prédications

- Exemple b.1 Je vous traiterais comme mon fils.
 comme si vous étiez mon fils/comme mon fils vous traite
- Exemple b.2 Le magistrat juge les enfants coupables.
 les enfants coupables sont jugés/les enfants sont jugés coupables
- Exemple b.3 Je l'ai quitté joyeux.
 je suis joyeux/il est joyeux

III.3.2. Implications pour l'analyse automatique

Lorsqu'il analyse un texte, un analyseur doit pouvoir détecter les **ambiguïtés effectives**. Dans ce cas, il n'est bien sûr pas en mesure de choisir seul la bonne interprétation. L'analyseur peut aussi manifester trois types de comportements parasites vis-à-vis de l'ambiguïté. Il peut : ne pas détecter une ambiguïté effective, ne pas savoir résoudre une ambiguïté virtuelle, et produire une **ambiguïté artefact**.

III.3.2.1. Ambiguïtés effectives non détectées

L'analyseur peut être incapable de repérer une ambiguïté effective en langue. Il peut aussi être incapable de repérer complètement l'ambiguïté effective (il ne détecte pas toutes les interprétations).

Dans ce cas, nous dirons que l'analyseur manifeste un défaut de couverture de la langue.

III.3.2.2. Ambiguïtés virtuelles non résolues

L'analyseur peut aussi être incapable de résoudre automatiquement une ambiguïté virtuelle en langue. Il considère donc cette ambiguïté virtuelle comme une ambiguïté effective.

Ce défaut pourrait être assimilé à un défaut de couverture. Il nous semble cependant que le niveau où se situe ce défaut n'est pas de l'ordre de la langue. Le problème est plutôt de l'ordre de l'« interprétation ». L'analyseur ne dispose pas d'informations suffisamment précises sur le contexte.

Si la phrase « Pierre a fait porter des chocolats à Lucie. » est virtuellement ambiguë pour un lecteur, dans les deux contextes que nous proposons ci-dessous, l'ambiguïté disparaît :

- Exemple 1 : Comme elle venait d'accoucher, Pierre a fait porter des chocolats à Lucie.

Exemple 2 : Comme il avait les bras encombrés, Pierre a fait porter des chocolats à Lucie.

Pour que l'ambiguïté virtuelle soit résolue, il faudrait que l'analyseur puisse inférer que « les chocolats sont un cadeau à Lucie qui vient d'accoucher ; elle reçoit donc les chocolats » (pour l'Exemple 1), et que « comme Pierre a les bras encombrés, il ne peut pas porter les chocolats ; il les fait donc porter par Lucie » (pour l'Exemple 2).

III.3.2.3. Ambiguïtés artefacts

Comme le précise [Chanod, J.-P. 1993] : « la description informatique des mécanismes langagiers n'est pas neutre. Elle produit ses propres interprétations, hors du champ d'étude de la linguistique traditionnelle. Les descriptions informatiques interagissent et produisent de nombreux effets de bord ou analyses indésirables. ... [le parasitisme computationnel] se manifeste par la production d'analyses indésirables, inappropriées résultant de l'application de règles tout à fait fondées linguistiquement par ailleurs. »

Plus la description de la langue est fine, plus l'analyseur produit des représentations multiples. Il s'ensuit que l'analyseur va produire des structures qui représentent des **ambiguïtés artefacts**, c'est-à-dire des ambiguïtés parasites qui n'existent pas en langue.

Pour rendre ce phénomène concret, imaginons que le concepteur d'un analyseur morpho-syntaxique veuille affiner la description des groupes nominaux pour capturer le fait qu'un groupe nominal peut être constitué de deux noms ou groupes nominaux juxtaposés (comme dans « visite éclair », « modèle quatre portes », etc.). Sans autre précaution, cette manipulation va produire des catastrophes comme pour la phrase « Le valet dessert la terrasse » dans laquelle « le valet dessert » sera éventuellement un groupe nominal, « la » un pronom et « terrasse » un verbe (« elle est terrassée par le valet dessert ») !

III.3.3. Spécification formelle de l'ambiguïté

Jusqu'à présent, nous avons dit qu'une phrase est ambiguë. En fait, on peut presque toujours réduire la localisation d'une ambiguïté à une partie de la phrase. Nous allons formaliser cela [Blanchon, H. and Boitet, C. 2004b, Boitet, C. and Tomokio, M. 1995].

Prenons par exemple la phrase suivante :

(1) **Do you know where the international telephone services are located?**

Le fragment souligné contient une ambiguïté d'attachement qui peut être représentée par deux squelettes [Black, E., *et al.* 1993] :

[international telephone] services / international [telephone services]

Cependant, il n'est pas suffisant de considérer cette séquence isolément. Prenons comme exemple la phrase suivante :

(2) **The international telephone services many countries.**

L'ambiguïté a disparu ! Dans la pratique, il est très fréquent que l'ambiguïté relative à un fragment apparaisse, disparaisse puis réapparaisse lorsque l'on augmente le contexte du fragment. Ainsi, pour définir proprement une ambiguïté, il faut considérer le fragment à l'intérieur d'une phrase et clarifier l'idée que le fragment utile est le plus court fragment sur lequel l'ambiguïté puisse être observée.

De manière formelle, on peut donc dire qu'un fragment **F** présente une ambiguïté de degré **n** ($n \geq 2$) dans une phrase **U** s'il possède **n** représentations différentes qui peuvent être utilisées pour produire une représentation complète de **U**. Pour être le support de l'ambiguïté, **F** doit être minimal relativement à l'ambiguïté considérée. Cela signifie que **F**, et les **n**

représentations qui lui sont associées, ne peut être réduit à un fragment strictement plus petit F' , et ses n sous-représentations associées, sans perdre la première propriété.

Dans l'exemple (1), le fragment "the international telephone services", associé à ses deux représentations the [international telephone] services / the international [telephone services], n'est pas minimal car il peut être réduit au fragment "international telephone services", associé à ses deux représentations [international telephone] services / international [telephone services], qui est minimal.

Nous proposons la définition formelle suivante :

Une ambiguïté A de degré n ($n \geq 2$) relative à un système de représentation R , peut être formellement définie comme :

$A = (U, F, \langle S_1, S_2, \dots, S_m \rangle, \langle s_1, s_2, \dots, s_n \rangle, m \geq n)$ où :

- U est une phrase complète, appelée le contexte de l'ambiguïté.
- F est un fragment de U , habituellement, mais non nécessairement connexe, le *support* de l'ambiguïté.
- les S_1, S_2, \dots, S_m sont les représentations distinctes de U dans R , et les s_1, s_2, \dots, s_n leurs sous-parties représentant F telles que $\forall i, j ; s_i \neq s_j$.
- Condition de minimalité :

Quel que soit F' un fragment de U strictement contenu dans F , et s'_1, s'_2, \dots, s'_n ($n \leq m$) les parties respectives de s_1, s_2, \dots, s_n correspondant à F' , il existe alors au moins une paire s'_i, s'_j ($i \neq j$) telle que $s'_i = s'_j$.

Le *type* de l'ambiguïté A dépend de la différence qui caractérise les s_i . Il doit être défini relativement à chaque R particulier.

Figure III-2 : Définition formelle d'une ambiguïté

Nous pouvons associer la représentation graphique suivante à la définition précédente.

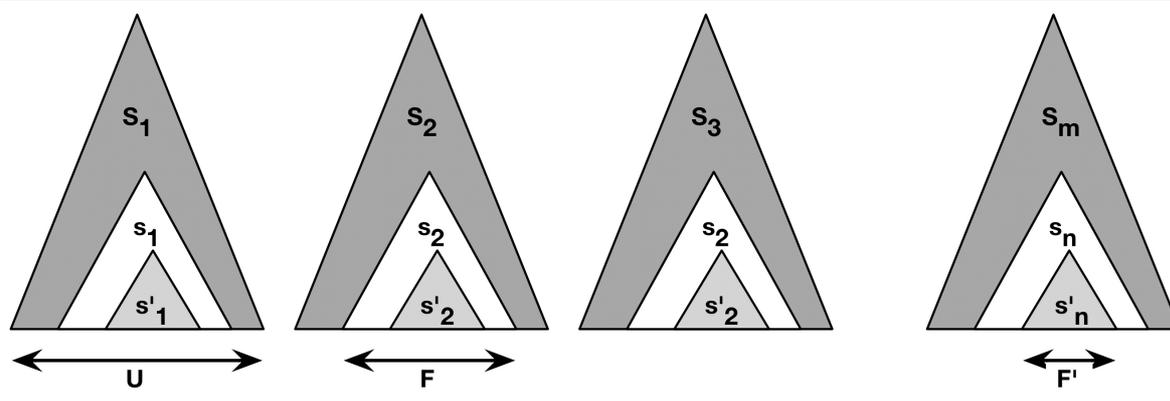


Figure III-3 : Représentation graphique d'une ambiguïté

On peut voir que S_2 et S_3 partagent la même représentation de F ce qui signifie que S_2 et S_3 diffèrent ailleurs que sur F .

Chapitre IV

Revue des différentes situations de communication interpersonnelle

Pour terminer cette partie, nous passons en revue différentes situations de communication interpersonnelle que nous caractérisons selon cinq axes de classification : la taille du domaine, le degré de planification du contenu, le délai de réception du message, le médiateur de la communication, et la nécessité d'une interaction avec l'émetteur pour assurer une transmission de qualité du message produit.

Nous sommes intéressés par la taille du domaine, le type de planification du message, l'intervalle de temps écoulé entre son émission et sa réception, le type de médiateur utilisé et le degré d'interaction avec l'émetteur ou le médiateur souhaitable pour assurer une bonne réception du message.

Pour les trois types de médiation possibles (pas de médiateur, médiateur humain, système de traduction), nous proposons d'examiner trois types de communication : la communication écrite, la communication orale monodirectionnelle (exemple : le discours, ou la conférence), et la communication orale bidirectionnelle (le dialogue).

IV.1. Communication monolingue non médiatisée

IV.1.1. Communication écrite

Un auteur rédige un document destiné à des lecteurs partageant sa langue ; il n'y a donc pas besoin de médiateur pour assurer la communication. Le domaine du document est normalement fermé, et la mise en forme du contenu exige habituellement certains efforts de planification.

Les lecteurs accéderont normalement au document produit par l'auteur après un certain délai (pas d'unité de temps). Ceux qui ont des problèmes de compréhension du contenu n'ont pas d'autre choix que d'essayer de rentrer en contact avec l'auteur afin de lui demander des explications et des éclaircissements. L'interaction avec l'auteur afin d'atteindre une communication de qualité peut donc être très variable.

Cette situation est présentée dans la Figure IV-1.

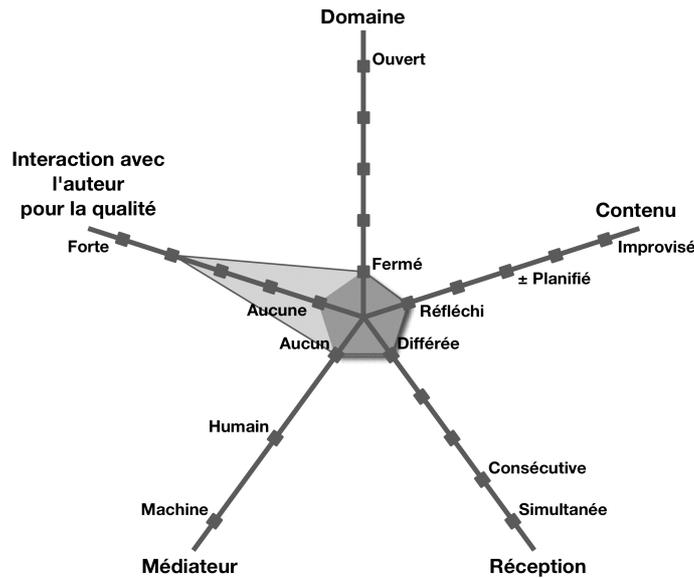


Figure IV-1 : Communication interpersonnelle monolingue écrite réfléchi

IV.1.2. Communication orale d'un orateur

La communication écrite est monodirectionnelle : un auteur s'adresse à des lecteurs qui ont en général très peu d'interaction avec lui. À l'oral, la situation équivalente est celle que l'on rencontre par exemple lors de conférences, allocutions ou cours magistraux.

Un orateur prononce un discours (zone sombre Figure IV-2) adressé à un auditoire qui parle sa langue. Ce discours concerne un sujet particulier (domaine fermé) et il a probablement été soigneusement préparé (contenu réfléchi).

Les auditeurs reçoivent le message en temps réel. Ceux qui ont des problèmes de compréhension du contenu n'ont, en général, pas d'autre choix que d'essayer de poser des questions à l'orateur lorsque celui-ci aura terminé. L'interaction avec l'orateur prendra alors une forme plus libre, on arrive à une situation de dialogue semi-formel (zone claire Figure IV-2).

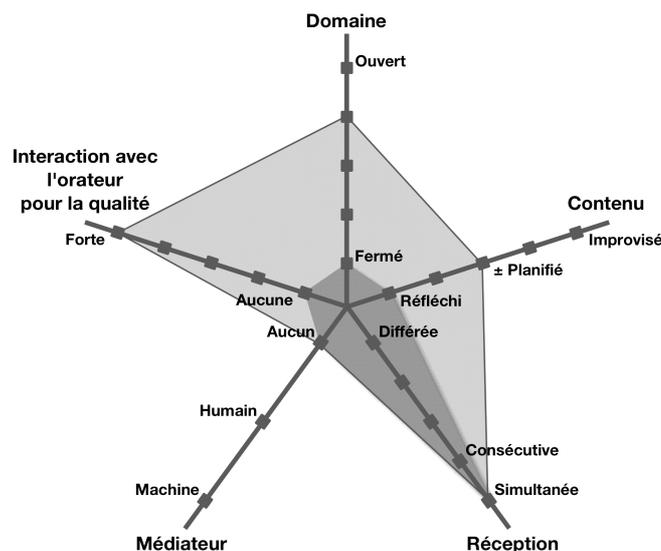


Figure IV-2 : Communication orale monolingue d'un orateur

IV.1.3. Communication orale entre interlocuteurs

En situation de dialogue, on rencontre deux situations extrêmes : le dialogue formel et le dialogue informel.

En dialogue formel — négociation commerciale, discussion diplomatique — (zone sombre Figure IV-3), le domaine du dialogue est prévu avant la rencontre et donc limité. Les interlocuteurs ont un langage policé et réfléchi. La réception est instantanée et les interlocuteurs peuvent se demander mutuellement des précisions et vérifier qu'ils comprennent bien ce que dit leur partenaire.

Par contraste, en dialogue informel — rencontre impromptue, discussion de couloir — (zone claire Figure IV-3), le domaine du dialogue est complètement ouvert ; il n'est pas prévu et dépend des circonstances de la rencontre. Le langage utilisé est donc « naturel » et l'on va retrouver les phénomènes décrits dans la section III.2. La réception est instantanée et les interlocuteurs peuvent se demander mutuellement des précisions et vérifier qu'ils comprennent bien ce que dit leur partenaire.

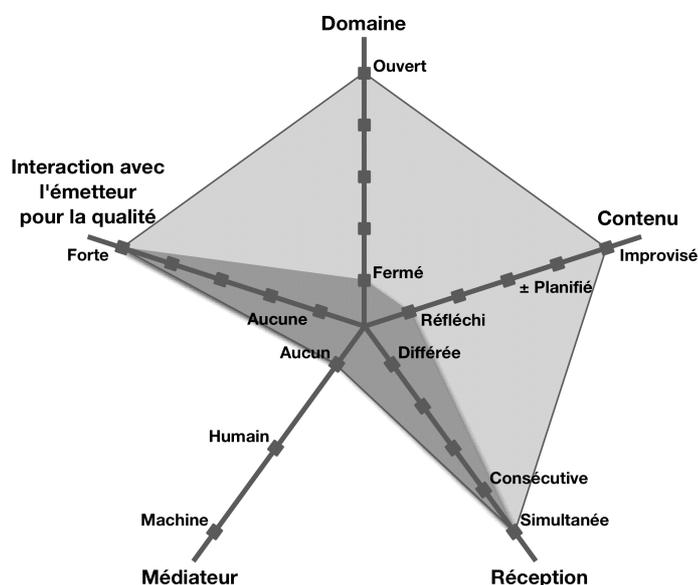


Figure IV-3 : Communication orale monolingue entre interlocuteurs

Ces deux situations ont leur équivalent à l'écrit. L'échange formel se fait plutôt par courrier postal ou électronique. L'échange informel peut se faire par les canaux précédents ou en temps presque réel en tchat.

IV.2. Communication multilingue médiatisée par l'homme

Nous reprenons ici les mêmes familles de situations. Maintenant, la communication n'est plus monolingue, mais multilingue. Le message en langue source est traduit en langue cible par un interprète humain.

IV.2.1. Communication écrite traduite via un traducteur humain

Le document source possède les propriétés que nous avons proposées au IV.1.1.

Un traducteur doit maintenant produire un document en langue cible en restant fidèle au contenu du document source. En fonction des problèmes de compréhension du contenu

qu'il rencontre, le traducteur devra interagir, plus (zone claire Figure IV-4) ou moins (zone sombre Figure IV-4), avec l'auteur afin de lui demander des explications et des éclaircissements.

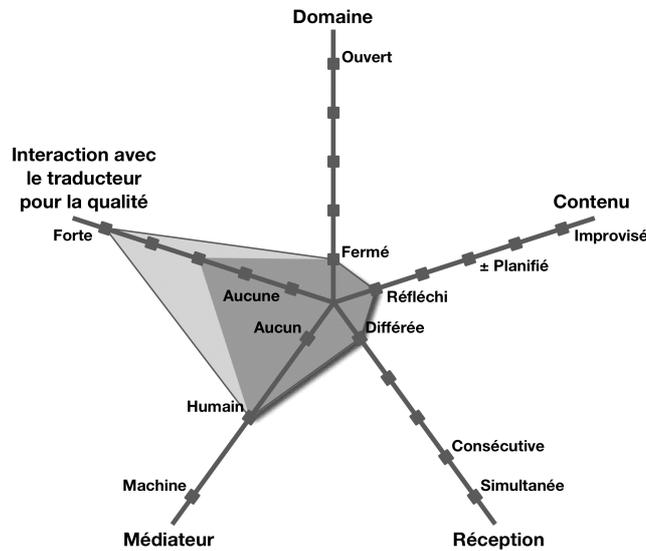


Figure IV-4 : Communication écrite traduite via un traducteur humain

IV.2.2. Communication orale d'un orateur traduite via un interprète humain

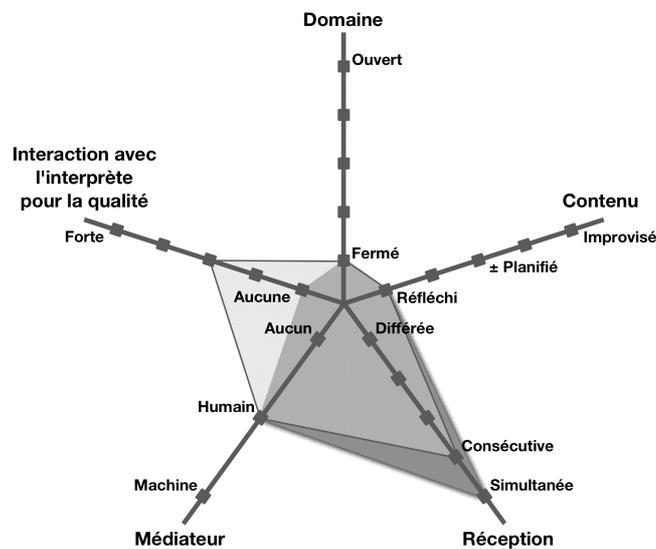


Figure IV-5 : Communication orale d'un orateur traduite via un interprète humain

La traduction du discours peut être simultanée (zone sombre Figure IV-5) ou consécutive (zone claire Figure IV-5). Le discours a bien sûr les mêmes propriétés que ci-dessus (IV.1.2).

Si la traduction est simultanée (zone sombre Figure IV-5), l'interprète ne peut pas interagir avec l'orateur afin de lui demander de préciser la pensée ; il prend la responsabilité des choix d'interprétations qu'il effectue.

Si la traduction est consécutive (zone claire Figure IV-5), l'interprète peut éventuellement prendre la responsabilité de demander des précisions à l'orateur afin de faire la meilleure traduction possible.

Comme en communication monolingue, une interaction semi-formelle avec l'assistance pourra avoir lieu à la fin du discours de l'orateur.

IV.2.3. Communication orale entre interlocuteurs via un interprète humain

La communication orale entre interlocuteurs via un interprète humain est plutôt une situation qui est préparée, elle est donc plutôt de type formel⁵⁶ et réfléchi⁵⁷ (zone sombre Figure IV-6). Le domaine sera donc plutôt fermé et la traduction sera alternativement simultanée et consécutive en fonction des besoins d'interaction de l'interprète avec le locuteur à des fins de clarification [Oviatt, S. and Cohen, P. 1991].

Il existe cependant des situations plus informelles pour lesquelles le domaine est plus ouvert et le discours des interlocuteurs est moins planifié (zone claire Figure IV-6). On peut citer par exemple des conversations transcontinentales sur des croisements entre animaux (chiens, chevaux, ...) ou des conversations de politesse de parents avec des parents étrangers chez qui ils ont envoyé un enfant.

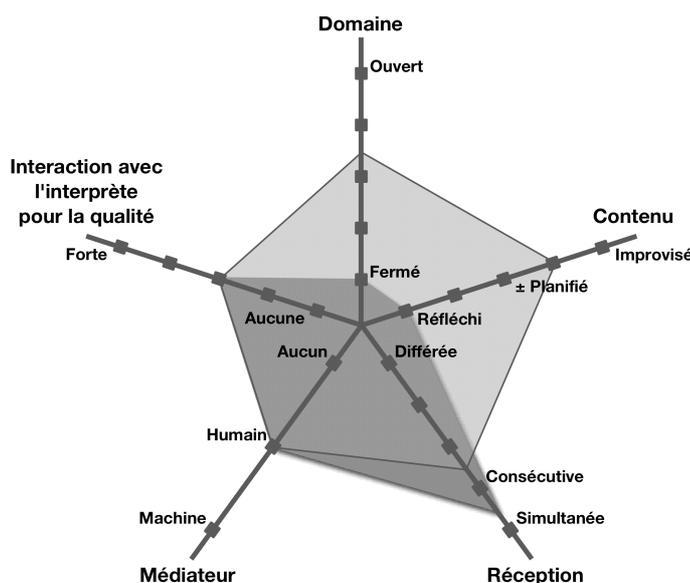


Figure IV-6 : Communication orale entre interlocuteurs via un interprète humain

IV.3. Communication multilingue médiatisée par la machine

Examinons finalement la communication multilingue pour laquelle le message en langue source est traduit en langue cible par un système de TA.

⁵⁶ Nous verrons dans le Chapitre IX que l'on peut imaginer des situations pour lesquelles la rencontre entre les interlocuteurs peut être fortuite.

⁵⁷ Les situations seront les mêmes que celles que nous avons proposées dans la section IV.1.3. Ici, les interlocuteurs ne parlent pas la même langue.

IV.3.1. Communication écrite traduite via un système de TA

Nous l'avons dit, la TA peut produire d'excellents résultats, et ce sans intervention humaine, lorsque le domaine du document est fermé (zone sombre Figure IV-7).

En revanche, dès que le domaine s'ouvre (devient plus important), la qualité de la traduction décroît s'il n'y a pas intervention humaine (zone claire Figure IV-7), qu'il s'agisse de pré-édition, d'interaction médiane, ou de post-édition.

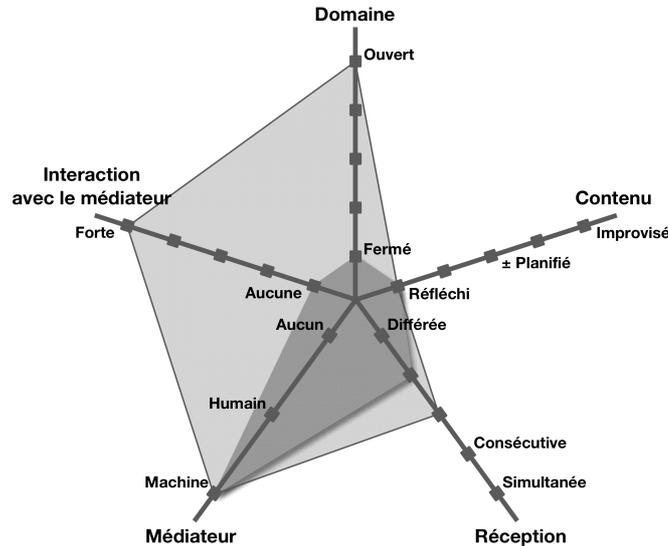


Figure IV-7 : Communication écrite traduite via un système de TA

IV.3.2. Communication orale d'un orateur traduite via un système de TA

Pour traduire le discours d'un orateur, on ne peut pas envisager que le système puisse requérir l'assistance de l'orateur (Figure IV-8). Parce qu'un long monologue est bien plus complexe que de simples tours de parole, cette tâche semble être hors de portée pour l'instant et nous ne connaissons pas de travaux dans cette direction.

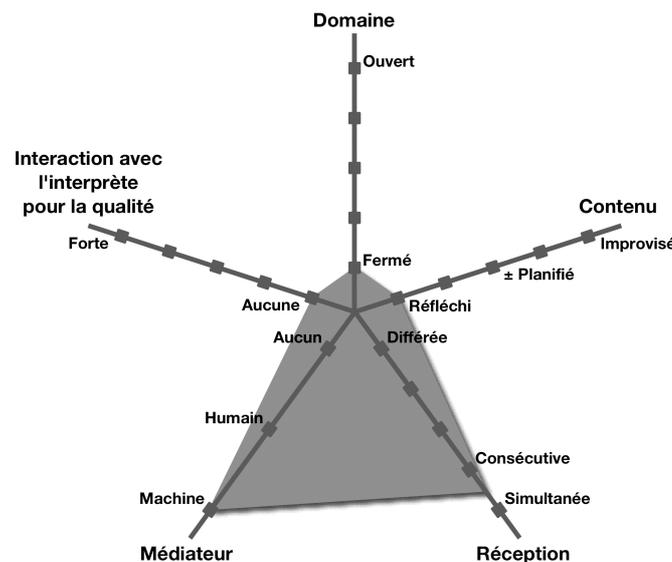


Figure IV-8 : Communication orale d'un orateur traduite via un système de TA (encore utopique)

IV.3.3. Communication orale entre interlocuteurs via un système de TA

En communication orale finalisée entre interlocuteurs sur un domaine très fermé avec un discours réfléchi (zone sombre Figure IV-9), les tours de parole peuvent être assez courts, et comme nous le verrons (Chapitre XI et Chapitre XII), il est éventuellement possible de se passer d'une gestion complexe du contexte.

Par contre, si l'on veut traduire des conversations sur des domaines plus ouverts et en permettant aux locuteurs de moins planifier leur discours, il nous semble que l'on ne pourra pas se passer de faire intervenir le locuteur afin qu'il assiste le processus de traduction (zone claire Figure IV-9).

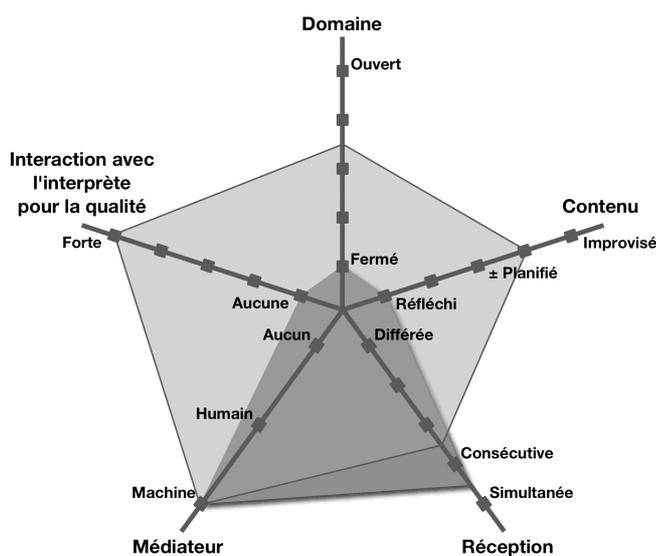


Figure IV-9 : Communication orale entre interlocuteurs via un système de TA

Conclusion

Parmi les différentes situations que nous avons évoquées, nous nous intéresserons à partir de maintenant à la **communication multilingue médiatisée par la machine** dans deux situations qui s'opposent selon trois dimensions : la modalité de communication, la planification du message, et la taille du domaine du discours.

Modalité de communication	Écrit	Oral
Degré de planification	Réfléchi	Spontané
Taille du domaine	Tout terrain	Spécialisé

Table I : Caractéristiques des situations de communication interpersonnelle médiatisée par la machine sur lesquelles nous travaillons

La seconde colonne de la Table I caractérise la **communication interpersonnelle écrite, réfléchie, et tout terrain** médiatisée par la machine. La troisième colonne caractérise la **communication interpersonnelle orale, spontanée, et spécialisée** médiatisée par la machine.

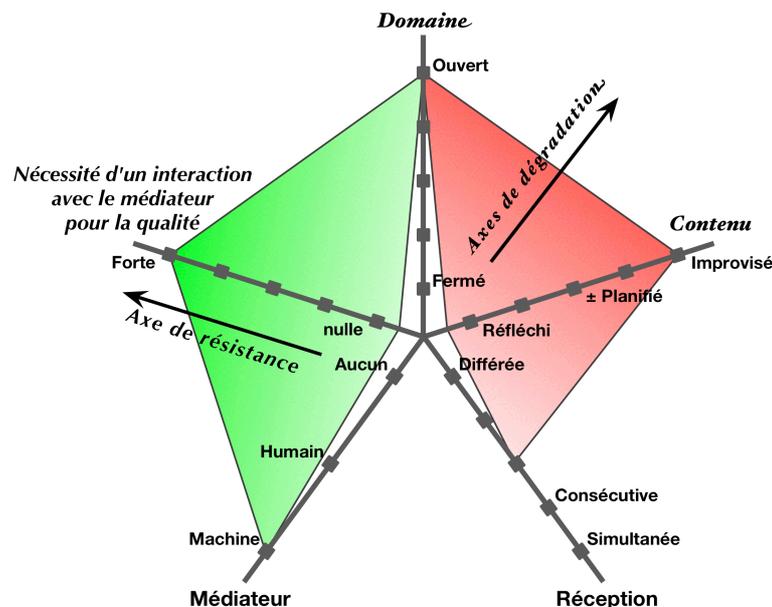


Figure I : Axes de dégradation et axe de résistance en communication interpersonnelle médiatisée par la machine

Dans la classification à cinq branches que nous avons proposée, on peut distinguer deux axes de dégradation de la qualité des systèmes : la planification et la taille du domaine. Pour résister à la dégradation de la qualité, la seule voie actuellement envisageable est de permettre à l'utilisateur d'interagir avec le système afin de l'aider.

En **communication interpersonnelle écrite, réfléchie, et tout terrain**, l'utilisateur peut aider le système de traduction au moyen de dialogues de désambiguïsation.

En **communication interpersonnelle orale, spontanée, et spécialisée**, l'utilisateur peut aider le système de traduction en corrigeant les erreurs du module de reconnaissance de la parole. Travailler en domaine restreint permet de concevoir des analyseurs capables de lutter, dans une certaine mesure, contre le bruit (performatif) et contre le flou (syntaxique) d'un message spontané.

Deuxième Partie

**Communication interpersonnelle
écrite réfléchie tout terrain**

Introduction

Les analyseurs sont conçus pour tirer le meilleur parti possible des informations présentes dans le texte afin de désambiguïser automatiquement le plus grand nombre d'ambiguïtés virtuelles et de produire le plus petit nombre possible d'ambiguïtés artefacts. Le plus souvent, ils ne produisent qu'une seule analyse.

Chaque fois que l'analyseur dispose d'un nombre suffisant d'indices pour faire un choix, ce choix est effectué. Ainsi, à un instant donné, l'analyseur peut avoir construit une représentation multiple qui sera réduite à une représentation unique dès qu'il aura pu prendre une décision. Si un certain degré d'incertitude persiste à la fin, le choix final sera fait de façon heuristique.

Dans l'approche que nous avons adoptée dans le cadre du projet LIDIA, l'analyseur n'utilise pas d'heuristique pour faire de choix lorsque les « connaissances » et les « règles » qu'il met en œuvre ne lui permettent pas de choisir. C'est l'auteur qui choisit à la place du système, après l'analyse, lors d'une session de désambiguïstation interactive. La maquette LIDIA-1, présentée aux chapitres V et VI, nous a permis d'étudier les contraintes informatiques, ergonomiques et linguistiques d'un système de Traduction Automatique Fondé sur le Dialogue (TAFD).

Les progrès réalisés en désambiguïstation interactive depuis 1994 sont présentés dans le chapitre VII. L'étude de corpus permet de trouver les classes d'ambiguïtés auxquelles l'analyseur sera confronté, puis de proposer pour chacune une description et une manière de produire une question. Nous avons fait cela pour le français et pour l'anglais. Nous avons aussi vérifié, au moyen de deux expériences, que la manière de produire des questions que nous proposons permet d'obtenir des questions compréhensibles par un auteur non spécialiste. C'est un point très important, car il faut, *in fine*, que le système soit utilisable.

Nous avons aussi apporté plusieurs améliorations à la première architecture informatique que nous avons proposée et implémentée. Ces améliorations permettent d'accéder aux services de TAFD au moyen de clients qui ne nécessitent d'autres ressources qu'une connexion Internet. Ces clients sont des éditeurs de documents augmentés des fonctionnalités d'accès aux services de TAFD à travers le réseau.

Au cours de ce travail, nous nous sommes aussi rendu compte que les informations obtenues pendant la session de désambiguïstation interactive d'un document source pourraient être utiles aux futurs lecteurs dudit document. En effet, ces informations, qui donnent le sens effectif (voulu par l'auteur) des segments ambigus, pourraient être ajoutées au document

source pour produire un Document Auto-Explicatif (DAE). Un visualiseur de DAE devrait montrer les segments ambigus du document, et lorsqu'il le souhaite, le lecteur pourrait demander à voir, sous une forme intelligible, l'interprétation à retenir. Cette idée est présentée en détail au chapitre VIII.

Chapitre V

Amélioration de la qualité en TA

Améliorer la qualité en TA peut se faire selon trois types d'approche :

1. Mettre en œuvre des heuristiques que l'on peut classer en trois catégories :
 - heuristiques psycholinguistiques,
 - heuristiques linguistiques,
 - heuristiques apprises sur le corpus de documents à traduire.
2. Contraindre la langue par le vocabulaire et la syntaxe utilisable dans les textes, ainsi que les sens possibles des mots et des constructions. C'est l'approche par langue contrôlée (langage contrôlé).
3. Faire intervenir l'utilisateur, de préférence l'auteur, dans le cadre de sessions de désambiguïsation interactive.

Après avoir brièvement passé en revue les deux premières approches, nous nous concentrons sur la troisième en évoquant quelques travaux. Nous présentons ensuite l'approche proposée par le GETA dans le cadre du projet LIDIA.

V.1. TA sans interaction

L'approche par langue contrôlée, si elle a connu certains succès, reste confinée à des domaines de spécialité dont la niche applicative est très étroite ; on a affaire à des systèmes sur mesure. Les projets KBMT-89 et KANT (cf. infra) montrent les limites de cette approche.

Si l'on ne peut espérer une qualité parfaite, on peut mettre en œuvre des heuristiques et espérer qu'en cas de difficulté le système fera les bons choix.

V.1.1. Heuristiques psycholinguistiques

On regroupe ici abusivement les analyseurs basés sur la théorie « Principes et Paramètres » de la grammaire générative chomskienne (*principle-based parsing*) et les analyseurs qui se disent fondés sur les modèles psycholinguistiques.

Pour la première famille, le lecteur peut consulter [L'haire, S., *et al.* 2000, Lin, D. 1994, Merlo, P. 1995] et [Berthouzoz, C. and Merlo, P. 1995, Berthouzoz, C. and Merlo, P. 2000] qui ajoutent une étape de désambiguïsation statistique (cf. V.1.3). Pour la seconde

famille, [Doan-Nguyen, H. 2003] dresse un état de l'art très complet des modèles et des expériences passées.

V.1.2. Heuristiques sémantiques

Il s'agit ici de permettre à l'analyseur d'utiliser des « connaissances » sémantiques pertinentes pour la tâche.

Ce genre de technique est en particulier utilisé pour la désambiguïsation de la polysémie lexicale. Le lexique est organisé sous forme de réseau. Les mots d'un énoncé sont excités et le sens le plus probable de chacun d'entre eux doit émerger.

Il existe une très vaste littérature sur le sujet. Par exemple, [Veronis, J. and Ide, N. M. 1990] utilisent des réseaux de neurones, [Hirst, G. 1988] utilise des mots « polaroids » et [Lafourcade, M., *et al.* 2002, Lafourcade, M. and Standford, E. 1999], des vecteurs conceptuels. Pour un état de l'art du domaine, le lecteur peut consulter [Hirst, G. 1987] et [Jurafsky, D. and Martin, J. H. 2000b].

V.1.3. Heuristiques adaptées à un corpus donné

Avec cette approche, les règles de désambiguïsation sont apprises par observation de régularités sur le corpus. Elles seront d'autant plus efficaces que le corpus contient des documents homogènes syntaxiquement.

Ces techniques sont très couramment utilisées, voir par exemple [Gala Pavia, N. 2003] et [Jurafsky, D. and Martin, J. H. 2000a].

V.2. TA interactive : nouveaux développements

La TA interactive est un courant qui est né dans les années 60 [Kay, M. 1973, Kay, M. 1980]. Dans [Blanchon, H. 1994a], nous avons dressé un état de l'art très complet de ce courant de recherche jusqu'en 1994. Nous avons identifié trois classes de systèmes : les systèmes pionniers, les systèmes pour lesquels l'interaction est ajoutée a posteriori comme béquille et les systèmes pour lesquels l'interaction est intégrée comme ingrédient de base au départ. Nous faisons ici un bref rappel bibliographique sur les projets et systèmes étudiés, avant d'examiner des travaux plus récents.

V.2.1. Bref rappel historique

V.2.1.1. Expériences pionnières

MIND (1963-73)

Dans la fin des années 60 et le début des années 70, l'un des grands projets en linguistique informatique fut le projet MIND (Management of Information through Natural Discourse) à la Rand Corporation. Ses principaux maîtres d'œuvre furent Martin Kay et Ronald Kaplan. MIND [Hutchins, W. J. 1986, p 296, Kay, M. 1973] devait être une boîte à outils du TALN pour réaliser différents types de systèmes, depuis des testeurs de grammaires jusqu'à des systèmes de traduction automatique en passant par des systèmes de question-réponse. Un module de désambiguïsation interactive était prévu.

ITS (1973-81) et ALPS (1980-)

Le projet ITS visait le développement d'un système interactif pour traduire des textes en anglais vers le français, l'espagnol, le portugais, l'allemand et le chinois [Melby, A. K. 1978, Melby, A. K., *et al.* 1980]. En 1979, une évaluation menée par les commanditaires du projet a montré qu'ITS avait peu de possibilités commerciales [Melby, A. K. 1981]. À cette date, une partie de l'équipe créa cependant la société ALPS qui poursuivit le développement d'un système de traduction interactif [Bateman, R. 1985, Good, R. L. 1988, Sibley, J. 1988, Tenny, M. D. 1985, Weaver, A. 1988, Zirkle, A. T. 1990], l'équipe d'Alan Melby se réorientant vers des applications de type « poste de travail du traducteur humain » [Melby, A. K. 1982, Melby, A. K. 1987].

V.2.1.2. Désambiguïsation interactive comme béquille *a posteriori*

Dans KBMT-89 et LMT, la désambiguïsation interactive n'a pas été proposée *a priori* comme une composante des systèmes à développer, mais *a posteriori*, pour pallier les lacunes de ceux-ci.

KBMT-89 (1987-89)

L'objectif du projet, financé par IBM et conduit à l'Université Carnegie Mellon, était la réalisation d'une maquette d'un système de traduction fondé sur les connaissances pour traduire des manuels d'installation et de maintenance de micro-ordinateurs de l'anglais vers le japonais et inversement. Les spécifications furent développées par Jaime Carbonell et Masuru Tomita [Carbonell, J. G., *et al.* 1981, Carbonell, J. G. and Tomita, M. 1985, Carbonell, J. G. and Tomita, M. 1987]. La maquette finale, appelée KBMT-89 [Brown, R. D. 1989, Brown, R. D. and Nirenburg, S. 1990, Goodman, K. and Nirenburg, S. 1991, Nirenburg, S. 1989, Nirenburg, S. and al. 1989], fit l'objet d'une démonstration en février 1989.

LMT (1989-)

LMT (Logic-based Machine Translation) [McCord, M. 1989a, McCord, M. 1989b] est un système de traduction automatique expérimental développé par M. McCord au centre de recherche Thomas J. Watson d'IBM pour tester l'utilisation de la programmation logique en TALN. Entre 1989 et 1991, IBM a développé un environnement, LMT, pour engager des coopérations avec d'autres centres de recherche IBM en Allemagne, Espagne, Israël et au Danemark. Ces coopérations ont produit plusieurs résultats [Rimon, M., *et al.* 1991], dont un composant de désambiguïsation interactive [Ben-Ari, D., *et al.* 1988].

V.2.1.3. Désambiguïsation interactive comme ingrédient de base

DLT (1982-89)

Distributed Language Translation (DLT) fut un projet de Traduction Automatique conduit par la société de services informatiques Buro voor Systeemontwikkeling (BSO) aux Pays-Bas. Les études préliminaires commencèrent en 1979 sous la direction de Toon Witkam. Une étude de faisabilité fut soutenue par la Commission des Communautés Européennes entre 1982 et 1983 [Witkam, A. P. M. 1983]. En 1985, BSO obtint un contrat de 6 ans avec le ministère néerlandais de l'économie. L'objectif premier était la mise au point pour 1987 d'un prototype anglais-français produisant des traductions de qualité grâce à un processus de désambiguïsation interactive [Papagaaij, B. C. 1986, Sadler, V.

1989, Schubert, K. 1986] et l'obtention d'un système commercial pour 1993⁵⁸. L'objectif du projet DLT était de construire un système de traduction par pivot⁵⁹ entre les langues européennes (anglais, français, allemand, italien) [Schubert, K. 1988, Witkam, T. 1988].

N-Tran (1984-87)

N-Tran fut un projet, développé au Centre for Computational Linguistics de l'université de Manchester, et soutenu par International Computers Limited et le Science and Engineering Research Council dans le cadre du programme Alvey de recherche sur les systèmes intelligents fondés sur la connaissance (Intelligent Knowledge-Based Systems). Le projet, intitulé « lire et écrire en japonais sans connaître cette langue » fut financé pour une première période de deux ans d'octobre 1984 à octobre 1986 [Johnson, R. and Whitelock, P. 1985, Tsujii, J.-I. 1986], puis reconduit pour un an jusqu'en octobre 1987. Avec N-Tran [Whitelock, P. J., *et al.* 1986, Wood, M. M. 1989, Wood, M. M. G. and Chandler, B. 1988], on souhaitait permettre à un anglais monolingue, rédacteur de documents techniques, de produire simultanément des versions en anglais et en japonais. Un prototype a finalement été présenté en décembre 1987 [Chandler, B., *et al.* 1987].

À la suite de ces travaux, un bon nombre de chercheurs ont eu l'idée de concevoir des systèmes où des textes assez stéréotypés dans des domaines fixes peuvent être générés à l'aide d'une interaction avec l'utilisateur. Le système de l'UMIST [Jones, D. and Tsujii, J. 1990, Somers, H. and Jones, D. 1992, Somers, H. L., *et al.* 1990] qui générait de la correspondance commerciale en français, en allemand et en espagnol, le système de Saito et Tomita [Saito, H. and Tomita, M. 1986] qui produisait des textes en japonais, et le système de Zaki et Noor [Zaki, A. B. and Noor, H. 1991], permettait de composer des lettres en anglais et en malais. Enfin, il y a des produits commerciaux de ce type, comme Ambassador sur le Macintosh.

JETS (1985-)

Les travaux sur le système JETS [Maruyama, N., *et al.* 1988, Tsutsumi, T., *et al.* 1993] ont commencé en 1985 dans les laboratoires d'IBM à Tokyo. Le système a été testé en interne et chez des clients. JETS offre à l'utilisateur un environnement interactif d'analyse du japonais [Maruyama, H., *et al.* 1990]. Le système pose des questions pour lever des ambiguïtés lors des phases d'analyse morphologique, syntaxique et sémantique lorsqu'il ne peut résoudre seul les ambiguïtés. Les réalisateurs se sont concentrés sur la traduction de textes japonais simples, tels que les annotations dans les dessins techniques et les « notices de changement de conception » (design change notice).

ITS-2 (1990-1994)

Le système ITS-2 (Interactive Translation System 2) est un prototype de système interactif de traduction de phrases, développé au début des années 1990 [Laenzlinger, C. and Wehrli, É. 1991, Wehrli, É. 1990, Wehrli, É. 1992, Wehrli, É. 1993, Wehrli, É. 1994, Wehrli, É. and Ramluckun, M. 1993] au Laboratoire d'Analyse et de Technologie du Langage (LATL) de l'Université de Genève. Il accepte des phrases prises dans un sous-ensemble de la grammaire du français pour les traduire en anglais. Le projet poursuit plusieurs objectifs, aussi bien théoriques que pratiques. Il s'agit de montrer l'apport de l'approche interactive au problème de la traduction automatique, ainsi que les avantages d'un système basé sur le modèle grammatical chomskyen du Gouvernement et du Liage.

⁵⁸ À partir de 1989, les progrès ont été minimes, le projet ayant été "gelé" (communication personnelle de Toon Witkam).

⁵⁹ Une partie de l'espéranto muni de marques cachées (parenthèses spéciales)

V.2.2. KANT

Le projet KANT⁶⁰ (90-) s'inscrit dans le prolongement du projet KBMT. À l'origine [Mitamura, T. 1991, Nyberg, E. and Mitamura, T. 1992], le projet visait à concevoir et implémenter un système qui permette la rédaction en anglais contrôlé de documents techniques, produits par Caterpillar, et leur traduction, à la demande, vers plusieurs langues cibles sans post-édition ni intervention humaine pour la désambiguïsation. Il s'agit donc de traduire des documents appartenant à un sous-langage contraint par le domaine et la langue. Pour atteindre ces objectifs, l'architecture linguistique choisie est interlingue [Leavitt, J., *et al.* 1994]. Pour ce faire, le système aidera l'auteur à pré-éditer (normaliser) les documents à traduire (cf. *infra*).

À partir de 1994, les publications sur KANT réintroduisent explicitement la question des ambiguïtés non résolues par le système et proposent un module de désambiguïsation interactive [Baker, K., *et al.* 1994a, Baker, K., *et al.* 1994b]. Deux types d'ambiguïtés sont concernées, les ambiguïtés lexicales et les ambiguïtés structurales. Pour les ambiguïtés lexicales, l'auteur doit choisir le sens désiré. Pour les ambiguïtés structurales, les différentes structures sont explicitées et l'auteur doit choisir. Par exemple, pour la phrase "Fasten the assembly with the lever.", le dialogue suivant est proposé à l'auteur :

"What with modifies is ambiguous"
AUTHOR CHOICE: "fasten with"
AUTHOR CHOICE: "assembly with"

Pour les acteurs du projet, la désambiguïsation interactive n'est pas satisfaisante car elle diminue la productivité des auteurs de documents et, plus grave encore, ceux-ci font parfois de mauvais choix lors de la désambiguïsation. Dans un corpus choisi [Mitamura, T., *et al.* 1999], 29% des attachements prépositionnels ne sont pas résolus automatiquement. En utilisant deux types d'heuristiques (attachement minimal et patrons [id.]), les résultats obtenus avec le système fonctionnant en mode purement automatique ont été identiques aux résultats obtenus avec désambiguïsation interactive.

Méthode utilisée	Trad. correctes	Trad. incorrectes	Total
Désambiguïsation interactive	2442 (89%)	306 (11%)	2748 (100%)
Attachement minimal	2290 (83%)	458 (17%)	2748 (100%)
Attachement minimal + patrons	2449 (89%)	299 (11%)	2748 (100%)

Table V-1 : Comparaison des taux de succès en désambiguïsation automatique et en désambiguïsation interactive avec KANT

En 2000, [Nyberg, E. and Mitamura, T. 2000] ont réorganisé l'architecture informatique et réimplémenté les algorithmes du noyau de KANT avec KANTOO. Afin de diminuer la charge cognitive du rédacteur, [Mitamura, T., *et al.* 1999] ont aussi proposé et implémenté un module de réécriture du document source afin que celui-ci soit conforme. La nouvelle architecture proposée est décrite dans la Figure V-1.

L'étape de réécriture doit fournir un diagnostic [Mitamura, T., *et al.* 2003a, Mitamura, T., *et al.* 2003b] pertinent et compréhensible par l'auteur afin qu'il puisse intervenir de manière profitable. Nous proposons ci-dessous (Figure V-2 à Figure V-7) des exemples de boîtes de dialogues proposées à l'auteur par le système.

⁶⁰ Page d'accueil du projet : <http://www.lti.cs.cmu.edu/Research/Kant/>

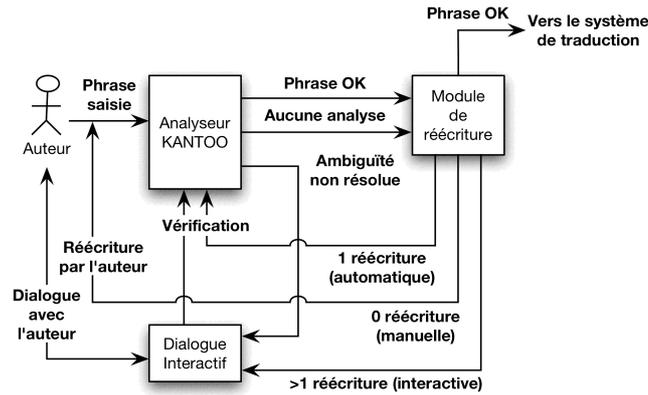


Figure V-1 : KANTOO et le module de réécriture

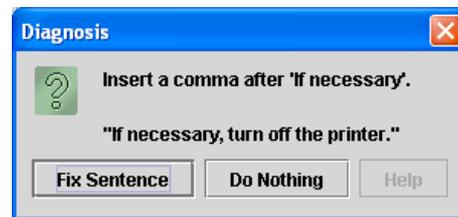
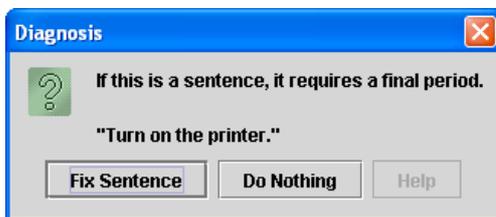


Figure V-2 : Réécriture de la ponctuation



Figure V-3 : Explication de la fonction syntaxique d'une proposition



Figure V-4 : Ajout d'un article manquant

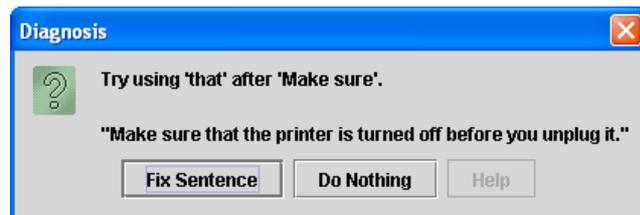


Figure V-5 : Explication de la subordination

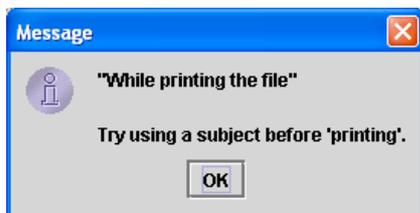


Figure V-6 : Réécriture des gérondifs

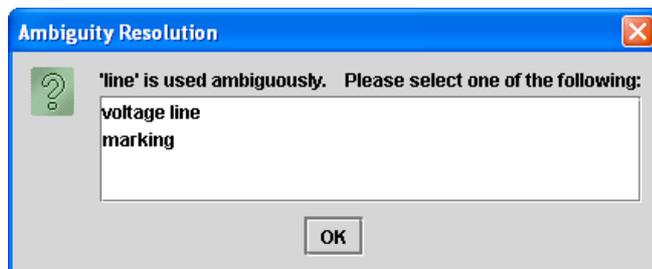


Figure V-7 : Résolution d'une ambiguïté de polysémie

V.2.3. TRANSTYPE

Le projet TRANSTYPE⁶¹ vise le développement d'un nouveau type d'outil interactif pour assister les traducteurs humains qui sont bilingues. Le concept proposé correspond à de la TA Interactive médiatisée par le texte cible [Foster, G., *et al.* 1997]. Le système envisagé observe l'utilisateur à mesure que celui-ci tape une traduction, tente constamment d'anticiper ses intentions, et propose des « suites » aux portions de texte déjà tapées [Langlais, P., *et al.* 2000a]. L'utilisateur est libre d'accepter ces propositions telles quelles, de les modifier ou simplement de les ignorer.

La contribution du système se fait à trois niveaux : 1) lorsque le traducteur a déjà choisi une formulation particulière, le système peut en accélérer substantiellement la frappe ; 2) lorsque le traducteur suspend la frappe pour rechercher une formulation adéquate, la machine peut lui offrir une suggestion ; et 3) lorsque le système détecte ce qu'il croit être une erreur de la part du traducteur, il peut émettre un avertissement.

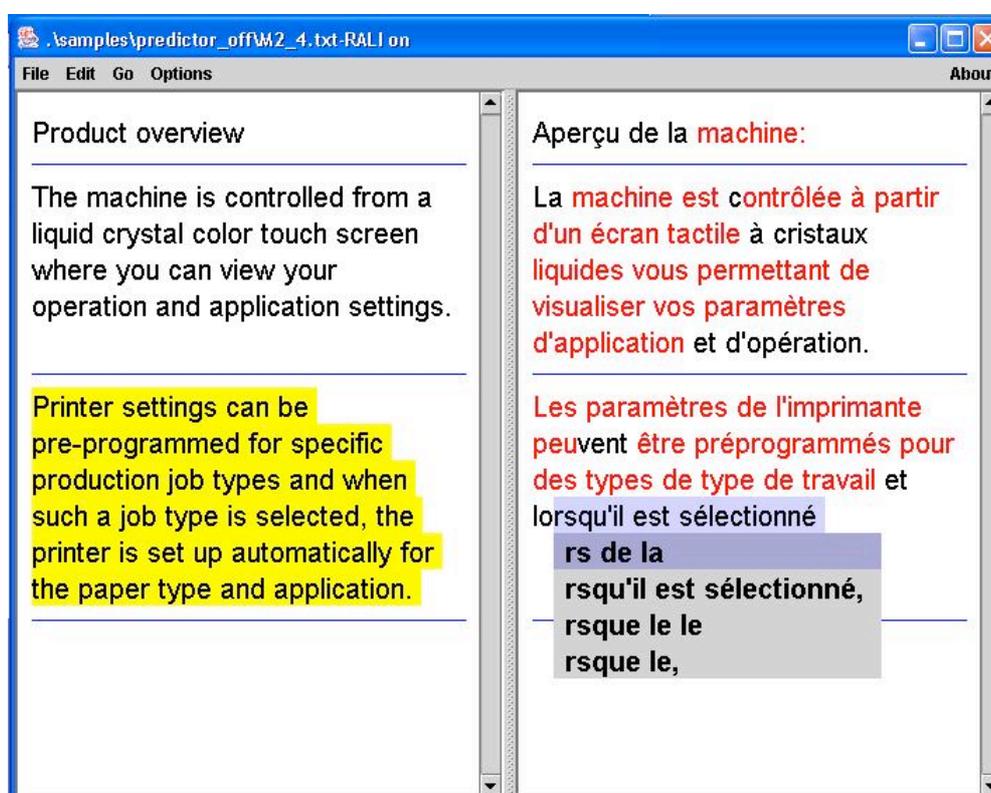


Figure V-8 : Capture d'écran de TransType en action

L'innovation de l'approche réside dans le mode d'interaction entre l'utilisateur et le système de traduction automatique (TA) sur lequel TRANSTYPE se base. Habituellement, les systèmes de TA interactive demandent l'assistance de l'utilisateur lors de l'analyse du texte source, de façon à améliorer la qualité de la traduction machine. Pour ses promoteurs, le texte cible constitue un meilleur site d'interaction que le texte source. Ils espèrent proposer un outil plus naturel et utile.

Le projet essaie de surmonter plusieurs défis technologiques. Premièrement, le système doit tenir compte non seulement du texte source, mais également de la portion déjà produite de la traduction. Il ne doit retenir que les propositions qui ont un degré de confiance suffisant. Son comportement doit donc s'adapter au contexte, et exploiter au

⁶¹ Page d'accueil du projet : <http://www-rali.iro.umontreal.ca/ProjetTransType.fr.html>

maximum la contribution de l'utilisateur. Le système doit également opérer en temps réel, c'est-à-dire qu'il doit être en mesure de « suivre » l'utilisateur. En outre, le système doit présenter ses propositions de façon claire à l'utilisateur et lui fournir les moyens d'exploiter efficacement celles-ci, sans pour autant opposer d'obstacles à l'exercice de la frappe.

Les résultats initiaux sont des plus prometteurs : le programme fonctionne en temps réel et il pourrait épargner à l'utilisateur près de 60% des frappes [Langlais, P., *et al.* 2001, Langlais, P., *et al.* 2000b]. En raison de la nature des modèles utilisés, il est possible d'envisager le développement rapide d'extensions à d'autres paires de langues. C'est ce dernier objectif que vise une nouvelle phase du projet TRANSTYPE2⁶² [Esteban, J., *et al.* 2004, Valderrábanos, A., *et al.* 2003] lancé en mars 2002 pour une durée de 36 mois. Six couples de langues sont proposés depuis l'anglais vers le français, l'espagnol et l'allemand ; et vice versa. La modalité d'entrée sera l'écrit ou l'oral.

En adoptant presque la même idée, NEC [Yamabana, K., *et al.* 1997] avaient proposé une idée similaire pour aider des Japonais à rédiger des documents anglais. Dans une première étape, le système produit un texte mixte qui contient une traduction des mots pleins en anglais, alors que les mots outils demeurent inchangés (b). L'utilisateur peut alors sélectionner un mot pour voir, dans un menu, les autres traductions possibles et en choisir une meilleure si nécessaire. Il peut aussi, grâce à un autre menu, changer la flexion du mot proposé. Lors de cette étape, le système prédit le segment du texte qui doit être traduit. L'utilisateur peut alors appeler le système de traduction qui produit une phrase en anglais (c).

- a) 私は彼に論文を渡した
- b) I は he に paper を give た
- c) I gave him a paper

Figure V-9 : Étapes de traduction interactive, pour un traducteur non professionnel, proposée par NEC

V.3. Contribution pour aller plus loin : LIDIA

Pour aller plus loin, le GETA a proposé le concept de Traduction Automatisée Fondée sur le Dialogue (TAFD) pour auteur monolingue [Boitet, C. 1990b]. En 1994, dans [Boitet, C. and Blanchon, H. 1994], nous précisons les opportunités et nécessités de la TAFD⁶³.

V.3.1. Motivations

Les recherches et développements en TAFD sont motivés par la limitation des paradigmes actuels, par l'importance croissante des langues nationales dans le contexte de l'internationalisation, et par de récents progrès méthodologiques et technologiques.

⁶² TransType2 est un projet cofinancé par l'UE dans le cadre du 5^{ème} PCRD et par le Canada. La page d'accueil du projet se trouve à : <http://tt2.sema.es/Index.htm>.

⁶³ Le texte de cette section est une adaptation d'un extrait de [Boitet, C. and Blanchon, H. 1994].

V.3.1.1. Limitation des paradigmes actuels

La TAFL est très bien adaptée à la TAO du veilleur. Des systèmes portables de qualité tout à fait convenable pour l'usage visé commencent à se répandre au Japon, à des prix abordables (environ 5 M¥ pour un système logiciel et matériel complet⁶⁴).

Par contre, elle est loin de pouvoir répondre à tous les besoins en TAO du réviseur. Outre le fait qu'elle demande évidemment autant de révisions que de langues cibles, elle reste trop chère pour des usages légers. En effet, selon les chiffres donnés par les producteurs de systèmes de TA [JEIDA 1989], la création *ex nihilo* d'un système opérationnel de TAFL coûte entre 200 et 300 hommes-années, avec des développeurs spécialisés, et l'adaptation d'un système de TAFL existant à un nouveau domaine et à une nouvelle typologie de textes coûte de l'ordre de 5 à 10 hommes-années. Un utilisateur n'a intérêt à s'équiper d'un tel système que s'il doit traduire de gros flux de textes homogènes et informatisés, comme des manuels d'utilisation ou de maintenance⁶⁵. Adapter un système disponible à des besoins ponctuels n'est pas non plus une solution viable⁶⁶.

D'autre part, une condition essentielle de succès de la TAO du réviseur est de constituer une équipe de développement et de maintenance des linguiciels (dictionnaires, grammaires) qui soit en liaison constante avec l'équipe de révision, et si possible avec les auteurs des documents à traduire⁶⁷. C'est ce qu'a réussi la PAHO (Pan American Health Association), avec ses systèmes ENGSPAN et SPANAM [Vasconcellos, M. and León, M. 1988]. On peut faire un parallèle avec les systèmes experts, qui peuvent être développés par des tierces parties, mais qui doivent ensuite être totalement maîtrisés par leurs utilisateurs, seuls à même de les faire évoluer de façon adéquate.

En ce qui concerne la T AFC [Carbonell, J. G. and Tomita, M. 1985, Nirenburg, S. 1989, Nyberg, E. and Mitamura, T. 1992], elle est totalement inapplicable en TAO du veilleur, et elle est moins applicable que la TAFL en TAO du réviseur car, tant qu'il faudra construire les "ontologies" spécialement pour la traduction, elle restera beaucoup plus onéreuse, et ce pour un résultat guère meilleur.

V.3.1.2. Importance croissante des langues nationales et de l'internationalisation

De plus en plus, nous désirons rédiger dans notre langue, et transmettre nos textes à l'étranger, qu'il s'agisse de messages électroniques, de lettres, d'articles, de manuels techniques, voire de livres. Contrairement à ce que d'aucuns prédisaient il y a une

⁶⁴ Station Sparc, lecteur optique, système avec 80.000 termes, dictionnaire utilisateur, et options de personnalisation.

⁶⁵ En prenant l'hypothèse d'un système coûtant 1 MF (400 KF de base et 600 KF de spécialisation au vocabulaire et au type de texte) et d'un amortissement sur 2 ans, il faut un flux de 10.000 p/an (en comptant 10%/an de maintenance, 60 F/page de coût machine, et 100 F/page de révision, contre 150 F/page de traduction et 70 F/page de révision pour la méthode manuelle, soit 60 F/page de gain pour amortir 1,2 MF). À coût machine nul, il faudrait encore 5.000 p/an. Une autre façon de comparer est de dire qu'il faut de 5 à 10 hommes-année pour spécialiser un système, et 8 pour traduire et réviser 10.000 pages (par exemple, on peut le faire en un an avec 6 traducteurs et 2 réviseurs travaillant 1700 h/an).

⁶⁶ Ce serait comme réoutiller une usine pour produire quelques dizaines de voitures. En effet, sans compter la saisie optique ou manuelle, entraînant toujours un coût important de vérification, ni la maintenance, ni même l'achat du système de base, mais seulement sa spécialisation (600 KF) et les coûts de traduction et de révision, on arrive avec les hypothèses précédentes à 632, 680, 760 et 920 KF pour 200, 500, 1.000 et 2.000 pages, contre 44, 110, 220 et 440 KF pour la méthode classique manuelle, soit environ 14,5, 6, 3,5 et 2 fois plus, respectivement (chiffres de 1991).

⁶⁷ Dans le "contre-rapport ALPAC" du JEIDA [JEIDA 1989] comme au MTS-II à Munich en août 1989, Fujitsu reconnaissait clairement avoir fait une erreur en distribuant largement ATLAS-II : seules étaient en effet rentables les traductions effectuées chez Fujitsu, soit pour sa documentation, soit dans le cadre d'un contrat avec la CEE (Japinfo) concernant la veille technologique et ne demandant donc qu'une révision minimale.

cinquante d'années, l'internationalisation croissante ne s'est pas accompagnée d'une uniformisation linguistique vers l'anglais, mais au contraire d'un renforcement considérable de l'usage scientifique et technique de langues déjà importantes de ce point de vue, comme le japonais, et d'une promotion volontariste de bien d'autres, comme le malais-indonésien ou l'arabe, pour les amener au même niveau. À notre sens, et même si l'anglais continue de se renforcer en tant que *volapük* scientifique moderne, cette évolution ne fera que se renforcer, les langues étant, pour reprendre une expression de C. Hagège dans un article paru dans *Le Monde* début 1990, les "drapeaux des identités nationales".

Il ne s'agit pas seulement de politique, mais d'efficacité. Dans les projets coopératifs européens (Esprit, Eureka), par exemple, la communication est gênée par la nécessité de lire et d'écrire en anglais. Pour la grande majorité des participants, lire en anglais pose des problèmes de compréhension et prend beaucoup de temps. Quant à écrire, si même c'est envisageable, le résultat est souvent difficile à comprendre, voire illisible. De grandes sociétés comme Thomson chiffrent à plusieurs dizaines de MF par an les pertes dues aux problèmes de communication multilingue.

Les trois types de TAO "classique" ne peuvent évidemment répondre à ce nouveau besoin. En effet, la TAO du veilleur, sans préédition ni postédition, ne peut donner une qualité suffisante, et la TAO du réviseur comme la TAO du traducteur s'adressent par définition à des spécialistes au moins bilingues, et non à des rédacteurs supposés ne connaître aucune des langues cibles, ou au plus une, et ce imparfaitement.

V.3.1.3. Progrès méthodologiques et technologiques

L'idée de la TAFD date des années soixante [Kay, M. 1973], et a été incorporée à plusieurs maquettes ou prototypes dans les années soixante-dix et quatre-vingt à l'université Brigham Young (Povo, Utha) [Balkan, L. 1992, Melby, A. K. 1982, Melby, A. K. 1991, Melby, A. K., *et al.* 1980, Sibley, J. 1988, Tenny, M. D. 1985, Weaver, A. 1988], à l'université Carnegie Mellon (Pittsburgh, Pensylvanie) [Carbonell, J. G. and Tomita, M. 1985, Saito, H. and Tomita, M. 1986, Tomita, M. 1984, Tomita, M. 1985, Tomita, M. 1986], et dans les Universités de Manchester et Sheffield (Angleterre) [Chandler, B., *et al.* 1987, Johnson, R. and Whitelock, P. 1985, Wood, M. M. 1989, Wood, M. M. G. and Chandler, B. 1988]. Si ces travaux n'ont pas donné lieu à des systèmes utilisables en pratique, c'est à notre avis que les dialogues devaient être conduits par des spécialistes⁶⁸, que la couverture linguistique était trop limitée, et que l'on ne disposait pas encore d'environnements interactifs conviviaux.

La méthodologie s'est affinée ces dernières années. Tout d'abord, l'utilisateur envisagé n'est plus un spécialiste, mais un rédacteur, ou plutôt un *auteur* [Huang, X. 1990a, Huang, X. 1990b, Maruyama, H., *et al.* 1990, Maruyama, N., *et al.* 1988, Sadler, V. 1989, Schubert, K. 1988, Somers, H. and Jones, D. 1992, Somers, H. L., *et al.* 1990, Tsutsumi, T., *et al.* 1993, Wehrli, É. 1991, Wehrli, É. 1992, Wehrli, É. 1993, Witkam, A. P. M. 1983, Witkam, T. 1988]. Nous préférons ce dernier terme. D'un côté, en effet, « auteur » est moins restrictif que « rédacteur » : un auteur est quelqu'un qui veut créer un texte, et peut le faire en l'écrivant, en le dictant, ou encore en le construisant interactivement. De l'autre, « auteur » est plus restrictif que « rédacteur », « locuteur », ou « commentateur », car « auteur » désigne quelqu'un qui désire créer un produit final "propre", alors que les autres termes peuvent renvoyer à des personnes désirant seulement produire un message écrit ou parlé de façon

⁶⁸ ITS [Melby, A. K., *et al.* 1980] demandait même *plusieurs* intervenants, un linguiste spécialiste du système pour l'analyse, et un bilingue pour chaque langue cible.

« spontanée », en vue d'une communication immédiate, et non disposées à conduire un dialogue éventuellement lourd pour rendre leur message « propre »⁶⁹.

D'autre part, l'informatique personnelle a fait des progrès gigantesques. On dispose maintenant d'ordinateurs personnels très puissants et bon marché, d'environnements conviviaux, de l'intégration du multimédia, et d'outils de télécommunication permettant le recours à des serveurs. Enfin, les techniques et outils de génie logiciel modernes (essentiellement la programmation par objets) permettent de construire des systèmes complexes et interactifs bien plus rapidement et sûrement que par le passé.

V.3.2. Critères de choix de l'approche par dialogue

Nous proposons quatre critères pour choisir la TAFD :

- la qualité visée doit être élevée, et la révision impossible ou très coûteuse ;
- le contexte doit être fortement multilingue (1→n, comme pour la dissémination de documentation technique, ou n↔n, comme dans des projets internationaux) ;
- l'entrée ou le domaine ne doivent pas être trop contraints ou contrôlés (sinon, mieux vaut utiliser la TAFL ou la TAFC) ;
- les utilisateurs doivent être prêts à participer à des dialogues de normalisation et de désambiguïsation.

Dans toute situation, il faut de plus pouvoir rendre les dialogues acceptables, en les laissant à l'initiative de l'utilisateur, en lui fournissant des moyens de les contrôler et de les réduire (en jouant sur des paramètres, en insérant directement des marques de désambiguïsation, etc.), et en lui laissant si possible le choix entre plusieurs média.

V.3.3. Situations traductionnelles adaptées à la TAFD

V.3.3.1. Entrée écrite

Parmi les situations favorables avec entrée écrite, on peut mentionner :

- la traduction de volumes relativement faibles de documentation technique en plusieurs langues, typiquement 5.000 à 8.000 pages à distribuer sur un Disque Optique Numérique, par exemple dans les 9 langues officielles de la CEE/UE (et peut-être dans d'autres aussi, comme le russe, l'arabe, le japonais, ou le chinois).
- la diffusion d'information dans plusieurs langues (sur la circulation, sur la météo, ou dans des congrès, des manifestations sportives, des situations d'urgence...), qui demande une sortie orale aussi bien qu'écrite ;
- l'échange télématique de notes et de documents de travail dans des projets internationaux.

V.3.3.2. Entrée orale

Comme l'état de l'art en reconnaissance de parole ne permet pas de traiter à la fois un grand vocabulaire, de la parole continue, et un locuteur arbitraire, il semble n'y avoir que très peu de situations favorables à la TAFD avec entrée orale :

⁶⁹ « Propre » signifie ici conforme à une certaine grammaire (permettant éventuellement des constructions incorrectes, pourvu qu'elles soient attestées) et ne comportant pas de parties « réflexives », ou « automodificatrices », si fréquentes dans la parole et même l'écriture spontanée (au moins manuscrite), comme des hésitations, des faux départs, des reprises, des répétitions, des corrections, des abréviations arbitraires (apocopes), etc.

- la production de commentaires ou de résumés à partir de scènes visuelles et auditives (par exemple, pour le sous-titrage d'émissions de télévision en langues étrangères) ;
- l'interprétation de dialogues bilingues très contraints, tels que les appels téléphoniques de politesse entre parents d'enfants échangés entre familles pour apprendre les langues, ou l'assistance téléphonique à des voyageurs étrangers (consultation médicale, réservation...). Ici, le dialogue doit être le plus réduit possible, et une combinaison entre TAFD et TAFC (analogue à l'architecture finale de KBMT-89 [Nirenburg, S. and al. 1989]) semblerait indiquée.
- la production de présentations vidéo multilingues.

V.3.3.3. Création dans une langue d'un message dans une autre

Il y a enfin beaucoup de situations intéressantes où le message source n'est créé que pour vérifier le contenu du (ou des) message cible. Le message est créé par interaction avec le système. C'est par exemple le cas de lettres officielles ou formelles, qui ont des structures très différentes dans différentes cultures. Somers & Tsujii [Somers, H. L., *et al.* 1990] ont ici proposé le terme de « traduction sans texte source », mais il serait plus exact de parler de génération interactive multilingue que de traduction.

V.3.4. Situation dans la classification proposée

La TAFD vise donc à produire des traductions de qualité en domaine ouvert grâce à une forte interaction entre le système et l'auteur du document à traduire.

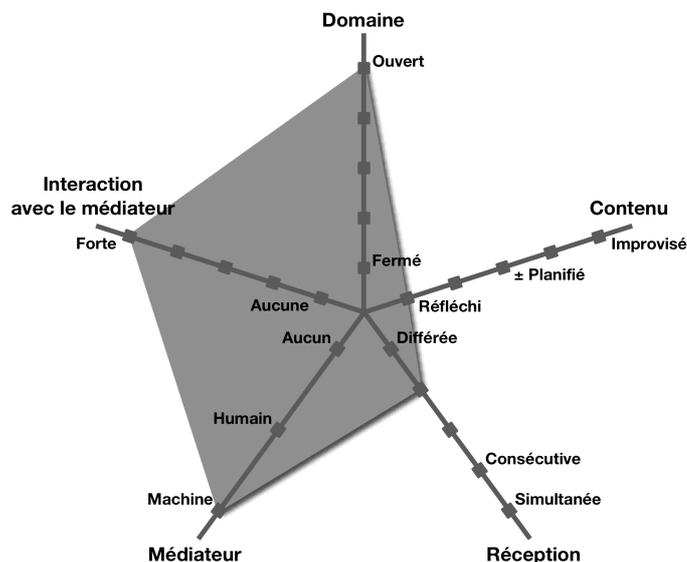


Figure V-10 : Caractéristiques de la TAFD

Chapitre VI

LIDIA-1 (89-95)

Les maquettes LIDIA-1 — il y en a eu deux — sont les fruits de nos premiers efforts vers la TAFD « pour tous ». Nous présentons ici les grandes lignes qui ont influencé leur conception⁷⁰. Le processus de désambiguïsation interactive sera développé dans le chapitre suivant.

VI.1. Architecture linguistique et voies intermédiaires nouvelles⁷¹

Le fait que la TAFD suppose un dialogue avec l'auteur permet d'envisager de nouvelles possibilités au niveau des traitements linguistiques. Il ne s'agit pas de solutions radicalement nouvelles, mais, comme souvent dans un domaine technique, de nouveaux compromis, de voies intermédiaires nouvelles, avec ça et là des innovations intéressantes.

Dans la majorité des situations adaptées à la TAFD, il faut un système de couverture lexicale et grammaticale très large. D'où des questions théoriques importantes :

- Sachant qu'on n'obtient de bons résultats que sur des langages restreints, comment construire une base de connaissances linguistiques utilisable comme une union de sous-langages ? Est-il possible de séparer les aspects grammaticaux et lexicaux ? Nathalie Denos, lors de son stage DEA [Denos, N. 1993] a commencé à aborder cette question.
- *Comment atteindre la large couverture lexicale nécessaire ?* Typiquement, un système de TAFL contient de 3.10^4 à 3.10^5 termes, en 2 langues. Le cas de METEO (3.10^3 termes) est atypique, à cause de son domaine très restreint [Chandioux, J. 1988]. Mais un système de TAFD visant le grand public et non restreint à un domaine particulier demandera de 3.10^5 à 3.10^6 termes, et ce en plusieurs langues !
- Dans des situations fortement multilingues, l'approche par interlingua est séduisante. Mais, *comment surmonter les difficultés d'ingénierie rencontrées dans la construction d'un grand lexique interlingue* par les récents projets japonais ATLAS, PIVOT, EDR, et CICC ?
- Il est crucial que des non-spécialistes puissent facilement comprendre les questions du système, éventuellement lui demander les raisons de certaines questions, et comprendre ses réponses. Une question importante (et nouvelle) est donc de trouver

⁷⁰ Pour une description complète des maquettes LIDIA-1, le lecteur peut consulter [Blanchon, H. 1994a] pour la première maquette et [Boitet, C., et al. 1995] pour la seconde.

⁷¹ Le texte qui suit est une adaptation d'un extrait de [Boitet, C. and Blanchon, H. 1994].

comment rendre la base de connaissances linguistiques d'un système de TAFD accessible à un utilisateur naïf.

Bien que les réponses proposées ici ne soient ni complètes ni définitives, elles sont basées sur une longue pratique de la TAFL et sur une étude approfondie de beaucoup de systèmes de TA passés ou présents, en particulier ceux qui utilisent une étape de désambiguïsation interactive.

VI.1.1. Transfert multiniveau à acceptions, propriétés et relations interlingues

Les systèmes modernes de TAO sont fondés sur un *transfert sémantique*, le passage par un *interlingua*, ou un *transfert multiniveau*. « Transfert sémantique » est un terme introduit par les Japonais pour désigner exactement l'approche du CETA entre 1960 et 1970, que B. Vauquois, reprenant un terme dû à Shaumjan, appelait approche par « pivot hybride » : la structure interface source fournie au transfert contient des éléments lexicaux propres à la langue source, et des propriétés et relations interlingues (traits sémantiques, cas profonds, traits abstraits de nombre, aspect, modalité, détermination...). Dans un véritable « interlingua », les éléments lexicaux doivent eux aussi être interlingues (« concepts » s'il y a une véritable ontologie, « acceptions » sinon).

Le *transfert multiniveau*, au sens de Vauquois [Vauquois, B. 1988, Vauquois, B. and Boitet, C. 1985] diffère du transfert sémantique en ce que, outre les attributs et relations interlingues, on garde des attributs et des relations spécifiques à la langue source (classe syntagmatique, genre, nombre, détermination, temps, mode, fonction syntaxique...). Cela donne des « filets de sécurité », et permet de faire interagir les niveaux. La Figure VI-1 présente un exemple de structure multiniveau.

En TAFD, nous proposons de rajouter aux représentations multiniveaux un niveau lexical, celui des *acceptions interlingues*, sans aller jusqu'à introduire des concepts, puisqu'il faudrait alors construire une ontologie. La base lexicale multilingue sous-jacente (BDLM) contiendra alors un dictionnaire monolingue pour chaque langue traitée par le système, et un dictionnaire interlingue pour les acceptions interlingues. Chaque acception interlingue a une image dans chaque dictionnaire monolingue, avec une définition appropriée dans la langue correspondante, utilisée lors de la désambiguïsation interactive du sens⁷².

Qu'on nous permette d'insister sur le fait que le terme « interlingue » ne signifie pas pour nous « indépendant de toutes les langues », mais « intermédiaire entre les langues connues du système ». Par exemple, si le système travaille avec le français, l'anglais et le russe, il y aura une seule acception pour « mur » en tant qu'objet concret. Dès qu'on ajoutera l'allemand ou l'italien, il faudra ajouter les raffinements « mur vu de l'extérieur » (Mauer, muro) et « mur vu de l'intérieur » (Wand, parete).

VI.1.2. Encodage d'informations linguistiques pouvant provenir d'une prédiction

Les syntagmes figés spéciaux sont transformés en des occurrences spéciales, de façon à aider l'analyse et la traduction. Par exemple, « Cacher les bulles » donnera

⁷² On peut aussi *expliquer* à l'auteur pourquoi une telle question est posée, et même montrer les mots en question dans les autres langues. L'introduction d'aspects d'auto-apprentissage dans ce genre de systèmes les rendrait plus acceptables par les utilisateurs potentiels.

&FXN_Cacher_les_bulles, qui pourra être traité par une sous-grammaire morphologique appropriée.

Après la désambiguïstation lexicale, nous pourrions attacher à chaque occurrence le numéro du sens dans la BDLM, par exemple **glace.1** pour « glace à manger » et **glace.2** pour « miroir ». Mais cela n'est pas très lisible et interdit la prédiction directe. Nous permettons donc d'ajouter au numéro de sens un fragment de la définition qui donne la distinction, la « clé sémantique », d'habitude un autre mot ou terme, ce qui donne **glace.1=aliment** ou **glace.2=miroir**. Comme en général plus d'une clé sémantique peut être associée à une même acception, par exemple **glace.1=a_manger** ou **glace.1=dessert**, il faut aussi permettre à l'utilisateur d'entrer **glace=dessert**. Dans ce cas, le système devrait consulter la BDLM, trouver que l'acception de « glace » la plus proche de « dessert » est le sens 1, et transformer **glace=dessert** en **glace.1=dessert** ou même en **glace.1=aliment**.

Dans la maquette LIDIA, les annotations concernant les informations grammaticales relatives aux mots, comme la catégorie morphosyntaxique (verbe, nom...), le nombre, le genre, le temps, le mode, etc., sont attachées aux occurrences de façon analogue.

Le dernier type d'annotation concerne les structures (arbres) produites par l'analyseur. Pour délimiter les groupes syntagmatiques, on peut utiliser des parenthèses spéciales, comme **{&rel...}** pour une proposition relative⁷³, ou simplement **{...}** si la catégorie syntagmatique n'est pas connue ou trop ésotérique pour les utilisateurs naïfs (par exemple, "groupe adjectival" ou "groupe cardinal"). Pour représenter un lien anaphorique, on pourrait ajouter au pronom une copie de son référent. En cas d'élision, on rajouterait des occurrences « cachées » (**centrale .&eld=inertielle**).

D'autres informations grammaticales et sémantiques peuvent être attachées aux terminaux et aux non-terminaux. Par exemple, « Devant cette somme, il ne rend pas sa glace. » pourrait avoir la représentation intermédiaire suivante (avant d'avoir fini la désambiguïstation) :

```
{ {&grd,cause *devant.&vrb cette somme.2&nf , } il ne rend.2 pas=ne sa glace.1 }
```

qui lèverait l'ambiguïté sur « devant » (verbe/préposition) et « rendre » (vomir/restituer/faire devenir).

Le point principal est que la vue du système d'annotations dont dispose l'auteur concerne tous les niveaux de description linguistique, mais doit être incomplète à chaque niveau, parce qu'aucune notion non familière ne doit apparaître. Par exemple, « verbe » est une notion familière pour presque tout adulte instruit, mais pas « verbe modal ». Au niveau des fonctions syntaxiques, « sujet », « objet » et « complément » sont familiers, mais sans doute pas « attribut », « épithète », « tête » (ou « gouverneur »). Il en va de même au niveau des cas profonds. Prenons par exemple les trois phrases suivantes.

```
Un processus de traduction en ARIANE-G5 se compose d'une suite de trois étapes
(analyse, transfert et génération). Chaque étape est constituée d'une suite de différentes
phases de traitement. Chaque phase est relative à l'emploi d'un LSPL précis.
```

Voici une vue simplifiée de la structure arborescente obtenue après la prédiction optionnelle, l'analyse et la désambiguïstation interactive (le texte des trois phrases est en gras afin de faciliter la lecture). On verra plus loin une vue légèrement différente.

```
{ { Un.&art processus.&n,suj { de.&prep traduction.&n,comp { en.&prep ARIANE-
G5.&np,comp } } se.&refl compose.&v,phvb { d' .&prep une.&art suite.&n,obj1 {
de.&prep trois.&card étapes.&n,comp { (.&lp analyse.&n,app ,.&ponc
transfert.&n,coord et.&cjcoord génération.&n,coord ) .&rp } } } ..&ponc } { {
```

⁷³ Ou bien on ajoute "rel" à l'information attachée à sa tête, comme dans l'exemple suivant ("compose.&v,phvb").

Chaque.&art étape.&n,suj } est.&v,aux constituée.&v,phvb { d'&prep une.&art suite.&n,comp { de.&prep { différentes.&adj,epit } phases.&n,comp { de.&prep traitement.&n,comp } } ..&ponc } { { Chaque.&art phase.&n,suj } est.&v,phvb { relative.&adj,atsubj { à.&prep l'&art emploi.&n,obj1 { d'&prep un.&art LSPL.&np,comp { précis.&adj,epit } } } } ..&ponc }

Ces questions de codage interne ne sont pas secondaires, comme on le pense trop souvent. Elles sont importantes en TAO classique, et encore plus en TAFD, puisqu'il faut qu'un utilisateur non-spécialiste puisse y accéder (éventuellement à travers une vue simplifiée). C'est d'ailleurs un défi pour les linguistes d'organiser les informations linguistiques pour permettre de telles vues.

VI.1.3. Étapes du processus de traduction

Une phrase du texte en langue source est d'abord analysée pour produire une structure *mmc-source* (multisolution, multiniveau⁷⁴, concrète⁷⁵). Cette structure *mmc* est alors utilisée pour construire un arbre des questions qui seront posées à l'auteur.

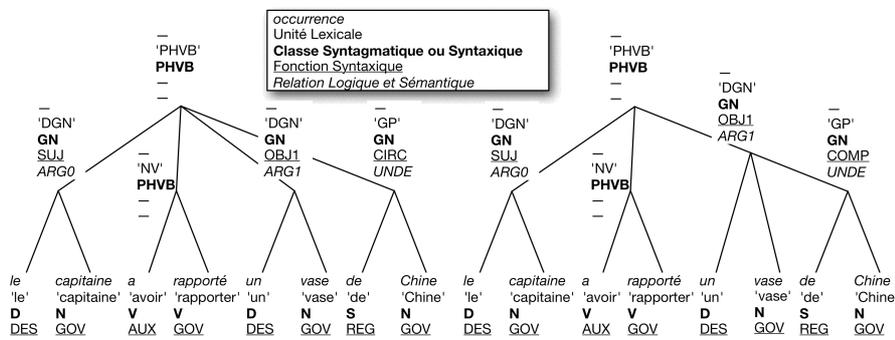


Figure VI-1 : Exemple de structure mmc

À l'issue de l'étape de désambiguïsation interactive, le système obtient la structure *umc-source* (unisolution, multiniveau, concrète) non ambiguë choisie par l'auteur. Cette structure *umc* est ensuite abstraite en une structure *uma-source* (unisolution, multiniveau, abstraite⁷⁵).

Un composant de transfert lexical et structural produit maintenant une structure *gma-cible* (génératrice, multiniveau, abstraite). Une structure *gma* est plus générale et génératrice qu'une structure *uma* car les niveaux de surface (fonctions syntaxiques, catégories syntagmatiques ...) peuvent ne pas être renseignés. Dans ce cas, ce sont des préférences du transfert qui les instancieront. On peut faire la même chose en retournant vers le français par rétro-traduction.

L'étape de sélection de paraphrase produit une structure *uma-cible* qui est homogène à la structure qui serait produite en analysant puis en désambiguïsant interactivement le texte cible qui va être généré. Le processus de traduction se termine avec les générations syntaxique et morphologique.

Après l'étape de sélection de paraphrase dans une langue cible, nous avons aussi implémenté une rétro-traduction vers le français.

La Figure VI-2 montre un diagramme fonctionnel des processus que nous venons de décrire.

⁷⁴ La structure contient trois niveaux d'interprétation linguistique : le niveau des classes syntaxiques et syntagmatiques, le niveau des fonctions syntaxiques, et le niveau des relations logiques et sémantiques.

⁷⁵ Une représentation « concrète » d'un texte est telle qu'on retrouve le texte représenté grâce à un parcours canonique de la structure (mot des feuilles pour un arbre de constituants, parcours infixé pour un arbre de dépendances). Sinon, la structure est dite « abstraite ».

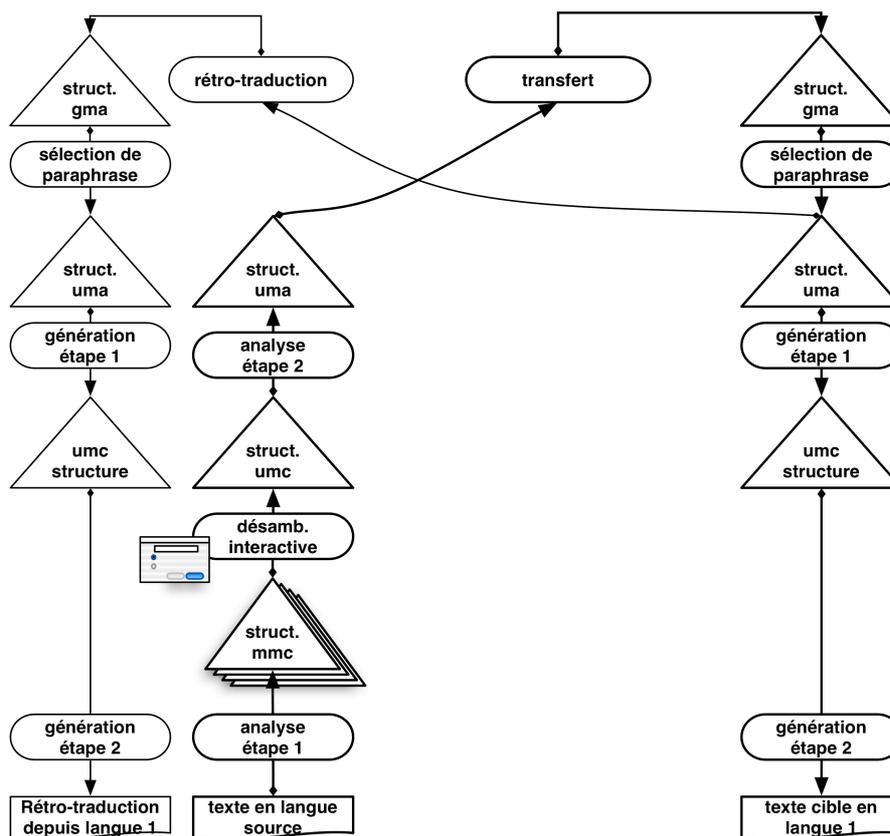


Figure VI-2 : Architecture linguicielle de LIDIA

VI.2. Architectures informatiques

L'architecture informatique d'un système est intimement liée à des contraintes ergonomiques que l'on se fixe pour celui-ci.

VI.2.1. Aspects ergonomiques

Les aspects ergonomiques sont bien sûr importants en TAO classique. En TAO du veilleur ou du réviseur, il y a ainsi une notion de qualité *apparente*, liée à la présentation. La même traduction paraît bien meilleure si elle est formatée comme un texte que si elle est présentée phrase par phrase. De même, un réviseur accepte plus volontiers des choix portant sur de petits groupes de mots que sur des phrases complètes. Les constructeurs de systèmes ont aussi beaucoup travaillé sur les éditeurs bilingues, sur les outils de paramétrage, et sur les aides à la création de dictionnaires.

En TAFD, l'ergonomie est un aspect encore plus crucial, et les choix ergonomiques influencent directement l'architecture de tout le système. Les choix principaux sont les suivants :

- Le système doit-il tourner sur des ordinateurs personnels bon marché ?
- Doit-il fonctionner en temps réel, ou l'asynchronisme est-il préférable ?
- Une architecture avec serveur(s) est-elle possible ?
- Comment les dialogues doivent-ils être organisés ? Est-il nécessaire et/ou possible de les conduire dans un environnement multimédia et multimodal ? Par exemple,

peut-on améliorer l'efficacité et la convivialité des dialogues de désambiguïsation grâce à l'utilisation de synthèse de parole et d'interactions graphiques ?

Si l'on vise vraiment la "TAO pour tous", il est impératif de pouvoir utiliser les systèmes de TAFD sur des ordinateurs personnels, ou sur des stations de travail de bas de gamme. Pour les autres questions, nous répondons : asynchronisme, distribution, interaction non-préemptive.

VI.2.1.1. Asynchronisme

Nous avons choisi une organisation asynchrone, analogue à celle de CRITIQUE [Richardson, S. and Braden-Harder, L. 1993], pour des raisons ergonomiques et pratiques. D'abord, une organisation en temps réel est nécessaire seulement si nous souhaitons que l'utilisateur réponde immédiatement aux questions posées par le système, ce que nous ne voulons pas. Ensuite, il serait contre-productif de poser au rédacteur des questions « à la volée » sur un texte non achevé. Enfin, si l'on prend en compte le degré de complexité inhérent à n'importe quel système de TAFD général et de bonne qualité, nous ne pouvons espérer ni une exécution en temps réel, ni une implantation des ressources nécessaires (base lexicale, linguiciel) sur le genre de matériel envisagé.

VI.2.1.2. Architecture distribuée

Une exécution en temps réel n'est pas non plus possible dans le cadre de l'architecture distribuée que nous préconisons, car (a) l'utilisation d'un serveur puissant pour prendre en charge tout ou partie des traitements non-interactifs de traduction est préférable au stockage de gros systèmes sur chaque station de travail, même si l'on envisage une exécution en tâche de fond ; et (b) on peut ainsi mieux résoudre les problèmes de maintenance des ressources linguistiques. Les utilisateurs peuvent alors bénéficier de manière transparente de mises à jour des serveurs d'analyse, de désambiguïsation, de transfert et de génération utilisés.

VI.2.1.3. Désambiguïsation interactive multimodale non-préemptive

Il nous paraît essentiel de permettre à l'auteur de décider quand il souhaite entamer le dialogue avec le système. En d'autres termes, le système propose et l'utilisateur dispose. En effet, les précédents essais en TAFD nous semblent avoir échoué en partie à cause du caractère modal des dialogues de désambiguïsation.

VI.2.2. Différentes implémentations

La situation que nous avons considérée est la production de documents multilingues sous la forme de documents HyperCard. HyperCard est un environnement de production de documents hypertextes, appelés « piles », dont les pages sont des « cartes ». Les cartes contiennent différents types d'objets, dont des champs textuels. Les maquettes LIDIA-1 permettent de traduire vers l'anglais, l'allemand, et le russe une pile en français qui présente, en contexte, des phrases comportant les ambiguïtés que nous avons choisi de traiter.

VI.2.2.1. LIDIA-1.3270

La première version de la maquette LIDIA-1 [Blanchon, H. 1994a] était implémentée comme illustré dans la Figure VI-3.

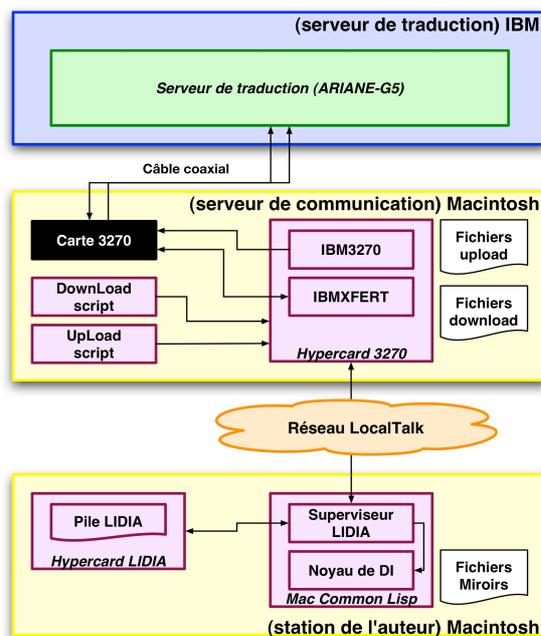


Figure VI-3 : Architecture informatique LIDIA-1 (3270)

La station de rédaction de l'auteur hébergeait l'environnement de rédaction (HyperCard) qui communiquait, au moyen d'AppleEvents⁷⁶, avec un environnement de programmation en Lisp dans lequel étaient programmés un superviseur de traitement et un module de désambiguïsation interactive.

Une autre machine hébergeait une carte d'émulation de terminal 3270, relié par un câble coaxial avec le serveur de traduction, ainsi que les outils logiciels pour gérer un terminal. La station de rédaction et le serveur de communication communiquaient aussi par AppleEvent. Il s'agissait ici de demander des traitements et de récupérer les résultats via une console.

VI.2.2.2. LIDIA-1.mail

La maquette LIDIA-1.mail [Boitet, C., *et al.* 1995] (Figure VI-4) est composée de trois modules : le module de rédaction essentiellement basé sur HyperCard, le module de traduction qui fait appel à ARIANE-G5 via le réseau et à Eudora, et enfin le module de désambiguïsation dont l'élément principal est le noyau de désambiguïsation interactive.

Ces trois modules communiquent encore par AppleEvents, à travers des interfaces de communication : client de rédaction, serveur de désambiguïsation, serveur de communication. La liaison entre ces trois modules, et donc l'orchestration générale de la maquette est assurée par un coordinateur.

La communication entre la partie Macintosh (rédaction, désambiguïsation) et la partie IBM 9221 (traduction) se fait par échanges de courriers électroniques, alors qu'elle se faisait par connexion directe. Cette approche permet l'utilisation d'outils standard disponibles sur les deux types de systèmes et offre un protocole de communication simple et particulièrement robuste.

⁷⁶ Les AppleEvents sont un protocole de communication entre applications propriétaire développé par Apple.

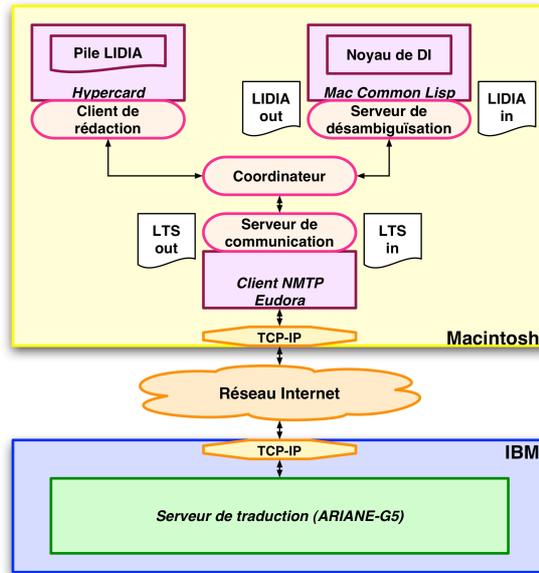


Figure VI-4 : Architecture informatique de LIDIA-1 (Mail)

En 1998, Élie Prudhomme [Prudhomme, É. 1998] a réalisé une interface d'accès via un navigateur Web aux services LIDIA. Faute de pouvoir mettre en place une interface de rédaction satisfaisante dans un tel environnement, cette piste a été abandonnée.

VI.3. Démonstration

La pile de démonstration, intitulée « LIDIA les histoires », contient deux types de cartes, des cartes d'histoires et des cartes de phrase.

Une carte d'histoires (Figure VI-5) rassemble plusieurs histoires qui partagent une même phrase, ambiguë hors contexte. La ou les ambiguïtés sont solubles par un humain, qui comprend le contexte global, mais ne le seraient pas par un système expert.

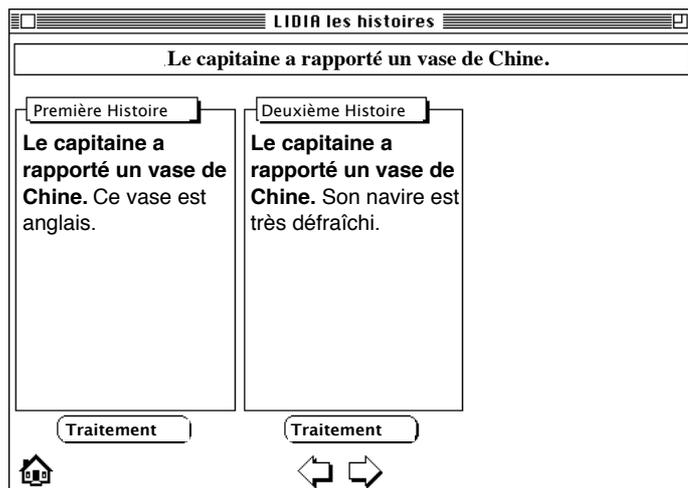


Figure VI-5 : Carte d'histoires concernant « Le capitaine à rapporté un vase de Chine. »

Pour les besoins de la démonstration, chaque histoire est présentée dans une carte de phrase (Figure VI-6) sur laquelle le contexte de la phrase ambiguë peut être caché ou non.

Pour que les histoires soient traduites, on demande la traduction des champs des cartes de phrases.

Comme souhaité plus haut, on n'est jamais interrompu par une question. Les objets se contentent d'indiquer qu'ils attendent l'intervention de l'utilisateur afin que la traduction se poursuive, et l'utilisateur décide quand intervenir et à quelle question répondre.

L'utilisateur peut activer les traitements LIDIA les plus fréquents grâce à une palette d'outils. La première ligne d'outils (Figure VI-6), considérée de gauche à droite, permet de traduire l'objet sélectionné, et de voir la progression des traitements, les annotations, et la rétrotraduction en français. La seconde ligne permet de se déplacer parmi les cartes qui composent le document.

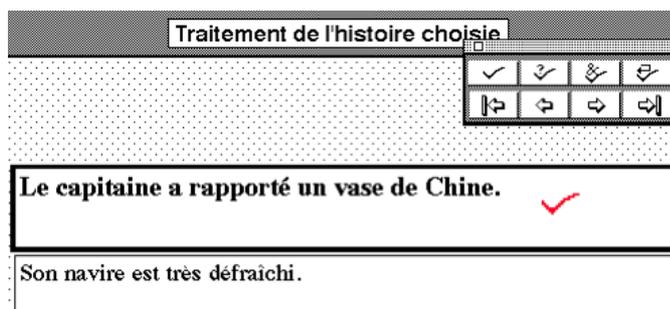


Figure VI-6 : Sélection d'un champ textuel à traduire sur une carte de phrase

Un bouton de suivi apparaît sur l'objet sélectionné. Si l'utilisateur clique sur ce bouton (? - Figure VI-7), une fenêtre flottante apparaît (Figure VI-8).



Figure VI-7 : Demande de l'état d'avancement des traitements

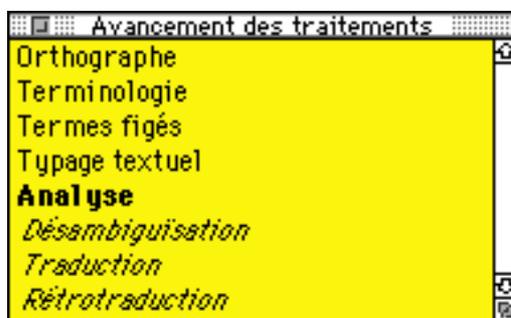


Figure VI-8 : Fenêtre flottante de suivi de l'état des traitements

Le traitement en cours est distingué en gras, les traitements déjà effectués sont en style normal, et les traitements à venir sont en italiques. Ici, le système est en train d'effectuer l'analyse de la phrase choisie.

Après l'analyse, un bouton (? !! - Figure VI-9) apparaît sur l'objet à traduire si son contenu est ambigu et nécessite donc une désambiguïsation interactive.

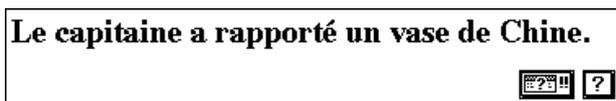


Figure VI-9 : Signalement de questions de désambiguïsation en suspens

Lorsqu'il décide de résoudre les ambiguïtés concernant un objet particulier, l'utilisateur clique sur ce bouton et les questions sont proposées comme ci-dessous, à l'aide de « rephrasages » simples.

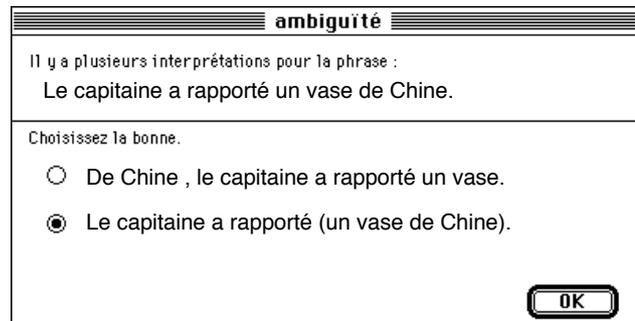


Figure VI-10 : Désambiguïssation structurale avec LIDIA-1

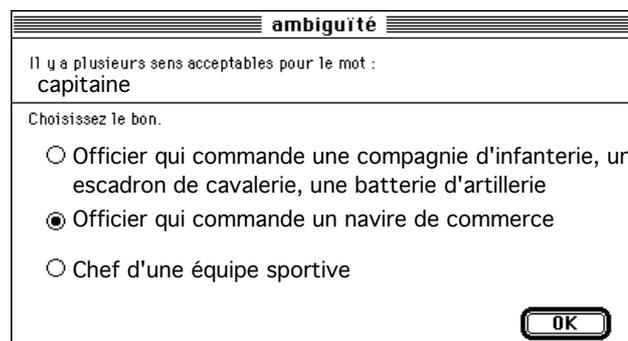


Figure VI-11 : Désambiguïssation de polysémie avec LIDIA-1

Le processus de désambiguïssation est terminé pour cette phrase.

Sur demande, l'auteur peut obtenir une vue de la forme annotée du texte. Celle qui est produite ici montre, pour l'analyse finalement retenue, les groupes syntagmatiques, accompagnés de leurs fonctions syntaxiques, ainsi que la classe syntaxique de chaque occurrence (Figure VI-12).



Figure VI-12 : Forme annotée de l'analyse désambiguïssée

On voit qu'il n'est pas déraisonnable d'imaginer que des utilisateurs expérimentés pourraient insérer eux-mêmes de telles marques, évitant ainsi certains dialogues. De plus comme le propose [Bernth, A. and Gdaniec, C. 2001], l'auteur pourrait apprendre à reconnaître les constructions syntaxiques qui donnent lieu à l'apparition d'une ambiguïté et essayer de les éviter pour avoir à répondre à moins de questions.

Pour vérifier les traductions produites par le système dans chacune des langues cibles, nous pouvons demander une « rétro-traduction ». Pour la seconde interprétation de la phrase de l'exemple et pour l'allemand, nous obtenons le texte de la Figure VI-13.

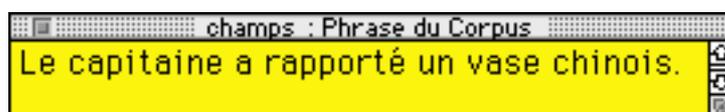


Figure VI-13 : Rétrotraduction permettant de vérifier la traduction

La Figure VI-14 montre la traduction de la phrase « Le capitaine a rapporté un vase de Chine. » dans le contexte des deux histoires préparées.



Figure VI-14 : Traduction en allemand d'une carte d'histoires

Chapitre VII

Progrès en désambiguïsation interactive, LIDIA-2 (94-03) et LIDIA-3 (04-)

MIDDIM est un projet de trois ans qui a eu lieu dans le cadre du "Memorandum of Understanding" conclu entre le CNRS et ATR en 1991. Il a été préparé en 1992-93 lors de l'année sabbatique de Ch. Boitet à ATR et a été signé en juin 1993. Le but du projet était d'étudier les techniques de désambiguïsation interactive multimodale, dans le cadre de la TA Fondée sur le Dialogue de documents écrits (GETA) et de l'interprétation automatisée de dialogues finalisés (ATR), téléphoniques ou multimodaux.

Dans le cadre de ce projet, des échanges de chercheurs entre le GETA et ATR ont été réalisés chaque année. Ces séjours ont permis de préciser les objectifs du projet [Boitet, C. 1993], de réaliser une base de données de documents annotés en ambiguïtés [Boitet, C. 1994, Boitet, C. and Tomokio, M. 1995, Boitet, C. and Tomokiyo, M. 1995], et faire des expériences de désambiguïsation interactive dans le cadre de télécommunications interprétées [Fafiotte, G. and Boitet, C. 1994, Fafiotte, G. and Boitet, C. 1996]. Ce projet a été une part importante de mon séjour post-doctoral à ATR (1994-95) durant lequel j'ai travaillé sur les ambiguïtés en situation de dialogue oral.

VII.1. Étude sur les ambiguïtés en situations de dialogue oral finalisé⁷⁷

Au cours des travaux sur les interfaces multimédia pour les systèmes de traduction de parole conduits à ATR, des corpus de dialogue spontané ont été collectés. Certains sont des conversations monolingues, d'autres sont bilingues médiatisés par un interprète. Nous avons examiné ces données pour dresser un inventaire grossier des types d'ambiguïtés qui apparaissent dans des dialogues spontanés. Une meilleure appréhension des ambiguïtés qui apparaissent véritablement doit permettre d'anticiper plus précisément la nature des difficultés que ces ambiguïtés vont poser aux systèmes de traduction de parole.

De plus, grâce aux caractéristiques de certaines des données collectées, nous avons été en mesure d'étudier les autocorrections des locuteurs, en particulier en ce qui concerne les parties ambiguës. On peut alors soulever la question suivante : les locuteurs vont-ils spontanément désambiguïser les ambiguïtés de leur discours si on leur en donne

⁷⁷ Le texte de cette section est une adaptation en français de [Fais, L. and Blanchon, H. 1996].

l'opportunité ? En effet, si les locuteurs désambigüisent eux-mêmes leurs discours avec un minimum de guidage et de formation, cela pourrait modifier profondément la manière par laquelle nous abordons la gestion de la présence d'ambigüités dans le matériau à traduire.

Nous devons tout d'abord préciser quelques éléments sur la méthode d'investigation employée. Premièrement, les ambigüités ont été collationnées à la main. Cela signifie que leur détection est subordonnée aux mêmes subtils processus cognitifs qui permettent aux humains de désambigüiser automatiquement une vaste majorité des ambigüités qu'ils rencontrent dans leurs interactions langagières quotidiennes. Lorsqu'une personne dit "I am going to the bank", il y a rarement une raison qui pousserait son interlocuteur à lui demander si elle veut dire "the side of a river" ou "the financial institution". Ainsi, bien que nous ayons fait tout notre possible pour découvrir toutes les ambigüités, nous ne pouvons pas garantir que nos propres processus cognitifs ne sont pas intervenus en nous empêchant d'en voir certaines.

Pour faire notre inventaire [Fais, L. and Blanchon, H. 1996, Fais, L. and Loken-Kim, K.-H. 1994], nous nous sommes fondés sur notre idée de ce qu'un analyseur linguistique assez simple pourrait produire sans avoir accès à une certaine connaissance du monde ou à une représentation du contexte pour l'aider à choisir une représentation unique du texte à analyser. Lorsqu'il nous a semblé qu'une approche assez simple dans l'esprit de l'analyse automatique produirait plus d'une analyse, nous avons décidé de comptabiliser une ambigüité.

VII.1.1. Corpus d'étude

Afin de tester l'interface multimédia du système de traduction de dialogues oraux finalisés développé à ATR-ITL (Interpreting Telecommunications Research Laboratories), trois expériences d'enregistrement de conversations spontanées ont été réalisées.

Lors de la première expérience (C1, Table VII-1), les sujets, jouant le rôle d'un client, devaient se rendre sur les lieux d'une conférence à laquelle ils devaient assister en engageant une conversation avec un agent de la conférence. Les clients et les agents étaient tous deux de langue maternelle anglaise. Cette situation sera par la suite nommée Communication Interpersonnelle Orale Monolingue (**CIO Monolingue**).

Lors des deux expériences suivantes [Fais, L. and Loken-Kim, K.-H. 1995, Fais, L., *et al.* 1995], un client anglophone interagissait avec un agent japonophone afin d'obtenir un itinéraire et une réservation d'hôtel.

Dans la première de ces deux expériences (C2, Table VII-1), la conversation était médiatisée par un interprète humain traduisant de l'anglais vers le japonais et vice-versa. Cette situation sera par la suite nommée Communication Interpersonnelle Orale Bilingue Médiatisée par un Humain (**CIO-MH Bilingue**).

Dans la seconde (C3, Table VII-1), la conversation était médiatisée par un système de traduction automatique simulé (technique du Magicien d'Oz) par deux interprètes, un pour chaque direction de traduction. Cette situation sera par la suite nommée Communication Interpersonnelle Orale Bilingue Médiatisée par un Magicien d'Oz (**CIO-MMOz Bilingue**).

Dans ces deux dernières expériences, ce sont des interprètes professionnels, entraînés à l'imitation d'un système de traduction automatique, qui assuraient la traduction. La situation C3, en introduisant une certaine complexité dans le dialogue, nous a permis d'observer si et quand les locuteurs désambigüisent leurs propres énoncés (cf. infra).

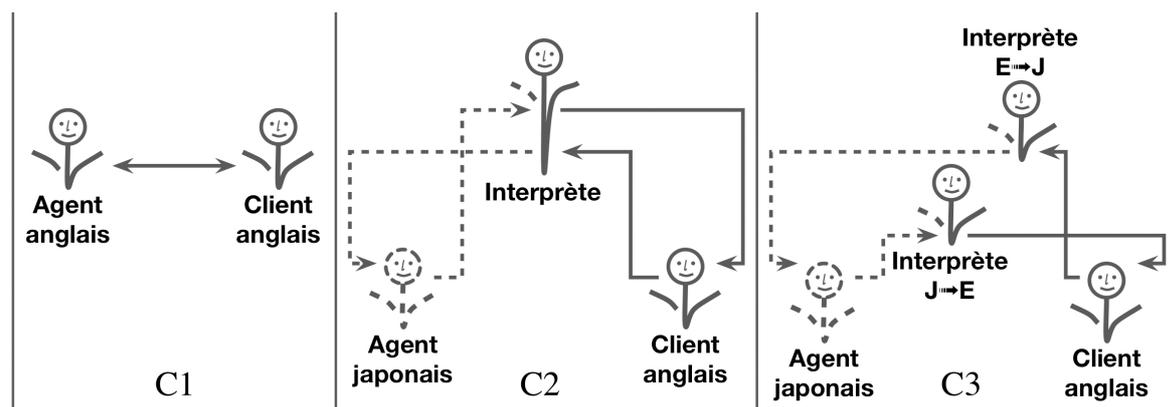


Figure VII-1 : Situations expérimentales de collectes de corpus avec EMMI (interactions)

Dans chaque expérience, les sujets ont interagi dans deux modes : via un poste téléphonique ordinaire et via un environnement informatique multimédia (EMMI) permettant aux sujets d'échanger texte, parole et dessin [Loken-Kim, K.-H., *et al.* 1993].

Corpus	Situation	Sujets anglais	Sujets japonais	Tâche	Nb. de mots (anglais)
C1	CIO Monolingue	12	0	Demande d'itinéraire	12 342
C2	CIO-MH Bilingue	9	5	Demande d'itinéraire + Réservation d'hôtel	9 513
C3	CIO-MMOz Bilingue	10	10	Demande d'itinéraire + Réservation d'hôtel	12 636

Table VII-1 : Situations expérimentales de collecte de corpus avec EMMI (données factuelles)

VII.1.2. Corpus d'ambiguïtés

Nous avons recensé les ambiguïtés sur les transcriptions des corpus C1, C2, et C3. Une liste représentative des ambiguïtés collectées est donnée en Annexe I. En particulier, dans le cas des ambiguïtés de polysémie, nous avons rencontré beaucoup d'instances de la même ambiguïté, un seul exemplaire est proposé en annexe. Nous n'avons pas considéré l'indétermination de la référence pronominale, ni les ambiguïtés dues aux ellipses.

Pour la classification, nous avons utilisé les catégories qui nous ont servi de base de travail à ATR (cf. Table VII-2 et Table VII-3). Ces catégories sont motivées par le besoin de disposer de patrons syntaxiques stables pour leur reconnaissance automatique.

Nous avons évoqué plus haut la capacité des interlocuteurs à désambiguïser « automatiquement » les énoncés afin que les ambiguïtés ne forment pas un obstacle à la communication. Il nous a semblé intéressant de compiler aussi les cas de véritable incompréhension entre les interlocuteurs, et d'examiner si une ambiguïté en était la raison ou non. Nous avons trouvé trois occurrences de cette situation, une dans chaque corpus. La première est due à la cooccurrence d'une ambiguïté de classe syntaxique et à une mauvaise réception du mot "does". Lorsque le client demande à l'agent "where does the bus stop...?" l'agent entend "where is the bus stop", et une longue discussion est nécessaire pour dissiper le malentendu.

Voici l'extrait du dialogue en question⁷⁸.

C(lient): OK and out of curiosity where does the bus stop (on) on this map {is it} closer than the [train] station
 A(gent): {/breath intake/} the bus stop is actually down this way (it's) it's a little bit off the map to the south
 C: [oh] the number five bus
 A: yes (number) and thi [Akiba, Y., *et al.*] number five bus makes its first stop at thi International Conference Center
 C: [oh] but what I mean is the bus stop for the conference center is closer to the conference center than Keage station {where}
 A: {no} the bus stop is off the map
 C: +oh+
 A: +off the+ map
 C: so I'd have to walk that far
 A: yes {(if yu)}
 C: {after the bus}
 A: yes
 C: so there isn't a bus that stops on that road
 A: well the [ah] bus number five will take you all the way to thi International Conference Center (is {is that})
 C: {right in fr}ont of it
 A: yes it will stop at the {Conference Center}
 C: {oh right tha}t's what I +meant+
 A: +will take+ you {right to} the conference center
 C: {right} oh +OK right+
 A: +that's correct+

L'ambiguïté impliquée dans l'incompréhension de la seconde expérience (C2) est une ambiguïté de subordination.

Interpreter: [ah] You have to go to number five bus stop number five
 C : Bus stop number five or bus number five

L'ambiguïté impliquée dans l'incompréhension de la troisième expérience (C3) est due à une ellipse, ambiguïté que nous n'avons pas répertoriée dans le cadre de cette étude.

C: [um] How much by bus?
 Wizard: Five hundred yen
 C: No, no, no, how much time by bus?

Bien sûr, comparé à l'incidence effective des ambiguïtés que nous avons relevées dans le corpus, les occurrences de vraie incompréhension semblent rares. Mais nous ne devons pas oublier que ces conversations se déroulent entre humains qui mettent en œuvre, non seulement leur connaissance linguistique, mais aussi leur connaissance du monde. Dans le cas de la traduction automatique, à l'opposé, non seulement la contribution de connaissances externes a ses limites (voir l'exemple de KANT pour s'en persuader), mais encore la qualité de la traduction peut crucialement dépendre du bon choix parmi différentes interprétations, choix que les humains effectuent très facilement. Ainsi, les systèmes de TALN rencontrent bien plus de difficultés que les humains, comme cela a déjà été dit.

⁷⁸ Brève note sur la transcription : les crochets marquent les pauses remplies telles que [um] ; les parenthèses marquent les faux départs tels que "(it's) its a little bit off the map" ; les barres de division marquent les bruits qui ne sont pas de la parole, tels que /breath intake/ ; les accolades marquent les bruits ou la parole qui sont concomitants aux bruits ou à la parole de l'autre locuteur. Lors d'un risque de confusion entre les éléments simultanés, des signes + sont utilisés pour permettre les appariements.

VII.1.3. Observations

La Figure VII-2 donne le nombre moyen d'ambiguïtés pour 100 mots dans les trois corpus. Nous pouvons faire plusieurs commentaires sur ces résultats. Les ambiguïtés sont plus fréquentes en **CIO Monolingue** alors qu'elles le sont significativement moins dans les configurations **CIO-MH Bilingue** et **CIO-MMOz Bilingue** ($p < 0,007$). Ce phénomène est peut être similaire à ce que l'on observe à propos des faux départs, hésitations et auto-corrrections. Les locuteurs semblent être moins prompts à en faire quand ils interagissent avec la machine [Suhm, B., *et al.* 1994].

Ainsi, les locuteurs confrontés à une situation de communication complexifiée par l'interprétation pourraient essayer de parler plus distinctement en essayant, entre autres, d'éviter des formulations qui seraient ambiguës. Cependant, cela n'est qu'une simple conjecture ; nous n'avons pas de moyen, à partir de ces données, de dire si les locuteurs tentent activement d'éliminer les ambiguïtés dans les deux dernières expériences.

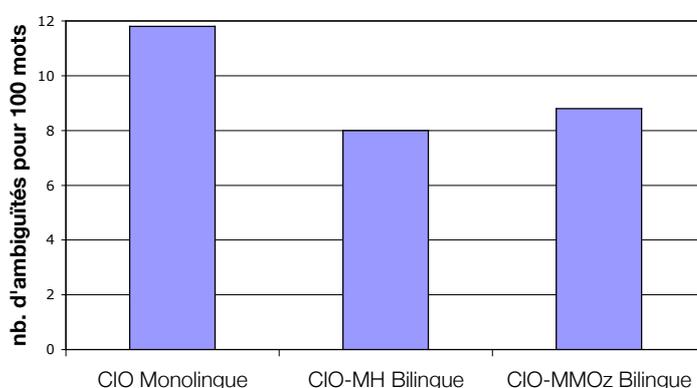


Figure VII-2 : Ambiguïtés pour 100 mots dans les trois situations d'expérimentation EMMI

Si ces chiffres fournissent une idée générale de la situation, une répartition des ambiguïtés (Figure VII-3), selon leur type (cf. Table VII-2, p. 123) pour chacune des expériences, fournit des informations intéressantes.

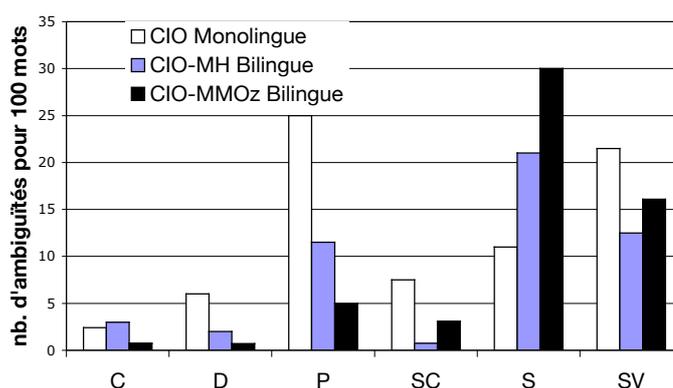


Figure VII-3 : Ambiguïtés de chaque type pour 100 mots dans les trois situations d'expérimentation EMMI

La tendance générale, un plus grand nombre d'ambiguïtés dans le corpus **CIO Monolingue** que dans les corpus **CIO-MH Bilingue** et **CIO-MMOz Bilingue**, est confirmée pour toutes les catégories, avec une exception marquée pour les ambiguïtés de subordination (S).

La fréquence des ambiguïtés de subordination est augmentée par la présence disproportionnée de la séquence "international conference center" qui est la destination dans la tâche de recherche d'itinéraire. Cette séquence constitue à elle seule 66% des ambiguïtés de subordination dans le corpus **CIO Monolingue**, 47% dans le corpus **CIO-MH Bilingue**, et 62% dans le corpus **CIO-MMOz Bilingue**. Néanmoins, si l'on exclut cette séquence, le corpus **CIO Monolingue** présente encore un nombre d'ambiguïtés de subordination inférieur aux deux autres (Figure VII-4).

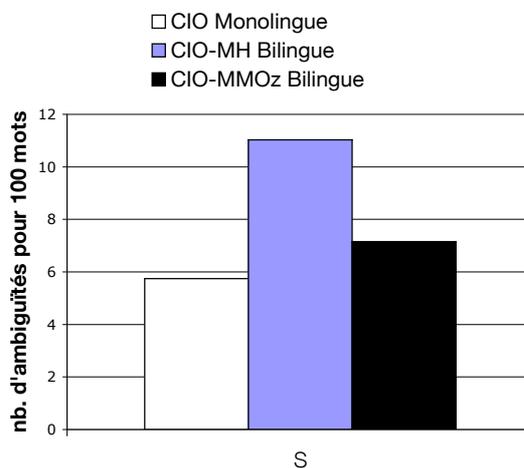


Figure VII-4 : Ambiguïtés de subordination corrigées pour 100 mots dans les trois situations d'expérimentation EMMI

Cela semble être un effet secondaire de la nature des tâches à réaliser lors des expériences. Dans les collectes **CIO-MH Bilingue** et **CIO-MMOz Bilingue**, les clients devaient non seulement réserver un hôtel, mais encore obtenir un itinéraire. Le nombre disproportionné d'ambiguïtés de subordination dans les deux dernières collectes est lié à la tâche de réservation hôtelière (voir l'annexe I pour des exemples).

VII.1.4. Conclusions

VII.1.4.1. Premières idées pour alléger la désambiguïsation interactive

Un examen rapide des types d'ambiguïté rencontrés dans les corpus permet de proposer des suggestions afin de réduire le nombre de certaines ambiguïtés. La grande fréquence de certaines ambiguïtés suggère qu'un module de désambiguïsation automatique doit être personnalisable. Dans ce cas, les utilisateurs devraient pouvoir choisir l'interprétation préférée d'un mot — pour la polysémie — ou d'une séquence de mots — pour une subordination — dès la première question la concernant, afin de ne plus être interrogés sur celle-ci ultérieurement.

De plus, l'utilisation du contexte doit pouvoir aider à résoudre automatiquement certaines ambiguïtés de classes syntaxiques, telles que dans la séquence suivante :

Interpreter: Would you prefer to go by bus or taxi?
 Client: Taxi

Une référence à la classe syntaxique de "taxi" dans le tour de parole de l'interprète permet de connaître la classe syntaxique de "taxi" dans le tour de parole du client.

Ces idées, si elles étaient implémentées, pourraient réduire la fréquence de deux des trois classes d'ambiguïté les plus souvent constatées (subordination et polysémie, cf. Figure VII-3).

La troisième classe la plus fréquente, ambiguïté de subordination avec un verbe, ne bénéficierait pas des stratégies précédemment évoquées. Ce type d'ambiguïté est l'un des types d'ambiguïtés le plus difficile à résoudre. Il est clair que pour ce type d'ambiguïtés, et pour d'autres, consulter le locuteur est nécessaire à chaque fois pour ne pas commettre d'erreur.

VII.1.4.2. « Simple répétition » : une stratégie de désambiguïsation ?

Pour l'expérience **CIO-MMOz Bilingue**, afin de faire croire aux sujets qu'ils interagissaient vraiment avec un système de traduction automatique, les compères (magiciens) devaient simuler l'incompréhension lorsque le tour de parole du sujet était trop long, pas assez « lisse », ou trop complexe. Dans ces situations, le compère demandait au sujet de répéter ("please repeat"). Les sujets utilisaient alors un grand éventail de stratégies pour clarifier leur tour de parole [Fais, L., *et al.* 1995] ; ici nous nous concentrons sur les tours de parole ambigus.

Dans le corpus **CIO-MMOz Bilingue**, il y a 161 demandes de clarification de la part des compères dont 42 (26%) mettent en jeu une ou plusieurs ambiguïtés listées dans l'annexe I. Ces 42 tours de parole représentent 44 ambiguïtés. 22 (50%) de celles-ci disparaissent dans le tour de parole suivant la demande de répétition de la part du compère. On observe quand même que sur 17 tours de parole (39%), soit l'ambiguïté demeure (15 tours, 24%), soit elle est remplacée par une autre ambiguïté (2 tours, 5%). Nous avons aussi observé des cas (5 tours, 11%) où le tour de parole avant répétition ne contient pas d'ambiguïté, mais en contient après.

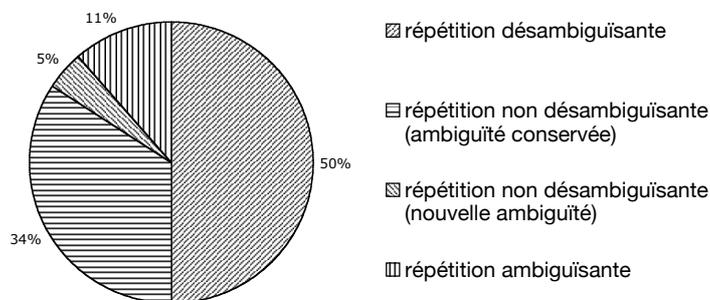


Figure VII-5 : Répartition des effets de la répétition sur l'ambiguïté

Nous ne suggérons pas que les ambiguïtés doivent être résolues simplement en demandant au locuteur de répéter. Les sujets ont montré une certaine frustration à devoir répéter leurs tours de parole, même à une fréquence assez faible (1,7 répétitions pour 10 tours). Si on leur avait demandé de répéter chaque tour de parole ambigu, cela aurait nécessité une répétition un tour sur deux et ils n'auraient peut-être pas terminé la tâche. Cependant, comme dernier recours, la répétition est éventuellement une stratégie de désambiguïsation. Il est encourageant que l'on puisse réduire de moitié le nombre des ambiguïtés avec une telle stratégie.

Toutefois, cela peut être simplement dû au hasard. Ainsi, nous n'avons aucune preuve que les sujets se soient rendus compte que la difficulté du « système » à comprendre leur tour de

parole était due à une ambiguïté et qu'ils aient alors cherché à l'éliminer. Au mieux, dans quelques cas, il semble que les sujets ont compris que quelque chose n'allait pas dans la formulation de leur tour de parole. Ils ont alors pris la décision de le reformuler si différemment que l'ambiguïté a disparu. Voici quelques exemples de stratégies :

C: Yes, so **one single room** from October [um] twenty-fourth through the twenty-eighth

WOZ: Please repeat (S)

C: OK, [um] yes, a single room for one person

C: [ah] (l) I think I'll **pass on** breakfast

WOZ: Please repeat (PH)

C: [um] [ah] (no) No thank you, (l) I don't want breakfast

C: [ah] (is it thi) Is it a straight walk or should I take a taxi or **bus**?

WOZ: Please repeat (SC)

C: (what's the) What's the best way to get there?

Il y a aussi des expressions ambiguës qui font partie intégrante du contexte de la conversation, les sujets ne peuvent pas les changer. On peut citer par exemple "International Conference Center" ou "check in". En fait, l'un des sujets a introduit une ambiguïté dans sa répétition d'un tour de parole à propos du premier exemple. Il a dû penser qu'en étant plus spécifique, il serait plus clair.

C: {l;l;l} am at Kyoto Station and I need to get some directions to a conference in Kyoto. Can you help me?

WOZ: Please repeat

C: Hello, /ls/ I want to find out how to get to thi **International Conference Center in Kyoto** (S)

Un autre sujet a créé exactement la même situation avec le second exemple.

C: [ah] When [ah] can I check in? (it's [ah]) It's morning now

WOZ: Please repeat

C: When can I **check in** to the hotel, what time? (PH)

WOZ: Please repeat

C: What time can I **check in** to the hotel? (PH)

De manière générale, les sujets tendent à effacer dans leur reformulation des segments complets de leur tour de parole original. Cette stratégie est dominante dans le corpus. Certaines fois, cette simplification supprime l'ambiguïté.

C: [ah] Yeah I'd like to **check out** October thirtieth (PH)

WOZ: Please speak slowly

C: [ah] Until the thirtieth

À l'extrême, les sujets peuvent aussi changer complètement le libellé de leur tour de parole, et ce changement peut supprimer l'ambiguïté.

C: But that **says** Keage. Is Keage Keitsu? Are they the same? (P)

WOZ: Please repeat

C: [um] I don't see Keitsu on the window. Where is Keitsu? (can you sh)

Comme on le voit, les cas où l'ambiguïté est réintroduite concernent des situations pour lesquelles nous avons dit que le problème devait être résolu par un réglage du système, qui doit quand même dépendre de la tâche en cours dans le domaine considéré. Pour "check in", on peut décider de n'admettre que le sens « remplir sa fiche de renseignement » dans le

contexte de la réservation hôtelière et refuser l'interprétation "check (in to the hotel)", au risque que le système « comprenne » parfois mal.

VII.2. Nouvelle architecture pour la DI⁷⁹

Dans la nouvelle architecture que nous avons proposée [Blanchon, H. 1995a, Blanchon, H. 1995b, Blanchon, H. 1995c], un module de désambiguïsation est constitué de deux composants : un moteur de désambiguïsation et un logiciel de désambiguïsation. Le moteur est le noyau du module et il est indépendant de la langue et du corpus à traiter. Ce composant sera utilisé pour tous les modules de désambiguïsation à développer. Par contre, le logiciel dépend de la langue et des ambiguïtés présentes dans le corpus à traiter. Le moteur peut être associé à différents logiciels pour composer des modules de désambiguïsation particuliers. La figure suivante donne une idée de l'architecture générale proposée.

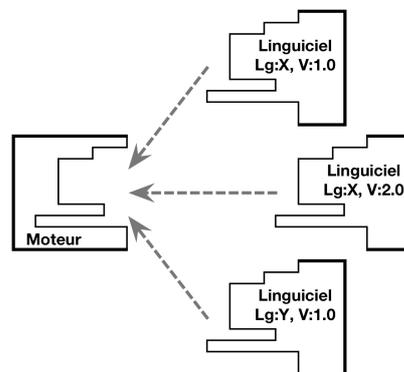


Figure VII-6 : Moteur et logiciels pour des modules de désambiguïsation interactive

Une instance de module de désambiguïsation propre à une application donnée est l'association d'un logiciel particulier et du moteur.

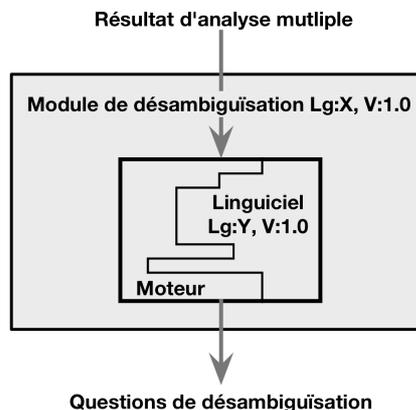


Figure VII-7 : Instance d'un module de désambiguïsation particulier

Idéalement, nous devrions fournir au concepteur d'un module de désambiguïsation un ensemble d'outils lui permettant de décrire :

- les ambiguïtés à résoudre,
- la formation des items de la question qui permettront leur résolution,

⁷⁹ Le texte de cette section est une adaptation de différents articles, en particulier [Blanchon, H. 1995a, Blanchon, H. 1995b, Blanchon, H. 1995c]

- la méta-présentation des questions (invite, mise en relief du segment ambigu, l'entête de la liste des items) ;
- la stratégie de découverte des ambiguïtés, si plusieurs sont présentes en entrée ;
- la modalité à utiliser pour résoudre les ambiguïtés (question écrite ou orale) ;
- la façon dont les questions doivent être préparées (toutes à la fois, une par une).

Ces descriptions forment le logiciel.

Le moteur utilisera alors le logiciel pour réaliser le processus de désambiguïsation interactive. Il doit donc fournir les services suivants :

- une méthode de reconnaissance des ambiguïtés décrites dans le logiciel,
- un ensemble d'opérateurs à utiliser pour décrire la formation des items de question ;
- un langage de description d'automates qui permette de décrire l'ordre dans lequel les ambiguïtés seront résolues ;
- un ensemble de classes de dialogue correspondant aux différentes modalités proposées ;
- des méthodes de présentation des questions ;
- des méthodes implémentant différentes stratégies de préparation et de présentation des questions.

Dans le cadre de la réalisation de la maquette LIDIA-1, une première version d'un module de désambiguïsation pour le français a été réalisée [Blanchon, H. 1992b, Blanchon, H. 1994b]. Ce premier module ne respectait pas complètement les principes que nous venons de proposer. Nous nous concentrons ici sur les résultats de notre travail de réingénierie en CLOS (Common Lisp Object System), effectué à ATR-ITL (1994-95).

VII.2.1. Proposition⁸⁰

Dans le cadre du projet LIDIA, l'analyseur produit une structure d'arbre décoré. Les ambiguïtés sont donc décrites en termes de propriétés de cette structure en utilisant des patrons d'arbres.

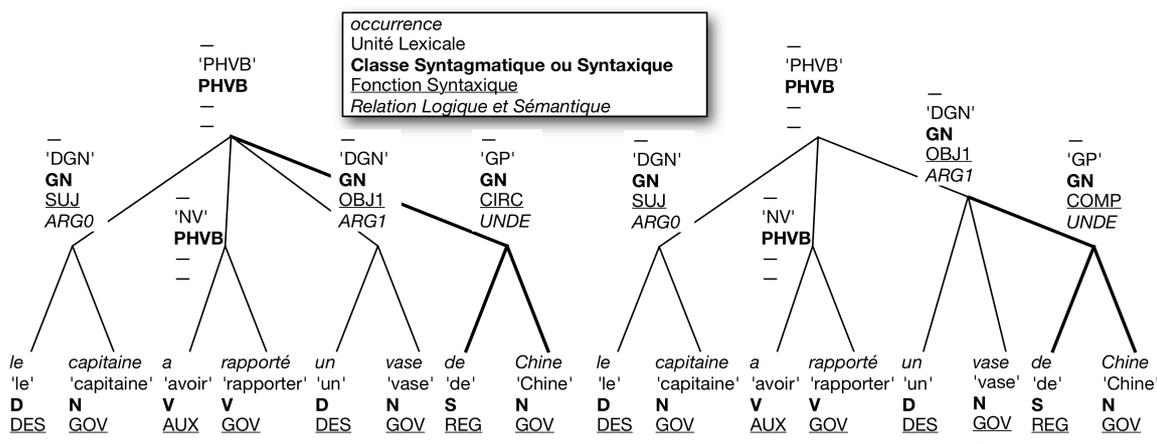


Figure VII-8 : Analyses produites pour la phrase «Le capitaine a rapporté un vase de Chine.»

⁸⁰ Les exemples sur le français que nous donnerons dans cette section sont issus des travaux de Katty Grasson durant sa première année de thèse sur la réingénierie du module de désambiguïsation de la maquette LIDIA-1 avec les idées que nous avons proposées lors de notre séjour à ATR.

Si une phrase est ambiguë, l'analyseur produit une représentation (solution) pour chacune des interprétations possibles. Il est possible de distinguer des familles de solutions qui, pour un même segment du texte, auront des représentations différentes. Dans la Figure VII-8, on observe que pour « Le capitaine a rapporté un vase de Chine. », l'analyseur propose deux solutions dont les différences sont marquées.

Une vision abstraite, par « forêt », de ces deux analyses est donnée Figure VII-9.

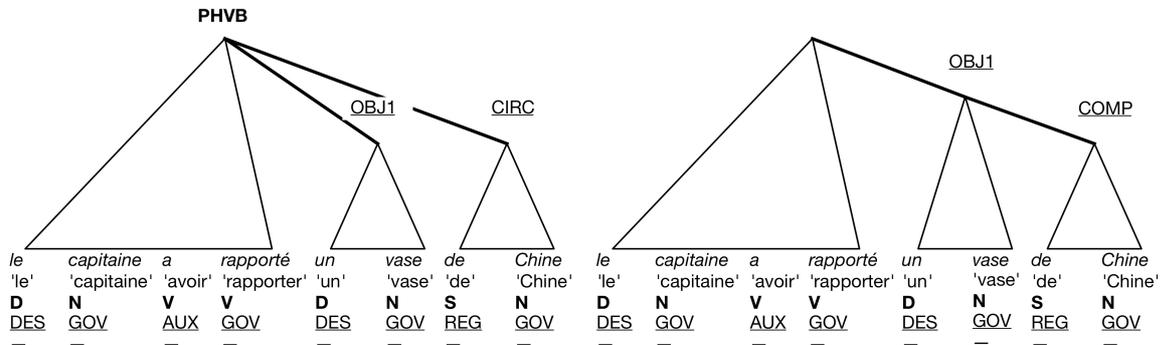


Figure VII-9 : Vision abstraite par forêt des analyses de la phrase « Le capitaine a rapporté un vase de Chine »

Cette ambiguïté peut finalement être décrite et généralisée par le couple de patrons suivants :

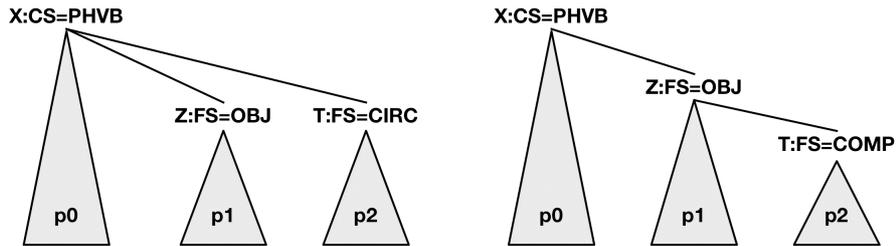


Figure VII-10 : Faisceau de patrons définissant le type d'ambiguïté présent dans la phrase « le capitaine a rapporté un vase de Chine. »

Un patron décrit une famille d'arbres avec des variables de nœud et des variables de forêt. Les variables de nœud sont ici **X**, **Z** et **T** ; les variables de forêt sont **p0**, **p1** et **p2**. On peut observer que les variables de nœud sont contraintes : la classe syntagmatique de **X** doit être **PHVB**, la fonction syntaxique de **Z** doit être **OBJ**, et la fonction syntaxique de **T** doit être **CIRC** pour le patron de gauche et **COMP** pour le patron de droite. La valeur (le contenu) des variables de forêt n'est pas contrainte dans la définition du patron.

Un tel ensemble de patrons est appelé **faisceau** dans notre terminologie. Les patrons d'un faisceau partagent donc un ensemble de variables de nœud et de variables de forêt. Nous montrerons plus loin que ces mêmes patrons sont en quelque sorte synchronisés sur leurs variables de forêt dont les instances devront vérifier un certain nombre de propriétés.

Un type d'ambiguïté d'une classe d'ambiguïtés particulière est donc défini par la cooccurrence d'un certain nombre de patrons membres d'un certain faisceau. Pour l'ambiguïté de notre exemple, le faisceau comporte les deux patrons suivants :

```
((?is ?x node-prop-equal-p 'CS 'PHVB)
 (?* ?p0)
 ((?is ?y node-prop-equal-p 'FS 'OBJ)
 (?* ?p1))
 ((?is ?z node-prop-equal-p 'FS 'CIRC)
 (?* ?p2)))
```

Figure VII-11 : Patron 1 de définition d'une ambiguïté de structure argumentaire du verbe de type 3

```
((?is ?x node-prop-equal-p 'CS 'PHVB)
 (?* ?p0)
 ((?is ?y node-prop-equal-p 'FS 'OBJ)
 (?* ?p1)
 ((?is ?z node-prop-equal-p 'FS 'COMP)
 (?* ?p2)))
```

Figure VII-12 : Patron 2 de définition d'une ambiguïté de structure argumentaire du verbe de type 3

Ces patrons sont des objets de type `lidia-patterns`, nommés `sav-t3-1` (Figure VII-11) et `sav-t3-2` (Figure VII-12). Ils sont utilisés pour définir un faisceau qui décrit l'ambiguïté de structure argumentaire du verbe (`sav`) de type 3 (`t3`).

```
(defvar *sav-t3* (make-instance 'pattern-beam ;objet de type faisceau de patrons
 :beam-name 'sav-t3* ;nom du faisceau
 :beam-value (list *sav-t3-1* *sav-t3-2*)) ;liste des patrons du faisceau
```

Voyons maintenant comment s'effectue la reconnaissance d'une ambiguïté.

VII.2.1.1. Reconnaissance d'une ambiguïté : définition

Un type d'ambiguïté est décrit au moyen d'un faisceau de patrons qui sont des schémas d'arbres. Le langage de description de patrons est celui proposé dans [Norvig, P. 1992].

De manière formelle, une phrase **S** pour laquelle l'analyseur produit **s** solutions **Sol_i** présente l'ambiguïté décrite par le faisceau **B** constitué de **b** patrons **P_j** si et seulement si :

- le nombre de solutions (**s**) est supérieur ou égal au nombre de patrons (**b**),
- pour chaque solution, il existe un unique patron qui s'apparie avec elle,
- chaque patron s'apparie avec au moins une solution,
- la distance **fd** entre les valeurs des variables de forêt des appariements est nulle.

Soit :

```

Une phrase S pour laquelle l'analyseur produit s solutions Soli présente l'ambiguïté
décrite par le faisceau B constitué de b patrons Pj si et seulement si :

- b < s ; (condition 0)
-  $\forall i, \exists !j / \text{match.p}(\text{Sol}_i, P_j)=t$  ; (condition 1)
-  $\forall j, \exists i / \text{match.p}(\text{Sol}_i, P_j)=t$  ; (condition 2)
-  $\forall i, i', \forall j, j', \text{ si } \text{match.p}(\text{Sol}_i, P_j)=t \vee \text{match.p}(\text{Sol}_{i'}, P_{j'})=t \text{ alors}$ 
   $\text{fd}(\text{match.binding}(\text{Sol}_i, P_j), \text{match.binding}(\text{Sol}_{i'}, P_{j'}))=0$  (condition 3)

```

Figure VII-13 : Définition de la reconnaissance d'une ambiguïté

La distance **fd** entre deux appariements `match.binding(Soli, Pj)`, `match.binding(Soli', Pj')` est nulle, si et seulement si :

- la couverture de chaque variable de forêt, exceptée la dernière, est la même dans les deux appariements, et
- pour la dernière variable de forêt, la couverture est la même, ou bien, l'une des couvertures est préfixe de l'autre.

La couverture (**coverage**) d'une variable de forêt correspond au mot des feuilles, réduit aux occurrences du texte analysé, du sous-arbre auquel correspond cette variable.

Ainsi :

$$\begin{array}{l}
 \text{Soit } ?p_{j,k} \text{ (resp. } ?p_{j',k}), 1 \leq k \leq p, \text{ les } l \text{ variables de forêt utilisées dans le patron } P_j \text{ (resp. } P_{j'}). \\
 \text{fd}(\text{match.binding}(S_i, P_j), \text{match.binding}(S_i, P_{j'}))=0 \\
 \Leftrightarrow \\
 \forall k, 1 \leq k < p, \text{ coverage}(?p_{j,k})=\text{coverage}(?p_{j',k}) \\
 \wedge (\text{coverage}(?p_{p,k})=\text{coverage}(?p_{p',k}) \\
 \vee \\
 (\text{prefix-p}(\text{coverage}(?p_{p,k}),\text{coverage}(?p_{p',k}))=t \\
 \vee \\
 \text{prefix-p}(\text{coverage}(?p_{p',k}),\text{coverage}(?p_{p,k}))))
 \end{array}$$

Figure VII-14 : Définition de la distance entre les ensembles de variables de forêt de deux patrons

Nous allons voir qu'en pratique, nous ne chercherons à vérifier que les *conditions 1* et *2*, la *condition 3* étant vérifiée par effet de bord de l'algorithme d'appariement.

VII.2.1.2. Reconnaissance d'une ambiguïté : implémentation

VII.2.1.2.1. Aperçu

En pratique la reconnaissance d'un type d'ambiguïté est réalisée par une fonction qui prend en entrée un faisceau ($(P_1 P_2 \dots)$) et un ensemble de solutions numérotées (**ESN**). Un **ESN** est une liste de couples indice, solution ($((1 \text{ Sol}_1) (2 \text{ Sol}_2) \dots)$). Cette fonction rend une liste de quatre valeurs :

- un booléen (t ou nil) qui marque si le faisceau a pu être apparié avec l'ensemble de solutions. Sinon, les autres éléments de la liste ne sont pas significatifs,
- le nom du faisceau reconnu,
- une liste de triplets utilisés ensuite pour construire les items de dialogue des questions de désambiguïsation,
- une liste d'**ESN** qui, si nécessaire, serviront à construire d'autres questions.

Le noyau du module de détection d'une ambiguïté remplit une matrice dans laquelle l'appariement des patrons du faisceau sur les différentes solutions se fait de façon contrôlée afin d'assurer une bonne propriété de la distance des variables de forêts par défaut lors des essais d'appariement.

VII.2.1.2.2. Synchronisation de l'appariement des patrons

L'appariement des patrons d'un faisceau avec les solutions est exécuté en faisant un parcours des solutions en profondeur d'abord. À chaque itération, l'invariant comprend l'indice (**n**) du nœud à partir duquel une tentative d'appariement est effectuée.

Nous disons que l'appariement des patrons est « synchronisé » car toutes les solutions ont le même nombre de nœuds dans la partie déjà examinée (Figure VII-15) et chacune des ambiguïtés présentes dans les différentes solutions consomme le même nombre de nœuds dans chaque solution. Nous sommes donc assurés qu'une ambiguïté donnée aura pour racine, dans les différentes solutions, des nœuds ayant le même indice dans une numérotation des arbres en profondeur d'abord.

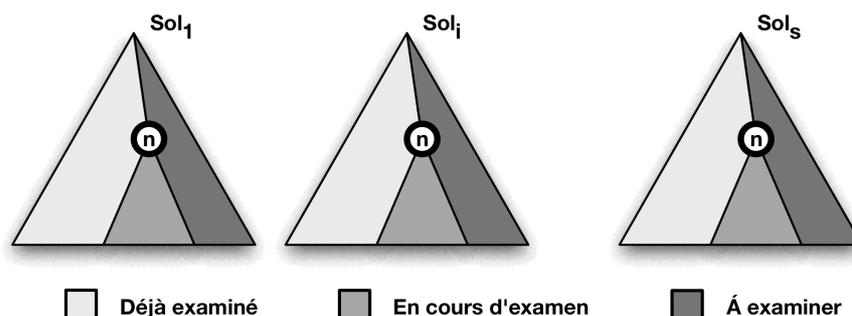


Figure VII-15 : Synchronisation de l'appariement des patrons avec les solutions

Pour détecter une ambiguïté d'un certain type décrite par un faisceau **B**, il suffit de trouver le plus petit indice **n** qui permette de vérifier les conditions d'appariement (*condition 1* et *condition 2*). La proposition inter-appariement (*condition 3*) est alors nécessairement vérifiée.

VII.2.1.2.3. Construction d'une matrice d'appariements

La recherche du bon indice **n** se fait en construisant, pour chaque **i** ($1 \leq i \leq nb$ de nœuds de chacune des solutions) une matrice comportant autant de lignes que de solutions et autant de colonnes que de patrons dans le faisceau. Chaque cellule de la matrice contient deux valeurs : la valeur booléenne de `match.p(Soli, Pj)`, et la valeur de `match.binding(Soli, Pj)` qui est une liste de paires comportant le nom de la variable du patron et la valeur prise par celle-ci lors de l'appariement.

solution	patron	P ₁	...	P _j	...	P _b
Sol ₁						
...						
Sol _i				<code>match.p(Sol_i, P_j)</code> , <code>match.binding(Sol_i, P_j)</code>		
...						
Sol _s						

Pour un indice **n**, l'appariement du faisceau avec l'ensemble des solutions vérifie la *condition 1*, si une seule cellule de chaque ligne a la valeur de `match.binding(Soli, Pj)` positionnée à vraie. La *condition 2* est vérifiée si aucune colonne n'est vide. La *condition 0* est vérifiée avant que ne commence l'appariement. La *condition 3* est vérifiée par effet de bord de l'algorithme.

Lorsque l'on a finalement trouvé un bon indice **n**, toutes les données nécessaires à la préparation d'une question sont disponibles.

VII.2.1.3. Production de la question concernant une ambiguïté

Afin de préparer le rephrasage de l'interprétation relative à un patron donné, une méthode de rephrasage est associée à ce patron.

Cette méthode de rephrasage fait des manipulations de surface sur les variables de forêt du patron instanciées lors de l'appariement. Ces manipulations sont décrites au moyen d'un ensemble d'opérateurs sur les chaînes associées aux instances des variables.

Agreement produit la forme fléchie d'un adjectif selon des contraintes de genre et de nombre.

exemple : Agreement("noirs",(("gn" [...] (gnr fem nbr plu)) ([...])) -> "noires"

Text	produit le mot des feuilles selon l'occurrence du texte correspondant à l'arbre passé en paramètre.
Subject	produit le texte du sujet de l'arbre passé en paramètre.
<u>Note</u> :	Il existe le même genre de fonction pour chaque fonction syntaxique : VerbalGroup, Circ, Comp, Gadj...
Coord	produit le texte de la conjonction de coordination de l'arbre passé en paramètre.
But_Coord	produit le texte de l'arbre passé en paramètre privé de la conjonction de coordination.
But_Sub	produit le texte de l'arbre passé en paramètre privé de la conjonction de subordination.
But_Det	produit le texte de l'arbre passé en paramètre privé du déterminant.
Distribute	distribue un groupe adjectival sur des groupes nominaux et assemble les groupes ainsi formés avec une conjonction de coordination.
<u>Note</u> :	La distribution doit être faite en réalisant l'accord en genre et en nombre entre le groupe adjectival et les groupes nominaux. Il faut pour cela faire appel à un module de génération morphologique.

exemple : Distribute((A, B, C), ou ,(1, 3), (2, 3)) -> A C ou B C

Substitute remplace une préposition ambiguë par une proposition non ambiguë (dans le contexte) en fonction de certaines propriétés : fonction syntaxique ou relation logico-sémantique.

exemple : Substitute("de", #Objet_1) -> "à propos de"

Project appliqué à une feuille d'une structure *mmc*, projette l'occurrence suivie de sa classe syntaxique.

Bracket met entre parenthèses le texte de ses arguments.

ex : Bracket("classeurs" ,"noirs") -> "(classeurs noirs)"

Definition produit le sens d'un lemme étant donné un numéro de sens.

Les méthodes de rephrasage pour les patrons de l'ambiguïté définie ci-dessus sont les suivantes :

Text(?p2) , Text(?p0) Text(?p1)

Figure VII-16 : Méthode de rephrasage associée au patron sav-t3-1 (Figure VII-11)

Text(?p0) Determiner(?p1) Bracket(But_Det(?p1), Text(?p2))

OU

Text(?p0) Bracket(Text(?p1), Text(?p2))

Figure VII-17 : Méthodes de rephrasage pouvant être associées au patron sav-t3-1 (Figure VII-12)

Pour la phrase « Le capitaine a rapporté un vase de Chine. », nous produisons alors les items suivants :

- « de Chine, le capitaine a rapporté un vase. » pour le patron **sav-t3-1** et la méthode associée (Figure VII-16).
- « le capitaine a rapporté (un vase de Chine). » ou « Le capitaine a rapporté un (vase de Chine). » pour le patron **sav-t3-2** et les méthodes associées (Figure VII-17).

Finalement, pour l'ambiguïté structure argumentaire du verbe de type 3, nous avons les définitions suivantes :

```
(defvar *sav-t3-1* (make-instance 'pattern ;objet de type patron
                               :pattern-name '*sav-t3-1* ;nom du patron
                               :pattern-value '((?is ?x node-prop-equal-p 'CS 'PHVB) ;schéma d'arbre
                                               (?* ?p0)
                                               ((?is ?y node-prop-equal-p 'FS 'OBJ)
                                                (?* ?p1))
                                               ((?is ?z node-prop-equal-p 'FS 'CIRC)
                                                (?* ?p2)))
                               :pattern-method #'item-production-method)) ;méthode de rephrasage

(defmethod item-production-method ((pattern-name (eql '* sav-t3-1*)) binding)
  ;méthode générique contrainte par la valeur de son premier paramètre
  ;utilisation de la commande de formatage pour produire le texte de rephrasage
  (format nil "~A,~A ~A."
    (string-trim '#\space) (apply #'text (cdr (assoc '?p2 binding))))
    (string-trim '#\space) (apply #'text (cdr (assoc '?p0 binding))))
    (string-trim '#\space) (apply #'text (cdr (assoc '?p1 binding))))))
```

```
(defvar *sav-t3-2* (make-instance 'pattern ;objet de type patron
                               :pattern-name '*sav-t3-2* ;nom du patron
                               :pattern-value '((?is ?x node-prop-equal-p 'CS 'PHVB) ;schéma d'arbre
                                               (?* ?p0)
                                               ((?is ?y node-prop-equal-p 'FS 'OBJ)
                                                (?* ?p1))
                                               ((?is ?z node-prop-equal-p 'FS 'COMP)
                                                (?* ?p2)))
                               :pattern-method #'item-production-method)) ;méthode de rephrasage

(defmethod item-production-method ((pattern-name (eql '* sav-t3-2*)) binding)
  ;méthode générique contrainte par la valeur de son premier paramètre
  ;utilisation de la commande de formatage pour produire le texte de rephrasage
  (format nil "~A ~A."
    (string-trim '#\space) (apply #'text (cdr (assoc '?p0 binding))))
    (apply #'bracket (string-trim '#\space) (apply #'text (cdr (assoc '?p0 binding))))
    (string-trim '#\space) (apply #'text (cdr (assoc '?p1 binding))))))
```

Dans la maquette du français, nous avons six types d'ambiguïté de structure argumentaire du verbe, donc six faisceaux de description. Ces six faisceaux sont regroupés pour des raisons de stratégie de désambiguïstation dans une pile de faisceaux. Cette pile est définie comme suit :

```
(defvar *sav-beam-stack* (make-instance 'beam-stack ;objet pile de faisceau
                                     :beam-stack-name '*sav-beam-stack* ;nom de la pile
                                     ;liste des faisceaux de la pile
                                     :beam-stack-value (list *sav-t1* *sav-t2* *sav-t3* *sav-t4* *sav-t5* *sav-t6*)))
```

VII.2.1.4. Vue globale du processus de DI

Un module de désambiguïstation reçoit donc en entrée une structure *mmc* et produit en sortie un arbre de questions. Par itérations successives, les ambiguïtés présentes dans la structure *mmc* sont découvertes en créant des partitions de l'ensemble des solutions de la

structure. Lorsque les parties sont toutes réduites à une solution, le processus de production de l'arbre de questions est terminé. Le schéma de la Figure VII-18 illustre ce processus.

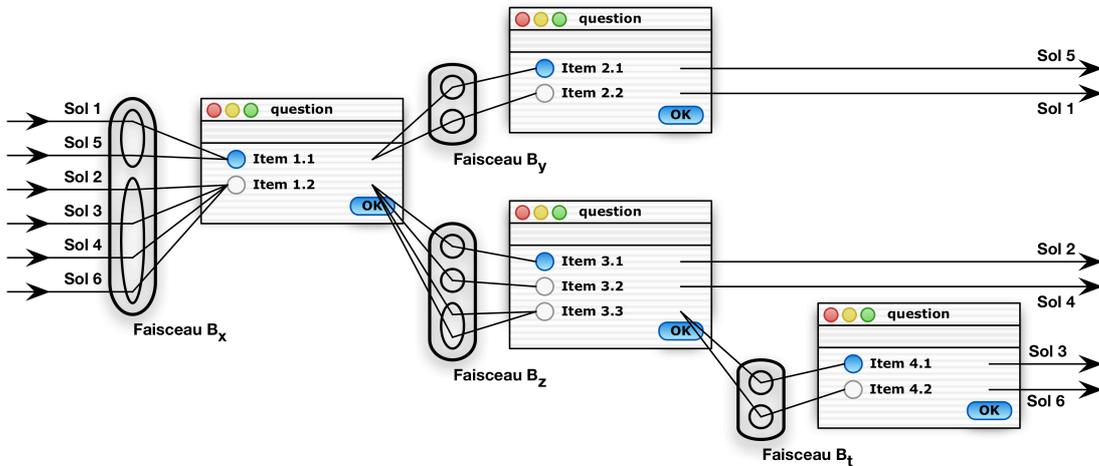


Figure VII-18 : Construction d'un arbre de questions (vue externe)

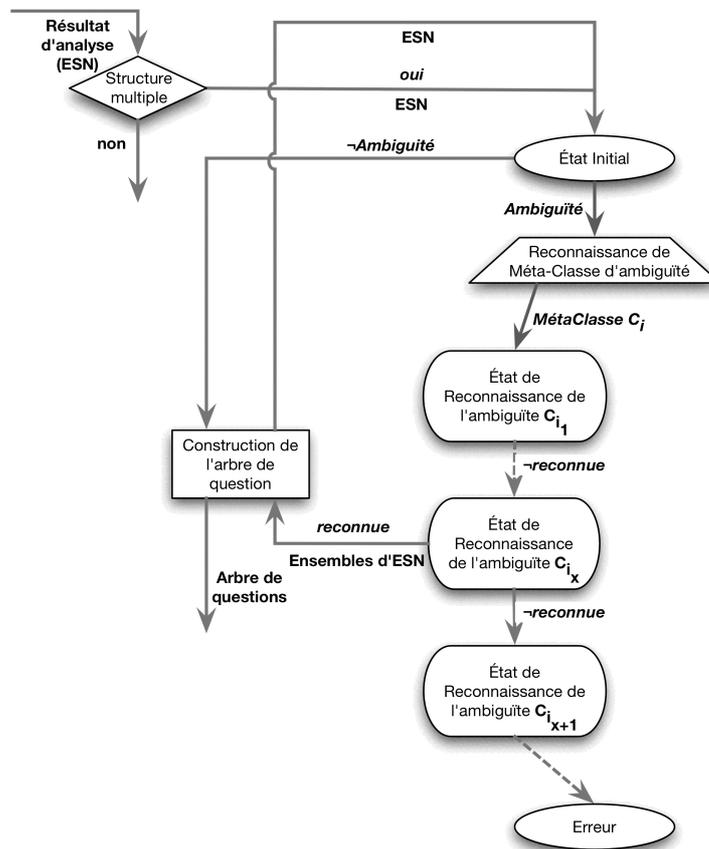


Figure VII-19 : Construction d'un arbre de questions (vue interne)

Les itérations successives se font en parcourant un automate qui comporte deux types d'états (Figure VII-19) : des états de reconnaissance de méta-classes d'ambiguïtés et des état de reconnaissance de classes d'ambiguïté. Nous avons proposé quatre méta-classes d'ambiguïtés :

- les ambiguïtés de polysémie,
- les ambiguïtés de classe syntaxique,

- les ambiguïtés de géométrie,
- les ambiguïtés de décoration.

Les ambiguïtés de polysémie ne sont pas reconnues par des faisceaux de patrons. Il suffit en effet de consulter le mot des feuilles pour découvrir des occurrences pour lesquelles l'analyseur n'a pas pu choisir le sens (nœuds contenant plus d'un numéro de sens). Les trois autres classes d'ambiguïtés sont décrites au moyen de faisceaux de patrons.

Un état de reconnaissance d'une classe d'ambiguïtés cherche à apparier l'un des faisceaux de description d'un type d'ambiguïtés qui appartient à une pile de patrons. La définition de piles de patrons permet de regrouper des ambiguïtés d'une même classe qui peuvent s'exprimer de différentes façons, comme par exemple les ambiguïtés de structures argumentaires du verbe.

VII.2.2. Illustration

Nous avons implémenté un premier module de désambiguïsation interactive pour l'anglais avec la nouvelle architecture que nous venons de décrire. Ensuite, Katty Grasson a effectué une réingénierie du module de désambiguïsation en respectant les règles d'implémentation que nous avons proposées.

VII.2.2.1. Désambiguïsation de l'anglais (ATR)

Notre idée de départ était de traiter les ambiguïtés de l'oral [Blanchon, H. 1996, Blanchon, H. 1997] dans un contexte de traduction de parole. Pour ce faire, nous avons effectué l'étude de corpus dont les résultats ont été présentés dans la section VII.1. Les types d'ambiguïtés repérés ne constituent pas un inventaire complet des ambiguïtés de l'anglais. Il ne s'agit donc pas d'un module de désambiguïsation exhaustif — celui de LIDIA est bien plus important comme nous le verrons ensuite —, mais d'une première mise en œuvre de la nouvelle architecture que nous avons proposée. L'annexe IV propose des exemples d'ambiguïtés traitées avec le module développé à ATR.

VII.2.2.2. Désambiguïsation du français (LIDIA)

Le module de désambiguïsation développé dans le cadre de nos travaux sur LIDIA est lui, plus représentatif des types d'ambiguïtés du français [Blanchon, H. 1994b, Blanchon, H. 1994c, Blanchon, H. 1996]. L'annexe V fournit des exemples des ambiguïtés traitées.

VII.2.3. Évaluation en utilisabilité, utilité sur l'anglais⁸¹

Nous avons conduit deux expériences [Blanchon, H. and Fais, L. 1996a, Blanchon, H. and Fais, L. 1997] afin de vérifier la compréhensibilité du libellé des questions de désambiguïsation que nous sommes capables de produire avec la méthode proposée et implémentée.

VII.2.3.1. Objectifs

Nous cherchions alors à vérifier que notre proposition permet de produire des questions de désambiguïsation auxquelles les auteurs peuvent répondre sans faire d'erreur. [Mitamura, T., *et al.* 1999] font en effet remarquer que les utilisateurs répondent parfois

⁸¹ Le texte de cette section est une adaptation en français des articles et rapports [Blanchon, H. and Fais, L. 1996a, Blanchon, H. and Fais, L. 1996b, Blanchon, H. and Fais, L. 1996c, Blanchon, H. and Fais, L. 1997]

mal. Cependant, les services que nous proposons dans LIDIA, la visualisation de la structure annotée et la rétrotraduction constituent deux « filets de sécurité » éventuels. Idéalement, nous devrions conduire des expériences permettant à des auteurs de désambiguïser leurs propres documents. Ce genre d'expérience est malheureusement extrêmement coûteux. Dans les expériences que nous avons réalisées, tous les sujets désambiguïsent le même document.

VII.2.3.2. Évaluation pilote

Nous avons conduit une première expérience [Blanchon, H. and Fais, L. 1996b] qui visait à évaluer la compréhensibilité des questions de désambiguïsation. Cependant, comme nous allons le montrer, cette expérience introduisait un biais qui n'a pas permis d'évaluer effectivement la compréhensibilité des questions.

VII.2.3.2.1. Données et protocole

Pour cette première expérience, nous avons construit artificiellement un texte contenant 35 phrases ambiguës. Les phrases ambiguës ont été sélectionnées parmi un corpus de conversations spontanées collectées à ATR [Fais, L. and Blanchon, H. 1996]. Le texte⁸² est constitué de deux histoires différentes. Les 35 phrases ambiguës sont distribuées également dans sept classes d'ambiguïtés définies dans le module de désambiguïsation interactive.

Pour chaque catégorie, le texte propose deux phrases à interprétation facile et trois phrases à interprétation difficile. Par interprétation facile, nous entendons l'interprétation la plus fréquente ou la plus saillante d'une séquence donnée de mots, celle qui vient immédiatement à l'esprit. Au contraire, une interprétation difficile correspond à une interprétation inhabituelle, bien que possible, d'une séquence donnée de mots. Par exemple, l'interprétation facile de la phrase "I want to check in the hotel" est "I want to register at the hotel" ; l'interprétation difficile est "I want to investigate the hotel".

Les codes (cf. Table VII-2) présents dans le texte complet repris en Annexe II donnent l'abréviation du type d'ambiguïté, ainsi que son interprétation à retenir **E** pour naturelle (Easy) et **H** pour difficile (Hard).

Une copie du texte complet était remise aux sujets afin qu'ils le lisent et le comprennent. Pour l'expérience, ils étaient assis devant un écran sur lequel une fenêtre dotée d'un ascenseur vertical affichait une partie du texte. Ils étaient alors invités à lire le texte à haute voix lentement et attentivement. Lorsqu'ils avaient fini de prononcer une phrase ambiguë, l'expérimentateur déclenchait l'affichage d'une question de désambiguïsation sur leur écran (cf. Figure VII-20).

Vingt-deux sujets, répartis en deux groupes de onze, ont été recrutés parmi les chercheurs de ATR ayant l'anglais comme langue maternelle. Au premier groupe de sujets, nous avons proposé des questions libellées comme un humain aurait pu le faire pour lever l'ambiguïté. Les questions sont donc formulées dans un style naturel et relativement verbeux. Au second groupe de sujets, nous avons proposé des questions telles qu'un processus de désambiguïsation interactive l'aurait fait. Les sujets pouvaient soit sélectionner une reformulation désambiguïsante parmi deux ou trois, soit répondre qu'ils étaient incapables d'en choisir une.

⁸² Le texte de l'évaluation pilote, ainsi que le libellé des questions est fourni en Annexe II.

Abréviation	Type	Exemple
D	Décoration	please exit by exit 14 - exit through Exit 14 - exit in proximity of Exit 14
PH	Verbe à particule	Go over it carefully - pass over - review
CO	Coordination	western bed and bath - western bath - not western bath
S	Subordination sans verbe	a train ticket vendor - train ticket - ticket vendor
P	Polysémie	story - narrative - floor
SV	Subordination avec verbe	You will see many taxis waiting - the taxis are waiting - you are waiting
SC	Classe syntaxique	You can take a bus or taxi - you can taxi - you can take a taxi

Table VII-2 : Classes d'ambiguïtés (base de travail) et exemples tirés de l'expérience pilote

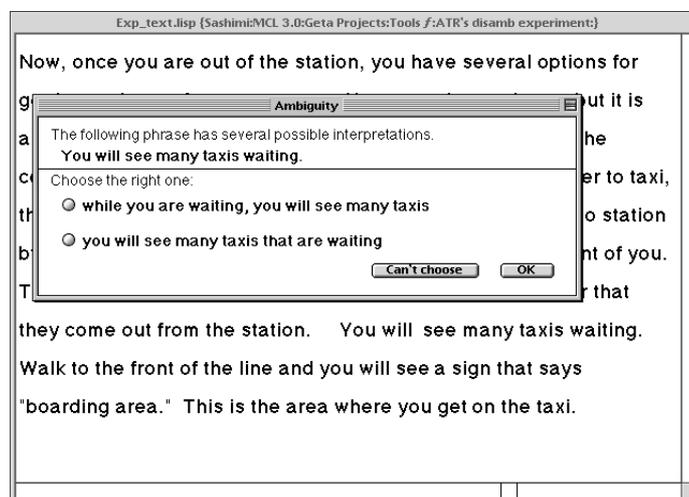


Figure VII-20 : Écran du sujet avec présentation d'une question de désambiguïsation

VII.2.3.2.2. Résultats

Chaque classe d'ambiguïté était représentée cinq fois dans le texte, deux phrases de chaque classe ayant une interprétation facile et trois une interprétation difficile. La Figure VII-21 illustre la différence pour les deux types d'interprétation. Comme on pouvait s'y attendre, les sujets ont donné de meilleures réponses aux questions concernant les phrases d'interprétation facile.

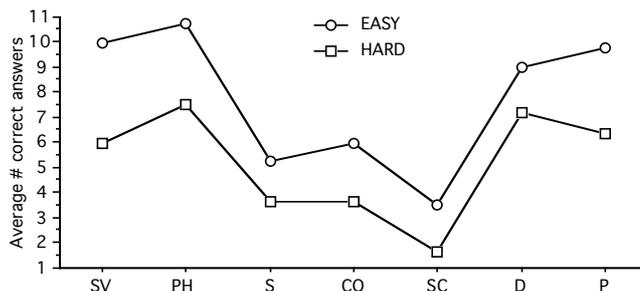


Figure VII-21 : Nombre de réponses correctes pour les interprétations faciles et difficiles pour la première expérience en compréhensibilité des questions de désambiguïsation

Comme les sujets pouvaient choisir de répondre ou non à chaque question posée, nous avons aussi détaillé les bonnes réponses, les mauvaises réponses et les refus de choix.

La Figure VII-22 et la Figure VII-23 montrent comment les sujets ont réagi aux deux types de questions de désambiguïsation (humaine, machine) que nous leur avons proposées. Tout d'abord, la seule classe pour laquelle il y a une différence marquée entre les deux types de questions est la classe des ambiguïtés de classe syntaxique. Ces ambiguïtés sont donc plus faciles à résoudre avec les questions humaines qu'avec les questions du système. Cela se vérifie non seulement en examinant les bonnes réponses (Figure VII-22), moins nombreuses, mais aussi les mauvaises réponses (Figure VII-23), plus nombreuses.

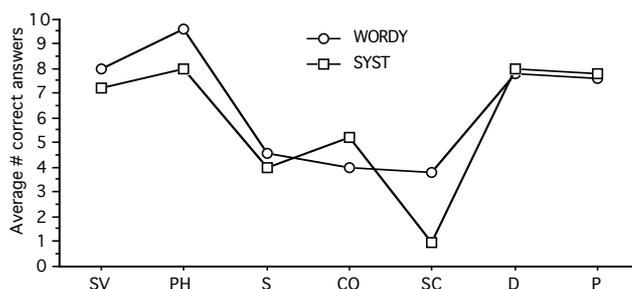


Figure VII-22 : Nombre de réponses correctes pour les questions humaines et machines pour la première expérience en compréhensibilité des questions de désambiguïsation

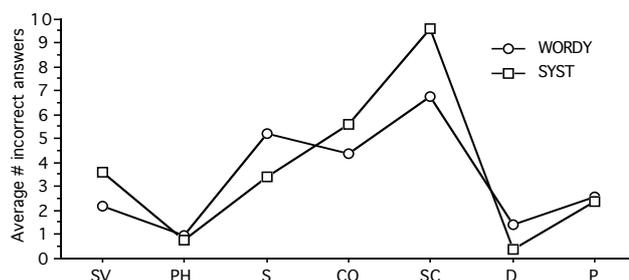


Figure VII-23 : Nombre de réponses incorrectes pour les questions humaines et machines pour la première expérience en compréhensibilité des questions de désambiguïsation

Les classes d'ambiguïté semblent se ranger dans deux catégories en fonction de la manière dont les sujets ont été capables de retrouver la bonne interprétation. Les ambiguïtés de subordination avec verbe, verbe à particule, décoration et polysémie reçoivent un nombre relativement élevé de bonnes réponses, ainsi qu'un nombre relativement faible de mauvaises réponses. Par contre, les sujets ont été moins efficaces pour la résolution des ambiguïtés de subordination, de coordination et de classe syntaxique.

Les situations pour lesquelles les sujets n'ont pas été capables de répondre sont aussi intéressantes. La Figure VII-24 permet d'observer une différence importante entre les questions humaines et les questions machines pour les verbes à particule, la subordination et la coordination. Pour les deux premiers types d'ambiguïtés, les questions machines étaient plus difficiles à comprendre et les sujets n'ont pas pu répondre dans un plus grand nombre de cas. Cependant, pour les ambiguïtés de coordination, les résultats sont inversés. Les sujets ont pu répondre plus aisément aux questions machines.

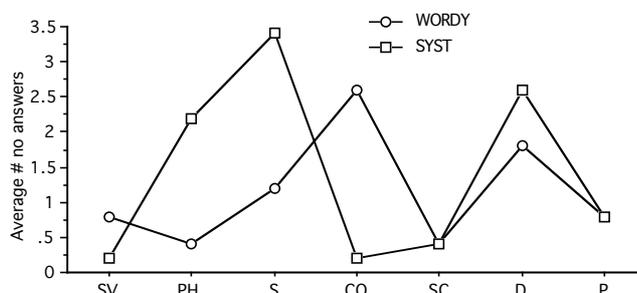


Figure VII-24 : Nombre de questions restées sans réponse pour les questions humaines et machines pour la première expérience en compréhensibilité des questions de désambiguïsation

VII.2.3.2.3. Commentaires

En résumé, on n'observe une réelle différence dans la performance des sujets en fonction du type de question que pour les ambiguïtés de classe syntaxique. Dans ce cas seulement, les sujets répondent mieux aux questions humaines. De manière générale, et quel que soit le type de question, les sujets ont mieux répondu aux questions concernant une subordination avec verbe, un verbe à particule, une décoration et une polysémie qu'aux questions concernant une subordination, une coordination et une classe syntaxique. Les sujets semblaient plus confiants dans leur capacité à répondre à une question humaine (cette capacité étant mesurée par le nombre de réponses non données), à l'exception notable des ambiguïtés de coordination.

À l'issue de l'expérience, les sujets devaient remplir un questionnaire à propos de la difficulté de la tâche en fonction du type d'ambiguïté à résoudre. Afin de leur rappeler le type d'ambiguïté à résoudre, le questionnaire contenait deux exemples de question pour chaque type. Pour répondre aux questions, les sujets se sont presque toujours reportés aux exemples du questionnaire, car ils étaient incapables de généraliser à la classe d'ambiguïté concernée.

En un certain sens, cela n'est pas étonnant car, au sein d'une même classe, il y avait de grosses différences en termes de difficulté de capture de l'interprétation. Être interrompu pendant la lecture n'était pas un gros problème. Nous suspectons que cela est en partie dû au fait que cela était posé comme règle de l'expérience.

Au cours de cette expérience et en discutant avec les sujets, nous avons appris un certain nombre de choses qui ont permis d'améliorer l'expérience suivante. Les sujets se sont souvent plaints du manque de naturel du texte. Il y a à cela deux raisons. Premièrement, le texte comprend des extraits de vrais dialogues oraux transcrits dans le contexte de phrases rédigées. Deuxièmement, afin de conduire les lecteurs vers les sens difficiles, le texte explicite de manière forcée la situation (qui n'est donc pas naturelle). La rédaction semble donc lourde et maladroite.

Nous avons réalisé qu'utiliser des phrases difficiles était une erreur. Cela rend le texte trop maladroit et la tâche plus difficile pour les sujets en brouillant l'objectif de celle-ci : comment

les lecteurs peuvent-ils répondre correctement en fonction de différentes formes de questions. Nous avons aussi changé l'agencement des fenêtres sur l'écran afin que les sujets puissent viser le texte afin de confirmer leur interprétation de phrase ambiguë. Les sujets se sont plaints de ne pas pouvoir faire cela dans l'expérience pilote. C'est aussi un obstacle supplémentaire dans l'accomplissement de la tâche.

Bien que la possibilité de ne pas répondre ait permis d'obtenir des informations intéressantes, elle a empêché l'émergence d'une tendance claire dans la réponse aux questions. Pour cette raison, choisir de ne pas répondre n'a plus été possible dans la seconde expérience.

VII.2.3.3. Seconde évaluation

La seconde évaluation (avril 1996) [Blanchon, H. and Fais, L. 1996c] a pris en compte les recommandations proposées lors de l'évaluation pilote.

VII.2.3.3.1. Données et protocole

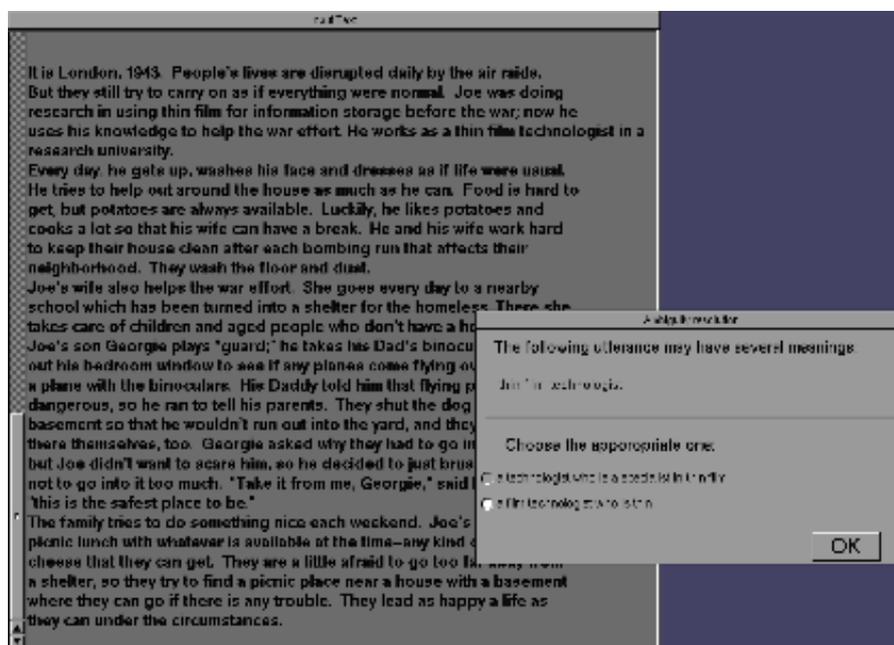


Figure VII-25: Interface de lecture et de désambiguïsation sous forme de boîte de dialogue de la seconde expérience en compréhensibilité des questions de désambiguïsation

Pour cette évaluation, nous avons utilisé 60 sujets répartis en quatre groupes. Ces sujets ont été sélectionnés par une société de services dont le cahier des charges précisait que les sujets devaient avoir l'anglais pour langue maternelle et ne pas être familiers avec les applications du TALN. L'expérimentateur leur présentait la tâche comme l'évaluation d'un système de compréhension de la langue naturelle orale. Ils devaient donc lire un texte⁸³ à haute et intelligible voix en marquant une pause à la fin de chaque phrase afin que le système puisse finir son calcul de compréhension, et poser éventuellement une question (Figure VII-25).

Le texte est constitué de trois histoires différentes. Il comporte 35 phrases ambiguës, à interprétation facile, distribuées également dans sept classes d'ambiguïté retenues lors de l'expérience pilote (Table VII-3).

⁸³ Le texte de la seconde évaluation ainsi que le libellé des questions est fourni en Annexe III.

Abréviation	Type	Exemple
D	Décoration	The lamb did look ready to eat – the lamb will eat – the lamb will be eaten
PH	Verbe à particule	Take it from me – remove it from me – believe me
CO	Coordination	medieval food and drink – medieval drink – not medieval drink
S	Subordination sans verbe	He is a thin film technologist – thin film – thin technologist
P	Polysémie	bank – river side – financial institution
SV	Subordination avec verbe	He brought a bottle of wine from France – a bottle of French wine – a bottle, brought from France, of wine
SC	Classe syntaxique	She washes her face and dresses – she washes her dresses – she dresses

Table VII-3 : Classes d'ambiguïtés (base de travail) et exemples tirés de la seconde expérience

L'Annexe III propose le texte complet de l'évaluation, le libellé des questions, ainsi qu'un tableau qui donne le nombre de bonnes réponses à chaque question pour chaque groupe de sujets.

Comme pour l'expérience pilote, deux ensembles de questions ont été préparés : questions humaines et questions machines. Nous avons aussi utilisé deux modalités pour proposer les questions : des boîtes de dialogue textuelles et des questions orales préenregistrées. Nous avons utilisé cette nouvelle modalité, d'une part pour correspondre à la modalité d'entrée (oral) et d'autre part car l'oral est une modalité intéressante de désambiguïstation [Lehiste, I. 1973, Lehiste, I., *et al.* 1976, O'Shaughnessy, D. 1989, Streeter, L. A. 1978]. À la fin d'une question orale, le sujet pouvait demander à réécouter la question. En combinant le type de question et la modalité, nous avons donc quatre groupes de sujets.

VII.2.3.3.2. Résultats

Questions écrites humaines (groupe HE)

Nous avons observé un fort taux de succès pour tous les sujets excepté pour une question sur une ambiguïté de coordination dans la troisième phrase du paragraphe suivant :

Joe's wife also helps the war effort. She goes every day to a nearby school which has been turned into a shelter for the homeless. There she takes care of **children and aged people who don't have a home.**

pour laquelle nous proposons les interprétations suivantes :

children and aged people who don't have a home

- children who don't have a home and aged people who don't have a home (choix 1)
- any kind of children, and aged people who don't have a home (choix 2)

Nous attendions une réponse qui attache la proposition subordonnée **“who don’t have a home”** à **“children”** et **“aged people”**. Nous avons obtenu sept réponses correctes (choix 1) et huit réponses incorrectes (choix 2). Il semble que, dans ce cas précis, les sujets n’ont pas utilisé le texte pour choisir une interprétation, mais ont fondé leur décision sur leurs propres convictions. L’un des sujets interrogés à propos de sa réponse à cette question nous a répondu « bien sûr on aiderait tous les enfants ». De plus, les erreurs ne sont pas corrélées avec le sexe (cinq hommes et trois femmes).

Questions orales humaines (groupe HO)

Pour l’ambiguïté discutée ci-dessus, nous avons un résultat excellent dans cette situation (une seule erreur). Ce résultat peut s’expliquer par le fait que la modalité orale oblige plus ou moins les sujets à répondre à la volée. Lorsque la première interprétation est choisie, le sujet est moins réceptif à la seconde.

Par contre, pour la seconde phrase du paragraphe suivant :

The trip took him all day, and he was exhausted when he got home. He called Tom and **told him he had had a terrible time that day.**

pour laquelle nous proposons les interprétations suivantes :

he told him he had had a terrible time that day

- he told him on the same day, that he had had a terrible time (choix 1)
- he told him that that day had been terrible (choix 2)

Nous attendions une réponse qui attache **“that day”** à **“a terrible time”**. Nous avons obtenu huit réponses correctes (choix 2) et sept réponses incorrectes (choix 1). Pour la même question, les résultats sont excellents dans les autres situations. Il était peut être délicat de comprendre la différence entre les propositions à l’oral.

Questions écrites du système (groupe ME)

Nous avons observé un fort taux de succès pour tous les sujets, excepté pour les quatre phrases suivantes.

(I) Pour l’ambiguïté de la seconde phrase du paragraphe suivant :

Bill had to go to the library to get a book. But he didn’t want to have to pay **the cost of the taxi to the library.**

pour laquelle nous proposons les interprétations suivantes :

the cost of the taxi to the library

- the cost of the taxi towards the library (choix 1)
- the cost of the taxi for the library (choix 2)

Nous attendions la réponse **“towards the library”**. Nous avons obtenu dix réponses correctes (choix 1) et cinq réponses incorrectes (choix 2).

(II) Pour l’ambiguïté de la seconde phrase du paragraphe suivant :

She didn’t have a date to go to the festival, though, so she wondered what to do. Finally she thought she might call up her **Old English professor** and see if he wanted to go with her.

pour laquelle nous proposons les interprétations suivantes :

Old English professor

- (Old English) professor (choix 1)
- old (English professor) (choix 2)

Nous attendions la réponse “**Old English**”. Nous avons obtenu aussi dix réponses correctes (choix 1) et cinq réponses incorrectes (choix 2).

Par contre, nous avons un réel problème avec les ambiguïtés des deux phrases suivantes.

(III) Pour l’ambiguïté de la seconde phrase du paragraphe suivant :

It was early, so some of the stalls weren’t quite open, but they found one that advertised a special French style of mutton. The professor admitted that he was quite fond of **the French school of cooking** and so they went to see if they could get some mutton stew.

pour laquelle nous proposons les interprétations suivantes :

- French school of cooking
- (French school) of cooking (choix 1)
 - French (school of cooking) (choix 2)

Nous attendions la réponse “**(French school) of cooking**”. Mais nous avons obtenu sept réponses correctes (choix 1) et neuf réponses incorrectes (choix 2).

(IV) Pour l’ambiguïté de la seconde phrase du paragraphe suivant :

Other people were also strolling along the river. Some were eating cakes and **some were eating apples**.

pour laquelle nous proposons les interprétations suivantes :

- some were eating apples
- some eat apples (choix 1)
 - some are (eating apples) (choix 2)

Nous attendions la réponse “**some eat apples**”. Nous n’avons obtenu que quatre réponses correctes (choix 1) et onze réponses incorrectes (choix 2).

L’expérimentateur a très souvent demandé aux sujets ayant mal répondu pour les phrases (III) et (IV) qu’ils lui précisent le sens des deux phrases. Aucun des sujets n’avait mal compris. Ce qui signifie que la question était mal posée, car s’écartant du texte original. Ceci est un argument pour les rephrasages « rustiques ». Nous y reviendrons.

Questions orales du système (Groupe MO)

Dans ce cas la situation s’améliore pour les phrases (I) — plus que trois mauvaises réponses — et (II) — plus que quatre mauvaises réponses. Par contre, la situation se dégrade encore pour les phrases (III) — treize mauvaises réponses contre huit précédemment — et (IV) — quinze mauvaises réponses contre onze précédemment. Pourtant, pour chacune des deux dernières phrases, tous les sujets sauf deux étaient d’accord avec le sens attendu.

Pour la résolution de l’ambiguïté de la phrase (III), la méthode de production du dialogue utilisée est strictement la même que pour les autres ambiguïtés du même type. Le problème d’interprétation de la question est peut-être lié, dans ce cas bien précis, à la façon dont les gens se représentent « le style de cuisine française ».

En ce qui concerne la résolution de l’ambiguïté de la phrase (IV) nous avons, par erreur, utilisé une méthode un peu différente de celle que nous avons utilisée pour résoudre les autres ambiguïtés du même type. Nous aurions dû avoir :

- some were eating apples
- some (were eating) apples (choix 1)
 - some were (eating apples) (choix 2)

Peut-être que cette présentation aurait donné de meilleurs résultats. Nous avons aussi constaté, après réécoute de la question concernant la phrase (IV) que la forme prosodique de la seconde interprétation (choix 2) était similaire à la forme prosodique (intonation, placement des pauses) de la phrase qui signifie que les pommes sont mangées. Cela explique pourquoi cette forme prosodique, du choix 2, qui signifie à l'écoute le choix 1 a été unanimement préférée.

VII.2.3.3.3. Analyse

Les données collectées nous ont permis de faire trois types d'analyse à partir des réponses aux questions de désambiguïsation, du comportement des sujets, et des réponses à un questionnaire.

Analyse statistique

Nous cherchons d'abord à évaluer si la différence de performance entre les groupes **Mx** et les groupes **Hx** est ou non statistiquement significative.

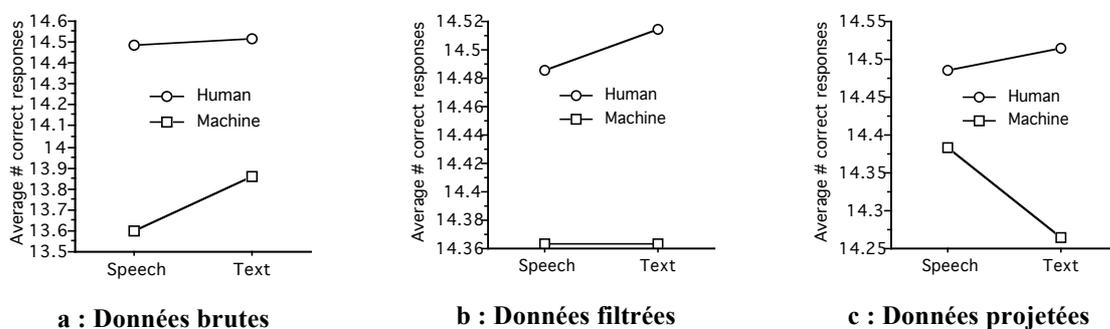


Figure VII-26 : Analyse des données de la seconde évaluation en compréhension des questions de désambiguïsation interactive

En utilisant les 35 phrases ambiguës du texte, la différence de performance entre les groupes **Mx** et **Hx** est significative ($p < 0,05$; Figure VII-26.a). Ce score dépend très fortement des résultats obtenus pour les phrases (III) et (IV). Il semble que les performances des sujets soient meilleures avec les questions écrites qu'avec les questions orales, mais nous ne pouvons proposer aucune conclusion définitive, puisque la différence entre les groupes **xE** et **xO** n'est pas significative.

À cause des questions que nous avons soulevées sur les dialogues de désambiguïsation proposés pour les phases (III) et (IV), nous avons aussi analysé les résultats obtenus en écartant ces deux phrases (Figure VII-26.b). Dans ce cas, il n'y a plus de différence significative entre les performances des groupes **Mx** et **Hx**. Il semble encore que les performances des sujets soient meilleures avec les questions écrites qu'avec les questions orales, mais nous ne pouvons proposer aucune conclusion définitive, puisque la différence entre les groupes **xE** et **xO** n'est pas significative.

Nous avons finalement examiné les données en excluant la phrase (III) et en ajustant les réponses de la phrase (IV) en fonction de notre perception de ce qu'auraient été les réponses avec une question correctement posée (Figure VII-26.c). Ici encore, il n'y a pas de différence significative entre les performances des groupes **Mx** et **Hx**. Les performances des sujets sont différentes entre les questions écrites et les questions orales, mais nous ne pouvons proposer aucune conclusion définitive, puisque la différence entre les groupes **xE** et **xO** n'est pas significative.

Analyse de comportement

Le comportement des sujets peut être évalué selon certains paramètres pertinents pour la tâche.

La vitesse de lecture ainsi que la longueur des pauses entre les phrases est très différente d'un sujet à l'autre. L'expérimentateur n'a jamais demandé aux sujets rapides de ralentir, il les a plutôt conduits à réaliser que leur lecture était trop rapide en déclenchant une question après que le sujet avait commencé la lecture de la phrase suivante. Les sujets s'adaptèrent finalement en réduisant leur vitesse de lecture afin de ne plus être interrompus. Pour les groupes **xO**, nous avons remarqué une adaptation marquée. La faible vitesse d'élocution des questions orales a sans doute influencé la vitesse de lecture des sujets.

Les groupes **xO** avaient la possibilité de demander à réécouter la question. Cependant, cette option a été très rarement utilisée (environ 3,5% des cas) :

- 17 fois pour le groupe **HO** sur 525 questions, soit 3,24%,
- 19 fois pour le groupe **MO** sur 525 questions, soit 3,6%.

Cela peut signifier que la vitesse d'élocution utilisée pour poser une question permettait aux sujets de bien comprendre et différencier les interprétations proposées.

Le temps mis par les sujets pour répondre aux questions est plus court chez les groupes **xO**, qui n'avaient bien sûr pas à lire la question. Chez les groupes **xE**, les temps de réponse varient de cinq à vingt secondes sans que l'on puisse distinguer une différence significative entre les groupes **HE** et **ME**. Les questions du système ne sont pas plus longues à interpréter.

Enfin, les sujets ont tous été à l'aise avec la tâche et confiants dans leurs réponses. Ils ont tous déclaré que la tâche était simple à accomplir.

Analyse du questionnaire

Chaque sujet a rempli un questionnaire post-expérimental. Nous présentons ici quelques-uns des résultats brièvement commentés.

À propos de la difficulté des questions, seulement 15% des sujets ont trouvé qu'il était difficile de répondre (Table VII-4).

Groupe	Très simple	Plutôt simple	Simple	Assez simple	Déroutant simple
HE	2	2	4	8	1
HO	6	3		7	
ME	4	3	4	3	6
MO	3	3	3	8	2

Table VII-4 : Facilité des questions

Les propositions des sujets pour rendre les questions plus facilement compréhensibles sont fournies Table VII-5.

Dans les groupes **xO**, trois sujets ont dit qu'ils auraient préféré que les questions fussent posées sous forme écrite. Ce nombre est certes faible, mais il montre que la modalité pourrait être laissée au choix de l'utilisateur.

Pour les questions orales, les méta-informations de chaque question doivent être les plus courtes possible ; il faut aussi plus utiliser l'intonation.

Pour les questions écrites, des méta-informations typographiques peuvent être utilisées (parenthèses et gras).

Interrogés sur la stratégie qu'ils ont mis en œuvre pour répondre aux questions, les sujets disent utiliser le plus souvent le contexte et la substitution des propositions dans la phrase complète (Table VII-6).

Groupe	Proposition	Num.
HE	Plus de contexte	1
	Une meilleure distinction entre les choix	2
HO	Une voix moins monotone	1
	Des questions plus courtes	5
	Des choix écrits	1
	Être interrogé sur mon propre texte	1
ME	Plus de contexte	3
	Des choix plus verbeux	2
	Des parenthèses + du gras pour l'emphase	1
	Un rephrasage proposé par l'utilisateur	1
	Utilisation de marqueurs grammaticaux	1
MO	Des exemples de phrase au lieu d'intonation et pauses	1
	Des questions plus courtes	3
	Utilisation d'intonation	1
	Mise à disposition d'explications plus détaillées	2
	Choix sur l'écran avec des parenthèses	2

Table VII-5 : Recommandations des sujets

Groupe	Stratégie	Nb.
HE	- Utilisation du contexte	8
	- Substitution dans la phrase	4
HO	- Utilisation du contexte	7
	- Substitution dans la phrase	2
ME	- Utilisation du contexte	9
	- Substitution dans la phrase	3
MO	- Utilisation du contexte	4
	- Substitution dans la phrase	2

Table VII-6 : Stratégie adoptée pour répondre aux questions

Groupe	Prédictibilité des questions & reconnaissance de patrons
HE	3
HO	3
ME	2
MO	5

Table VII-7 : Prédictibilité des questions

Sans que nous ne le leur ayons demandé, certain sujets nous ont dit qu'au fur et à mesure qu'ils progressaient dans leur lecture et la réponse aux questions, les questions devenaient prévisibles en fonction de certains patrons syntaxiques (Table VII-7). Même si nous aurions aimé qu'ils fussent plus nombreux à faire cette remarque, cela indique quand même qu'il y a vraiment des patrons discernables par un non initié, même après un bref entraînement.

VII.2.3.3.4. Commentaires

Si l'on tient compte des problèmes posés par les phrases (III) et (IV), il n'y a donc pas de différence significative entre les performances des sujets, que ce soit en fonction du style

des questions (humain, système), ou en fonction de leur modalité (orale, écrite). Le premier résultat est essentiel au succès d'un module de désambiguïsation interactive. Nous avons aussi constaté que les sujets sont capables d'interpréter les questions sous forme humaine, mais nous ne pensons pas que des questions produites par un système de DI puissent être aussi naturelles. Rappelons que l'une des contraintes que nous nous imposons pour la production automatique de question est de ne pas produire des rephrasages (interprétations) qui puissent être ambigus. Il est donc tout aussi essentiel que les utilisateurs puissent comprendre les questions produites par le système. Les résultats que nous rapportons ici montrent que c'est en effet le cas.

Nous avons aussi cherché à savoir laquelle des modalités permet d'avoir des questions plus compréhensibles. C'est une question de conception, et cela concerne seulement les modalités de sortie que le système doit être capable de manipuler. Les résultats que nous proposons, ainsi que les commentaires de certains des sujets qui auraient préféré des questions écrites plutôt qu'orales, suggèrent que l'une des caractéristiques d'un module de désambiguïsation interactive peut être de proposer plusieurs modalités en sortie et de permettre à l'utilisateur de choisir la modalité qu'il préfère. Notre étude montre que les deux modalités sont également acceptables et « déchiffrables ».

Bien que l'option « répéter la question » n'ait pas été très utilisée, il est tout de même nécessaire de la prévoir pour le cas où l'utilisateur n'aurait pas bien compris la question après la première écoute. D'autres suggestions proposées par les sujets sont facilement implémentables. Par exemple, un contexte plus grand peut être proposé dans la question. Cela facilite la stratégie de réponse mise le plus souvent en œuvre (remplacement du segment ambigu par ses rephrasages dans le contexte de la phrase). Le contenu des questions orales peut aussi être rendu plus compact et leur énonciation plus rapide. Comment mieux utiliser l'intonation et la prosodie reste une question ouverte intéressante.

Dans le futur, il serait aussi très intéressant de faire une telle expérience en utilisant, pour chaque sujet, un support textuel qu'il aurait lui-même produit. Cela permettrait une meilleure analyse de la méthode de désambiguïsation que nous proposons. Une telle expérience serait bien plus longue et coûteuse que les expériences que nous avons réalisées.

VII.3. LIDIA-2 et LIDIA-3

En 2000, nous avons fait un bilan critique de l'architecture informatique des maquettes LIDIA-1 qui a fait apparaître les problèmes suivants :

- L'environnement de rédaction, HyperCard, après une vague de grand succès chez les utilisateurs Macintosh, tombait en désuétude et était inévitablement voué à l'oubli. Cet environnement interdisait aussi, *de facto*, de diffuser largement une interface d'accès aux services LIDIA.
- La communication entre les modules, assurée par des AppleEvents, interdisait de faire coopérer des modules qui n'étaient pas développés sur des plates-formes de type Macintosh. Elle empêchait aussi de déporter le module de désambiguïsation sur une autre machine située sur un sous-réseau distinct de celui de l'environnement de rédaction.
- L'utilisation du courrier électronique pour assurer la transmission de commandes et de données entre le serveur de rédaction et le serveur de traduction était trop lourde.
- Aucun résultat intermédiaire de traitement n'était conservé pour une réutilisation ultérieure éventuelle.

VII.3.1. Nouvelles propositions

L'expérience C-STAR II (Chapitre X) nous a montré l'intérêt de faire communiquer des modules développés sur des plates-formes différentes avec le protocole Telnet, au moyen d'un serveur de communication. Il semblait aussi raisonnable de développer un environnement d'accès aux services LIDIA portable d'une plate-forme à une autre.

VII.3.1.1. Nouvelle architecture informatique

Dans la nouvelle architecture que nous avons proposée, tous les composants logiciels qui interviennent dans la chaîne de traitement LIDIA communiquent à travers un serveur de communication auquel ils sont connectés. Nous avons identifié :

- un client LIDIA qui sert d'environnement de rédaction et de supervision de la traduction des documents,
- un serveur de traduction qui assure les traitements d'analyse et de génération,
- un serveur de désambiguïsation.

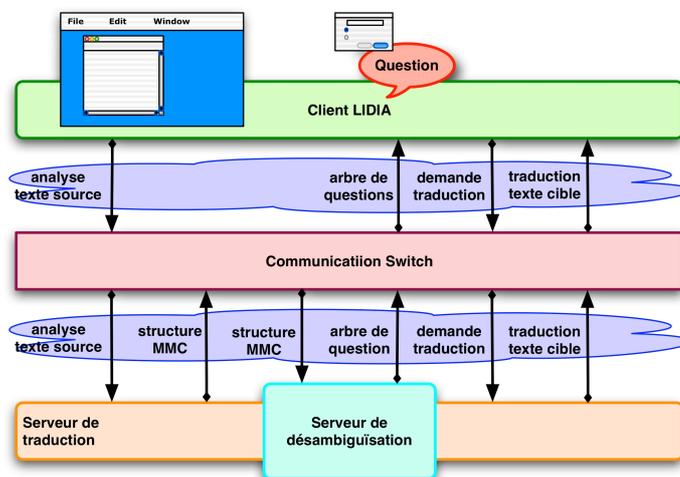


Figure VII-27 : Architecture informatique de LIDIA-2

Le processus de traduction d'un segment de document source se fait comme suit :

- 1) Le client LIDIA envoie un message de demande d'analyse au serveur de communication qui le renvoie au serveur de traduction dans un format adapté.
- 2) Le serveur de traduction traite cette demande, et produit un résultat (une structure mmc) qui est transmis au serveur de traduction qui le renvoie au serveur de désambiguïsation.
- 3) Le serveur de désambiguïsation produit un arbre de désambiguïsation (vide ou non) qui est transmis au serveur de communication qui le transmet au client de rédaction (après transcodage, translittération et reformatage).
- 4) S'il y a des questions à poser à l'auteur, le client le signale. Lorsque l'auteur a répondu aux questions, ou s'il n'y a pas de question, le client de rédaction transmet une demande de traduction au serveur de communication.
- 5) Le serveur de traduction traite la demande de traduction, et produit un résultat qui est transmis au client LIDIA.

La Figure VII-27 illustre les différentes étapes du processus de traduction.

Une première maquette de cette nouvelle architecture a été réalisée par Mohammed Belkhatir [Belkhatir, M. 2001] avec un client LIDIA réalisé en langage Tcl/Tk.

Nous souhaitons que ce client LIDIA puisse gérer un document contenant le texte original augmenté, pour chaque phrase, de l'arbre des questions, d'une trace des réponses fournies par l'auteur, et des traductions produites dans les langues cibles disponibles. Cela s'est avéré trop lourd à implémenter avec le langage Tcl.

Avec Ghislain Gressard [Gressard, G. 2001], nous avons choisi d'utiliser un fichier XML pour représenter le document et de réaliser un nouveau client LIDIA en Java. Une seconde version de la maquette LIDIA-2 a été produite en 2003 [Gressard, G. and Delaval, G. 2003] et [Blanchon, H. and Boitet, C. 2003, Blanchon, H. and Boitet, C. 2004a, Blanchon, H. and Boitet, C. 2004b].

VII.3.1.2. Utilisation d'un document XML

Le document XML produit par LIDIA-2 est manipulé par l'API DOM. Lorsque l'auteur ouvre un document existant, sa syntaxe est vérifiée avec l'API SAX.

Ce document contient un en-tête (`<description>`, Figure VII-28) et un contenu (`<content>`, Figure VII-29) — le texte — sous forme de paragraphes (`<paragraphe>`) et de phrases (`<phrase>`), entrées par l'utilisateur, qui sont enrichies par les informations produites par les différentes étapes de traitement.

Pour chaque phrase du texte, le contenu comprend la langue source, le texte de la phrase, l'arbre des questions, ainsi que la ou les traductions obtenues dans les langues cibles choisies. L'arbre des questions est une représentation *à la Lisp* de l'arbre des questions produit par le module de désambiguïsation, enrichie par une trace du chemin suivi par l'utilisateur lors de la désambiguïsation effective.

```
<description>
  <titre><![CDATA[A trip to Tokyo]]></titre>
  <langue><![CDATA[ENG]]></langue>
  <auteur>
    <prenom><![CDATA[herve]]></prenom>
    <nom><![CDATA[blanchon]]></nom>
  </auteur>
</description>
```

Figure VII-28 : Descripteur d'un document LIDIA-2

```
<phrase source="ENG" stamp="51054803544695">
  <original><![CDATA[ I will show you where you are located right now.]]></original>
  <question>
    <reformulation choix="NON"><![CDATA[I will show you (where you are located right now).]]>
      <analyse><![CDATA[...]]></analyse>
    </reformulation>
    <reformulation choix="OUI"><![CDATA[right now, I will show you where you are located.]]>
      <analyse><![CDATA[...]]></analyse>
    </reformulation>
  </question>
  <traduction cible="FRA"><![CDATA[Je vais tout de suite vous montrer où vous êtes.]]>
</traduction>
</phrase>
```

Figure VII-29 : Extrait d'un fichier LIDIA-2 pour une phrase

VII.3.2. Première mise en œuvre (LIDIA-2)

LIDIA-2 est une version en Java de l'environnement d'accès aux services de TADF. Dans cette version, nous avons utilisé le désambiguiseur de l'anglais [Blanchon, H. 1995c] directement utilisable dans la nouvelle architecture d'intégration des composants que nous avons mise en œuvre. Nous en proposons une rapide démonstration.

L'auteur personnalise d'abord son environnement. Il peut alors créer un nouveau document ou ouvrir un document existant. La fenêtre du document est divisée en deux sections : la partie supérieure est la fenêtre d'édition, la partie inférieure affiche des informations relatives à l'état du traitement du document.

Lorsque l'auteur demande l'analyse du document (Figure VII-30), les phrases ambiguës sont colorées en brun et les phrases non ambiguës (la première phrase) en vert. On peut voir, que le texte contient sept phrases ambiguës et une phrase qui ne l'est pas.

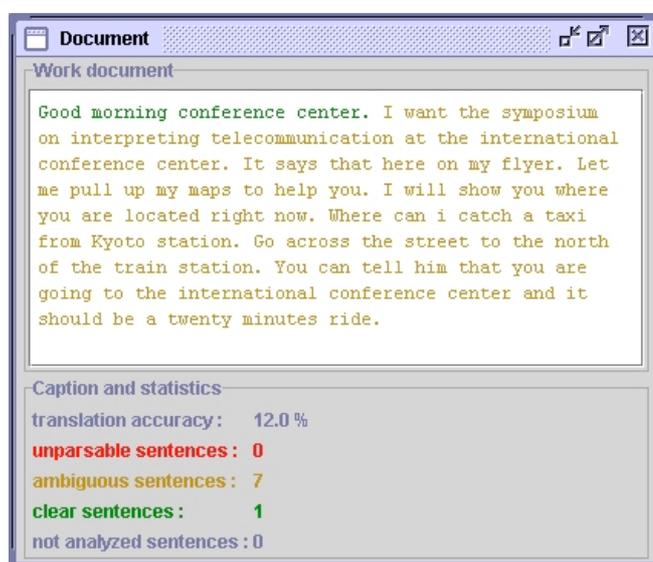


Figure VII-30 : Fenêtre de document LIDIA-2 (après analyse)

Quand l'auteur clique deux fois sur une phrase ambiguë, le dialogue de désambiguïsation relatif à cette phrase est activé. L'ordre des questions correspond à un parcours depuis la racine jusqu'à une feuille dans l'arbre des questions.

Par exemple, lorsque l'auteur choisit la phrase "I want the symposium on interpreting telecommunications at the international conference center", une première question (Figure VII-31) lui est proposée.

La partie inférieure de la fenêtre lui montre qu'il répond à la première question et qu'il devra ensuite répondre à au plus une autre question. À tout moment, il peut aussi arrêter la session de désambiguïsation (bouton Suspend). Lorsqu'il a répondu à une question, il passe à la question suivante avec le bouton Next.

En cours de session (Figure VII-32), l'auteur peut aussi revenir à la question précédente (bouton Previous), ou alors recommencer la session (bouton Reinitialize). Lorsqu'il a répondu à toutes les questions (Figure VII-33), il peut clore la session (bouton Close). Il doit répondre à toutes les questions en une seule fois.

Lorsque le texte du document est traduit, le fichier LIDIA contient un bitexte augmenté d'une trace du parcours de l'auteur dans l'arbre de questions. Ce bitexte peut être affiché (Figure VII-34) et filtré pour obtenir des documents monolingues.

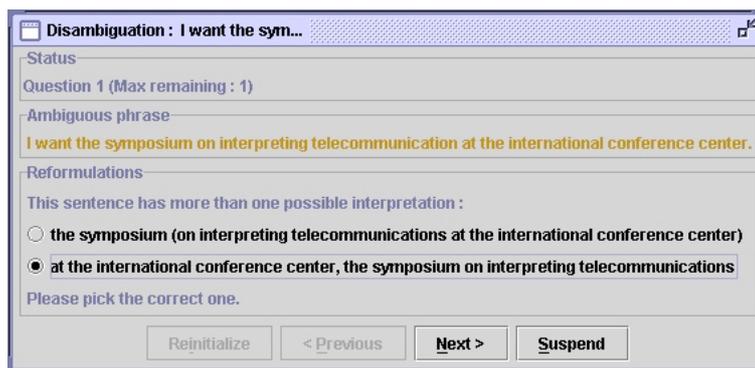


Figure VII-31 : Première question de désambiguïsation de la phrase "I want the symposium on interpreting telecommunications at the international conference center"

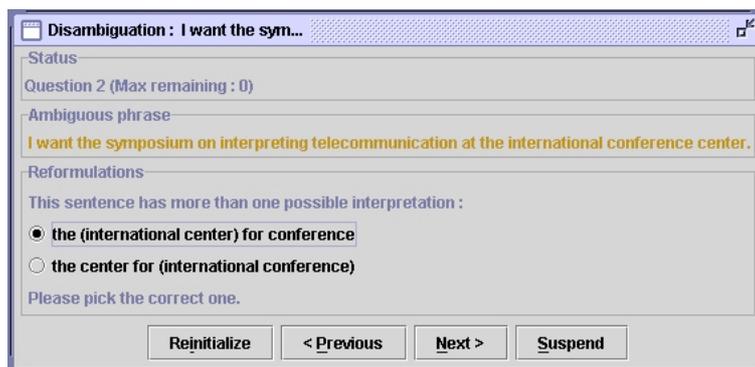


Figure VII-32 : Seconde question de désambiguïsation de la phrase "I want the symposium on interpreting telecommunications at the international conference center"

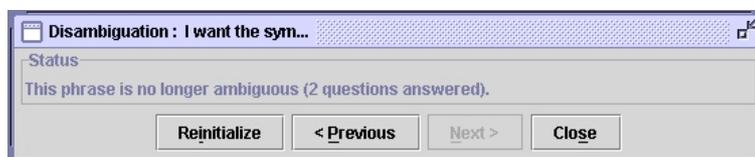


Figure VII-33 : Fin de la session de désambiguïsation de la phrase "I want the symposium on interpreting telecommunications at the international conference center"

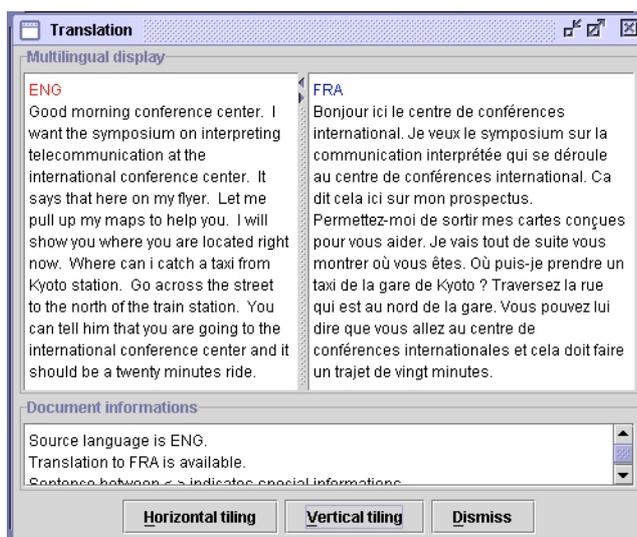


Figure VII-34 : Affichage d'un bitexte

VII.3.3. Vers une meilleure implémentation et de nouveaux scénarios d'usages (LIDIA-3)

Dans la maquette LIDIA-2, l'éditeur de document est un microéditeur XML. Il nous a donc semblé nécessaire d'aller plus loin en intégrant les services LIDIA dans un vrai éditeur XML extensible et de transformer le document LIDIA en un document actif [Quint, V. and Vatton, I. 1994]. La proximité de l'équipe WAM à l'INRIA Grenoble nous a conduit à choisir la plate-forme Amaya⁸⁴ pour intégrer les services LIDIA.

Amaya permet de naviguer sur Internet et est un outil d'édition de documents WYSIWYG : on visualise l'aspect du document final au fur et à mesure de l'édition. La représentation interne d'un document est structurée et conforme à une DTD (Document Type Definition). Les fonctionnalités offertes par Amaya peuvent être étendues grâce à quatre langages. En particulier, un développeur peut définir une nouvelle barre de menu, le contenu des menus et les fonctions qui sont appelées par ces menus (**langage A**). On peut aussi définir de nouvelles DTD avec un langage de déclaration de structures de schéma (**langage S**).

Nicolas Romier [Romier, N. 2004] a travaillé sur cette nouvelle version d'un client LIDIA dite LIDIA-3. Comme nous avons rencontré des problèmes d'implémentation des services LIDIA, et que le temps imparti à ce stage était relativement bref (8 semaines), nous nous sommes concentrés sur la définition de différents modes d'interaction entre un utilisateur et un document LIDIA en établissant différents scénarios. Nous avons identifié quatre types d'utilisateurs (et donc de modes): l'auteur, le lecteur (cf. Chapitre VIII), le traducteur et l'expert.

Ce travail d'implémentation reste donc à terminer. Il faudra aussi intégrer dans la nouvelle architecture les modules d'analyse, de désambiguïsation interactive, de transfert et de génération (vers l'anglais, l'allemand et le russe) développés pour la maquette LIDIA-1. Cela nous permettrait d'avoir une plate-forme d'expérimentation plus riche.

VII.4. Perspectives de recherche

VII.4.1. Permettre une désambiguïsation interactive incomplète

Dans le contexte de vraies applications, l'analyseur rencontrera un grand nombre d'ambiguïtés. Il est donc possible que, pour certaines phases, l'arbre des questions ait une telle profondeur que l'auteur n'accepte de répondre qu'aux questions cruciales.

Supposons, par exemple, qu'une phrase de longueur N possède k^N interprétations et que les descripteurs d'ambiguïté sont constitués en moyenne de p patrons. $(k/p) \cdot N$ questions désambiguïseraient complètement cette phrase. Si par exemple $(k/p)=1/2$, il y aurait alors 120 questions pour une page de 240 mots. Bien qu'il ne faille pas plus de 10 minutes pour répondre à toutes ces questions⁸⁵, si chaque réponse prend 5 secondes, l'auteur peut vouloir consacrer moins de temps à la désambiguïsation interactive.

⁸⁴ <http://www.w3.org/Amaya/>

⁸⁵ Ce temps doit être comparé aux mesures de temps de travail fournies par les traducteurs professionnels. Pour une page et pour chaque langue cible, il faut compter 1 heure pour produire une première traduction et 20 minutes de postédition.

En d'autres termes, étant donnée une structure *mmc*, quelques réponses à des questions de désambiguïisation et, certaines préférences de l'utilisateur, le système devrait être capable de faire des choix et de produire une traduction unique ou alors une représentation factorisée explicitant les différentes traductions possibles en langue cible. Afin d'implémenter une telle stratégie, il est nécessaire que les modules utilisés puissent mettre en œuvre des techniques heuristiques de désambiguïisation automatique ou soient capables de manipuler des structures ambiguës.

VII.4.2. Certifier le sens

À partir du degré de complétion de la désambiguïisation d'une phrase, et en prenant en compte la « crucialité pour la traduction » des ambiguïtés non résolues, il est sans doute possible de calculer un « niveau de certification du sens » associé à la traduction, et de le calculer au niveau des paragraphes, des sections, etc., jusqu'au document lui-même.

VII.4.3. Désambiguïisation par analogie

La technique de désambiguïisation que nous avons proposée nous oblige à décrire explicitement une ambiguïté comme la cooccurrence de différents patrons (sous forme d'arbres) et à associer à chaque patron une méthode de production d'un item de question. Il semble possible de se passer de cette description explicite en ayant recours aux idées développées par Yves Lepage autour du concept d'analogie.

Une analogie [Lepage, Y. 2004] est « une conformité⁸⁶ de rapports énoncée par « A est à B ce que C est à D » et notée $A : B :: C : D$. Au seul niveau des chaînes de caractères, des exemples d'analogies sont : **elle est infirmière : tu es infirmière :: elle est professeur : tu es professeur** ou : **venir : il vient :: prévenir : il prévient**. Une équation analogique consiste en la donnée des trois premiers termes seulement **venir : il vient :: devenir : x**. »

Si V est un alphabet et δ la distance d'édition canonique (insertion et suppression seulement) alors $\forall (A, B, C, D) \in (V^*)^4$:

$$A : B :: C : D \Rightarrow \begin{cases} \delta(A, B) = \delta(C, D) \\ \delta(A, C) = \delta(B, D) \\ \forall a, |A|_a + |D|_a = |B|_a + |C|_a \end{cases}$$

Équation VII-1 : Propriétés des termes participant à une équation analogique

où $|A|_a$ est le nombre d'occurrences de a dans A .

On peut faire jouer l'analogie dans plusieurs espaces à la fois en partant de deux espaces alignés. On impose que les analogies du premier espace soient conservées par homomorphisme entre structures analogiques [Lepage, Y. 2003].

Si A, B , et C sont des chaînes du premier espace et U, V, W des chaînes du second espace respectivement alignées, alors lorsque l'on résoud dans le premier espace l'équation analogique $A : B :: C : x$ on résoudra aussi l'équation analogique $U : V :: W : y$. Cela est illustré par la Figure VII-35.

⁸⁶ [Lepage, Y. 2001] définissait l'analogie comme une égalité de rapports notée $A : B = C : D$, or il n'y a pas transitivité entre équations analogiques : $A : B = C : D$ et $C : D = E : F$ n'implique pas $A : B = E : F$. Pour plus respecter l'usage, le terme « égalité » a été remplacé par le terme « conformité » ; le signe « = » a été remplacé par le signe « :: ».

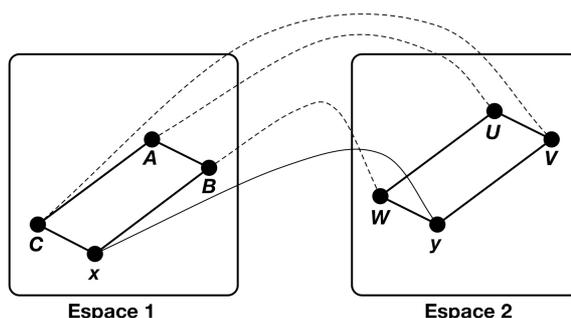


Figure VII-35 : Homomorphisme entre structures analogiques

Si l'**Espace 1** est celui des phrases du français, et l'**Espace 2** celui des phrases du japonais ; les phrases du français étant alignées avec leur traduction en japonais, on peut faire de la traduction par analogie entre le Français et le Japonais. Pour obtenir la traduction en japonais (**y**) de la phrase en français **x**, il faut trouver trois phrases françaises (**f1**, **f2**, **f3**) en relation d'analogie avec **x**. On transpose et on résoud alors l'analogie obtenue avec les phrases japonaises en utilisant les traductions **j1**, **j2**, **j3** de **f1**, **f2**, **f3** et on obtient **y**, la traduction de **x** en japonais.

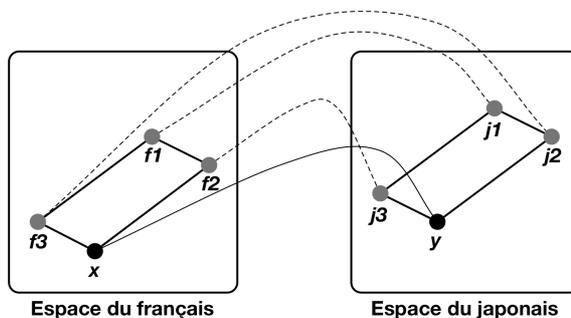


Figure VII-36 : Traduction par analogie

Pour le cas qui nous intéresse, on se propose d'appliquer la même approche pour la production des questions de désambiguïsation. Les espaces que nous considérons alors sont l'espace du français ambigu et l'espace du français désambiguïsé.

Il y aura autant d'espaces du français ambigu \mathbf{EFA}_i que de classes d'ambiguïtés. Ces différents espaces peuvent être distingués, par exemple, au moyen d'une Machine à Vecteurs de Support (SVM). Chacun de ces espaces du français ambigu est associé à un espace du français désambiguïsé \mathbf{EFNA}_i de la manière suivante : une phrase ambiguë $\mathbf{x} \in \mathbf{EFA}_i$ est mise en correspondance avec des énoncés non ambigus $\mathbf{y}_i \in \mathbf{EFNA}_i$ ($2 \leq i \leq n$). n correspond au nombre de patrons que nous utilisons pour décrire l'ambiguïté de type i ; les \mathbf{y}_i correspondent aux items de dialogues que nous produirions pour résoudre l'ambiguïté de type i de l'énoncé \mathbf{x} .

Pour produire la question de désambiguïsation correspondant à une phrase \mathbf{p} qui possède une ambiguïté de la classe i , il faut d'abord **trouver** les termes des équations analogiques qui utilisent \mathbf{p} et \mathbf{x} dans l'espace \mathbf{EFA}_i . En utilisant les trois énoncés correspondant dans \mathbf{EFNA}_i aux trois membres de \mathbf{EFA}_i utilisés dans les équations analogiques dans \mathbf{EFA}_i on va alors **résoudre** autant d'équations analogiques dans \mathbf{EFNA}_i . Les inconnues de chacune de ces nouvelles équations analogiques seront les items de dialogue qui permettent de désambiguïser la phrase \mathbf{p} . Ce processus est illustré par la Figure VII-37 dans le cas où $n=2$.

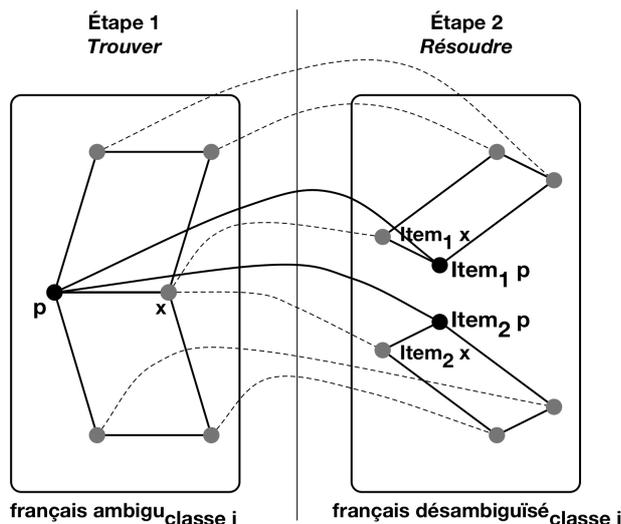


Figure VII-37 : Désambiguïstation par analogie - vue abstraite

L'arête **px** commune aux deux rectangles analogiques dans le français ambigu exprime le fait que aussi bien **p** que **x** sont communs à deux paradigmes. Lorsque Catherine Fuchs [Fuchs, C. 1996] dit que l'ambiguïté appartient au système de la langue, c'est parce qu'il y a une grande collection de points sur une droite, tels que **p** et **x** ici, qui sont à l'intersection de plusieurs paradigmes, deux ici, représentés par les deux rectangles.

Plus concrètement, pour la classe des ambiguïtés de coordination verbale (Annexe V – V.2.1.), une phase modèle de ce type d'ambiguïté en français ambigu doit être associée à deux paraphrases distinctes en français non ambigu.

français ambigu	français non ambigu
Il atteint la grange et la ferme.	Il atteint la grange et il la ferme.
Il atteint la grange et la ferme.	Il atteint la grange et il atteint la ferme.

Lorsque le module de désambiguïstation par analogie doit résoudre l'ambiguïté de la phrase « il observe la photo et la classe », il doit d'abord détecter que l'ambiguïté de cette phrase est du type de l'ambiguïté de la phrase « il atteint la grange et la ferme ». Il pourra alors produire les deux items de dialogue attendus comme le montre la figure suivante.

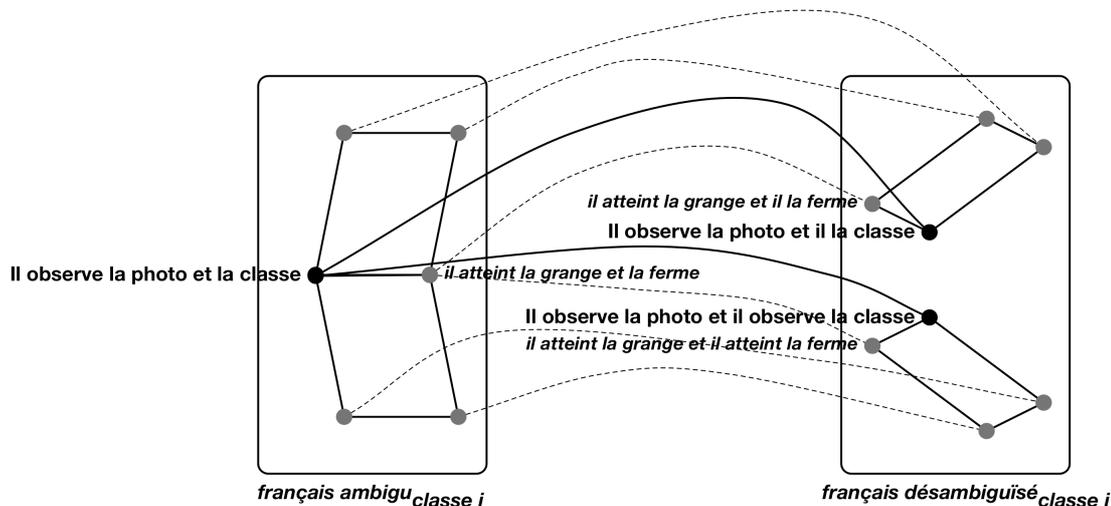


Figure VII-38 : Désambiguïstation par analogie - vue concrète

Si elle est élégante, cette approche pose plusieurs questions. La plus difficile nous semble la suivante : que doivent contenir les espaces du français ambigu afin que l'on puisse trouver les termes d'une équation analogique quel que soit p ? Lorsque l'on aura répondu à cette question, on pourra alors définir la correspondance entre chacun des espaces du français ambigu et leurs espaces respectifs du français désambiguïsé.

Chapitre VIII

Quand la qualité ne suffit pas : les Documents Auto-Explicatifs

Même quand c'est parfait, ce n'est pas
suffisant !

Au cours du développement de la maquette LIDIA-1 [Boitet, C. and Blanchon, H. 1994, Boitet, C. and Blanchon, H. 1995], nous avons naturellement été conduits à l'idée que les informations obtenues par le système lors de la phase de désambiguïsation interactive pourraient être conservées afin d'enrichir le document avec le sens qu'il véhicule. Un Document Auto-Explicatif (DAE) contient le texte du document ainsi que sa représentation linguistique, profonde ou multiniveau, désambiguïsée et augmentée d'une mémoire des ambiguïtés qu'il contient et d'une trace du processus de désambiguïsation.

Avec le concept de DAE, nous proposons une manière, tout-à-fait nouvelle et potentiellement très utile, d'accéder au contenu d'un document et de l'utiliser. Un lecteur de DAE pourrait « cliquer » sur un segment textuel marqué comme ambigu et obtenir une présentation des différentes interprétations possibles avec un marquage de celle qui doit être retenue. Ainsi, un DAE rédigé dans un langage non contrôlé serait moins ambigu qu'un texte rédigé dans un langage contrôlé, qui n'est certes pas ambigu pour la machine, mais le reste très souvent pour un humain.

Nous verrons qu'il est même possible d'assurer l'univocité du document produit en langue cible en produisant un DAE en langue cible, sans intervention humaine cette fois.

VIII.1. Proposition

Au cours de la traduction, ou après l'analyse suivie de désambiguïsation interactive uniquement, les informations nécessaires à la construction d'un DAE en langue source sont conservées. En effet, la production d'un tel DAE ne nécessite pas l'exécution de chacune des étapes du processus de traduction ; il suffit d'une étape d'analyse multiple suivie d'une étape de désambiguïsation interactive.

Nous proposons aussi de produire des DAE en langue cible sans intervention humaine. L'étape de génération produit une structure intermédiaire équivalente à une structure

d'analyse désambiguïsée (*umc*). Il suffit alors, pour créer un DAE en langue cible, de faire une analyse multiple (*mmc*) de chaque phrase effectivement générée, puis de construire un arbre des questions concernant ces phrases.

Sachant que l'on connaît la bonne structure *umc* à retenir, on pourra ensuite calculer automatiquement les réponses aux questions de désambiguïsation qui permettent de la retenir. On disposera alors des informations nécessaires pour produire un DAE en langue cible, et ce sans aucune intervention humaine : le système aura ainsi effectué « la désambiguïsation que l'auteur aurait pu faire lui-même en langue cible » (en supposant qu'il la connaisse et ait la patience de la faire une seconde fois).

Le schéma suivant illustre les changements à effectuer dans l'architecture linguicielle LIDIA.

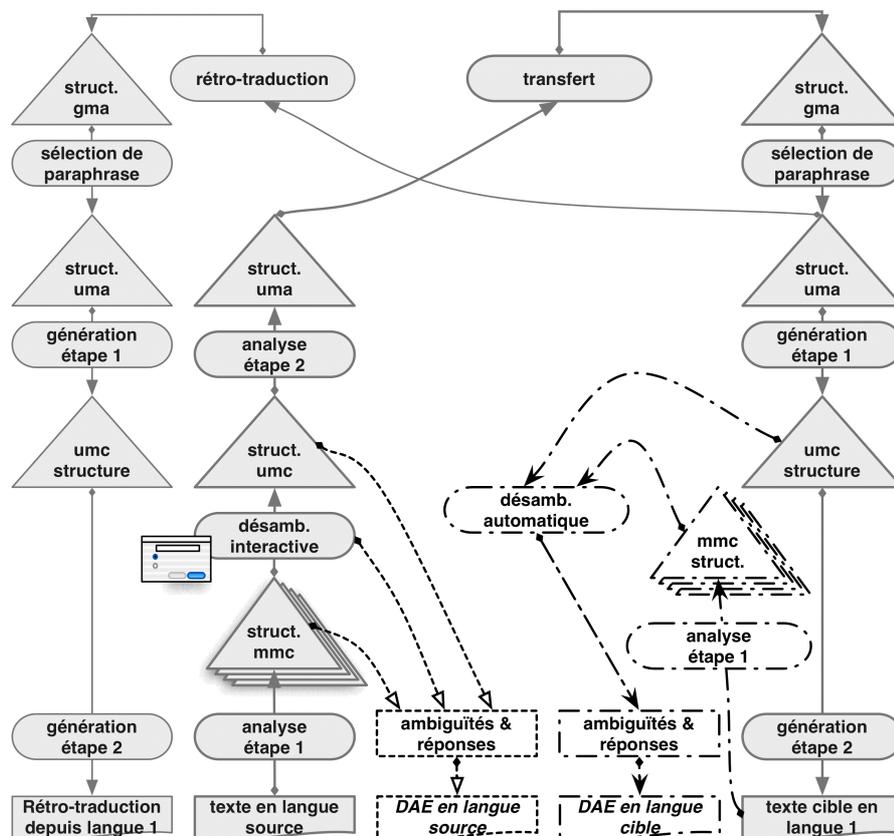


Figure VIII-1 : Architecture linguicielle de LIDIA avec intégration de la construction de DAE

VIII.2. Première mise en œuvre dans le cadre de LIDIA-2

Les travaux de Ghislain Gressard ont été poursuivis, par Eugénie Schonek [Schonek, E. 2003] et Abir Gorayeb [Ghorayeb, A. 2002] dans le cadre de son stage de DEA.

Un Document Auto-Explicatif est un document LIDIA-2 filtré dans lequel, pour chaque phrase, on ne retient que le parcours de l'auteur dans l'arbre de questions. Il s'agit du texte non grisé de la Figure VIII-2.

```

<phrase source="ENG" stamp="51054803544695">
<original><![CDATA[ Let me pull up my maps to help you.]]></original>
<question>
<reformulation choix="NON"><![CDATA[Let me pull up (my maps to help you).]]>
<analyse><![CDATA[...]]></analyse>
</reformulation>
<reformulation choix="OUI"><![CDATA[to help you, let me pull up my maps.]]>
<analyse><![CDATA[...]]></analyse>
</reformulation>
</question>
<traduction cible="FRA"><![CDATA[Je vais tout de suite vous montrer où vous êtes.]]>
</traduction>
</phrase>
    
```

Figure VIII-2 : Extrait d'un document auto-explicatif LIDIA-2

Un tel visualiseur de DAE doit permettre à un lecteur de lire le contenu du document et d'appréhender le « sens exact » de ce que l'auteur a voulu dire. À cette fin, le lecteur doit être prévenu que certains segments du texte peuvent avoir plusieurs interprétations (« sens »). Il faudra donc que le visualiseur puisse marquer les segments ambigus. Le lecteur pourra alors désigner un segment ambigu et demander à voir une explication.

Il n'est pas possible pour l'instant (cf. VIII.3.1) de localiser précisément les segments ambigus. Avec le premier visualiseur de DAE, pour obtenir les informations relatives aux différentes interprétations possibles d'une phrase, le lecteur doit cliquer deux fois sur son texte. Une boîte de dialogue apparaît alors. Elle permet au lecteur de naviguer dans l'arbre des questions en voyant les rephrasages sélectionnés par l'auteur (\Rightarrow ... \Leftarrow) lors de la désambiguïsation, comme le montre la Figure VIII-1.

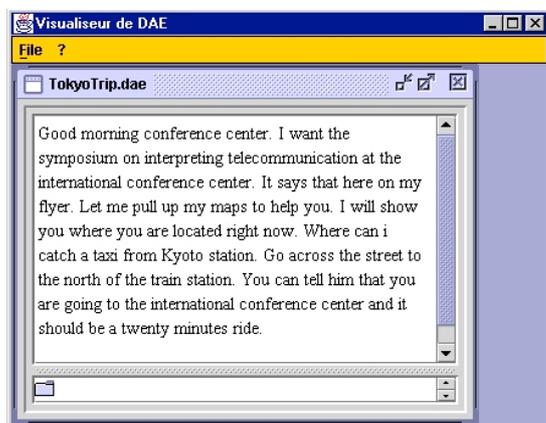


Figure VIII-3 : Interface du visualiseur de DAE

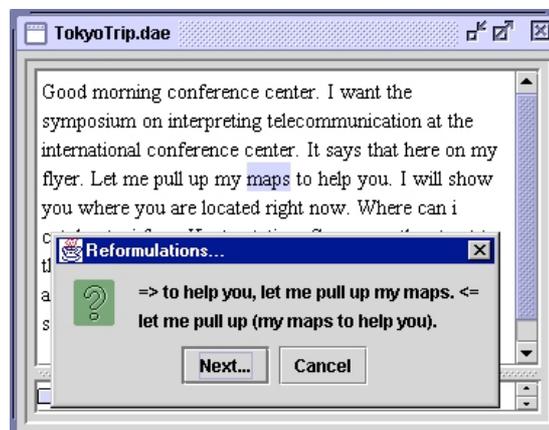


Figure VIII-4 : Affichage de la lecture à retenir pour la phrase sélectionnée

VIII.3. Perspectives de recherche

VIII.3.1. Présenter le support de l'ambiguïté

Pour améliorer l'interface de présentation d'un DAE, il faudrait que les supports des ambiguïtés soient indiqués (Figure VIII-5) afin que le lecteur puisse identifier précisément les segments qui posent un problème d'interprétation. Or les patrons que nous utilisons actuellement (cf. VII.2.1.1) n'utilisent pas la notion de support de l'ambiguïté. En effet, ils

capturent très souvent un segment plus grand que le support afin de permettre un rephrasage compréhensible de l'ambiguïté.

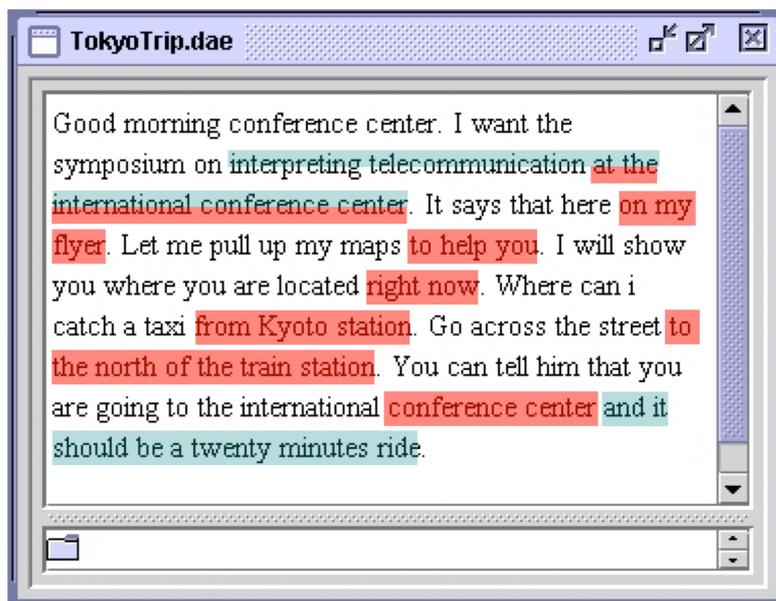


Figure VIII-5 : Présentation des supports d'ambiguïté d'un DAE

On peut même envisager d'utiliser des couleurs différentes pour identifier chacun des types d'ambiguïté.

Il faudrait donc que le processus de désambiguïtation interactive fournisse cette information. Deux voies sont possibles.

Dans une première approche, on peut choisir d'attacher la description du support de l'ambiguïté à chaque faisceau. Le support peut, en effet, être défini à partir des variables utilisées dans les patrons.

La seconde approche, proposée dans [Boitet, C. and Tomokio, M. 1995 1995a, Boitet, C. and Tomokiyo, M. 1995 1995a], oblige à changer la description des ambiguïtés, en utilisant uniquement leur support. Cependant, pour faire des rephrasages du type que ceux que nous produisons actuellement, il faudrait décrire les informations supplémentaires à utiliser lors de la fabrication des items de dialogue.

Nous avons déjà fait des études dans ce sens. Vincent Frey [Frey, V. 1996] a commencé à étudier la construction automatique de descripteurs d'ambiguïté à partir d'un corpus d'arbres multiresolution. Les informations manquant pour le rephrasage pourraient alors être retrouvées au moyen d'heuristiques [Irgadian, F. 2002].

VIII.3.2. Créer des DAE en langue cible

Montrons maintenant comment l'architecture que nous proposons permet aussi de produire des DAE en langue cible. Comme l'étape de génération produit une structure intermédiaire équivalente à une structure d'analyse désambiguïtée (*umc*), il suffit de faire une analyse multiple (*mmc*) des phrases effectivement générées, puis de construire un arbre des questions concernant ces phrases.

Sachant que l'on connaît la structure *umc* à retenir, on peut calculer automatiquement les réponses aux questions de désambiguïtation à la place d'un lecteur en langue cible. On pourra donc produire un DAE en langue cible sans intervention humaine.

Atteindre cet objectif est cependant difficile en pratique puisqu'il faut disposer d'un analyseur multiple dans chacune des langues traitées (source et cibles). Nous espérons construire un prototype complet implémentant cette idée grâce à des coopérations internationales.

Conclusion

L'architecture informatique que nous avons proposée pour les maquettes LIDIA-2 et LIDIA-3 nous semble adaptée aux exigences de la diffusion de la TADF pour tous. L'utilisation d'un document XML permet de représenter de façon homogène le contenu textuel, les informations collectées lors de la désambiguïsation interactive, et les traductions dans plusieurs langues cible.

Nous avons montré que des questions intéressantes se posent encore sur le thème de la désambiguïsation interactive. Pour faire une étude « grandeur nature » de la réalisation d'un module de désambiguïsation utilisant les patrons et/ou l'analogie, il faut avoir accès à un analyseur à large couverture qui produise des sorties multiples. Nous espérons pouvoir réaliser prochainement des expériences avec Mark Seligman (Spoken Translation Inc.⁸⁷) et PT de Linguatéc sur la désambiguïsation par patrons, et avec Yves Lepage (ATR-SLT) et l'analyseur de l'anglais Link Grammar⁸⁸ sur la désambiguïsation par analogie.

Nos travaux en TADF et en désambiguïsation interactive nous ont permis de proposer le concept de Documents Auto-Explicatifs. Ce concept n'est pas né au sein de la communauté du document numérique, parce qu'elle n'envisage pas, jusqu'à présent, la possibilité de faire une analyse complète des énoncés, mais s'intéresse principalement à l'édition et au fromatage des documents, et, depuis l'avènement de XML, à l'annotation des documents, mais cela sans que leurs auteurs ne soient impliqués. Même si nous n'en sommes qu'à nos premiers pas, et si nous nous sommes rendu compte qu'il faudrait revoir le processus de désambiguïsation pour pouvoir présenter le support de l'ambiguïté dans un DAE, nous avons prouvé que notre proposition est implémentable. Une telle **aide à l'accès au sens** représentera une innovation majeure, impossible, et même inimaginable, avant l'avènement de l'informatique, et des traitements puissants et conviviaux qu'elle permet, en particulier en traitement des langues.

⁸⁷ <http://www.spokentranslation.com/>

⁸⁸ <http://www.link.cs.cmu.edu/link/>

Troisième Partie

**Communication interpersonnelle
orale spontanée spécialisée**

Introduction

Lorsqu'on examine la littérature en traduction de l'oral, on est d'abord surpris par le fait que les auteurs utilisent indistinctement plusieurs termes pour nommer la même chose, ou le même terme pour désigner des choses différentes. On rencontre ainsi : traduction de la langue orale (spoken language translation), traduction de dialogue oral (spoken dialogue translation), traduction orale (spoken translation), traduction de parole (speech translation), traduction parole à parole (speech-to-speech translation).

Dans ce document, nous utiliserons les termes

- **traduction de parole** pour parler des systèmes qui traduisent la parole dans un sens : entrée orale en langue source et sortie orale ou écrite en langue cible. Il s'agit par exemple de traduire des conférences, ou de diffuser de l'information.
- **traduction de dialogue** pour parler des systèmes qui traduisent la parole dans un dialogue, avec entrée orale et sortie orale dans les deux langues. Ce sont donc des systèmes bidirectionnels.

La traduction de dialogue a quinze ans cette année, est-ce l'âge de raison ? À notre avis... non ! Pourquoi ? Voici deux premiers éléments de réponse.

Comme nous allons essayer de le montrer, le domaine de la traduction de dialogue ignore le lien de parenté, voire même de filiation qui le lie au domaine, bien plus ancien, de la **communication orale personne-système informationnelle**. Ce domaine a beaucoup de choses à nous apprendre, mais le fossé qui le sépare de la traduction de parole s'élargit au lieu de se combler.

Le domaine de la traduction de parole a connu des projets ambitieux autour des années 2000 (C-STAR II, Verbmobil, NESPOLE!). Ces projets ont produit des résultats qui nous semblent très acceptables, mais aucune étude en utilisabilité n'a été réalisée à ce jour. Ainsi, nous ne savons pas dans quelle mesure les résultats produits peuvent être utiles pour des utilisateurs potentiels. En 2000, Alex Waibel (CMU) proposait l'objectif suivant aux partenaires du consortium C-STAR III : "We should eat our own dog food!"⁸⁹. Cet objectif, pourtant raisonnable pour qui veut faire des systèmes pour des utilisateurs, a été rapidement oublié au profit de la construction de corpus et de l'organisation de campagnes d'évaluation compétitive.

⁸⁹ Que l'on pourrait traduire par « Nous devons utiliser les systèmes que nous produisons ! »

Après avoir dressé un historique et analysé le domaine de la traduction de l'oral, nous présentons trois grands projets internationaux en traduction de dialogue oraux finalisés (C-STAR II, Verbmobil, et NESPOLE!). Le chapitre suivant est consacré à nos contributions au sein des projets C-STAR II et NESPOLE!. Dans C-STAR II, nous avons principalement animé l'équipe CLIPS++ et nous avons réalisé une première maquette d'un module d'analyse par patrons. Le projet NESPOLE! a été l'occasion d'implémenter des modules d'analyse et de génération pour les deux démonstrateurs.

Le dernier chapitre est consacré à notre contribution à la première campagne d'évaluation compétitive organisée par le consortium C-STAR III. Nous dressons ensuite une liste des problèmes auxquels nous pensons que le domaine est confronté. Nous terminons en proposant des pistes de recherche qui, si elles sont suivies, devraient permettre d'en résoudre certains. À cette occasion, nous donnons aussi les résultats de deux expériences visant à améliorer l'intégration des composants, en entrée, pour les systèmes de traduction de parole et de dialogue.

Chapitre IX

Historique et analyse du domaine

L'apparition de la traduction de dialogue s'est faite en trois étapes. Il y a d'abord eu l'étape de recherche autour des systèmes de communication orale et même multimodale entre personnes et systèmes d'information pour des domaines restreints. Les progrès parallèles en reconnaissance automatique de la parole et en traitement automatique de la langue écrite ont ensuite permis à la communauté de s'investir en traduction de parole. Il s'agit de traduire des énoncés produits par un utilisateur qui interagit avec un système informationnel ; donc encore sur des domaines très restreints. Une dernière étape de maturation des outils a permis d'envisager de s'intéresser à la traduction de dialogue sur des domaines plus ouverts.

Comme nous essaierons de le montrer, la communauté a perdu en quelque sorte la trace de ces origines. D'une part, les premiers obstacles à affronter à cause de l'élargissement du domaine ont mobilisé les énergies. Il s'agit d'abord de reconnaître la parole spontanée afin de pouvoir l'analyser. L'autre cause d'oubli des origines est liée à l'utilisation de méthodes statistiques maintenant très largement répandues dans le domaine.

IX.1. Communication orale entre personnes et systèmes d'information

Étrangement, on ne mentionne jamais les travaux sur les systèmes de communication orale entre personnes et systèmes d'information (COP-SI) dans la communauté scientifique qui s'intéresse à la traduction de dialogue. C'est sans doute un tort ! En effet, les systèmes de traduction de parole actuels, limités à un domaine et à certaines tâches, peuvent être considérés comme des systèmes d'accès à des systèmes non plus informatiques, mais humains.

Pour un état de l'art exhaustif des techniques, méthodes, et outils qui interviennent dans un système de communication entre personnes et systèmes informationnels, le lecteur peut consulter [McTear, M. F. 2002].

Nous présentons ici les approches qui d'une manière générale ne sont pas très éloignées des techniques utilisées en TA de parole, mais avec la gestion du dialogue en plus.

Le but de cette section est de donner une source d'inspiration pour le futur de la TAP.

IX.1.1. Stratégies de dialogue et domaines d'application

Un système de communication orale entre personnes et systèmes d'information (COP-SI) permet à un utilisateur d'interagir avec une application informatique telle qu'une base de données ou un système expert en utilisant la parole.

Un COP-SI doit :

- comprendre ce que lui demande l'utilisateur en utilisant un module de reconnaissance de la parole et un module de compréhension en contexte, puis
- verbaliser une réponse en utilisant un module de génération de la langue naturelle et un module de synthèse de la parole, et enfin
- engager un dialogue avec l'utilisateur.

Les COP-SI mettent en œuvre différentes stratégies de conduite du dialogue. Ils peuvent :

- garder l'initiative en conservant le contrôle de la conversation. Dans ce cas, l'utilisateur est très contraint ; à l'extrême, l'utilisateur est complètement passif, il répond uniquement à des questions du système.
- laisser l'initiative du contrôle de la conversation à l'utilisateur. Dans ce cas, c'est l'utilisateur qui contrôle la conversation ; à l'extrême, le système est complètement passif et répond uniquement aux questions de l'utilisateur.
- permettre des prises d'initiatives alternées entre le système et l'utilisateur. Dans ce cas, le contrôle de la conversation est partagé entre le système et l'utilisateur qui sont à tour de rôle initiateur et suiveur.

Les domaines d'application des systèmes de COP-SI sont :

- le renseignement sur des horaires de train et d'avion, avec éventuellement la possibilité de réserver.
- le renseignement météorologique.
- la gestion de comptes bancaires.
- le routage de communication téléphonique.

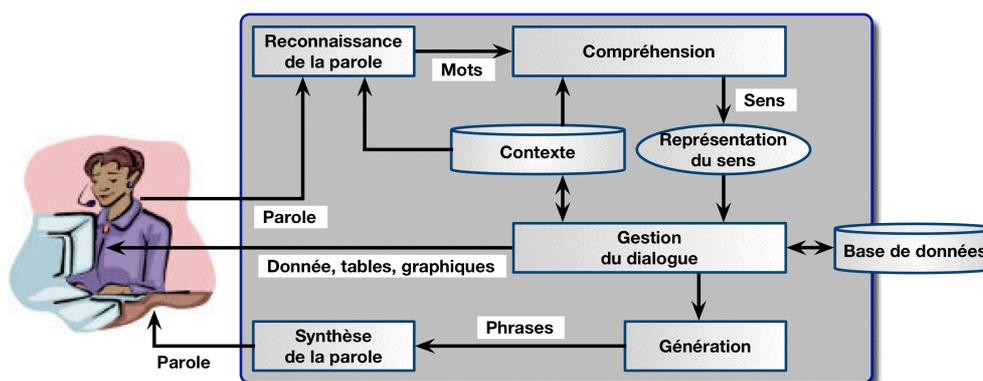


Figure IX-1 : Architecture et composants des systèmes de COP-SI

IX.1.2. Architectures et composants

L'architecture et les composants d'un COP-SI sont schématisés Figure IX-1 ci-contre.

IX.1.3. Galaxie GALAXY

Parmi les nombreux travaux sur les systèmes de COP-SI, ceux réalisés par l'équipe Spoken Language Systems du MIT⁹⁰ sont particulièrement intéressants⁹¹. En effet, ce groupe travaille depuis plus de quinze ans sur le sujet et a abordé différents aspects du domaine.

IX.1.3.1. GALAXY-I

Le premier système GALAXY [Goddeau, D., *et al.* 1994] capitalise les premiers efforts du groupe autour de trois systèmes VOYAGER (guide de ville et assistance au trafic urbain), ATIS (information et réservation de vol), et PEGASUS (informations sur le trafic aérien).

L'architecture est distribuée : un client accède, via Internet, à des serveurs spécialisés capables de répondre à un certain nombre de requêtes sur des domaines spécialisés.

IX.1.3.2. GALAXY-II

Dans le cadre du programme DARPA Communicator, l'architecture informatique de GALAXY a été revue pour donner naissance à GALAXY-II, dont l'architecture est devenue une référence pour le développement de systèmes de COP-SI [Seneff, S., *et al.* 1998, Seneff, S., *et al.* 1999].

La nouvelle architecture conserve la nature client-serveur de la première version. Un serveur de communication (HUB, Figure IX-2) assure la communication entre les différents modules.

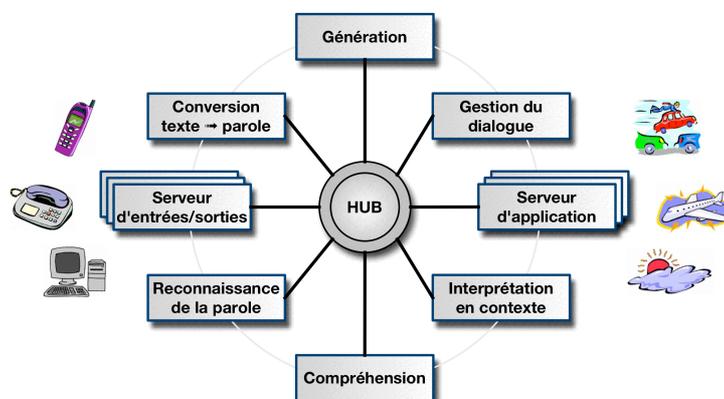


Figure IX-2 : Schéma de l'architecture GALAXY-II

Un formalisme de représentation sémantique est maintenant proposé aux développeurs de nouvelles applications dans le modèle GALAXY-II. Par exemple, pour la requête "Is there a flight from Boston to San Francisco Friday?", on obtient la représentation sémantique suivante :

⁹⁰ Plus d'informations sont disponibles à <http://www.sls.csail.mit.edu/sls/research/>

⁹¹ Bien sûr, beaucoup d'équipes travaillent sur ce domaine, mais notre but n'est pas d'être exhaustif.

Clause: EXIST
 Topic: FLIGHT
 Quantifier: INDEF
 Predicate: SOURCE
 Topic: CITY
 Name: Boston
 Predicate: DESTINATION
 Topic: CITY
 Name: San Francisco
 Predicate: TIME
 Topic: DATE
 Day: Friday

Figure IX-3 : Représentation sémantique en GALAXY-II

La galaxie GALAXY est composée des systèmes suivants :

- JUPITER : informations météorologiques mondiales (plus de 500 villes),
- MERCURY : information et réservation de vol,
- PEGASUS : information en temps réel sur le trafic aérien,
- VOYAGER : guide de ville et assistance au trafic urbain,
- ORION : agent personnel pour des services automatisés hors-ligne.

IX.2. Traduction de parole

Le premier pas vers la traduction de dialogue a été la traduction d'énoncés produits par des utilisateurs de COP-SI. Il ne s'agit donc pas de traduction de dialogue proprement dite. Cette démarche permet d'exploiter des corpus déjà collectés, transcrits et annotés dans le cadre de projets de COP-SI.

IX.2.1. Architecture et composants

L'architecture et les composants d'un Système de Traduction de Parole sont schématisés Figure IX-4.

Comme les énoncés à traduire sont sortis de leur contexte dialogique, celui-ci ne peut plus être exploité. Les énoncés sont donc traduits en exploitant les informations qu'ils contiennent.

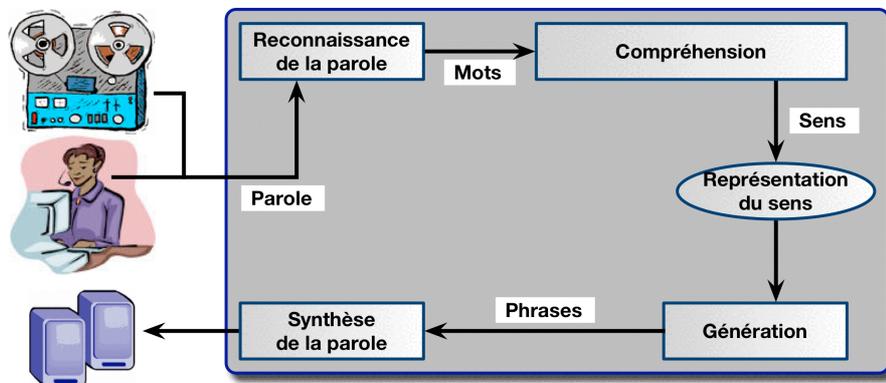


Figure IX-4 : Architecture d'un système de traduction de parole

IX.2.2. Spoken Language Translator

Le meilleur exemple de projet de traduction de parole est le projet SLT (1992-99) du SRI [Rayner, M., *et al.* 1993, Rayner, M., *et al.* 2000]. Il s'agit de traduire des énoncés produits par un utilisateur d'un COP-SI de renseignement et de réservation de vol : ATIS (Air Travel Information System). Il s'agit donc uniquement de traduire des requêtes, **et non pas d'une vraie situation de dialogue.**

IX.2.2.1. Architecture et composants

Le schéma suivant illustre les différentes étapes du processus de traduction.

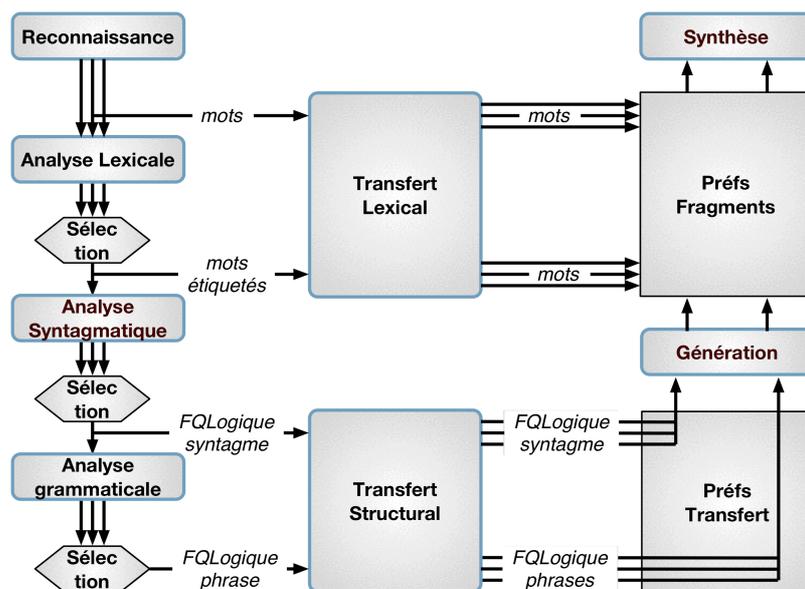


Figure IX-5 : Architecture linguistique de SLT

La traduction se fait via une étape de transfert sur différents niveaux de représentation du tour de parole : transfert lexical au niveau des occurrences, transfert lexical au niveau des occurrences étiquetées par des classes syntaxiques, transfert de syntagmes représentés sous une forme quasi logique, transfert de phrases représentées sous une forme quasi logique. Les formes quasi logiques sont soumises à un module de génération qui produit un texte en langue source. À chaque niveau, un module de sélection ne conserve que les hypothèses les plus probables.

Les niveaux de traduction inférieurs à un niveau donné servent de filet de sécurité pour le cas où ce niveau ne permettrait pas de produire une représentation. Cette méthode permet notamment de traduire des fragments.

Le modèle de langage du module de reconnaissance de l'anglais comporte 1751 mots, celui du suédois, 1265 mots.

IX.2.2.2. Évaluation

Les prototypes développés permettent une traduction bidirectionnelle entre l'anglais et le suédois (pour un client), ainsi que la traduction depuis l'anglais et le suédois vers le français. Il existe aussi une première version du système permettant de traduire du français vers l'anglais. Trois types d'évaluation ont été conduites. Nous nous intéressons ici à l'évaluation de la qualité de la traduction sans la synthèse.

IX.2.2.2.1. Données et protocole

Les systèmes de l'anglais vers le français et vers le suédois ont été évalués dans plusieurs versions sur un corpus de test non utilisé dans la phase de développement. Ce corpus comporte 400 tours de parole.

Les auteurs ont choisi une échelle de sept valeurs qu'ils regroupent finalement en trois catégories : traduction clairement utile (complètement acceptable, style non naturel, erreurs syntaxiques mineures), traduction à utilité mitigée (erreurs syntaxiques majeures, traduction partielle), traduction inutile (absurdité, mauvaise traduction).

Les performances des systèmes ont été évaluées sur les résultats de la reconnaissance. La qualité de l'hypothèse de reconnaissance produite a été mesurée avec un critère binaire (satisfaisante ou non) un peu similaire à ce qui a été fait dans le cadre du projet NESPOLE! (cf. section XI.3).

IX.2.2.2.2. Résultats

Dans le tableau suivant, nous présentons les résultats consolidés selon les trois catégories présentées ci-dessus.

Version du Système ⁹²	LFT	LexT	LPT	LFT	LFT ⁹³	LexT	LPT
<i>Langue cible</i>	<i>Swe</i>	<i>Swe</i>	<i>Swe</i>	<i>Fra</i>	<i>Fra</i>	<i>Fra</i>	<i>Fra</i>
Reconnaissance OK	80,2	76,2	74,8	71,0	100	67,5	68,0
Traduction utile	75,0	60,3	61,7	65,3	74,5	18,5	41,3
Traduction limite	16,0	24,8	23,3	7,0	14,4	40,8	30,3
Traduction inutile	9,0	14,9	14,9	27,8	11,1	40,7	28,5

Figure IX-6 : Résultat de l'évaluation SLT

IX.2.2.2.3. Commentaires

Les première et quatrième colonnes montre que le système complet produit 75% et 65% de traductions utiles vers le suédois et le français respectivement. Pour le système anglais → français, on observe un gain de 20% par rapport au système dans sa configuration LPT. Pour le couple anglais → suédois, le progrès n'est pas aussi franc. L'ordre des mots de l'anglais est plus proche de l'ordre des mots du suédois que du français, ce qui peut expliquer ces résultats. Pour cette même raison, les performances pour le suédois sont meilleures que pour le français.

IX.2.3. AT&T

Les laboratoires de recherche de AT&T ont conduit une expérience similaire avec leur système d'acheminement automatique de communications téléphoniques "How May I Help You" [Gorin, A. L., *et al.* 1997] qui a été proposé à des clients internes à partir de 2000 et à des clients externes deux ans plus tard. Depuis l'an 2000, une version multilingue est en cours de développement en utilisant une technologie fondée sur des transducteurs d'états finis [Bangalore, S. and Riccardi, G. 2001]. Actuellement, AT&T travaille sur des

⁹² LFT : système complet, LexT : transfert de surface uniquement, sans passage par la forme quasi-logique, LPT : transfert de surface et sur les formes quasi-logiques au niveau des syntagmes.

⁹³ Dans cette configuration, l'évaluation est faite sans la reconnaissance. On travaille sur les transcriptions (références) en entrée de traduction.

techniques pour la traduction en temps réel d'émissions radio-télédiffusées dans le cadre du projet Anuvaad⁹⁴.

IX.3. Traduction de dialogue

La traduction de dialogues est vraiment née en 1986 lors de la création du laboratoire ATR-ITL au Japon. Depuis, ce thème de recherche a attiré des équipes chaque année plus nombreuses. Nous avons choisi de présenter les activités du domaine par équipe de recherche et projet. Cet état de l'art ne se veut pas exhaustif, car les approches mises en œuvre sont trop nombreuses pour être toutes décrites, discutées et comparées. Le lecteur trouvera un nombre important de références bibliographiques pour une étude approfondie.

IX.3.1. États-Unis

Les deux centres de recherche les plus actifs en traduction de dialogue sont l'Université Carnegie Mellon à Pittsburgh et le laboratoire IBM T.J. Watson Research Center à Yorktown Heights.

IX.3.1.1. CMU et UKA

Deux équipes travaillent sur la traduction de dialogue ou de parole à CMU. La première équipe, jumelée avec une équipe de l'Université de Karlsruhe (UKA), est connue sous le nom ISL (Interactive Systems Labs) et est dirigée par le Professeur Alex Waibel. La seconde équipe est celle qui a développé le système de compréhension de la parole SPHINX.

IX.3.1.1.1. JANUS

Les versions du système de reconnaissance de la parole et de traduction JANUS, développées par l'ISL, ont été utilisées lors des différentes phases du projet C-STAR.

Le système JANUS-I [Woszczyzna, M., *et al.* 1993] permet de traduire une entrée orale en anglais ou en allemand vers l'anglais, l'allemand et le japonais. C'est ce système qui a été utilisé lors des démonstrations du premier consortium C-STAR.

JANUS-I intègre un module de reconnaissance de la parole des modules d'analyse et de génération. Trois modules d'analyse sont disponibles. Le module principal d'analyse est un analyseur syntaxique LR qui utilise une grammaire indépendante du contexte. S'il échoue, une analyse par patrons sémantiques, puis une analyse selon un modèle connexionniste sont alors essayés. La sortie du module principal est une f-structure qui est transformée en une représentation sémantique pivot.

Pour la génération, la représentation pivot est d'abord transformée en une f-structure cible qui est alors transmise à un module de génération GENKIT.

Tous ces modules sont décrits en détail dans [Woszczyzna, M., *et al.* 1993].

Deux préoccupations ont présidé au développement des systèmes JANUS-II [Waibel, A. 1996] et JANUS-III [Lavie, A., *et al.* 1997, Levin, L., *et al.* 2000b] utilisé pour les démonstrations C-STAR II : la robustesse et l'augmentation du vocabulaire du module de reconnaissance de la parole, et la robustesse de l'analyse vers un pivot.

⁹⁴ <http://www.research.att.com/~srini/Projects/Anuvaad/home.html>

Dans JANUS-II deux analyseurs sont utilisés. Le premier est un analyseur LR Généralisé basé sur l'analyseur proposé par [Tomita, M. 1987]. Cet analyseur saute les parties de l'entrée qui ne peuvent participer à une structure de phrase correcte. Le second analyseur essaie d'identifier les concepts sémantiques clés présents dans l'entrée et les éléments qui participent à leur structure sous-jacente. Cet analyseur est fondé sur des patrons.

Dans JANUS-III, l'analyseur utilise une grammaire indépendante du contexte représentée sous forme d'un réseau de transitions probabilistes. Il est implémenté sous forme d'analyseur par carte descendant.

Les analyseurs des systèmes JANUS-II et III produisent des représentations pivot telles que définies dans C-STAR II (cf. section XI.1.1). JANUS-III produit aussi des IF au format NESPOLE! (cf. section XI.2.1).

Récemment, un système de traduction de parole anglais-thaï utilisant l'IF pour le domaine médical a été présenté dans [Schultz, T., *et al.* 2004].

IX.3.1.1.2. DIPLOMAT et TONGES

Le projet DIPLOMAT [Frederking, R., *et al.* 1997] « vise à explorer le déploiement rapide, dans un délai de plusieurs jours à plusieurs semaines, de systèmes de traduction de parole ou de dialogue « portables » entre l'anglais et une autre langue. Ces systèmes doivent fournir des performances initiales raisonnables améliorables dans le temps. » De tels systèmes doivent, par exemple, permettre à des soldats anglophones de communiquer avec les populations locales lors de missions de maintien de la paix.

Pour atteindre cet objectif, [Frederking, R., *et al.* 1997] envisagent une approche multimoteur (certaines approches permettent d'obtenir des résultats plus rapidement que d'autres) avec sélection. L'interface utilisateur doit être différente pour celui que l'on aura entraîné (le soldat) et le novice [Brown, R. D. 1998].

Deux systèmes coopèrent pour fournir une traduction : un système de TA par transfert, et un système de TA par l'exemple [Frederking, R., *et al.* 2000].

L'utilisateur entraîné doit d'abord pouvoir corriger le résultat produit par le module de reconnaissance de la parole. Il doit aussi pouvoir invalider la traduction produite par le système si le sens de la rétro-traduction est trop éloigné du sens du message à traduire. Il doit aussi pouvoir choisir parmi plusieurs, la bonne rétro-traduction et modifier indirectement la traduction finalement proposée à son interlocuteur. L'utilisateur naïf, occasionnel, ne pourra que valider ou non la rétro-traduction produite par le système et donc valider ou non la traduction produite.

IX.3.1.1.3. Babylon et Speechlator

Dans le cadre du projet DARPA-Babylon, le groupe ISL a développé un système de traduction de dialogue pour des troupes envoyées sur des terrains d'opération. Dans le projet Speechlator [Waibel, A., *et al.* 2003a, Waibel, A., *et al.* 2003b], la paire de langues considérée est anglais-arabe égyptien. La méthode de traduction utilisée dans Speechlator, traduction statistique par pivot, est décrite dans [Kauers, M., *et al.* 2002].



Figure IX-7 : Speechlator, un système de traduction de dialogue anglais-arabe de CMU

IX.3.1.1.4. Traduction statistique

Bien sûr, ces équipes ont aussi une activité en TA statistique [Kauers, M., *et al.* 2002, Vogel, S., *et al.* 2004, Vogel, S., *et al.* 2003].

IX.3.1.2. IBM

Au centre de recherche IBM, deux groupes développent des systèmes de traduction de dialogue, le groupe de Yuqing Gao et le groupe de Young-Suk Lee.

Le premier groupe est à l'origine du système que l'on trouve parfois appelé MASTOR dans la littérature. Le second groupe a d'abord commencé par faire de la traduction de l'écrit avant de démarrer une activité en traduction de dialogue.

Le système MASTOR [Gao, Y., *et al.* 2004, Gu, L. and Gao, Y. 2003, Gu, L. and Gao, Y. 2004] sur le domaine de la protection de forces armées et l'orientation de blessés vers les services appropriés est le successeur d'un premier système développé par la même équipe sur le domaine du renseignement financier [Davies, K. and al. 1999]. La paire de langues considérée est anglais-chinois mandarin.

Ce système utilise une approche statistique qui utilise un pivot dont la représentation est un arbre syntaxico-sémantique [Gao, Y., *et al.* 2004, Gu, L. and Gao, Y. 2003] comme le montre la Figure IX-8.

Les différents étages du système de traduction sont les suivants :

- un module de compréhension qui utilise un analyseur sémantique statistique, fondé sur un arbre de décision, entraîné sur un corpus textuel annoté.
- un module de transfert qui produit un arbre de concepts dans la langue cible à partir du résultat du module de compréhension.
- un module de génération lexicale qui produit un texte en langue cible⁹⁵.

⁹⁵ Selon Ruhi Sarikara (communication personnelle le 07/10/2004), l'architecture linguicielle la plus récente du système utiliserait des transducteurs d'état finis.

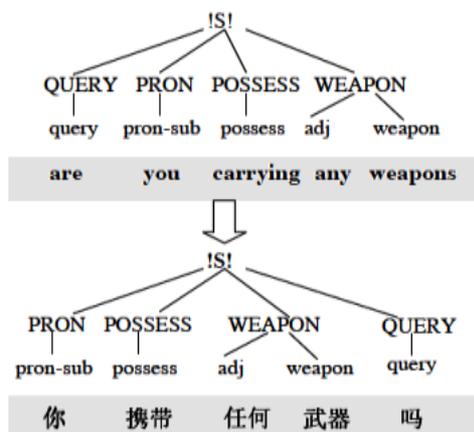


Figure IX-8 : Exemple de traduction anglais-chinois avec MASTOR

Récemment, le système a été porté sur un assistant numérique personnel [Zhou, B., *et al.* 2004, Zhou, B., *et al.* 2003] de type HP iPaq H3800 avec un système d'exploitation Linux.



Figure IX-9 : Interfaces du système MASTOR de IBM sur PDA

Plus récemment, [Lee, Y.-S. and Roukos, S. 2004] ont proposé et évalué un système purement statistique fondé sur un modèle de groupes [Tillmann, C. 2003].

IX.3.1.3. SRI

À la suite du projet SLT, SRI poursuit une certaine activité en traduction de dialogue dans le domaine médical en français-anglais [Rayner, M. and Bouillon, P. 2002a, Rayner, M. and Bouillon, P. 2002b], en anglais-japonais [Rayner, M., *et al.* 2003], et en anglais-pachtou [Precoda, K., *et al.* 2004]. Il faut cependant noter que le vocabulaire est extrêmement petit (200 mots).

IX.3.2. Europe

En Europe, les activités en traduction de parole se déroulent essentiellement dans le cadre de projets financés par l'Union Européenne. Pour ce qui est du CLIPS, nous détaillerons notre contribution dans les chapitres XI et XII.

IX.3.2.1. EuTRANS

Dans le cadre du projet LTR EuTrans (Example-based Understanding and Translation systems), deux approches ont été mises en œuvre pour faire de la traduction de parole en espagnol-anglais et en italien-anglais dans le cadre de l'interaction entre un client et un réceptionniste d'hôtel.

La première approche, proposée par les partenaires espagnols du projet (Université Polytechnique de Catalogne et Université Polytechnique de Valence), repose sur des techniques d'automates d'états finis [Amengual, J. C., *et al.* 1997, Amengual, J. C., *et al.* 2000, Casacuberta, F., *et al.* 2002, Pastor, M., *et al.* 2001].

La seconde approche, proposée par le partenaire allemand (RWTH, Aachen – équipe de Hermann Ney), est purement statistique [Casacuberta, F., *et al.* 2004].

L'équipe de Hermann Ney qui a participé au projet Verbmobil est une équipe en pointe en traduction statistique depuis la fin des années 90 [Ney, H. 2004, Tillmann, C., *et al.* 2000].

IX.3.2.2. Projets Européens en cours

Pour information, les projets Européens en cours qui s'intéressent à la traduction de parole sont PF-STAR⁹⁶ (Preparing future multisensorial interaction research), CHIL⁹⁷ (Computer in the Human Interaction Loop) et TC-STAR⁹⁸ (Technology and Corpora for Speech to Speech Translation).

IX.3.3. Asie

En Asie, le Japon, la Corée du Sud, et la Chine sont actifs dans le domaine de la traduction de parole. Ce domaine de recherche est d'ailleurs né au Japon en 1986.

IX.3.3.1. ATR - Japon

IX.3.3.1.1. ATR-Interpreting Telephony Research Laboratories et C-STAR I

Dans le cadre du projet ATR-ITL, deux systèmes de traduction de parole ont été réalisés : SL-TRANS [Morimoto, T., *et al.* 1992] et ASURA [Morimoto, T., *et al.* 1993].

La langue source de ASURA est le japonais ; l'anglais et l'allemand sont les deux langues cibles. La technique de traduction mise en œuvre est celle du transfert. L'analyse, commune aux deux langues cibles, utilise une stratégie fondée sur le lexique pendant laquelle la structure syntaxique et la structure sémantiques d'un énoncé sont construites à partir de constituants.

L'analyseur produit une structure de traits sémantique qui représente les contenus propositionnel et intentionnel de l'énoncé source. Lors du transfert (vers l'anglais ou l'allemand), les éléments du contenu propositionnel sont traduits en japonais, les éléments du contenu intentionnel sont convertis en acte de parole (tels que *inform*, *request*, *promise*, etc.). À partir de la structure produite lors du transfert, la génération (en anglais ou en allemand) construit un arbre syntaxique dont les feuilles constituent une phrase dans la langue cible visée.

⁹⁶ <http://pfstar.itc.it/>

⁹⁷ <http://chil.server.de/>

⁹⁸ <http://www.tc-star.org/>

IX.3.3.1.2. ATR-Interpreting Telecommunication Research Laboratories et C-STAR II

Pour le second projet ATR-ITL et pour C-STAR II, ATR-ITL a proposé le système MATRIX [Sumita, E., *et al.* 1999, Takezawa, T., *et al.* 1998] pour les paires de langues suivantes : japonais-anglais, japonais-coréen, japonais-allemand et japonais-chinois. Le module de traduction utilise une approche particulière du transfert qui exprime des correspondances entre des expressions de la langue source et des expressions de la langue cible à différents niveaux de représentation linguistique. Les expressions sont exprimées sous forme de patrons constitués de variables et de marqueurs de bornes de constituants tels que des mots fonctionnels.

Après les démonstrations C-STAR II, ATR a fait un gros effort d'évaluation des performances du système [Morishima, S. and Nakamura, S. 2002, Sugaya, F., *et al.* 1999, Sugaya, F., *et al.* 2000, Sugaya, F., *et al.* 2001]. Nous en avons déjà parlé dans la section II.3.1.1.2.

ATR a aussi collaboré avec le laboratoire de recherche de la société Matsushita afin de faire migrer MATRIX sur les plates-formes différentes : vidéotéléphone [Hiraoka, S., *et al.* 1997, Karaorman, M., *et al.* 1996], et assistant numérique [Matsui, K., *et al.* 2001].

IX.3.3.1.3. ATR-Interpreting Spoken Language Translation Laboratories et C-STAR III

Dans le cadre du laboratoire ATR-SLT, les méthodes développées pendant la deuxième phase de ATR-ITL ne sont pas définitivement abandonnées, le projet actuel se nomme C³ (Corpus Centered Computation). Il s'agit de placer le corpus au centre du développement des systèmes de traduction de dialogue. Comme il est dit dans [Sumita, E., *et al.* 2004] « les connaissances nécessaires à la tâche sont extraites des corpus, la qualité des traductions produites par les systèmes est évaluée par rapport aux corpus, la meilleure traduction parmi les traductions produites par différents systèmes est choisie par rapport aux corpus et les corpus eux-mêmes sont paraphrasés ou filtrés par des processus automatiques afin d'améliorer la qualité des données sur lesquelles les moteurs de traduction sont fondés ».

Dans ce contexte, deux techniques seront donc mises en œuvre : la traduction par l'exemple et la traduction statistique. ATR compte tirer le meilleur de ces deux approches en permettant au système de choisir la meilleure des traductions produites en les utilisant en parallèle.

Pour ce faire, ATR a développé deux systèmes utilisant des approches différentes de traduction par l'exemple et un système de traduction statistique. Les résultats complets de l'évaluation des performances des différentes approches sont développés dans [Sumita, E., *et al.* 2004].

IX.3.3.2. NEC - Japon

NEC a proposé le concept de traduction de conversation téléphonique pour la première fois à l'exposition Telecom'83 à Genève. Il a fallu attendre 1991 pour voir un premier prototype fonctionnant sur station de travail. Huit ans plus tard [Watanabe, T., *et al.* 2000] proposent un système de traduction bidirectionnel anglais-japonais pour le tourisme sur ordinateur portable. Ce système a finalement été porté sur un assistant numérique personnel de type Pocket PC 2002 [Ikeda, T., *et al.* 2002, Isotani, R., *et al.*

2002]. Ce système a été expérimenté en situation d'usage réel à l'aéroport de Narita (Tokyo).

La traduction proprement dite repose sur un formalisme de grammaires lexicalisées [Yamabana, K., *et al.* 2000] qui utilise un analyseur par carte ascendant.

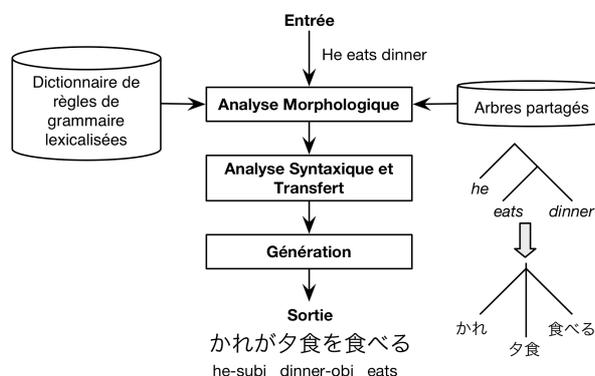


Figure IX-10 : Modèle de traduction de NEC

Le module de reconnaissance de la parole évalué sur 1 800 tours de paroles anglais et japonais a donné des taux de mots correctement reconnus de 95% pour l'anglais et 87% pour le japonais. Lors d'une évaluation sur 500 phrases choisies au hasard dans le corpus de conversation touristique de NEC, un juge bilingue devait classer les traductions dans trois catégories :

- **Good** : le sens est transmis et la syntaxe est correcte,
- **Understandable** : les informations principales sont transmises même si il y a des erreurs de syntaxe, et
- **Bad** : le juge ne peut comprendre le sens de la traduction ou si la traduction est un contresens.

Les résultats obtenus sont les suivants :

Paire de langue	Good	Understandable	Bad
japonais-anglais	74%	26%	10%
anglais-japonais	66%	22%	12%

Table IX-1 : Résultats d'évaluation subjective du système de NEC en 2002

Soit un taux de traductions acceptables de 90% pour la paire japonais-anglais et de 88% pour la paire anglais-japonais.

IX.3.3.3. ETRI – Corée du Sud

IX.3.3.3.1. Première expérience avec KT et KDD

Les recherches en traduction de parole ont commencé en Corée du Sud en 1991 sous l'impulsion de trois instituts de recherche ETRI (Electronics Telecommunication Research Institute, Taejon), KT (Korea Telecom) et KDD (Kokusai Denshin Denwa, Japon) afin de développer un système de traduction de dialogue entre le coréen et le japonais sur une tâche de réservation hôtelière. Les premières démonstrations ont eu lieu en 1995. Pour le coréen, ETRI était en charge de réaliser les modules pour un réceptionniste et KT pour un

client. Pour le Japonais, KDD devait réaliser les systèmes pour un réceptionniste et pour un client.

Le système coréen-japonais de ETRI [Lee, Y., *et al.* 1995] utilise un analyseur fondé sur les grammaires d'unification. Un module de transfert transforme ensuite le résultat d'analyse en une structure de traits en langue cible. Lors du transfert, les informations calculées par un gestionnaire de dialogue — qui prédit le tour de parole suivant et estime l'intention du locuteur en fonction de l'historique du dialogue —, et les informations disponibles comme connaissances du monde sont utilisées. Ainsi, par exemple, la traduction de l'équivalent coréen de « bonjour » peut être traduit différemment en japonais en fonction de l'intention du locuteur et de l'heure du jour.

Le système coréen-japonais de KT [Koo, M.-W., *et al.* 1995] met en œuvre un analyseur par carte avec une grammaire de dépendances. L'analyse en dépendance est intéressante pour le coréen qui est une langue dans laquelle l'ordre des mots est libre. Le transfert consiste simplement en une consultation de dictionnaire bilingue.

Les systèmes de KDD [Suzuki, M., *et al.* 1995] utilisent une méthode de traduction directe.

IX.3.3.3.2. C-STAR II

En tant que membre de C-STAR II, ETRI a d'abord proposé, en 1997, deux systèmes monodirectionnels coréen-japonais (par transfert) et coréen-anglais (par le pivot IF de C-STAR II) [Yang, J.-W. and Park, J. 1997].

Dans la suite du projet C-STAR II, ETRI a consacré ses efforts à la traduction par pivot [Lee, S.-J., *et al.* 2001, Park, J., *et al.* 1999] pour faire des démonstrations avec l'anglais et le français dans un premier temps, puis avec l'italien. Le système bidirectionnel japonais-coréen a été pris en charge par ATR.

IX.3.3.4. NLPR - Chine

Dernier venu dans le consortium C-STAR III, le NLPR (National Laboratory for Pattern Recognition, Pékin) a développé plusieurs maquettes de système de traduction de parole mono-directionnels en utilisant différentes approches.

[Zong, C., *et al.* 2000a] décrivent un système de traduction chinois-anglais qui utilise une analyse par patrons. La même année, en collaboration avec ATR le NLPR a aussi développé un système de traduction japonais-chinois [Zong, C., *et al.* 2000b] utilisant une technique de traduction fondée sur l'exemple proposée en 2002. Il s'agit de l'approche par « sablier » [Yamamoto, K. 2002]. Les énoncés à traduire sont paraphrasés pour essayer de converger vers des énoncés ou des segments d'énoncés présents dans une mémoire de traduction.

Depuis 2001, le NLPR utilise des approches multimoteurs. [Zong, C., *et al.* 2002] combinent les approches par patrons et par « sablier », avec une nouvelle approche interactive par mots clés et relations de dépendance entre ces mots clés. Lorsque l'analyseur n'est pas certain des mots clés ou des relations de dépendance, le système pose des questions au locuteur. [Zuo, Y., *et al.* 2004] combinent des approches par patrons, par pivot [Cao, W., *et al.* 2004] et statistique.

IX.3.4. Approches particulières

IX.3.4.1. Approches linguistiques

En 1996 [Seligman, M. 1996, Seligman, M. 1997], Mark Seligman rêvait d'un système de traduction sur Internet. En collaboration avec la société Linguistix Inc., il a pu réaliser un premier démonstrateur de traduction de parole bidirectionnel français-anglais en utilisant des composants disponibles dans le commerce [Seligman, M., *et al.* 1998]. Le système de traduction utilisé était le système de Linguistix développé pour faire de la traduction de tchat.

Ici, on souhaite utiliser des composants de dictée automatique à large couverture, des composants de traduction automatique à large couverture et de synthèse de la parole pour proposer des systèmes interactifs de traduction de parole. L'utilisateur doit pouvoir corriger les résultats de la reconnaissance, le système de traduction doit fournir une sortie multiple que l'utilisateur va aider à désambiguïser [Seligman, M. 2000]. Mark Seligman essaie de promouvoir et commercialiser cette approche dans le cadre de sa société Spoken Translation Inc.

Les systèmes de dictée automatique sont donc intéressants parce que la langue naturelle est utilisée à la fois pour la dictée (reconnaissance de la parole vers texte) et pour commander le système (demander des corrections, des insertions et suppressions de mots). Pour les systèmes de traduction, il faut trouver des systèmes qui produisent des sorties multiples ou alors convaincre leurs développeurs de modifier leurs systèmes.

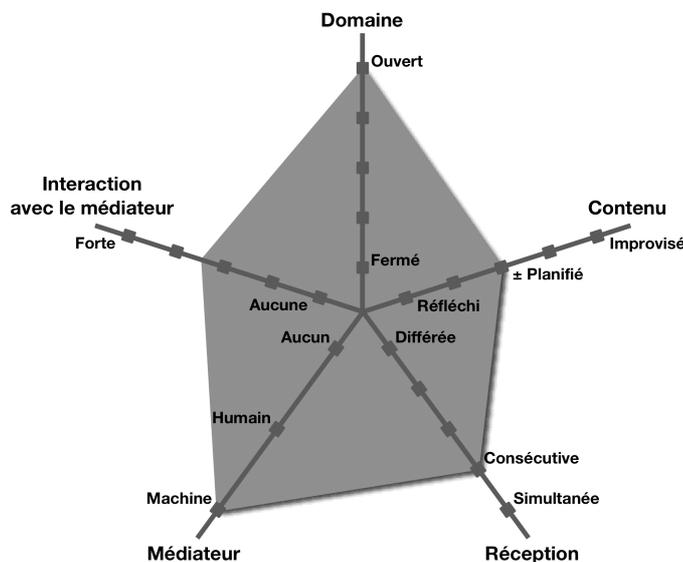


Figure IX-11 : Modèle de traduction proposé par Spoken Translation Inc.

IX.3.4.2. ERIM

Parallèlement à l'intégration du français dans des systèmes de traduction de dialogues, le GETA a développé plusieurs plates-formes [Fafiotte, G. 2004] dans le cadre des projets ERIM (Environnement Réseau pour l'Interprétariat Multimodal) et ChinFaDial (collecte de dialogues parlés spontanés français-chinois). Elles traitent de différents aspects de la communication orale spontanée bilingue non finalisée sur le Web : interprétariat humain à distance, collecte de données, intégration d'aides automatiques (serveur de TA de

Parole utilisant des composants du marché, interaction multimodale entre interlocuteurs, et prochainement aides en ligne aux intervenants, locuteurs ou interprètes).

La famille des plates-formes ERIM comprend, actuellement :

- ERIM-Interprète pour l'interprétariat multimodal à distance et base des autres plates-formes Figure IX-12.
- ERIM-Collecte pour la collecte de corpus de dialogues parlés spontanés bilingues traduits,
- ERIM-TA (vers un service de Traduction partiellement Automatique de Parole aidée par le locuteur, et banc d'essai de composants de TA de parole), et
- ERIM-Aides (vers des aides en ligne à la communication multilingue sur réseau).

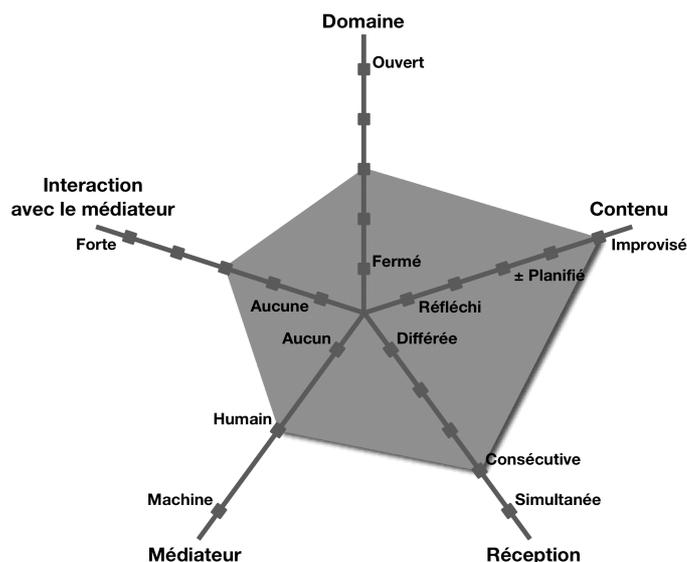


Figure IX-12 : Modèle de traduction de ERIM-Interprète

Grands projets en traduction de dialogues oraux finalisés

X.1. C-STAR

Le consortium C-STAR est une organisation souple qui réunit des groupes de recherche qui veulent bien travailler ensemble autour d'un thème commun : la traduction de dialogue. Le consortium n'est pas une source de financement. Les participants doivent trouver leurs propres sources de financement. Deux types de participations sont possibles, en tant que partenaire ou en tant qu'affilié. Les partenaires s'engagent à construire des systèmes complets qui leur permettent de participer à des démonstrations ou à prendre part à des activités communes (collectes de données, traduction de corpus, évaluation de systèmes). Les affiliés font de la recherche dans le domaine et participent à ses avancées, mais ne participent pas aux activités communes.

X.1.1. C-STAR I (1986-1993)

ATR (Advanced Telecommunication Research) a été créé en avril 1986. Il s'agissait alors d'une structure de droit privé, financée à 70% par le ministère des finances du Japon et à 30% par un consortium de 200 sociétés, au premier rang desquelles NTT et KDD. Les buts d'ATR ne sont pas seulement scientifiques. Il s'agit d'une opération de prestige, engagée dans le cadre de l'internationalisation, thème moteur officiel du Japon des années 80. C'est pourquoi ATR-ITL (Interpreting Telecommunications) s'est voulu le moteur d'un consortium international visant en particulier à réaliser des démonstrations spectaculaires et bien médiatisées à la fin de chaque projet.

C-STAR I est le nom du consortium qui a présenté les premières démonstrations publiques d'interprétation téléphonique en janvier 1993 dans le cadre du premier projet ATR-ITL. Les partenaires en étaient ATR, CMU (Pittsburgh, USA), Siemens (SNI, Munich), et l'université de Karlsruhe (UKA).

Le 28 janvier 1993, les partenaires ont fait une démonstration publique de leurs systèmes en conversation téléphonique. Un système de visioconférence permettait à chacun des interlocuteurs de voir ce qui se passait sur l'autre site.

Les systèmes utilisés pour les démonstrateurs sont décrits dans [Morimoto, T., *et al.* 1992, Morimoto, T., *et al.* 1993, Woszczyna, M., *et al.* 1993].

X.1.2. C-STAR II (1993-1999)

Devant le succès remporté par la première démonstration auprès du public et des bailleurs de fonds, le projet de ATR-ITL a été renouvelé et les fondateurs de C-STAR I ont décidé de poursuivre leur collaboration en essayant d'impliquer d'autres groupes de recherche.

X.1.2.1. Présentation

Dès 1993, ETRI (Electronics and Telecommunications Research Institute), un puissant institut coréen créé un peu à l'image de ATR a rejoint le nouveau consortium. L'IRST (Istituto per la Ricerca Scientifica e Tecnologica) à Trente en Italie (en 1994) et le groupe CLIPS++ (en 1996) sont aussi venus enrichir le nombre de langues disponibles. Le schéma suivant donne la liste et la distribution géographique des partenaires (en gras) et des affiliés.

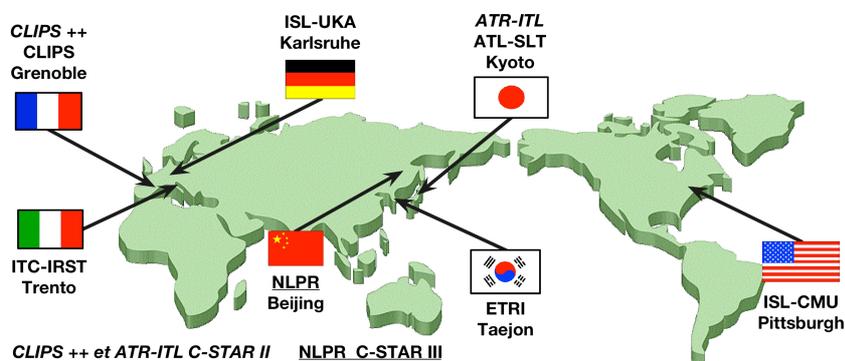


Figure X-1 : Partenaires de C-STAR II et de C-STAR III

Le domaine envisagé était celui du tourisme, et en particulier le renseignement sur les transports et la réservation, le renseignement sur les hôtels et la réservation, et le renseignement touristique.

Dans la grille de classification que nous avons proposée au chapitre IV, le projet C-STAR II peut être caractérisé comme le montre la Figure X-2.

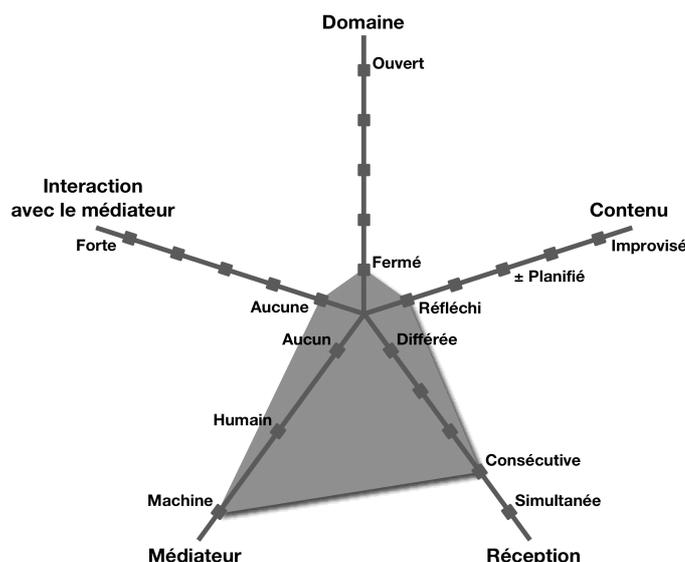


Figure X-2 : C-STAR II dans la classification à cinq branches proposée

X.1.2.2. Architecture Informatique

Les échanges entre les démonstrateurs sont de deux types : échanges audio et vidéo (assurés par des systèmes de visioconférence du commerce), et échanges concernant le processus de traduction proprement dit.

Les échanges audio et vidéo se font en connexion point à point si deux sites seulement souhaitent communiquer et via un pont si plusieurs sites souhaitent communiquer.

Les échanges concernant le processus de traduction se réduisent au minimum et consistent en des chaînes de caractères. Ces échanges se font au moyen d'un serveur de communication (CommSwitch) sur lequel se connectent les sites qui veulent communiquer. La connexion au serveur de communication se fait via le protocole Telnet. Cela permet à tous les sites connectés sur le serveur de communication de recevoir (d'entendre) ce qui est émis (dit) par l'un des sites.

Le serveur de communication inter-sites est appelé serveur de communication global.

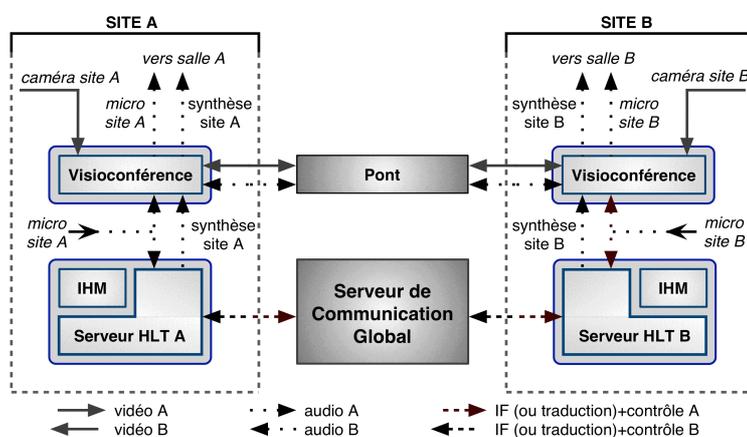


Figure X-3 : Architecture informatique de C-STAR II

X.1.2.3. Architectures linguistiques

Les démonstrateurs de ATR font de la traduction vers la langue cible (cf. section IX.3.3.1.2). Les autres sites communiquent entre eux en échangeant des structures « pivot » (le pivot est appelé IF pour Interchange Format). L'avantage de cette technique pour celui qui l'utilise est qu'elle permet d'avoir comme langues cibles toutes les langues passant par l'IF. Un partenaire qui passe par l'IF a aussi un générateur à partir de l'IF pour « écouter » les autres partenaires. Celui qui parle peut donc utiliser son propre générateur pour régénérer l'IF que le système a produit pour ce qu'il vient de dire. Ainsi, l'utilisateur peut voir ce que le système a compris pour peu que le générateur produise une paraphrase exprimant exactement le contenu de l'IF. Le schéma suivant illustre la situation pour le français.

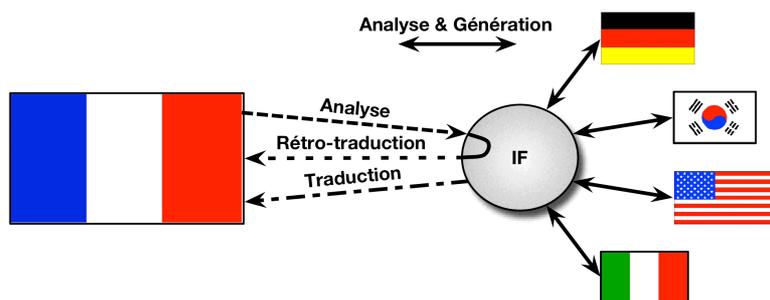


Figure X-4 : Avantages liés à l'utilisation d'un pivot pour la traduction

Les systèmes utilisés pour les démonstrations sont décrits dans :

- [Sugaya, F., *et al.* 1999, Takezawa, T., *et al.* 1998] pour ATR,
- [Lavie, A., *et al.* 1997, Waibel, A. 1996] pour CMU et UKA,
- [Blanchon, H. and Boitet, C. 2000, Blanchon, H., *et al.* 1999, Blanchon, H., *et al.* 2000, Boitet, C. 1997, Boitet, C. and Guilbaud, J.-P. 2000, Keller, E. and Zellner, B. 1998, Vaufreydaz, D., *et al.* 1999, Wehrli, É. and Wehrle, T. 1998, Wehrli, É., *et al.* 1999] pour le CLIPS++.
- [Koo, M.-W., *et al.* 1995, Lee, S.-J., *et al.* 2001, Lee, Y., *et al.* 1995, Park, J., *et al.* 1999, Suzuki, M., *et al.* 1995, Yang, J.-W. and Park, J. 1997] pour ETRI,
- [Angelini, B., *et al.* 1997, Cettolo, M., *et al.* 1999, Lazzari, G. 1999, Pianta, E. and Tovina, L. M. 1999] pour ITC-IRST.

X.2. Verbmobil

Le projet Verbmobil [Wahlster, W. 2000] (1993-2000) est un projet allemand cofinancé par le ministère fédéral allemand de l'éducation et de la recherche. Il a réuni jusqu'à trente-trois groupes de recherche et six firmes industrielles et reçu une aide financière de cent seize millions de Marks allemands (soixante millions d'Euros).

X.2.1. Présentation

Le système Verbmobil offre l'accès à la traduction de parole à des utilisateurs de téléphones cellulaires et permet des traductions bidirectionnelles entre l'anglais, l'allemand et le japonais. Les scénarios couverts sont : la prise de rendez-vous, l'organisation de voyage, la maintenance de PC à distance. Le second scénario a servi de support aux évaluations.

X.2.2. Architectures linguistiques

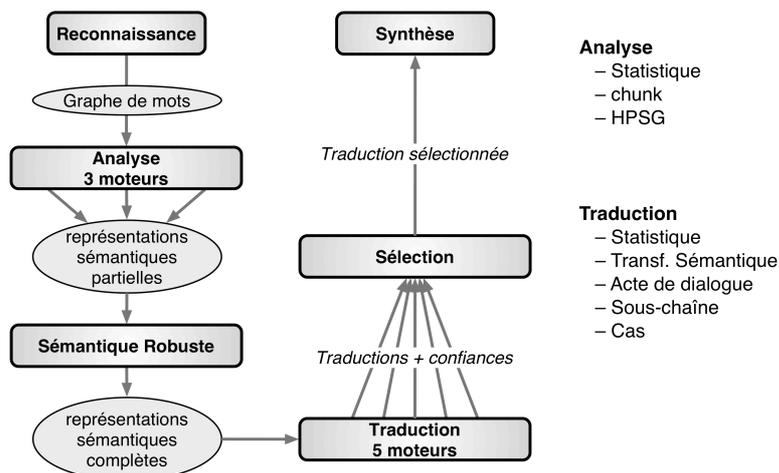


Figure X-5 : Architecture linguistique de Verbmobil

Trois modules sont utilisés en parallèle pour analyser un graphe de mots augmenté d'informations prosodiques. Les analyseurs produisent chacun une structure VIT (Verbmobil Interface Term) qui est souvent partielle. Ces différentes structures sont combinées pour former une représentation sémantique complète qui est ensuite transmise à cinq modules de traduction qui produisent un texte en langue cible auquel est associé un

niveau de confiance. Un module de sélection choisit alors la traduction à transmettre au module de synthèse de la parole.

X.2.3. Évaluation

Une importance particulière a été apportée à l'évaluation du système, tant en ce qui concerne la reconnaissance de parole proprement dite que la traduction. Nous présentons ici les résultats de l'évaluation conduite à l'automne 1999, ainsi que ceux de l'évaluation de masse de 2000.

X.2.3.1. Données et protocole

Afin de faire l'évaluation, des données en situation d'utilisation réelle du système ont été collectées. Elles ont permis d'évaluer l'utilisabilité du système (point de vue des utilisateurs), et la qualité linguistique des sorties du système (point de vue des développeurs).

En 1999, des utilisateurs de langue maternelle anglaise et allemande, observés par un superviseur, ont été conviés à utiliser le système afin de parler d'un rendez-vous. Les utilisateurs, ainsi que le superviseur, devaient évaluer le succès ou l'échec des interactions sur chaque thème abordé. Pour les quarante-cinq dialogues collectés, toutes les données produites par le système ont été enregistrées ; les tours de parole ont aussi été transcrits pour fournir des références.

Pour la qualité linguistique du système, les juges avaient à leur disposition une interface complète leur permettant de juger la traduction ainsi que des critères linguistiques sur les entrées et les sorties. La traduction est évaluée selon trois critères : équivalence traductionnelle (oui/non), compréhensibilité (oui/non), qualité (bonne, intermédiaire, mauvaise). L'entrée (langue source) et la sortie (langue cible) sont évaluées selon trois critères : bien formée syntaxiquement (oui/non), bien formée sémantiquement (oui/non), risque d'incompréhension (oui/non).

En 2000, une évaluation de masse, jugeant la qualité de traduction, a été conduite sur 5069 tours de parole en allemand et 4136 tours en anglais. Les juges devaient dire si oui ou non la traduction était « approximativement correcte » (la traduction d'un tour de parole est « approximativement correcte » si elle préserve l'intention du locuteur et les informations contenues dans le tour de parole).

X.2.3.2. Résultats de l'évaluation de masse

Taux de mots correctement reconnus		>50%	>75%	>80%
GER-ENG	<i>Nb tours concernés</i>	5069	3267	2723
	Sélection Automatique	57	66	68
	Sélection Manuelle	88	95	97
ENG-GER	<i>Nb tours concernés</i>	4136	3254	2291
	Sélection Automatique	53	58	60
	Sélection Manuelle	86	92	94

Table X-1 : Résultats de l'évaluation de masse dans *Verbmobil* en qualité de la traduction

La Table X-1 donne le pourcentage des traductions approximativement correctes en fonction de différents WAR (Word Accuracy Rate). La Sélection Manuelle permet de

mesurer la performance du système si le module de sélection de la meilleure traduction produite par les différents modules était parfait.

X.2.3.3. Commentaires

X.2.3.3.1. Évaluation fine (1999)

Au niveau du succès des dialogues, le taux de résolution des tâches du domaine est de 86,8%. Si l'on pondère le taux de succès de chaque tâche par sa fréquence, ce taux atteint 89,6%. Les tâches les plus fréquentes sont donc très souvent bien accomplies.

En ce qui concerne les caractéristiques linguistiques de la qualité de traduction, les évaluations ont montré que la qualité de la traduction est plus sensible à l'insertion qu'à la suppression d'information. En effet, si au moins 50% des éléments d'information du message source sont transmis, le cumul des marques **bon** et **intermédiaire** dépasse la marque **mauvais** dans l'évaluation de la qualité traductionnelle. À l'opposé, des insertions supérieures à 20% induisent une prédominance de la marque **mauvais**.

X.2.3.3.2. Évaluation de masse (2000)

Il n'est pas surprenant de constater que les performances du système croissent lorsque le WAR du module de reconnaissance de la parole croit. Nous verrons que les résultats annoncés sont proches des résultats du projet NESPOLE!.

Nous tenons aussi à souligner aussi que les performances potentielles du système sont tout à fait exceptionnelles (de 86 à 94 pourcent).

X.3. NESPOLE!

Le projet NESPOLE! (01/2000-01/2003) visait à capitaliser les efforts des partenaires européens et américains du consortium C-STAR II et aller plus loin en termes scientifiques. Les langues impliquées sont l'italien, le français, l'allemand et l'anglais. Les partenaires sont : ITC/IRST de Trente (Italie), ISL Labs. de UKA (Karlsruhe, Allemagne) et CMU (Pittsburgh, USA), Aethra (une société italienne spécialisée dans le domaine de la vidéoconférence), APT (une agence de tourisme dans la région du Trentin en Italie) et le laboratoire CLIPS (Grenoble, France).

X.3.1. Présentation

Le scénario NESPOLE! met en situation de dialogue un agent touristique italoophone localisé à Trente en Italie, et un client parlant anglais, français ou allemand qui peut être n'importe où. Les deux interlocuteurs utilisent chacun un terminal de communication le plus simple possible (PC équipé d'une carte son et d'un logiciel de vidéoconférence de type NetMeeting™). Les scénarios ont été définis grâce à des collectes de données à partir de situations réelles avec de vrais agents touristiques du Trentin.

Le client veut organiser un voyage dans la région du Trentin, et navigue sur le site Web de APT (l'agence de tourisme) pour obtenir des informations. Si le client veut en savoir plus, sur un sujet particulier, ou préfère avoir un contact plus direct, un service de traduction de parole en ligne lui permet de dialoguer, dans sa propre langue, avec un agent italien de APT. Une connexion NetMeeting™ est alors ouverte entre le client et l'agent, et la conversation médiatisée (avec service de traduction de parole) entre les deux personnes peut alors démarrer.

Dans le projet, l'accent était mis sur certains problèmes scientifiques en traduction automatique de parole : robustesse, extensibilité (extension de la couverture d'un domaine) et portabilité (passage d'un domaine à un autre).

Dans la grille de classification que nous avons proposée au chapitre IV, le projet NESPOLE! peut être caractérisé comme suit :

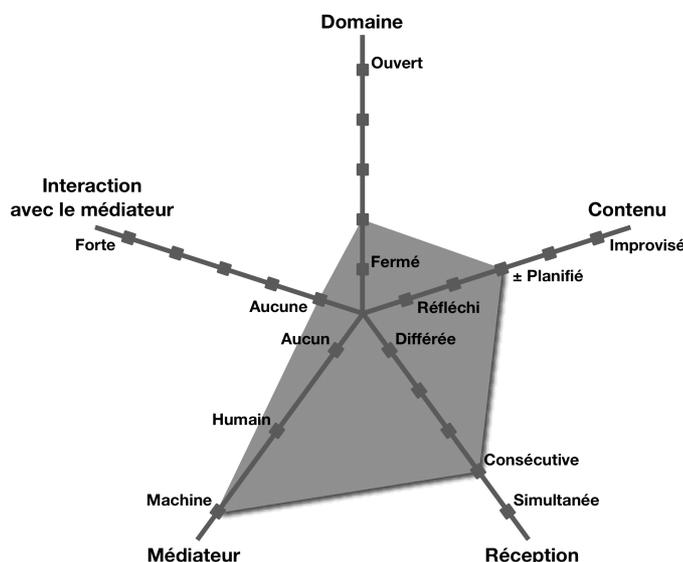


Figure X-6 : NESPOLE! dans la classification à cinq branches proposée

Par rapport au projet C-STAR II (Figure X-2), on peut remarquer que le domaine s'est élargi et que le contenu du dialogue est moins planifié.

X.3.2. Architecture informatique

Dans le cadre du projet C-STAR II, les travaux avaient uniquement concerné les modules de traitement des langues naturelles, sans qu'il y ait une véritable réflexion sur l'architecture informatique du système. Il en était résulté une interface utilisateur assez lourde, composée de plusieurs terminaux, chacun d'entre eux assurant une tâche particulière (vidéoconférence, reconnaissance, analyse, génération, synthèse).

L'architecture proposée par NESPOLE! prend en compte les récents développements dans la transmission de données audio et vidéo en utilisant le réseau Internet. Ainsi, l'architecture est pensée pour permettre au client ou à l'agent, voulant utiliser un service de traduction, de se servir d'un terminal de communication simple.

Tous les traitements liés à la traduction, invisibles pour l'utilisateur, sont alors réalisés de façon distante et distribués sur différents serveurs de traitement. La Figure X-7 montre l'architecture générale utilisée dans NESPOLE!, qui prend en compte d'une part, la localisation géographique des quatre serveurs de traduction de parole (TAP) de chaque langue et permet, d'autre part, un traitement symétrique des flux audio venant du client ou de l'agent.

Les terminaux client et agent utilisent le protocole classique de vidéoconférence H323, transmettant des flux audio (standards G711 ou G723.1) et vidéo (standard H261) compressés ainsi que des données (standard T120). Le serveur Nespole! (Global Nespole Server) est représenté sur la figure comme une entité unique, bien qu'il soit en réalité distribué entre différents sites, chacun correspondant à une des quatre langues du projet. En

effet, tout ou partie des modules de TAP de certains partenaires du projet étaient difficiles à transférer et à réinstaller sur un site unique, pour des raisons techniques ou de propriété intellectuelle.

Un élément important de cette architecture est le médiateur chargé d'interfacer les serveurs de TAP spécifiques d'une langue avec les terminaux clients. Toutes les communications entre Médiateur et Serveur Nespole se font par des sockets (assurant une relative indépendance par rapport au système d'exploitation utilisé à chaque extrémité) tandis que les communications entre Client (ou Agent) et Médiateur suivent le protocole H323.

Par exemple, si on suppose qu'un client parlant la langue X veut communiquer avec un agent parlant la langue Y, et si on note MY le médiateur pour le langage Y et SY le serveur de TAP pour le langage Y (même chose pour X), alors

- le client appelle MY,
- MY identifie l'appel comme provenant d'un terminal utilisant la langue source X et appelle ensuite le terminal de l'agent en langue Y,
- une connexion H323 pour la transmission des flux audio, vidéo et données est alors ouverte à la fois entre le terminal du client X et MY et aussi entre MY et le terminal de l'agent Y,
- MY est aussi connecté au serveur de TAP NESPOLE! pour y envoyer des signaux en langue source X reçus du terminal client, devant être traités, et en recevoir des signaux synthétiques en langue cible Y, afin de les rediriger vers le terminal de l'agent.

Cette architecture a l'avantage d'être complètement transparente pour l'utilisateur (agent ou client), mais pose cependant de nouveaux problèmes, les signaux audio traités par le serveur de TAP étant notamment dégradés car compressés pour être transmis sur le réseau IP.

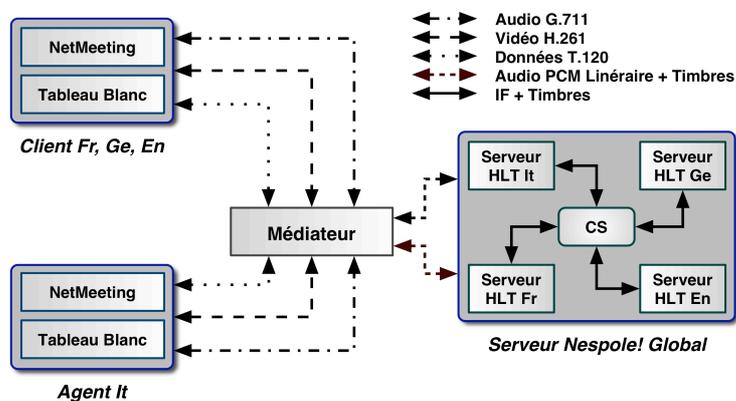


Figure X-7 : Architecture informatique de NESPOLE!

X.3.3. Interface utilisateur

L'interface utilisateur du démonstrateur NESPOLE! se répartit sur quatre fenêtres (Figure X-8).

La première fenêtre est celle de l'environnement de visioconférence NetMeeting ; elle permet aux utilisateurs de se voir et de se parler.

La seconde fenêtre (Nespole Monitor) permet à l'utilisateur de contrôler le système et propose différents retours d'information. La zone 2 (System hears:) affiche le résultat du module de reconnaissance vocale. Si l'utilisateur n'est pas satisfait, il peut soit corriger

scénario 2	Forfait tout compris
scénario 3	Séjour en été dans un parc
scénario 4	Circuit des lacs ou des châteaux
scénario 5	Événements et recherche de documentation

Les scénarios 1 et 2 ont été choisis pour le premier démonstrateur en « tourisme réduit » (2001). Les scénarios 2, 4 et 5 ont été abordés dans le second démonstrateur en « tourisme étendu » (2002).

Dans le cadre du projet, une importance particulière a été donnée à l'évaluation des deux démonstrateurs. Ces évaluations devaient permettre de mesurer à la fois les performances brutes des démonstrateurs, ainsi que les progrès accomplis entre le premier et le second. Nous y reviendrons dans le chapitre suivant.

Chapitre XI

Contributions à C-STAR II et NESPOLE!

Le groupe CLIPS++ a rejoint le consortium international C-STAR II en tant que partenaire en septembre 1996. Coordonné par le laboratoire CLIPS (France), notre groupe était composé de trois autres laboratoires : le LATL (Suisse), le LAIP (Suisse) et le LIRMM (France). La participation du CLIPS au consortium C-STAR II a été activement soutenue par la fédération IMAG via la labellisation du projet. Nous avons aussi reçu le soutien du centre de calcul de l'Université de Marseille et de IBM France, qui nous ont permis de faire tourner le module d'analyse du français sur des machines bien plus puissantes que celles dont disposait le GETA à cette époque.

XI.1. C-STAR II : première approche en groupe

À partir de janvier 1998, j'ai pris en charge la gestion et l'animation des activités liées au projet pour le groupe CLIPS++. Nous étions alors une équipe réduite :

- quatre personnes au CLIPS : Laurent Aublet-Cuvelier pour l'intégration des composants et l'interface du démonstrateur, Jean-Philippe Guilbaud pour l'analyse, Mohammad Akbar puis Dominique Vaufreydaz pour la reconnaissance de la parole, et moi-même ;
- deux personnes au LATL : Éric Wehrli et Thierry Etchégoyen pour la génération ;
- deux personnes au LAIP : Éric Keller et Brigitte Zelner pour la synthèse.

Nous avons adopté l'approche interlingue, déjà utilisée par 4 des partenaires de C-STAR II, dans laquelle l'interlingua, appelé IF, est un pivot sémantique spécialisé pour la tâche. Nous avons donc réalisé 4 modules : reconnaissance de la parole pour le français, analyse du français vers l'IF, génération du français à partir de l'IF, et synthèse orale du français. Ces modules coopèrent entre eux et avec les démonstrateurs des autres partenaires via un intégrateur.

XI.1.1. Langage pivot de C-STAR

L'IF [Corazza, A. 1999, Levin, L., *et al.* 2000a, Levin, L., *et al.* 1998, Tovina, L. M. 1998] est un pivot sémantique conçu pour s'affranchir des particularités des langues naturelles et permet de faire des traductions qui ne sont pas littérales, mais qui capturent l'intention de celui qui parle. Il est basé sur des actes de parole, des concepts et des arguments :

les actes de parole décrivent les intentions et les besoins de celui qui parle (give-information, introduce-self, ...).

les concepts définissent à propos de quoi les actes de dialogue sont exprimés (availability, train, flight, room, ...).

les arguments permettent d'instancier les valeurs des variables du discours (<room-type>, <time>, ...).

L'IF suivante : a:give-information+availability+room (room-type=(single & double), time= (week, md12)) correspond à un énoncé produit par un agent dont le contenu sémantique est « la semaine du 12, nous avons des chambres simples et doubles disponibles ».

L'IF utilisé dans le projet C-STAR II est définie au moyen de quarante-quatre actes de parole, soixante-huit concepts, cent dix-sept arguments et trente-quatre ensembles de valeurs. La documentation imprimée occupe dix-neuf pages, elle est accompagnée d'un manuel d'annotation de quatorze pages.

Pour un acte de parole ou un concept, la spécification propose la liste des concepts qui peuvent lui succéder, ainsi que la liste de ses arguments privilégiés.

Spécification de l'acte de parole give-information :

(give-information (x-reception x-addition x-completion connection inform time time-difference weather display budget possibility personal-data write search change temporal availability location address features contain price exchange-rate price-difference price-fluctuation payment spelling numeral telephone-number fax-number expiration-date preference pick-up drop-off purchase order click goto view reservation cancellation confirmation exchange send transfer) (<send-by>))

Spécification des concepts availability et room :

(availability (web-page restaurant meal departure arrival flight train event activity tour trip hotel attraction room person guide seat admission transportation taxi x-car-rental) (<time> <frequency> <duration> <destination> <origin> <location> <price> <for-whom> <what>))

(room nil (<hotel-name> <hotel-type> <hotel-style> <room-view> <room-type> <room-size> <contain> <bed-type> <room-location> <room-number> <room-name> <occupancy> <for-whom> <include>))

Pour un argument, la spécification donne la liste de ses valeurs possibles ; elles peuvent être atomiques (room) ou représentées par un ensemble nommé (*general-modifier*).

(room-type= room luxury modest matrimoniale twin double single family suite junior_suite senior_suite bedroom meeting conference smoking non-smoking western_style japanese_style *general-modifier*)

Il faut noter que la spécification ne précise pas explicitement comment les arguments peuvent être enchâssés. Pour « deux chambres » la représentation en IF est : room-type=(room, quantity=2). L'enchâssement des arguments se fait donc par accord implicite ou explicite entre les partenaires.

D'autres informations sont fournies en Annexe VI.

XI.1.2. Composants développés

Les modules de reconnaissance de la parole et d'analyse du français vers l'IF ont été développés au sein du CLIPS, respectivement par les équipes GEOD et GETA. Le module de génération de l'IF vers le français a été développé par le LATL. Le module de synthèse a été développé par le LAIP. Le LIRMM a pour sa part exploré la réalisation d'un module français vers IF avec une autre méthodologie.

XI.1.2.1. Reconnaissance de la parole : RAPHAEL

C'est un module [Vaufreydaz, D., *et al.* 1999] de reconnaissance multilocuteur de parole continue spontanée avec un vocabulaire centré sur le domaine du tourisme de 10K mots. Il est basé sur une architecture client-serveur, c'est à dire que le serveur de reconnaissance est mis en œuvre via des clients « légers » sur le réseau. Ce module est construit sur la boîte à outils de JANUS III en collaboration avec l'Université Carnegie Mellon (CMU).

Pour produire une transcription orthographique correspondant à une entrée vocale, RAPHAEL met en œuvre un modèle acoustique markovien indépendant du contexte entraîné sur 10 heures de parole continue (corpus BREF-80), et un modèle de langage stochastique entraîné sur un corpus de 140 millions de mots et optimisé pour la tâche de renseignement et de réservation touristique.

La transcription orthographique produite ne comporte aucune marque de ponctuation, aucune lettre capitale, les accords en genre et en nombres peuvent ne pas être marqués. Il peut aussi y avoir des insertions et/ou des effacements.

XI.1.2.2. Analyse du français vers l'IF

Ce module [Blanchon, H., *et al.* 1999, Boitet, C. and Guilbaud, J.-P. 2000] a été développé sous Ariane-G5, un générateur de systèmes de traduction qui intègre cinq langages spécialisés pour la programmation linguistique dans l'environnement VM/ESA/CMS. En 1999, il était opérationnel sur deux sites : à Grenoble (sur un IBM 9221-130 à 3,5 Mips), au CCSJ Marseille (sur un IBM 9672-R14 à 40 Mips). Lors de la démonstration du 22 juillet ce module fonctionnait à Montpellier au centre de calcul IBM (sur un IBM 9672-RX5 à 60 Mips).

Cet analyseur reçoit en entrée une transcription orthographique de l'énoncé oral. Parce qu'elle peut être mal formée, on ne peut pas mettre en œuvre un processus d'analyse du type de ceux mis en œuvre pour faire de l'analyse d'un texte écrit bien formé. De plus, l'analyseur doit être capable de faire l'analyse en composants de l'IF. Il s'agit donc en réalité d'une traduction et non d'une analyse. C'est pourquoi on a utilisé des étapes de transfert et de génération. Les étapes suivantes se succèdent :

- Analyse morphologique et lemmatisation des mots du texte ;
- Premières consultations de dictionnaires de transfert vers l'IF ;
- Analyse syntaxique pour la reconnaissance de structures grammaticales sémantiquement pertinentes : dates, quantités, numéros, prix ;
- Autres consultations de dictionnaires de transfert vers l'IF ;
- Génération syntaxique puis morphologique de l'IF.

En plus de ce module d'analyse en Ariane-G5, nous avons développé, en Tcl, un micromodule d'analyse qui permet d'aller très vite lors des cérémonies d'ouverture et de clôture (interaction avec les trois agents voyage en même temps, cf. infra XI.1.4.1.2) ; ce module couvre aussi certains types de tours de parole pour le cas où nous aurions eu des problèmes de communication avec l'analyseur situé à Montpellier.

XI.1.2.3. Génération de l'IF vers le français

Ce module [Wehrli, É. and Wehrle, T. 1998] est développé en partie sous GBGen qui est un outil de génération syntaxique à large couverture lexicale et grammaticale. C'est un outil déterministe basé sur une grammaire générative.

Il reçoit en entrée une IF qui est bien formée. On peut donc mettre en œuvre un processus très déterministe pour la génération. Pour cela, il faut et il suffit que l'IF reçue ait une structure couverte par le générateur. Le processus est mis en œuvre au moyen des étapes suivantes :

- Mise en correspondance entre une structure IF et une structure sémantique de GBGen ;
- Mise en œuvre des procédures de génération de GBGen pour produire une structure syntaxique ;
- Mise en œuvre de règles morphologiques de GBGen pour produire une forme orthographique correctement orthographiée et ponctuée.

XI.1.2.4. Synthèse du français

Le module du LAIP [Keller, E. and Werner, S. 1997, Keller, E. and Zellner, B. 1998] est un synthétiseur "text-to-speech" basé sur des règles. La synthèse se déroule en trois étapes : conversion texte vers phonèmes, génération de la prosodie, génération du signal.

La conversion texte vers phonèmes se fait avec des règles générales (540) et des règles spécialisées pour les nombres, les abréviations, les expressions figées, etc. Elle met en œuvre des dictionnaires généraux (7000 mots) et spécialisés (noms propres de la tâche). La génération de la prosodie met en œuvre des règles de type psycholinguistique et la génération du signal utilise la technique Mbrola de la faculté de Mons en Belgique.

XI.1.3. Démonstrateur du CLIPS++

XI.1.3.1. Intégration des composants

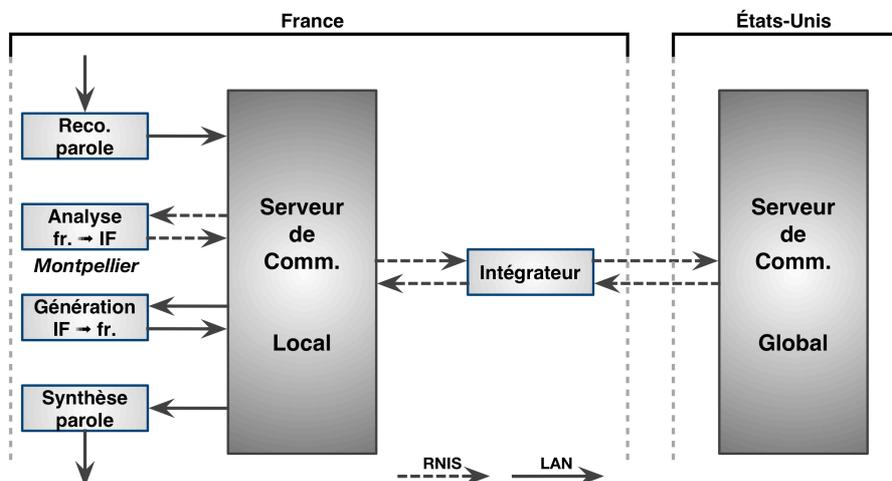


Figure XI-1 : Intégration des composants du démonstrateur C-STAR II du groupe CLIPS++

Pour des raisons de robustesse et de spécificité des ressources requises, le démonstrateur du groupe était distribué sur cinq machines :

- une machine de type PC pour le système de visioconférence ;
- une machine de type PC pour le serveur de reconnaissance de parole ;
- une machine de type IBM 9672-RX5, localisée à Montpellier, pour le serveur d'analyse ;
- une machine de type PC pour l'interface utilisateur et un serveur de communication local de secours ;

- une machine de type PC pour l'analyseur des sessions d'ouverture et de clôture, le générateur, le module de synthèse, le module d'intégration avec les autres démonstrateurs et le serveur de communication local.

XI.1.3.2. Interface

L'interface sur démonstrateur est distribuée sur deux écrans : le premier permet d'afficher la visioconférence (Figure XI-2) entre le client et les différents agents de voyage, le second est dédié à l'interface utilisateur proprement dite.

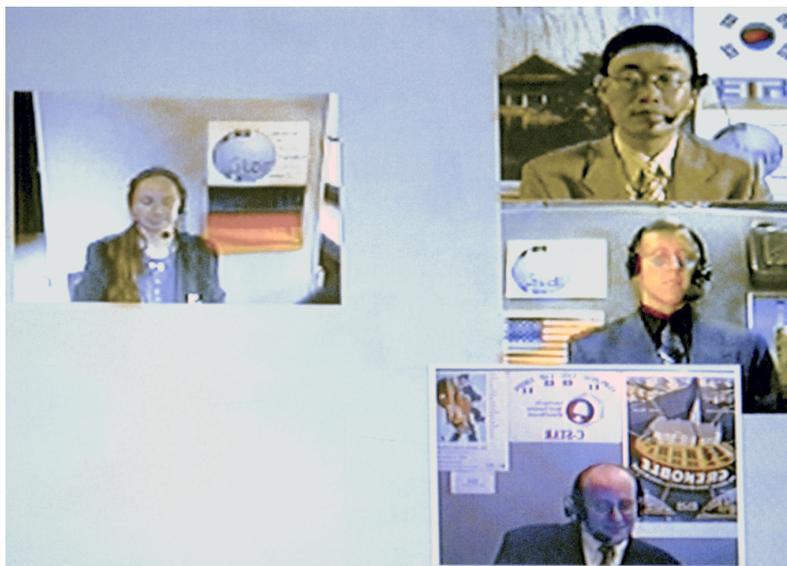


Figure XI-2 : Écran de visioconférence multisite du démonstrateur CSTAR II du groupe CLIPS++
L'écran d'interface utilisateur (Figure XI-3) est divisé en deux.

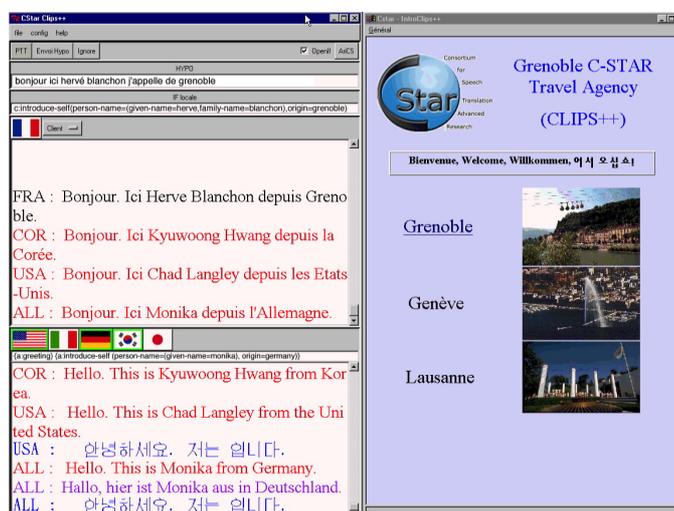


Figure XI-3 : Interface utilisateur du démonstrateur C-STAR II du groupe CLIPS++

À gauche se trouve l'interface du système de traduction proprement dit. La partie droite est consacrée à l'affichage des pages web que l'agent de voyage propose au client.

L'interface du système de traduction est détaillée Figure XI-4. Le bouton PTT est utilisé pour contrôler le module de reconnaissance de la parole. Lorsqu'il est enfoncé, le bouton devient orange, indiquant que le module de reconnaissance est prêt. Lorsque l'utilisateur commence à parler, le bouton devient vert, la reconnaissance de ce qui est dit commence et les résultats intermédiaires sont proposés à l'utilisateur en bleu.

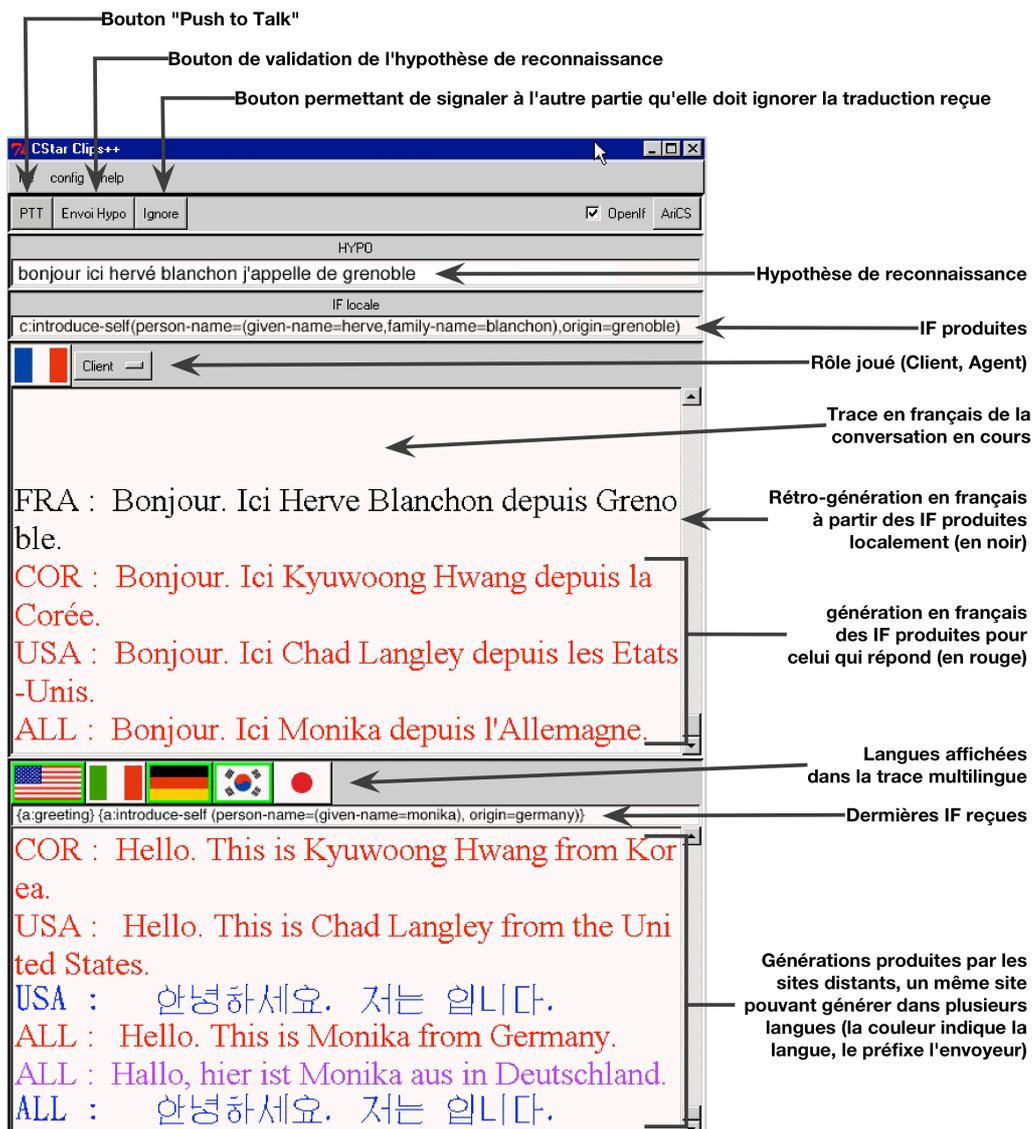


Figure XI-4 : Interface de contrôle du démonstrateur C-STAR II du groupe CLIPS++

Lorsque l'utilisateur se tait, le bouton devient de nouveau gris et le résultat final de la reconnaissance est affiché en noir. Nous aurions ainsi pu nous passer de la technique du "Push to Talk". Mais nous avons choisi de ne pas produire d'IF pour des reconnaissances trop éloignées du tour de parole d'origine. C'est une solution qui permet de gagner du temps, puisque, dans le démonstrateur, le module français vers IF est moins rapide que le module de reconnaissance de la parole. Nous utilisons donc un bouton de validation d'hypothèse de reconnaissance.

Lorsque l'utilisateur est un agent, l'interface comprend aussi un module qui permet de choisir une page web à envoyer au client. Les pages envoyées sont aussi affichées localement pour le contrôle visuel de l'agent. Un ensemble de pages Web, sur le transport, les hôtels et les activités touristiques, a été préparé par chaque partenaire. Ces pages sont disponibles dans la langue de chaque partenaire et dans la langue des partenaires avec lesquels il correspond. La Figure XI-5 montre une page web proposée au client français par l'agent de voyage coréen.



Figure XI-5 : Page web d'activité touristique proposée par l'agent coréen à un client français

XI.1.4. Démonstrations du CLIPS++

L'objectif du consortium était clairement orienté vers l'organisation de démonstrations afin de faire une publicité la plus large possible à nos travaux. Nous avons donc essayé d'en faire le plus possible.

XI.1.4.1. Démonstration finale

Un début de participation tardive et des moyens plus réduits que ceux de nos autres partenaires nous avaient incités à prévoir des démonstrations impliquant uniquement nos partenaires américain et allemand. Les premiers essais d'intégration des différents démonstrateurs, sur la base d'un scénario minimum, se sont déroulés en janvier et février 1999.

Le mois de mars a été consacré à la résolution de problèmes techniques concernant en particulier : la mise en place de la communication intersystèmes avec le protocole IP encapsulé dans une communication sous RNIS ; la vidéoconférence multipoint (très peu de systèmes de visioconférence multipoint à 384 Kilobit par seconde étaient homologués par France Telecom) ; l'interface du démonstrateur ; et l'intégration robuste des composants localement.

À cette phase de préparation technique ont succédé trois mois de préparation fine des modules de traitement linguistique (reconnaissance de la parole, analyse, génération). Les séances d'entraînement, hebdomadaires au début, sont finalement devenues à peu près quotidiennes au mois de juin.

Observant que nos résultats étaient devenus acceptables mi-mai, nous avons décidé de travailler aussi avec notre partenaire coréen. L'utilisation de l'IF a permis aux démonstrateurs français et coréens d'interopérer presque immédiatement à quelques problèmes de codages en IF près. Avec les décalages horaires, cela signifie que nous étions sur le « pont » de neuf heures du matin à dix heures du soir.

XI.1.4.1.1. Lieu, audience et couverture médiatique

La démonstration finale du 22 juillet 1999 s'est déroulée dans les locaux de la maison Jean Kuntzmann à l'institut IMAG devant 90 personnes parmi lesquelles des représentants des tutelles du CLIPS, des représentants de l'industrie et de centres de recherche (IBM, Xerox, Digigram, Azimut, FranceTelecom, Magellan Ingénierie, CNET, Spacio Guide), ainsi que de l'ANVAR.

Nous avons aussi fait des démonstrations privées de tout ou partie du système à Alcatel, IBM, Thomson, Texas Instrument, Schneider, Parrot, CyberStudio, EDF, Soprane, Grany Media.

La couverture médiatique observée est la suivante :

- Agences de Presse : AFP, AGRAP, Temps d'm, Agence REA, Editing Corporate.
- Presse écrite : Le Monde, Libération, Le Dauphiné Libéré, Science & Vie, Science & vie junior, Micro-Hebdo, Science & Vie Micro (dossier) et Wired.
- Télévision : France 3 et M6
- Radio : France Inter, France Info, Radio France Isère, BFM, Europe 2, RMC, RTL, RCF Isère.

XI.1.4.1.2. Scénario démontré

La fin de matinée du 22 a été consacrée à une session introductive à propos de C-STAR II, son histoire, les techniques mises en œuvre dans la traduction de parole, et le démonstrateur du groupe CLIPS++. La démonstration en elle-même a eu lieu à 14 heures.

Nous avons montré le scénario suivant :

Un client se connecte à une agence de voyage virtuelle avec des relais locaux en Corée du Sud, en Allemagne et aux États-Unis. Il salue d'abord les trois agents de voyage disponibles et leur demande des informations sur l'heure, le temps et la température dans leurs villes respectives. Dans cette phase, toutes les synthèses des agents de voyage étaient renvoyées dans le circuit de visioconférence. Ainsi, le client entendait en même temps les traductions en coréen, anglais et allemand de ce qu'il venait de dire. Les étapes suivantes se déroulaient avec la seule synthèse de l'agent interlocuteur audible.

Le client organise ensuite un voyage avec un collègue à Taejon en Corée du Sud. Il achète des billets d'avion pour Séoul, s'enquiert des moyens de transport possibles de Séoul à Taejon, réserve une chambre d'hôtel, demande des informations touristiques sur les environs de Taejon, et règle avec une carte de crédit.

Il prépare alors un voyage à New York avec sa femme. Il achète des billets d'avion pour New York, réserve une chambre d'hôtel, demande à voir un match de base-ball, et règle avec une carte de crédit.

Enfin, il organise un voyage à Heidelberg. Il achète un billet de train, cherche une chambre d'hôtel à un prix raisonnable, demande des informations touristiques sur les environs de Heidelberg, et règle avec une carte de crédit.

Pour les besoins de la démonstration, le client salue enfin nominativement les trois agents avec lesquels il a interagi.

XI.1.4.1.3. Résultats

Bien que nous ayons réalisé tardivement que le scénario général était assez répétitif, l'audience n'a fait aucun commentaire à ce sujet. Nous avons cependant bien veillé à utiliser des formulations différentes chaque fois que les mêmes thèmes étaient évoqués. La cérémonie d'ouverture, avec les synthèses distantes fonctionnant en parallèle, a été très bien reçue. Quelques erreurs ont su amuser les spectateurs. La démonstration a duré une demi-heure pendant laquelle les gens n'ont pas vu le temps passer.

La conversation avec l'agent coréen a eu le plus gros succès. L'écriture hangul (écriture coréenne) et une langue que l'on n'entend jamais, ont bien montré l'intérêt de la traduction

de parole lorsqu'il y a un message à échanger et pas de langue en partage pour le faire. En Europe, les gens sont moins sensibles à cet état de fait en ce qui concerne l'allemand ou l'anglais. La plupart des exemples repris par les médias concernaient des applications potentielles de la traduction de parole français-japonais ou français-coréen. Nous avons pourtant essayé de faire passer le message que, même si une langue est en partage, pour des choses qui dépassent le vocabulaire de tous les jours, une assistance est toujours nécessaire pour faire passer un message qui doit être précis.

La vitesse des systèmes n'a pas été commentée. Les temps de réponse des systèmes ont semblé raisonnables et acceptables. Dans le démonstrateur du CLIPS++, il fallait en moyenne 4 secondes pour faire du français vers l'IF. L'IF vers le français est plus rapide, de l'ordre de 1 à 2 secondes. L'échange de pages web contenant parfois des visites virtuelles des hôtels ou des musées a été aussi bien apprécié. Cela semble faire maintenant partie du décor et enrichir grandement l'interaction.

XI.1.4.2. Autres démonstrations organisées

XI.1.4.2.1. TEC-99 (Grenoble)

À l'occasion du salon TEC-99 qui s'est déroulé à Grenoble, nous avons effectué une démonstration dans les locaux de l'IMAG le 13 octobre 1999. Nous jouions le rôle d'un client et notre partenaire coréen celui d'un agent de voyage.

XI.1.4.2.2. IMAG (Grenoble)

Le 26 juin 2000, nous avons démontré le même scénario lors de la visite de Monsieur Vincent Courtilot alors conseiller spécial du ministre de l'éducation nationale, de la recherche et de la technologie.

XI.1.4.2.3. IBM (Paris)

Nous avons fait une démonstration similaire avec notre partenaire coréen le 27 octobre 1999 pour IBM France. Lors de cette démonstration, nous avons joué le rôle d'un client pour des agents allemand et coréen, et le rôle d'un agent pour un client coréen.

XI.1.4.3. Participations

XI.1.4.3.1. Démonstration finale de ETRI

Le 22 juillet 1999, nous avons pris part à la démonstration de ETRI en jouant, avec les partenaires américain et japonais, le rôle d'un agent de voyage. Comme pour le français-coréen, notre partenaire nous a transmis l'impact auprès du public de la démonstration coréen-français qui a même fait l'ouverture du journal télévisé du soir de la première chaîne de télévision publique Coréenne.

XI.1.4.3.2. TELECOM-99 (Genève)

Lors du salon Telecom-99 à Genève (9-16 octobre 1999), nous avons participé, avec l'IRST, à une série de démonstrations organisées par ETRI, en jouant le rôle d'un agent de voyage.

XI.1.5. Bilan

Cette première expérience en traduction de dialogues oraux s'est avérée très positive. Du point de vue scientifique, elle nous a permis de rejoindre une communauté à la pointe dans un domaine encore émergent à la fin des années 90.

Les problèmes scientifiques auxquels nous avons été confrontés sont la reconnaissance de la parole spontanée multilocuteur et l'analyse d'une entrée bruitée. Nous avons aussi résolu des problèmes technologiques au niveau de l'intégration des différents systèmes de traduction et au niveau de l'intégration des systèmes de visioconférence. En 1999, les systèmes à 384 kilobits par seconde étaient encore rares et leur intégration n'est pas allée sans poser quelques problèmes à nos fournisseurs de services.

L'objectif de produire des démonstrateurs communs a créé un véritable effet « boule de neige » qui a permis à notre petite équipe de tenir finalement un rôle plus qu'honorable et d'acquiescer la confiance de nos partenaires comme en témoigne notre participation aux démonstrations de ETRI. Les succès obtenus ont été sans aucun doute la clé du financement du projet NESPOLE! par l'Union Européenne.

XI.2. NESPOLE! : modules implémentés

Pour l'architecture du linguiciel, nous avons choisi de réutiliser l'approche par pivot qui avait montré ses avantages dans le cadre de C-STAR II.

XI.2.1. Langage pivot de NESPOLE!

Les principes de base n'ont pas changé, une représentation IF est toujours construite à partir d'un acte de parole, de concepts et d'arguments.

XI.2.1.1. Définition et couverture linguistique

L'IF utilisé dans le projet NESPOLE! est défini au moyen de soixante-quinze actes de parole (donc trente concernent le méta-dialogue), cent vingt-trois concepts, cent quatre-vingt-neuf arguments et deux cent soixante-seize ensembles de valeurs (totalisant 8 113 valeurs). La spécification est répartie en trois fichiers : spécification des actes de parole (2 300 lignes) et concepts (2 500 lignes), spécification des constructions des arguments (13 000 lignes), et spécification des listes de valeurs (5 600 lignes). On peut donc d'abord observer un changement d'échelle par rapport à l'IF de C-STAR II.

Les actes de parole sont définis avec leurs continuations (e.g. les concepts disposition, price, availability peuvent suivre give-information) et leurs arguments (des relations rhétoriques — e.g. cause=, conjunction=, disjunction= —, ou des arguments « ordinaires » — origin=, to-whom= —) autorisés.

DEF: give-information

continuations: ((+negation +accommodation) (+negation +action) (+negation +activity) (+negation +address) (+negation +admission) (+negation +airport) (+negation +arrival) (+negation +attraction) (+negation +availability) (+negation +balance) (+negation +begin-operation) +accommodation +action +activity +address +admission +airport +arrival +attraction +availability +balance +begin-operation +calculate +cancellation +change +checkin +checkout +click +completion +concept +connection +contain +currency +current-time +decrease +delay +departure +deposit +directions +display +disposition +tour +transportation +trip +understand +vehicle +view +weather +worsen +write)

arguments: (<after-rr=> <anti-condition=> <before-rr=> <besides=> <cause=> <co-occurrence=> <concessive=> <condition=> <conjunction=> <contrastive=> <dependency=> <destination=> <digression=> <disjunction=> <duration=> <e-time=> <end-of=> <end-of=> <exception=> <factuality=> <final=> <first=>

<focalizer=> <focus=> <for-whom=> <fourth=> <frequency=> <location=> <manner=> <middle-of=> <middle-of=> <origin=> <polarity=> <property=> <purpose=> <related-to=> <rhetorical=> <second=> <start-of=> <start-of=> <substitution=> <third=> <time=> <to-whom=> <while=> <with-whom=> <with=>)

comment: comment source -- give-information

comment: Any statement.

comment: Ex: "I am doing that." = give-information+action

comment: Compare to: "I will do that." = promise+action

comment: comment source -- *basic-info-sa*

comment: give-information and request-information cannot appear without a continuation concept.

Les concepts sont aussi définis avec leurs continuations (les concepts accommodation, room, activity peuvent suivre availability) et leurs arguments autorisés (e.g. for-whom, price, time sont des arguments autorisés de availability).

Les concepts qui n'ont pas de successeur (cf. +room page suivante) sont appelés des « concepts de focus » ou "focus concepts". Dans la suite du document, nous utiliserons la terminologie de l'IF.

DEF: +availability

continuations: (\$bottom\$ +accommodation +activity +admission +airport +arrival +attraction +concept +currency +departure +equipment +event +facility +information-object +meal +meeting +object +package +person +restaurant +room +seat +service +stay +tour +transportation +trip +vehicle)

arguments: (<beneficiary=> <destination=> <distance=> <duration=> <for-whom=> <frequency=> <location=> <origin=> <price-spec=> <price=> <provider=> <speed=> <time=> <via=> <with-whom=> <with=>)

comment: comment source -- *property-concepts*

comment: Property concepts: +availability, +occurrence, +existence, and +providing.

comment: +avail is used in the context of making reservations, to indicate that the agent can reserve something for a client.

comment: +providing is used where something is being made available by a specified provider, but not for reservations.

comment: +exist is used otherwise, even if the word "available" is present.

comment: Ex: +avail "Two double rooms are available."

comment: Ex: +providing "The hotel promotes/organizes/proposes tours for children."

comment: Ex: +exist "Shuttle service is available."

comment: Nespole usually produce +providing or +exist, the agent is info purveyor, and not performing the reservation task.

comment: Sentences like "you can reserve a seat for that performance" will be covered by: give-info+feasibility+reservation+...

comment: MAIN-PREDICATION

DEF: +room

continuations: (\$bottom\$)

arguments: (<bed-spec=> <destination=> <distance=> <duration=> <for-whom=> <frequency=> <location=> <meal-spec=> <occupancy=> <origin=> <room-spec=> <scenic-view=> <speed=> <telephone-number=> <time=> <via=> <with-whom=> <with=>)

comment: comment source -- *focus-concepts*

comment: MAIN-PREDICATION

comment: PRED-PARTICIPANT

DEF: room-spec=

Definition 1 (of 1)

:values

question relative pronoun *rooms* *room-names*

:relations

<accompanied-by=> <bed-spec=> <besides=> <called=> <consist-of=> <contain=> <contained-in=> <contrastive=> <cultural-style=> <dependency=> <destination=> <distance=> <duration=> <exclude=> <excluded-from=> <extension=> <for-whom=> <for=> <frequency=> <include=> <included-in=> <location=> <object-number=> <origin=> <per-unit=> <price-style=> <price=> <provider=> <purpose=> <related-to=> <smoking=> <specifier=> <time=> <whose=> <with-whom=>

:attributes

<age=> <audio-level=> <color=> <compass-point=> <focalizer=> <grouping=> <identifiability=> <language=> <manner=> <max-min=> <modifier=> <nationality=> <noise-level=> <normality=> <object-ref=> <operator=> <order-ref-ext=> <order-ref-int=> <plus-minus=> <polarity=> <portion=> <quantity=> <relative-region=> <sex=> <shape=> <side=> <size=> <temperature=> <type=>

:comments

```

shared comments from argument class -- @all
shared comments from argument class -- @qval
shared comments from argument class -- @spec
a double room with a double bed ==> room-spec=(double, bed-spec=(double,id=no))
Some buildings/campgrounds have special names for rooms/sites.
the Gertrude Stein room ==> (room-spec=name-gertrude_stein_room)
the Lincoln Bedroom ==> (room-spec=name-lincoln_bedroom)
shared comments from - room-spec=

```

Les arguments (cf. room-spec= ci-dessus) sont définis par une ou plusieurs définitions. Une définition comporte une valeur, et une collection de relations et d'attributs. La valeur peut être question⁹⁹, relative (la valeur est exprimée par un pronom relatif dans une proposition relative séparable¹⁰⁰), pronoun (la valeur est exprimée par un pronom démonstratif¹⁰¹), ou alors c'est une valeur dans un ensemble de valeurs (*room*, *room-name*). Les relations définissent des liens entre deux concepts (e.g. bed-spec, location, price pour room-spec). Les attributs définissent des liens entre un concept et un ensemble de valeurs (e.g. quantity, identifiability). Les attributs sont des arguments définis uniquement par une valeur et un ensemble d'attributs (pas de relation).

Les ensembles de valeurs sont définis en extension par une collection de valeurs atomiques ou d'ensembles de valeurs. Ils sont définis « à la demande » et ne cherchent pas à décrire une ontologie du domaine, ni une stricte hiérarchie ; ainsi par une suite d'inclusions, un même ensemble de valeurs **X** peut apparaître deux fois dans la définition d'un ensemble de valeurs **Y**.

```

DEF: *rooms*
bedroom camping_lot double_room family_room junior_suite matrimoniale room royal_suite senior_suite
single_room site sleeping_compartment suite tatami_room tent triple_room twin_room vehicle-cabin

```

L'IF permet de représenter les phénomènes linguistiques suivants :

- la négation,
- les modaux (disposition, éventualité, savoir indirect, possibilité, savoir direct, obligation),
- les fragments (pour les tours de parole incomplets ou interrompus sans prédicat),
- la coordination de phrases, d'ensembles,
- la détermination,
- les propositions relatives,
- les relations rhétoriques (exprimées explicitement),
- la question,
- les expressions temporelles,
- la comparaison,
- les ensembles.

Nous donnons d'autres exemples de la définition de l'IF de NESPOLE! en Annexe VII.

⁹⁹ Exemple : « Où y a-t-il une chambre disponible. » → request-information+availability+room (location=**question**, room-spec=(room, identifiability=no)).

¹⁰⁰ « Je veux la chambre | que vous proposez|. » est séparable → give-information+disposition+room(disposition=(desire,who=i), room-spec=(room, identifiability=yes)) + give-information+recommendation+object(object-spec=**relative**, who=you).

« La chambre qui est à Cavalese. » n'est pas séparable → give-information+concept(room-spec=(room, identifiability=yes, location=name-cavalese)).

¹⁰¹ « Il y une chambre disponible dans celui-ci. » → give-information+availability+room (location=**pronoun**, room-spec=(room, identifiability=no)).

XI.2.1.2. Analyse critique

Spécification complètement fondée sur les données et mouvante

L'IF ne veut pas décrire le domaine du tourisme à la façon d'une ontologie, sa spécification s'enrichit lorsque c'est nécessaire. C'est-à-dire lorsqu'un énoncé E qui appartient manifestement au domaine ne peut pas être représenté par la version courante de la spécification. Dans ce cas, la spécification est augmentée afin de pouvoir représenter le nouvel énoncé E.

Ainsi la spécification de l'IF a régulièrement changé au cours du projet. Sur une période de vingt mois entre janvier 2001 et août 2002, nous avons pu observer : 300 modifications dans la description des actes de dialogue effectuées en 100 étapes, 300 modifications dans la description des arguments effectuées en 100 étapes, et 730 modifications dans la description des valeurs effectuées en 200 étapes.

Il faut noter de plus que la métagrammaire utilisée pour décrire cette spécification a changé plusieurs fois ; la dernière fois, il faut le reconnaître, à notre demande. Nous avons, en effet, l'intention de produire un générateur IF vers français avec ARIANE-G5 de manière semi-automatique en analysant les fichiers de spécification et en construisant automatiquement des squelettes de dictionnaires et de grammaires ARIANE [Mazenot, S. 2002].

Structure plate

Une représentation IF est complètement plate. Ainsi, pour représenter la subordination dans l'énoncé « l'hôtel que vous proposez est trop cher », la spécification de l'IF [Levin, L., *et al.* 2003] propose la représentation suivante :

```
give-information+price+accommodation (accommodation-spec=(hotel, identifiability=yes),
price=(modifier=(expensive, intensity=over)))
+
give-information+recommendation+object (object-spec=relative, who=you)
```

Lors de l'étape de génération, il est impossible de produire un énoncé qui ressemble à l'énoncé d'origine. La première IF signifie « le prix de l'hôtel est trop cher », la seconde IF signifie « Vous recommandez ce dont nous venons de parler. »

Spécification sur-génératrice

Le dernier problème que l'on peut évoquer à propos de la spécification de l'IF est que, de manière interne (que l'on ne peut pas voir lorsque l'on consulte la spécification), les relations et les attributs sont définis sous forme de familles qui sont associées aux différents arguments. Ainsi, par exemple, l'argument room-spec est doté des attributs age, audio-level (qui décrit le volume d'un son), language, manner, relative-region, et sex.

Cet argument n'est qu'un exemple parmi tant d'autres. Nous n'avons malheureusement pas réussi à faire changer les choses en ce qui concerne ce phénomène. Ce phénomène a aussi été très pénalisant pour produire un générateur en Ariane-G5.

XI.2.1.3. Exemples comparatifs entre C-STAR II et NESPOLE!

Dans le tableau suivant, nous donnons les résultats de l'analyse en IF, en utilisant les analyseurs développés, de différents énoncés.

Énoncé	<client> et je voudrais deux chambres du dix au quinze septembre à trento
C-STAR II	{c:request-action+reservation+features+room+temporal (room-type=(room, quantity=2), time=(start-time=md10, end-time=(september, md15), location=trento))}
NESPOLE!	{c:give-information+disposition+room (conjunction=discourse, disposition=(desire, who=i), room-spec=(identifiability=no, quantity=2, bedroom, location=name-trento), time=(start-time=(md=10), end-time=(md=15, month=9)))}
Énoncé	<client> et je voudrais une chambre du dix au quinze septembre à trento
C-STAR II	{c:request-action+reservation+features+room+temporal time=(start-time=md10, end-time=(september, md15, location=trento))}
NESPOLE!	{c:give-information+disposition+room (conjunction=discourse, disposition=(desire, who=i), room-spec=(identifiability=no, quantity=1, bedroom, location=name-trento), time=(start-time=(md=10), end-time=(md=15, month=9)))}
Énoncé	<client> non merci beaucoup au revoir
C-STAR II	{c:negate} {c:thank} {c:closing}
NESPOLE!	{c:negate} {c:thank(manner=(intensity=intense))} {c:greeting (greeting=formal_goodbye)}

Table XI-1 : Comparaison des IF C-STAR II et NESPOLE! sur quelques exemples

XI.2.2. Analyse français vers IF

Pour l'analyse, nous avons choisi une approche fondée sur des patrons. Il s'agit principalement de faire une analyse par îlots, ou par segments, du résultat de la reconnaissance de la parole. Nous ne pouvons en effet pas nous appuyer sur une syntaxe générale forte des énoncés. En effet, l'étude des transcriptions des corpus collectés montre une fréquence assez élevée de fragments et des énoncés dont la syntaxe du français écrit est bousculée par les hésitations et les autocorrections.

Cette approche se situe dans la lignée des travaux sur l'analyse par patrons [Kitamura, M. and Murata, T. 2003, Zong, C., *et al.* 2000a] et de l'analyse par transducteur d'états finis [Amengual, J. C., *et al.* 1997, Bangalore, S. and Riccardi, G. 2001, Casacuberta, F., *et al.* 2002, Hetherington, L. 2004a, Mohri, M. 1997]¹⁰². Ces deux approches ne sont que deux façons de voir les mêmes opérations. Avec la première approche, on reconnaît d'abord un patron puis on manipule, transforme les variables instanciées pour produire une sortie. Avec la seconde approche, la reconnaissance et la transformation ont lieu lors du parcours du transducteur.

L'analyse s'appuie donc sur des segments d'énoncés bien formés. Les segments qui nous intéressent correspondent à la verbalisation d'informations pertinentes pour l'IF, en particulier des valeurs d'arguments. La description d'un segment d'information pertinent est implémentée sous forme d'expressions régulières. La fabrication des expressions régulières s'appuie sur les spécifications de l'IF et non sur des données collectées.

Pour le premier démonstrateur, nous avons collecté trente et un dialogues monolingues français avec une moyenne de 106,83 tours de parole par dialogue. La taille du vocabulaire est de 2070 lemmes. Ces données n'ont pas été annotées en IF. Pour le second démonstrateur, nous avons collecté dix-sept dialogues monolingues français avec une

¹⁰² Nous signalons au lecteur qu'il existe des boîtes à outils disponibles sur le Web, proposées par AT&T [Allauzen, C., *et al.* 2003, Mohri, M., *et al.* 2003] et le MIT [Hetherington, L. 2004b].

moyenne de 258,1 tours de parole. La taille du vocabulaire est de 2028 lemmes. Ces données ont été annotées en IF. Un extrait de dialogue annoté est donné en Annexe VIII.

L'implémentation de la reconnaissance des arguments a peu changé entre la première version de l'analyseur réalisé pour le premier démonstrateur NESPOLE! et celle du second démonstrateur. Par contre, des changements importants ont été apportés à l'architecture globale de l'analyseur.

XI.2.2.1. Analyseur du premier démonstrateur

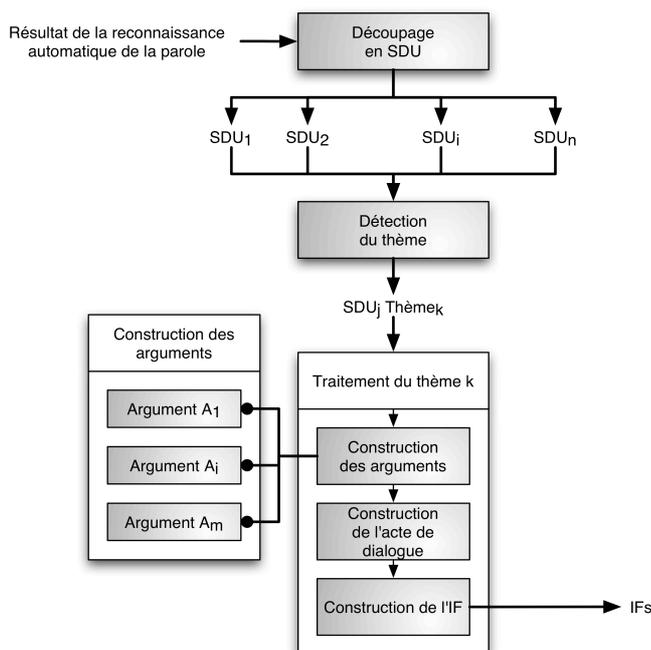


Figure XI-6 : Architecture de l'analyseur français vers IF du premier démonstrateur NESPOLE!

Pour le premier démonstrateur, l'analyse se déroule en quatre étapes. Le texte d'entrée est d'abord découpé en Unités Sémantiques de Dialogue¹⁰³ (SDU). Pour chaque SDU, un domaine est alors calculé. Selon le domaine, les arguments privilégiés autorisés et présents sont instanciés. Finalement, l'acte de dialogue est calculé en fonction des arguments en s'aidant aussi d'autres propriétés de la SDU. L'architecture du premier analyseur est donnée Figure XI-6.

XI.2.2.1.1. Segmentation en SDU

Pour découper la sortie textuelle du module de reconnaissance de la parole, nous utilisons des groupes simples et des articulations. Les groupes simples permettent de construire immédiatement un acte de parole sans continuation (DA). Ils sont classés en :

- affirmations (exemples : oui c'est ça, bien sûr),
- acquiescements (exemples : c'est d'accord, c'est ok, c'est bien, ok, oui),
- excuses (exemples : excusez-moi, désolé, pas de quoi),
- exclamations (exemples : c'est excellent, très bien, oh),
- salutations (exemples : bonjour, au revoir, bonne journée),

¹⁰³ Une SDU est un segment du texte à analyser qui peut être représenté par une unique IF. Un segment correspond donc normalement à une phrase.

- gestions du dialogue (exemples : allô, j'entends, j'écoute),
- remerciements (exemples : merci bien, merci), et quelques autres.

Les articulations sont les réalisations des arguments rhétoriques (conjonctions simples, syntagmes de conjonction) suivis d'un pronom ou d'un début de question (exemples : et donc je, et puis il, et j', donc on, est-ce que, quel est). Nous utilisons au total vingt-neuf expressions régulières pour faire la segmentation.

XI.2.2.1.2. Détection du domaine

Il s'agit de trouver soit un acte de parole terminal pour la SDU (Acknowledge, Affirm, Apologize, ApologyResponse, EndWait, Exclamation, Greeting, IntroduceSelf, Repeat, Reject, Negate, PleaseWait, RequestUnderstand, Thank, ThankResponse, Verify), qui sont traités sans arguments sauf pour Exclamation, Greeting, IntroduceSelf), soit un focus concept (Visit, Activity, Room, Accommodation, Attraction, Event, Admission, TransportLocation, Information, Locomotion, Plan, Price, Time, Distance, View). Ces derniers domaines correspondent à des arguments qui sont associés à des concepts terminaux dans la spécification de l'IF.

Une liste de mots et/ou d'expressions est associée aux actes de paroles terminaux (exemple : bonjour, salut, à bientôt, bonsoir, au revoir, enchanté, à plus tard, à plus, bonne journée pour greeting) et aux autres concepts (exemple : place de camping, salle de conférence, chambre double, chambre simple, suite pour room).

Lorsque c'est possible, un domaine est instancié, sinon le domaine est unknown. Ce dernier concerne des SDU sans domaine explicite (exemple : « 1 3 5 7 ») ou des concepts non encore traités. Pour chaque domaine, une fonction **domaine2lf** construit l'IF associée à la SDU considérée.

Lorsque plusieurs domaines peuvent être instanciés, celui qui est finalement choisi est celui qui apparaît le plus rapidement après le début de la SDU.

XI.2.2.1.3. Instanciation des arguments

Chaque fonction **domaine2lf** instancie les arguments autorisés pour l'acte de parole ou le concept associé. Par exemple dans le domaine room, le système essaie d'instancier room-spec, location, et duration qui sont des arguments autorisés.

Des fonctions **Argument2lf** construisent la valeur IF des arguments en essayant de trouver l'une des réalisations possibles de ceux-ci dans le texte d'entrée.

Cf. Annexe IX pour un exemple.

XI.2.2.1.4. Construction de l'acte de dialogue

L'acte de dialogue est le résultat de la concaténation de l'acte de parole, de l'attitude (disposition seulement), du prédicat principal et des participants du prédicat.

L'acte de parole est calculé en utilisant des informations sur la construction verbale (affirmation, question, rejet), la négation éventuelle de la construction verbale, les éventuelles vérifications ou demandes de vérification.

L'attitude est reconnue en utilisant des groupes syntagmatiques modèles. Ensuite, en fonction des arguments instanciés, le prédicat principal ainsi que les participants du prédicat sont calculés.

Finalement, l'IF est construite en concaténant le codage de celui qui parle (a: ou c:), l'acte de dialogue et les arguments mis entre parenthèses.

Cf. Annexe IX pour plus de détails.

XI.2.2.2. Critique de la solution adoptée

La couverture de ce premier analyseur est plutôt limitée. Pour les actes de paroles terminaux seize sont reconnus sur un total de trente-neuf (dont dix-neuf qui portent sur le dialogue). Pour les actes de parole non terminaux, six sont reconnus sur un total de dix-huit. Sur cent trois concepts, trente et un sont utilisés.

L'analyseur couvre un cinquième des arguments de niveau supérieur (trente-deux sur cent cinquante) sachant que les vingt et un arguments rhétoriques et huit attitudes ne sont pas traités. Sur le total des arguments, un quart est traité (soixante-treize sur deux cent quatre-vingt-un) sachant qu'il y a trente-cinq définitions synonymes.

La construction de l'acte de dialogues est très lacunaire à cause de la trop faible couverture des concepts. Pour les mêmes raisons, la construction des arguments est très limitée aussi.

XI.2.2.3. Analyseur du second démonstrateur

Au regard du bilan critique des résultats obtenus avec le premier analyseur et des corpus collectés pour le second démonstrateur, il était nécessaire de procéder à une augmentation importante de couverture de l'analyseur.

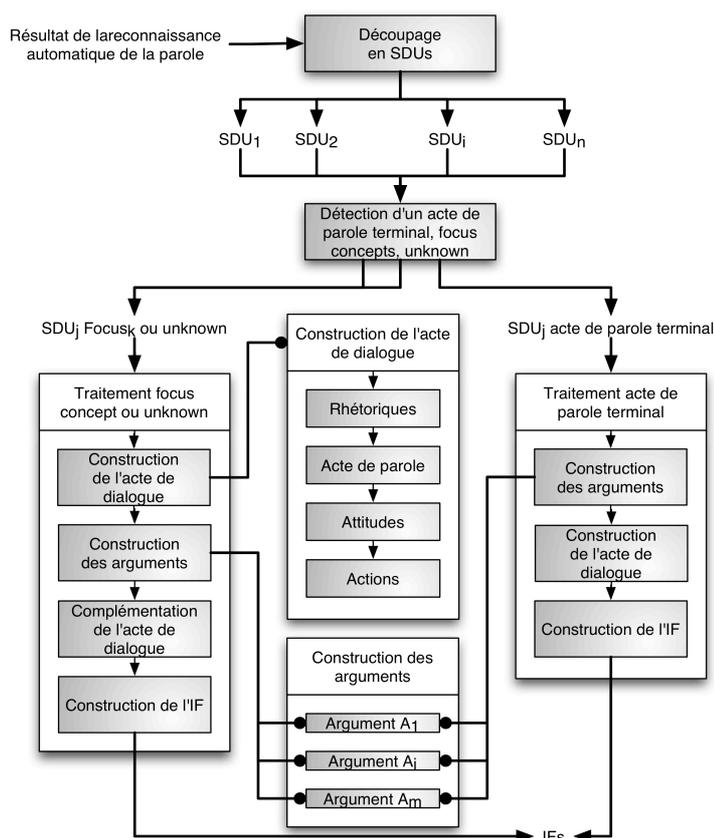


Figure XI-7 : Architecture du module d'analyse français vers IF du second démonstrateur NESPOLE!

Le corpus français collecté pour le second démonstrateur a été entièrement annoté en IF par Cécile Ponsard [Ponsard, C. 2002] qui a aussi participé à l'amélioration du module d'analyse. Une étude fine des données obtenues a révélé que les actes de dialogue étaient bien plus variés et complexes que dans les corpus collectés pour le premier démonstrateur.

Afin de réaliser l'augmentation de couverture requise, nous avons choisi de mieux modulariser les traitements de l'analyseur. Ainsi, la construction de l'acte de dialogue a été partagée au maximum. Nous avons aussi augmenté la couverture de l'analyseur en traitant une portion plus grande de la spécification de l'IF.

XI.2.2.3.1. Segmentation des SDU

La segmentation en SDU, tout en utilisant le même principe que pour le premier analyseur, met en jeu une stratégie plus fine. Nous gérons des cas de conjonction de subordination et de coordination interphrastique bien plus complexes. Nous utilisons maintenant trente-neuf expressions régulières pour cette tâche. Nous gérons aussi les appositions de phrase (coordinations et subordinations non marquées explicitement) au moyen d'un ensemble de dix-sept expressions régulières.

XI.2.2.3.2. Détection du domaine

Il s'agit de trouver un acte de dialogue terminal ou un focus concept (un concept terminal, sans successeur). Les actes de dialogue terminaux sont maintenant tous détectés, soit quarante-trois au total. Pour les focus concepts, nous en reconnaissons vingt-trois sur un total de trente-deux.

Des mots clés permettant de les sélectionner sont toutes les valeurs possibles prises par la valeur de leur argument privilégié, telles qu'elles sont proposées dans la spécification de l'IF. Par exemple, pour le concept *room*, l'ensemble des mots clés utilisés comporte vingt-trois valeurs.

Lorsque plusieurs domaines peuvent être instanciés, celui qui est finalement choisi est celui qui apparaît le plus rapidement après le début de la SDU. Si aucun domaine ne peut être instancié, le domaine est mis à la valeur *unknown*. Pour chaque domaine, une fonction **domaine2lf** construit l'IF associée à la SDU considérée.

Avec cet analyseur, lorsque l'on traite un domaine, on construit d'abord un préfixe de l'acte de dialogue avant d'instancier les arguments privilégiés du focus concept et de compléter éventuellement l'acte de dialogue par d'autres concepts et finalement le focus concept lui-même.

XI.2.2.3.3. Construction d'un préfixe de l'acte de dialogue et instanciation des arguments liés à celui-ci

La construction du préfixe de l'acte de dialogue se déroule en six étapes :

1. construction des arguments rhétoriques ;
2. construction de l'acte de parole ;
3. construction des attitudes ;
4. construction des actions définies comme telles dans la spécification (nous en traitons neuf sur vingt-cinq) ;
5. construction des concepts que l'on peut repérer grâce à des patrons très simples (+existence, +price, +reception, +in-operation, +understand, +providing) ;
6. construction du concept +action qui permet d'instancier des actions qui ne sont pas des concepts, et qui sont définies par le nom de leur "WordNet synset" (nous en reconnaissons vingt-six sur un total de cent-trente-cinq définies dans la spécification de l'IF).

XI.2.2.3.4. Instanciation des arguments liés au domaine et gestion des subordinations

Nous procédons de la même manière que dans le premier module pour l'instanciation des arguments. Cependant, nous nous intéressons seulement aux arguments préférés du focus concept en cours de traitement. De plus, nous essayons cette fois de construire des arguments qui prennent en compte la subordination.

L'IF de NESPOLE! permet en effet de gérer la subordination. Par exemple dans la phrase « Je voudrais deux chambres simples à Cavalese. », on a deux interprétations possibles : « à Cavalese » est subordonné à « deux chambres simples » (la bonne interprétation ici), ou « à Cavalese » est circonstant rattaché au verbe.

Si l'on applique strictement la spécification de l'IF, l'argument location qui représente « à Cavalese » doit donc être un sous-argument de room-spec et on obtient l'IF suivante :

```
{c:give-information+disposition+room(disposition=(desire, who=i),
room-spec=(identifiability=no, quantity=2, single_room, location=name-cavalese))}
```

On observe pourtant que les analyseurs produisent en général l'IF suivante :

```
{c:give-information+disposition+room(disposition=(desire, who=i),
room-spec=(identifiability=no, quantity=2, single_room), location=name-cavalese))}
```

qui, si on applique la spécification stricte de l'IF, représente l'interprétation suivante : « à Cavalese (circonstant), je voudrais deux chambres simples (dans quelle ville ?) ».

Si on génère ces deux IF, on obtient la même phrase en français : « Je désirerais 2 chambres simples à Cavalese. ». L'ambiguïté est toujours présente, elle le sera de même avec les générateurs de l'anglais et de l'italien. En allemand, on devrait obtenir "Ich möchte 2 Zimmer in Cavalese." et non pas "Ich möchte 2 Zimmer von Cavalese."

Cf. Annexe X pour un exemple.

XI.2.2.3.5. Complément de l'acte de dialogue

Lorsque les arguments préférés sont instanciés, on complète éventuellement l'acte de dialogue avec des concepts auxiliaires. Cf. Annexe X pour un exemple.

XI.2.3. Génération

Pour la génération, nous avons choisi de suivre deux voies en parallèle. La première voie devait nous permettre de produire, semi-automatiquement, un générateur, réalisé en ARIANE-G5, à partir des fichiers de spécification de l'IF. Sachant que cette expérience, ambitieuse d'un point de vue méthodologique et technique, était tout de même risquée, nous avons choisi d'explorer une voie plus directe à base de phrases à trous.

XI.2.3.1. Générateur du premier démonstrateur : fondé sur les données

Le générateur finalement utilisé pour le premier démonstrateur utilise un ensemble de phrases à trous associées à un acte de dialogue.

Le module reçoit en entrée la liste des IF produites pour un tour de parole. Pour chacune de ces IF, on vérifie d'abord si l'acte de dialogue est ou non un acte de parole terminal. Si l'acte de dialogue est un acte de parole terminal, on utilise une fonction de génération construite pour cet acte de parole.

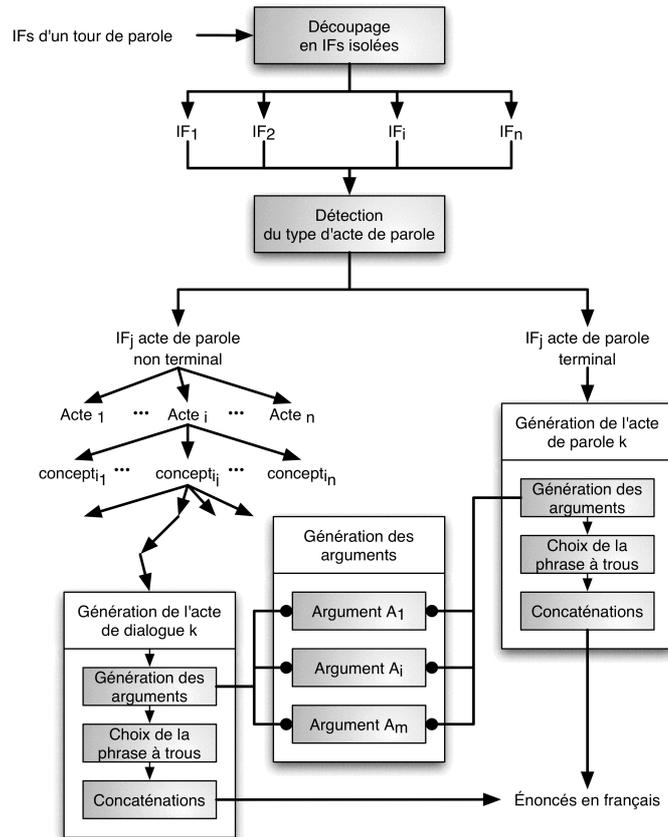


Figure XI-8 : Architecture du module de génération IF vers français du premier démonstrateur NESPOLE!

Si l'acte de parole n'est pas terminal, le système parcourt une forêt d'arbres construits comme suit : la racine des différents arbres est un acte de parole qui a pour fils les concepts qui peuvent le suivre ; à leur tour, les nœuds concepts ont pour fils les concepts qui peuvent les suivre. Lorsque l'on atteint une feuille (un acte de dialogue bien formé complet), on atteint aussi une fonction de génération d'un énoncé en français.

Les fonctions de génération associées aux actes de dialogue utilisent des fonctions de génération des arguments (une fonction par argument) et, en fonction des arguments instanciés, une phrase à trous que les valeurs d'argument viennent remplir est sélectionnée.

Par exemple, pour le tour de parole « je voudrais deux chambres simples à cavalese », l'analyseur du premier démonstrateur produit l'IF suivante :

```
c:give-information+disposition+room (disposition=(who=i, desire),
room-spec=(identifiability=no, quantity=2, single), location=(place-name=cavalese))
```

La fonction `DispositionRoom`, dans le contexte de l'acte de parole `give-information`, est utilisée par le générateur qui produit l'énoncé français suivant : « Je désire 2 chambres simples à Cavalese. » en utilisant la phrase à trous suivante :

```
if {$Disp!=0 && $RoomSpec!=0 && $AccomSpec==0 && $Duration==0 && $Price==0 && $Location!=0} {
  append the_result [services::IfDisposition2text $InWho $IDispo] " " [services::IfRoomSpec2Text $IRoomSpec]
  " " [services::IfLocation2Text $ILocation] " ."
```

Explication : on génère dans l'ordre

- le texte de l'argument `disposition` : `services::IfDisposition2text $InWho $IDispo]`
- un espace : " "
- le texte de l'argument `room-spec` : `[services::IfRoomSpec2Text $IRoomSpec]`

- un espace : " "
- le texte de l'argument location : [services::IfLocation2Text \$!Location]
- un point : "."

Sachant que **IfDisposition2text((who=i, desire))** produit « je désire » ; **IfRoomSpec2Text((identifiability=no, quantity=2, single))** produit « 2 chambres simples » et **IfLocation2Text((place-name=cavalese))** produit « à Cavalese ».

L'architecture générale du générateur du premier démonstrateur est donnée Figure XI-8.

XI.2.3.2. Critique de la solution adoptée

Cette méthode de génération, bien que simple à mettre en œuvre, a montré ses limites lors de la première évaluation (cf. section XI.3.2.3). En effet, pour pouvoir traiter toutes les IF acceptables selon la spécification, le nombre de feuilles de l'arbre de parcours des actes de paroles devrait être de l'ordre de 5.10^6 en couvrant des actes de dialogue pouvant contenir un acte de parole suivi de quatre concepts.

Pour développer ce module, nous avons mis en œuvre une approche fondée sur les données en utilisant des corpus d'IF pour des énoncés de client, puis des énoncés d'agent. Ainsi, nous avons pu réduire considérablement la profondeur (trois) et le nombre de feuilles de l'arbre.

- 50 actes de dialogue pour l'acte de parole give-information ;
- 35 actes de dialogue pour l'acte de parole request-information ;
- 1 acte de dialogue pour l'acte de parole request-suggestion ;
- 3 actes de dialogue pour des actions ;
- 1 acte de dialogue pour l'acte de parole offer.

Les phrases à trous doivent expliciter toutes les combinaisons possibles d'arguments, car l'absence d'un argument ne signifie pas qu'aucun texte ne doit être produit. Le nombre de combinaisons possibles, et donc le nombre de phrases à trous, est égal à 2^a où a est le nombre d'argument retenus. Pour la fonction `DispositionRoom`, nous avons retenu six arguments, soit 2^6 phrases à trous.

Nous avons donc abandonné cette piste, qui était destinée à parer rapidement la non disponibilité d'un générateur en ARIANE-G5.

XI.2.3.3. Générateur du second démonstrateur

Pour le second démonstrateur, nous avons repris l'idée d'implémenter le générateur en utilisant directement la spécification et sans nous préoccuper des données. Nous avons choisi cette fois de faire une génération concaténative. L'énoncé à produire n'est pas planifié statiquement au sein du module de génération. Il est construit dynamiquement lors d'un parcours de l'IF.

Le module reçoit en entrée la liste des IF produites pour un tour de parole. Pour chacune de ces IF, on vérifie d'abord si l'acte de dialogue contient ou non un acte de parole terminal. Si l'acte de dialogue est un acte de parole terminal, on utilise une fonction de génération construite pour cet acte de parole. Si l'acte de parole n'est pas un acte de parole terminal, la construction de la génération se déroule en quatre étapes. La Figure XI-9 illustre l'architecture de ce second module de génération.

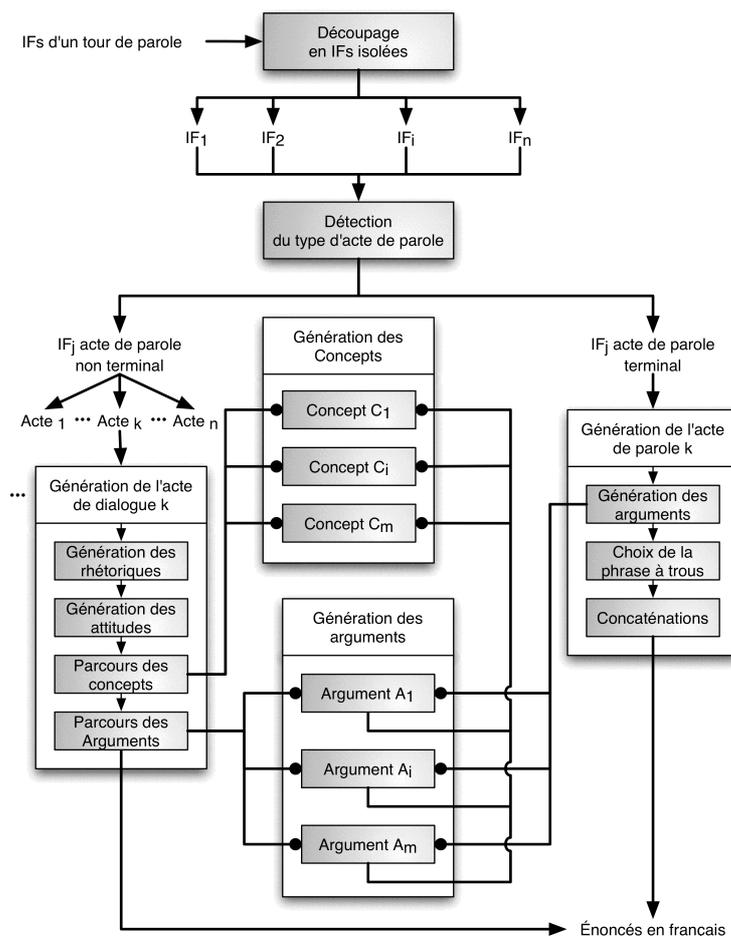


Figure XI-9 : Architecture du module de génération IF vers français du second démonstrateur NESPOLE!

XI.2.3.3.1. Génération d'un acte de parole terminal

Un acte de parole terminal est géré avec des phrases à trous car les arguments pertinents sont peu nombreux. Nous contrôlons aussi la présence d'un préfixe à l'acte de parole.

Par exemple, pour l'entrée « bonjour hervé blanchon à grenoble », l'analyseur produit les IF suivantes :

```
{c:greeting (greeting=good_day_opening)} {c:introduce-self (origin=name-grenoble, who=(given-name=name-herve, family-name=name-blanchon))}
```

greeting et introduce-self sont deux actes de parole terminaux. Cf. Annexe X Section X.2. pour le traitement de ces deux IF.

La génération en français des IF de ce tour de parole produit : « Bonjour ! Ici Hervé Blanchon depuis Grenoble. »

XI.2.3.3.2. Génération d'un acte de parole non terminal

Dans ce cas, la génération se déroule en plusieurs étapes :

1. Générer les arguments rhétoriques ;
2. Produire ou non un préfixe ou un suffixe de phrase en fonction de l'acte de parole ;
3. Générer les attitudes ;

4. Parcourir l'acte de parole en générant au fur et à mesure les arguments privilégiés des différents concepts rencontrés ;
5. Générer dans leur ordre d'apparition les arguments non consommés durant la phase précédente.

Pour l'énoncé suivant : « d accord et alors je voudrais une chambre du dix au quinze septembre à cavalese », l'analyseur produit les deux IF suivantes :

```
{c:acknowledge} {c:give-information+disposition+room(conjunction=discourse,
disposition=(desire, who=i), room-spec=(identifiability=no, bedroom), location=name-
cavalese, time=(start-time=(md=10), end-time=(md=15, month=9)))}
```

La trace de la génération pour les différentes étapes du processus donne :

1. et (\emptyset ; conjunction=discourse) +
2. \emptyset
3. je désirerais (+disposition ; disposition=(desire, who=i) +
4. une chambre simple (+room ; room-spec=(identifiability=no, bedroom)) +
5. à cavalese à partir du 10 jusqu'au 15 septembre. (\emptyset ; location=name-cavalese, time=(start-time=(md=10), end-time=(md=15, month=9))).

Pour produire finalement la phrase suivante : « Et je désirerais une chambre simple à Cavalese à partir du 10 jusqu'au 15 septembre. »

Pour l'IF :

```
c:request-verification-give-information+disposition+room(conjunction=discourse,
disposition=(desire, who=i), room-spec=(identifiability=no, bedroom), location=name-
cavalese, time=(start-time=(md=10), end-time=(md=15, month=9)))
```

L'étape 2 produit un **suffixe** correspondant à request-verification- : « Et vous désireriez une chambre à Cavalese à partir du 10 jusqu'au 15 septembre, **n'est-ce pas ?** ».

XI.2.4. Bilan sur l'IF

Pour ce qui est des résultats de l'évaluation des performances des modules, nous renvoyons le lecteur à la section suivante qui lui est consacrée. Nous nous concentrons ici sur l'IF.

L'expérience NESPOLE! nous a clairement montré que le changement d'échelle dans la couverture en conservant le même formalisme n'était pas une bonne idée. Pour prendre une image concrète : « nous avons réutilisé les matériaux et les outils que nous avons utilisés pour construire une barque (C-STAR II) afin de construire un paquebot transatlantique (NESPOLE!) ».

Il nous semble que pour augmenter la couverture, on ne peut pas se passer de représentations hiérarchiques pour représenter les phénomènes linguistiques et les ensembles de valeurs. On peut conserver les notions d'acte de parole et de concepts, mais il serait fort utile que ceux-ci dominent directement les arguments qui les concernent. Les arguments qu'on ne sait pas rattacher pourraient très bien être ramassés sous un nœud spécial de la structure. Il nous paraît aussi essentiel que la description ne soit pas surgénératrice comme c'est le cas actuellement.

Nous pensons aussi qu'il faut définitivement abandonner la construction d'un pivot à partir des données comme cela a été le cas tout au long du projet. Le dernier élément que nous évoquons ici est le manque de lisibilité et de compréhensibilité de la spécification. À notre avis, c'est l'un des éléments qui freinent la diffusion et l'utilisation de l'IF par des

groupes qui n'ont pas participé à son développement. Nous signalons ici que les évaluateurs du projet NESPOLE! avaient prédit un brillant avenir à l'IF !

XI.3. NESPOLE! : évaluation de la qualité des traductions

Dans le cadre du projet, une importance toute particulière était donnée à la validation des performances des démonstrateurs. Ces performances ont été mesurées en termes de qualité des traductions.

XI.3.1. Principes

La traduction bilingue est bien sûr l'intérêt premier d'un système de traduction que ce soit pour l'écrit ou l'oral. Nous avons montré que, dans le cadre de la traduction de l'écrit, la rétro-traduction permet à l'auteur de se faire une idée de ce que le système a traduit.

Pour la traduction de parole, la rétro-traduction (que nous appellerons ici **traduction monolingue**) revêt un intérêt très particulier car la mauvaise traduction d'un tour de parole peut être très préjudiciable au bon déroulement du dialogue. Nous avons évalué non seulement la **traduction bilingue**, mais aussi la traduction monolingue.

Afin de mesurer l'impact des erreurs de reconnaissance de la parole, les évaluations sont faites sur deux jeux de données, des transcriptions des tours de parole à traduire et les hypothèses de reconnaissance produites sur les mêmes tours de parole.

Enfin, il s'agit d'évaluation subjective.

XI.3.2. Premier démonstrateur en tourisme réduit¹⁰⁴

XI.3.2.1. Données et protocole

Quatre dialogues extraits de la base de données NESPOLE! [Burger, S., *et al.* 2001] ont été utilisés. Deux d'entre eux correspondent à une conversation entre un client qui veut organiser des vacances d'hiver dans le Val-di-Fiemme en Italie et un agent italien du bureau de tourisme APT. Les deux autres conversations concernent l'organisation de vacances d'été. Pour l'italien, il s'agit donc de tours de parole d'un agent. Pour les autres langues, ce sont des tours de clients.

Les signaux recueillis lors de la collecte de données servent d'entrée aux modules de reconnaissance vocale. Les transcriptions de ces signaux servent de référence pour la traduction, puisqu'elles simulent une reconnaissance sans erreur. Les tours de parole ont été segmentés manuellement en unités sémantiques de dialogue (SDU). Après avoir appliqué les modules de reconnaissance et/ou de traduction sur ces données, des juges humains évaluent la qualité de la traduction des SDU au sein d'un tour de parole en se référant aux SDU d'origine, comme cela est décrit dans la section suivante.

Pour la reconnaissance de la parole, nous avons utilisé la métrique standard en mots correctement reconnus (Word Accuracy Rate). Cependant, ce critère ne prend pas en compte le fait que certaines erreurs de reconnaissance peuvent avoir des conséquences plus ou moins importantes sur la qualité de la traduction produite par le système. Ainsi, nous

¹⁰⁴ Le texte de cette section est une adaptation en français de [Blanchon, H. 2004, Blanchon, H. and Besacier, L. 2004, Blanchon, H., *et al.* 2004a, Rossato, S., *et al.* 2002a, Rossato, S., *et al.* 2002b].

avons aussi vérifié si la sortie du module de reconnaissance peut être, ou non, considérée comme une paraphrase de la transcription manuelle du signal.

Pour chacun des jeux de test (vingt au total), nous avons utilisé trois juges ayant des compétences bilingues correspondant aux différents couples à évaluer. Les évaluations sont faites au niveau des SDU. Pour chaque SDU et sa version traduite, le juge doit évaluer la qualité de la traduction en donnant une note qui prend en compte deux types de jugements : (1) le sens du SDU original est-il préservé ? (2) la traduction est-elle fluide et grammaticale ? Si la réponse au premier jugement est négative, la note est BAD (sans tenir compte de la réponse à la seconde question). La seconde question permet de distinguer les notes PERFECT et OK.

Pour cette évaluation, les résultats ont été calculés en utilisant deux méthodes. Avec la première méthode (moyenne), on calcule le pourcentage des notes dans chacune des catégories pour chaque juge séparément, on calcule ensuite une simple moyenne arithmétique sur tous les juges.

Avec la seconde méthode, pour chaque SDU, le jugement finalement retenu est celui qui a obtenu le plus de suffrages de la part des juges (vote majoritaire). Si aucun vote majoritaire ne se dégage, la SDU est ignorée dans le comptage.

Avec les deux méthodes, les scores PERFECT et OK sont finalement additionnés pour obtenir une note d'acceptabilité (cf. Table XI-2).

XI.3.2.2. Résultats

RAP	WAR	71	62	64	77
<i>Hypos comme paraphrases</i>		65	66	68	70
Vote majoritaire		64	—	—	—
Traduction monolingue		<i>F-F c</i>	<i>E-E c</i>	<i>G-G c</i>	<i>I-I a</i>
<i>sur Refs/Hypos</i>		54/41	48/45	46/40	61/48
Vote majoritaire		51/38	—	—	60/44
Traduction bilingue		<i>F-I c</i>	<i>E-I c</i>	<i>G-I c</i>	
<i>sur Refs/Hypos</i>		44/34	55/43	32/27	
Vote majoritaire		38/31	—	—	
		<i>I-F a</i>	<i>I-E a</i>	<i>I-G a</i>	
<i>sur Refs/Hypos</i>		40/27	47/37	47/31	
Vote majoritaire		38/26	46/35	45/20	

Table XI-2 : Résultats de la première évaluation dans NESPOLE!

Le tableau précédent combine tous les résultats (en pourcentage) de traductions acceptables avec la méthode de la *moyenne*. Les résultats en vote majoritaire sont fournis quand ils ont été calculés.

XI.3.2.3. Commentaires

On peut constater que le vote majoritaire diminue systématiquement les scores de toutes les configurations. Lorsque c'est possible, nous commenterons uniquement les votes majoritaires qui ont été les seuls calculés pour la seconde évaluation.

Les résultats indiquent des pourcentages de traductions acceptables (sur les tours de clients et d'agents) dans une fourchette de 20% à 44% des SDU sur les hypothèses. Sur les références, les jugements sont, sans surprise, meilleurs dans une fourchette de 38 à 60 %. Pour le français, on atteint en monolingue sur les hypothèses un taux de 38% de traductions acceptables alors que l'on pouvait espérer 64%.

XI.3.2.3.1. Sur la traduction des tours de parole des clients

Pour la traduction bilingue depuis l'allemand, l'anglais et le français vers l'italien (sur des tours de clients seulement, et en moyenne), on observe très majoritairement une dégradation par rapport aux traductions monolingues des mêmes données. Cela signifie que l'on rencontre des problèmes d'accord sur le codage d'un contenu sémantique en IF (ce qui est équivalent à des problèmes d'accord sur le « sens » représenté par une IF donnée) et des problèmes de couverture.

Pour chaque site, on peut espérer que l'analyseur et le générateur soient en accord sur le sens des IF. On peut de même espérer aussi que le générateur soit conçu pour couvrir au mieux les IF qui sont produites par l'analyseur. Il semble que ces deux propriétés ne sont pas vérifiées par le générateur de l'italien.

XI.3.2.3.2. Sur les tours de parole de l'agent

Pour les tours de parole de l'agent (depuis l'italien), on observe que le système monolingue italien est meilleur que les autres systèmes monolingues.

En italien vers les autres langues, les trois générateurs reçoivent les mêmes IF en entrée. Leurs performances sont donc directement comparables. On peut remarquer que les systèmes ne se classent pas dans le même ordre en fonction du type de données (*Ref* ou *Hypo*). Nous n'avons pas eu le temps de dépouiller précisément les données pour expliquer ce phénomène. De même, les performances sont moins bonnes qu'en italien monolingue. Cela s'explique par les raisons évoquées ci-dessus pour rendre compte des différences de performance entre les systèmes monolingues et bilingues sur les tours de parole de clients.

XI.3.3. Second démonstrateur en tourisme étendu¹⁰⁵

XI.3.3.1. Données et protocole

Deux dialogues extraits de la seconde collecte NESPOLE! [Mana, N., *et al.* 2003] ont été utilisés. Ces dialogues couvrent des scénarios plus complexes non couverts par le premier démonstrateur, événements culturels, châteaux, lacs et forfaits.

Dans le principe, les différents types d'évaluation réalisés sont les mêmes que ceux de la première campagne soit : évaluation de la reconnaissance de la parole (WAR et hypothèse comme paraphrase), évaluation de la traduction monolingue sur quatre langues sur les données transcrites manuellement et sur les sorties du module de reconnaissance, évaluation de la traduction bilingue (6 couples) sur les données transcrites et les sorties de la reconnaissance (un exemple de fichier d'évaluation est donné en Annexe XI).

Afin de mesurer les progrès accomplis, nous avons utilisé les analyseurs et générateurs développés pour le premier démonstrateur sur toutes les configurations monolingues et bilingues sur les transcriptions uniquement (*Sur Refs (01)* dans la Table XI-8).

Nous avons changé notre procédure d'évaluation sur plusieurs points. Premièrement, nous avons abandonné l'échelle à trois valeurs utilisée lors de la première évaluation pour une nouvelle échelle à quatre valeurs fondée uniquement sur la préservation du sens et non sur la grammaticalité de la traduction. Deuxièmement, les juges choisis pour cette évaluation sont des élèves de dernière année d'école de traduction (DESS pour le français). Lors de la

¹⁰⁵ Cette section est une adaptation en français de [Blanchon, H. 2004, Blanchon, H. and Besacier, L. 2004, Blanchon, H., *et al.* 2004a].

première évaluation, les juges n'avaient pas de formation spécifique en traduction et les groupes étaient moins homogènes en terme de niveau en seconde langue.

Nous avons fait ces changements car le premier schéma de notation nous semblait poser plusieurs problèmes. Premièrement, bien que l'on ait demandé aux juges de faire deux évaluations (en sens et en fluidité), comme l'environnement d'évaluation leur présentait les trois options en même temps, il est possible que les juges aient été poussés à choisir la note médiane (Ok). Deuxièmement, les jugements de préservation du sens et de fluidité étaient faits en parallèle dans le même schéma d'évaluation. Troisièmement, la première question (le sens est-il préservé ?) était sévèrement interprétée : toute sorte de perte de sens rangeait la SDU dans la catégorie BAD.

Afin de résoudre ces problèmes, nous avons décidé d'utiliser une échelle de notes à quatre valeurs : VERY GOOD (toutes les informations sont présentes et faciles à comprendre), GOOD (toutes les informations importantes sont présentes), BAD (une ou plusieurs informations importantes ont été omises), VERY BAD (les informations importantes sont presque toutes absentes).

De plus, afin de rendre la tâche d'évaluation moins coûteuse, nous avons réduit l'évaluation à la préservation du sens. Nous n'avons pas demandé aux juges de juger la fluidité ou la grammaticalité. Finalement, pour pouvoir comparer les résultats de cette évaluation à la précédente, les notes 1 et 2 ont été additionnées comme acceptables. Les résultats complets de cette seconde évaluation sont donnés Table XI-8.

XI.3.3.2. Accord entre les juges et fidélité des mesures pour le français¹⁰⁶

Afin de vérifier la cohérence et la stabilité de notre protocole d'évaluation, il est important d'avoir une bonne mesure de l'accord des juges sur le jugement de la même sortie et de vérifier comment le jugement des juges est fidèle au cours du temps.

Les juges ont d'abord été entraînés sur le même jeu de données. Ils recevaient un ensemble d'instructions et un jeu d'essai. Ils devaient évaluer individuellement ce jeu d'essai et, à la fin, les notes étaient discutées en groupe afin de les faire converger vers les notes que nous attendions.

Dans le cadre de la tâche d'évaluation proprement dite, les juges disposaient tous de deux copies d'un même fichier de données témoins. Ils devaient évaluer ce fichier avant et après l'évaluation des fichiers de données à évaluer. Cela nous a permis de vérifier, premièrement, l'accord entre juges avant et après l'évaluation effective, et deuxièmement la fidélité du jugement de chacun au cours du temps.

Afin de mesurer l'accord des juges entre eux avant et après la tâche, nous avons fait trois catégories : (1) les juges ont le même jugement (unanimité), (2) les juges s'accordent sur deux jugements qui sont dans la même catégorie finale (vote majoritaire), (3) les juges ne s'accordent pas sur la même catégorie du vote finale (pas de vote majoritaire).

Nous avons obtenu les résultats suivants (en pourcentage) :

	(1) Unanimité	(2) Majorité	(3) Aucune majorité
Avant la tâche	71	28	1
Après la tâche	73	27	0

Table XI-3: Accord entre juges – It ou Fr vers Fr

¹⁰⁶ À notre connaissance, une telle étude n'a pas été faite pour les autres langues.

	(1) Unanimité	(2) Majorité	(3) Aucune majorité
Avant la tâche	88	15	0
Après la tâche	75	25	0

Table XI-4: Accord entre juges – Fr vers It

Du point de vue de la fidélité de chacun des juges dans le temps, nous avons fait trois catégories : (1) le juge fait le même jugement (fidélité), (2) le juge fait deux jugements différents qui restent dans la même catégorie (semi-fidélité), (3) le juge fait deux jugements différents qui ne sont pas dans la même catégorie (non fidélité).

Nous avons obtenus les résultats suivants (en pourcentage) :

	(1) Fidélité	(2) Semi-fidélité	(3) Non fidélité
Juge1	58	27	17
Juge2	83	13	6
Juge3	65	18	19

Table XI-5: Fidélité de chacun des juges au cours du temps – It ou Fr vers Fr

	(1) Fidélité	(2) Semi-fidélité	(3) Non fidélité
Juge4	83	13	6
Juge5	102	0	0
Juge6	58	27	17

Table XI-6: Fidélité de chacun des juges au cours du temps – Fr vers It

Nous avons noté qu'au cours du temps, les juges tendent à être plus sévères. Ainsi, pour les colonnes (2) et (3), il s'agit toujours de cas pour lesquels le jugement est dégradé (au sein de la classe acceptable dans colonne 2) ou, pire, passe de Acceptable à BAD ou à VERY BAD. Cela tendrait à prouver que les juges deviennent moins conciliants au fur et à mesure qu'ils progressent dans la tâche.

Afin de mesurer effectivement l'accord entre les juges, nous avons aussi calculé le coefficient Kappa pour les deux groupes de juges avant et après la tâche.

Groupe	Situation	Comparaison	Kappa
Juges Français It vers Fr et Fr vers Fr	Avant	J1 & J2	0.435
		J2 & J3	0.406
		J1 & J3	0.362
		Tous	0.394
	Après	J1 & J2	0.327
		J2 & J3	0.406
		J1 & J3	0.362
		Tous	0.368
Juges Italiens Fr vers It	Avant	J4 & J5	0.346
		J5 & J6	0.459
		J4 & J6	0.780
		Tous	0.503
	Après	J4 & J5	0.302
		J5 & J6	0.435
		J4 & J6	0.685
		Tous	0.463

Table XI-7 : Coefficients Kappa des juges pour l'évaluation du second démonstrateur NESPOLE!

Les coefficients Gamma sont tous très proches de 1. Cela implique que les juges sont cohérents sur la notion d'ordre sur les notes VERY GOOD, GOOD, BAD et VERY BAD. Dans le groupe italien, les juges J4 et J6 ont un excellent accord avant et après l'évaluation des fichiers de test ; J5 se distingue. Cependant, on peut noter que les scores Kappa diminuent

entre le début et la fin de l'expérience ; tous les évaluateurs ne sont pas victimes de la même usure. Pour le groupe français, la situation est moins tranchée ; le score Kappa diminue aussi pour le groupe, mais les accords entre juges sont fluctuants.

XI.3.3.3. Résultats

RAP	WAR	58	56	51	76
<i>Hypos comme paraphrase</i>		60	67	62	76
Traduction monolingue		<i>F-F c</i>	<i>E-E c</i>	<i>G-G c</i>	<i>I-I a</i>
<i>sur Refs (01)</i>		69	68	45	36
Refs/Hypos (02)		77/58	68/50	61/51	51/42
Traduction bilingue		<i>F-I c</i>	<i>E-I c</i>	<i>G-I c</i>	
<i>sur Refs (01)</i>		72	64	44	
Refs/Hypos (02)		77/58	70/50	x/x ¹⁰⁷	
		<i>I-F a</i>	<i>I-E a</i>	<i>I-G a</i>	
<i>on Refs (01)</i>		19	33	38	
Refs/Hypos (02)		37/33	33/30	45/38	

Table XI-8 : Résultats de la seconde évaluation du projet NESPOLE!

Le tableau précédent donne (en pourcentage) les résultats pour les traductions acceptables en utilisant un vote majoritaire.

XI.3.3.4. Commentaires

Les résultats indiquent des pourcentages de traductions acceptables (sur les tours de clients et d'agents) dans une fourchette de 42 à 48% des SDU sur les hypothèses. Sur les références, les jugements sont, sans surprise, meilleurs dans une fourchette de 51 à 77 %. Pour le français, on atteint en monolingue sur les hypothèses un taux de 58% de traductions acceptables alors que l'on pouvait espérer 60%.

XI.3.3.4.1. Sur la traduction des tours de parole des clients

Pour la traduction bilingue depuis l'allemand, l'anglais et le français vers l'italien (sur des tours de clients seulement), on n'observe pas de dégradation par rapport aux traductions monolingues des mêmes données. Cela signifie que le générateur italien sait générer avec les mêmes performances les IF produites par les analyseurs de l'allemand, de l'anglais et du français que les générateurs respectifs de ces différentes langues sources. Cela montre qu'il n'y pas de problème de codage d'un énoncé en langue naturelle en langue IF entre les sites. En particulier, pour le français, la dégradation entre la traduction monolingue et la traduction vers l'allemand était assez importante avec le premier démonstrateur.

XI.3.3.4.2. Sur les tours de parole de l'agent

Pour les tours de parole de l'agent (depuis l'italien), on observe une baisse sensible des performances entre le système monolingue italien et les systèmes bilingues vers l'allemand, l'anglais et le français. Dans ce cas, les trois générateurs ont en entrée les mêmes IF. Leurs performances sont donc directement comparables. On peut donc dire que le générateur allemand a les meilleures performances devant le générateur français et le générateur

¹⁰⁷ Les chiffres pour le second démonstrateur allemand-italien ne sont pas disponibles à cause d'un comportement inconsistant des juges à prendre en compte les pauses pleines. Les chiffres initialement fournis étaient 40/52 (les performances du système sur les hypothèses sont meilleures que sur les références !).

anglais. Pour le générateur du français, on peut dire qu'il s'agit d'un problème de couverture et non d'un problème d'accord sur le codage en IF.

Vers l'italien, les résultats depuis le français sont cette fois meilleurs que depuis l'anglais. Enfin, depuis l'italien, le générateur français affiche maintenant des performances comparables aux autres générateurs. Lors de la campagne d'évaluation de masse de 2000, pour un WAR inférieur à 75% les résultats montrent un taux de traductions acceptables de 75% pour le couple allemand-anglais et de 66% pour le couple anglais-allemand. Les niveaux de performance du second démonstrateur NESPOLE! sont sensiblement les mêmes.

XI.3.4. Mesure des progrès accomplis

Il est intéressant de noter que les traductions depuis l'italien, avec un taux de SDU acceptables comme paraphrases de l'entrée de 76% — bien supérieur aux autres langues —, sont significativement moins bonnes que les traductions vers l'italien. Cela est principalement dû aux caractéristiques des données d'évaluation. Les tours de parole des agents (traduits depuis l'italien) sont plus longs et complexes, et dans certains cas hors du domaine. Les tours de parole des clients sont en moyenne plus courts, moins complexes et dans le domaine.

Afin de quantifier ces différences et d'estimer leurs effets sur les résultats, nous avons demandé aux développeurs des systèmes de classer manuellement les SDU des données d'évaluation en trois catégories : (1) dans le domaine de couverture du premier démonstrateur, (2) dans le domaine de couverture du second démonstrateur, et (3) hors du domaine des démonstrateurs. Nous avons alors calculé les performances des systèmes sur ces trois groupes de SDU séparément.

Pour l'allemand, l'anglais et le français (tours de parole de clients), nous avons découvert qu'un nombre très petit de SDU appartenait aux groupes 2 et 3 (moins de cinq pour les trois langues). Ainsi les données sont presque toutes du domaine du premier démonstrateur.

Pour l'italien (tours de parole d'agents), par contraste, 13% des SDU sont classées dans le groupe 2 et 25% dans le groupe 3 (hors du domaine).

XI.3.4.1.1. Sur les tours de parole de clients

Dans leur version monolingue, les démonstrateurs de l'allemand, de l'anglais et du français ont donc à traiter des SDU du domaine du premier démonstrateur. Pour les systèmes français et allemand, on observe des améliorations de performance de 8 et 16 points respectivement entre le premier démonstrateur et le second. Cela signifie que la couverture du domaine des premiers démonstrateurs a été accrue.

Dans le cas des systèmes bilingues anglais-italien et français-italien, on observe aussi des augmentations de 5 et 6 points respectivement. Cela confirme que la couverture du domaine des premiers démonstrateurs a été améliorée.

XI.3.4.1.2. Sur les tours de parole de l'agent

La différence de performance sur les trois groupes de SDU est aussi très révélatrice. Pour le système monolingue italien, sur le premier groupe de SDU, nous observons une amélioration de 57% de traductions acceptables pour le premier démonstrateur à 63% pour le second démonstrateur sur les références. Cela indique une augmentation de couverture du second démonstrateur sur le domaine du premier démonstrateur.

Sur le second groupe de données, la différence est bien plus prononcée. Le premier démonstrateur n'atteint que 14% de traductions acceptables, alors que le second démonstrateur atteint un score de 39%. Bien sûr, le premier démonstrateur n'était pas conçu pour couvrir ce type de données, ce n'est donc pas surprenant.

Alors que le second démonstrateur a de bien meilleures performances sur ces données, ces performances n'atteignent pas le niveau atteint avec le premier démonstrateur. En italien-italien, on avait un taux de 60% de traductions acceptables sur les références avec le premier démonstrateur.

Le système italien n'a aucune couverture des SDU hors du domaine. Lorsqu'on exclut ces données, les performances du premier démonstrateur italien-italien sur les références a un taux de 49% de traductions acceptables alors que le second démonstrateur a un taux de traductions acceptables de 59% qui se rapproche alors des performances des seconds démonstrateurs monolingues allemand, anglais et français. De leur côté, les performances des seconds démonstrateurs depuis l'italien vers l'allemand, l'anglais et le français voient aussi leur performances augmenter pour passer respectivement à 45%, 41% et 53%.

Si l'on compare aussi les taux d'acceptabilité sur les références italiennes des premiers et seconds démonstrateurs, on se rend compte que, pour le français, le générateur progresse de 18% et le générateur allemand de 7%, le générateur anglais restant stable. Cela confirme bien l'inadaptation du premier générateur du français à traiter les IF produites par les Italiens et les progrès que notre nouvelle approche ont permis d'accomplir.

Chapitre XII

Contribution à C-STAR III et réflexions sur le domaine

Le consortium C-STAR III est né en septembre 2000. Les objectifs sont de construire un corpus multilingue commun sur le domaine du tourisme et d'organiser des campagnes d'évaluation compétitive sur celui-ci.

XII.1. Évaluation dans C-STAR III

XII.1.1. Première évaluation interne¹⁰⁸

En 2003, les partenaires ont conduit une expérience pilote en évaluation de systèmes de traduction de dialogue dans deux conditions sur la base du corpus BTEC [Kikiu, G., *et al.* 2004, Takezawa, T., *et al.* 2002]. Pour la première condition, nous avons choisi quatre langues sources (le coréen, le chinois, l'italien et le japonais) et une langue cible (l'anglais). Cinq systèmes ont participé à cette condition. Pour la seconde condition, le chinois était la langue source obligatoire et l'anglais la langue cible.

XII.1.1.1. Données et protocole

Les données de développement et de test des différents systèmes étaient constituées du corpus BTEC et de toute autre ressource monolingue. 500 phrases anglaises traduites dans toutes les langues sources (chinois, anglais et japonais) ont été utilisées comme données de test.

Pour la première condition, nous avons procédé, d'une part, à une évaluation subjective des résultats et d'autre part, à une évaluation objective en utilisant BLEU et NIST. L'évaluation subjective a été conduite selon les recommandations du NIST. Les scores BLEU et NIST ont été calculés sur les sorties brutes des systèmes.

Pour la seconde condition, l'évaluation a été réalisée avec BLEU et NIST uniquement.

XII.1.1.2. Résultats

La Table XII-1 donne les résultats obtenus par les différents systèmes dans la première condition.

¹⁰⁸ Ces résultats n'ont jamais été publiés officiellement, ils sont reproduits ici avec l'aimable autorisation des membres du consortium C-STAR III.

	Adéquation [0..5]	Fluidité [0..5]	BLEU [0..1]	NIST [0..∞[
S1 : I-E	4,00 (1)	3,76 (2)	0,6620 (1)	10,5706 (1)
S2 : J-E	3,92 (2)	4,03 (1)	0,5820 (2)	6,5565 (2)
S3 : K-E	3,01 (4)	2,81 (4)	0,3153 (4)	5,8889 (3)
S4 : C-E	2,59 (5)	2,30 (5)	0,2733 (5)	5,6830 (4)
S5 : C-E	3,21 (3)	3,74 (3)	0,5542 (3)	3,4013 (5)

Table XII-1 : Résultats de l'évaluation pilote interne du consortium C-STAR III : première condition

La Table XII-2 donne les résultats obtenus par les différents systèmes dans la seconde condition.

	BLEU [0..1]	NIST [0..∞[
S5	0.5542 (1)	3.4013 (3)
S6	0,3884 (2)	8.1383 (1)
S4	0.2733 (3)	5.6830 (2)

Table XII-2 : Résultats de l'évaluation pilote interne du consortium C-STAR III : seconde condition

Les mêmes données ont été soumises à des systèmes disponibles librement sur le Web. Pour le couple chinois-anglais, Systran est classé dernier pour BLEU et avant-dernier pour NIST. Pour les couples coréen-anglais et japonais-anglais, Systran est vraiment très loin pour BLEU. Enfin, pour le couple italien-anglais, les différents systèmes disponibles sur le Web (cinq au total) sont aussi bien moins bons que le système S1.

XII.1.1.3. Commentaires

On observe une incohérence entre les scores BLEU et NIST pour le système 5. Les sorties de ce système sont significativement plus courtes que les références, ce qui a une forte influence dans le calcul du score. À ce détail près, les classements des différents systèmes pour les deux mesures sont assez homogènes.

Nous avons aussi observé des différences importantes dans la forme des sorties des différents systèmes. Ces différences concernent : l'usage de majuscules ("Tokyo", ou "tokyo"), le rendu des numéraux (en lettres, ou chiffres séparés ou non), les abréviations ("OK" ou "okay"), les mots composés ("duty-free" ou "duty free"), l'utilisation de ponctuation ("juice, please" vs. "juice please"), les séparateurs de phrases, etc.

Sur les sorties obtenues, ATR a montré que cela peut induire des différences de $\pm 0,15$ points avec BLEU et $\pm 1,8$ points avec NIST. Ces différences ne sont pas négligeables. Pour les expériences futures, nous avons décidé qu'il faudrait normaliser les sorties des différents systèmes comme cela est déjà le cas en évaluation des systèmes de reconnaissance de la parole.

Dans tous les cas, les systèmes de traductions disponibles sur le Web ont des performances bien inférieures aux performances des systèmes développés par les partenaires. La comparaison est-elle juste ? Nous y revenons dans la section suivante.

XII.1.2. Contribution à la première campagne d'évaluation compétitive ouverte (IWSLT 2004)¹⁰⁹

En 2004, les membres du consortium C-STAR III ont décidé de proposer une campagne d'évaluation compétitive sur la base du corpus BTEC [Akiba, Y., *et al.* 2004]. Les couples de

¹⁰⁹ Cette section est une adaptation en français de [Blanchon, H., *et al.* 2004b]

langues choisis étaient japonais-anglais et chinois-anglais. Pour chacune de ces paires de langues, nous avons diffusé 20 000 paires de tours de parole du corpus BTEC pour que les participants puissent préparer leurs systèmes. Pour le couple japonais-anglais, deux conditions étaient proposées : données limitées (les données fournies uniquement), et données illimitées. Pour le couple chinois-anglais, trois conditions étaient proposées : données limitées, données additionnelles (ressources du LDC¹¹⁰ pour ce couple), et données illimitées.

Pour l'évaluation proprement dite, deux ensembles disjoints de phrases chinoises et japonaises ont été proposés aux compétiteurs. L'évaluation subjective respectait le protocole du NIST en adéquation et fidélité. Pour l'évaluation objective, nous avons choisi les mesures BLEU, NIST, GMT, mWER et mPER, pour lesquelles seize paraphrases en anglais de chacune des phrases chinoises et japonaises étaient disponibles.

En tant que partenaire du consortium non impliqué dans le développement d'un système de traduction japonais-anglais ou japonais-chinois, nous avons pensé qu'il serait intéressant de participer à cette évaluation afin de permettre une évaluation plus juste des systèmes commerciaux.

Nous avons choisi Systran car :

- les deux paires de langues sont disponibles,
- ces deux paires ont récemment été améliorées, bien que les efforts de Systran aient plus porté sur les couples anglais-chinois et anglais-japonais,
- Systran a mis à notre disposition la dernière version (v5) disponible dans sa version Premium Professional (Interface Windows, paramètres d'exécution modifiables, et possibilité d'utiliser des dictionnaires personnels),
- ces systèmes peuvent être considérés comme un bon compromis comparé à d'autres systèmes commerciaux pour ces couples de langues ; certains sont bien moins bons et certains sont bien meilleurs tels que ATLAS-II pour le couple anglais-japonais et ALT/JE pour le couple japonais-anglais.

XII.1.2.1. Systèmes utilisés et protocole

XII.1.2.1.1. Paire chinois-anglais

Nous avons soumis trois jeux de résultats pour la paire chinois-anglais avec les versions suivantes :

C_1	SYSTRAN web 5.0
C_2	SYSTRAN premium 5.0 avec ses propres dictionnaires
C_3	SYSTRAN premium 5.0 avec ses propres dictionnaires et un dictionnaire utilisateur

Le dictionnaire utilisateur de la version **C_3** contient 173 mots qui étaient inconnus pour la version **C_2** dans les 20 000 tours de parole chinois.

XII.1.2.1.2. Paire japonais-anglais

Nous avons soumis trois jeux de résultats pour la paire japonais-anglais avec les versions suivantes :

¹¹⁰ Linguistic Data Consortium : <http://ldc.upenn.edu/>

J_1	SYSTRAN web 5.0
J_2	SYSTRAN premium 5.0 avec ses propres dictionnaires
J_3	SYSTRAN premium 5.0 avec ses propres dictionnaires et un dictionnaire utilisateur

Le dictionnaire utilisateur de la version **J_3** contient 304 mots qui étaient inconnus pour la version **J_2** dans les 20 000 tours de parole japonais.

Le dernier jeu de résultats (**J_3**) a été révisé manuellement afin de produire des traductions acceptables en minimisant le nombre de changements. Sur les cinq cents traductions candidates, cinquante n'ont pas été modifiées, soit 10%.

J_4|J_3 révisé manuellement

XII.1.2.1.3. Normalisation des traductions produites

Les résultats de traduction devaient être normalisés comme suit, afin d'imiter le résultat d'un module de reconnaissance automatique de la parole :

- pas de lettre capitale,
- pas de marque de ponctuation,
- pas de mot composé (les « - » devaient être remplacés par un espace),
- chiffres et nombres transcrits en lettres.

XII.1.2.2. Résultats

XII.1.2.2.1. Résultats de l'évaluation objective pour la paire chinois-anglais

	BLEU	GMT	NIST	PER	WER
C_3	0.1620 1	0.5845 1	6.0061 1	0.5429 2	0.6581 2
C_1	0.1600 3	0.5802 3	5.9143 3	0.5423 1	0.6474 1
C_2	0.1620 1	0.5841 2	6.0039 2	0.5429 2	0.6581 2

Table XII-3 : Résultats de l'évaluation objective des systèmes Systran pour la paire chinois-anglais

XII.1.2.2.2. Résultats de l'évaluation objective pour la paire japonais-anglais

	BLEU	GMT	NIST	PER	WER
J_3	0.1320 1	0.5687 1	5.6476 1	0.5978 1	0.7304 1
J_2	0.1311 2	0.5672 2	5.6096 2	0.6012 2	0.7349 2
J_1	0.0810 3	0.5116 3	4.1935 3	0.7179 3	0.8726 3

Table XII-4 : Résultats de l'évaluation objective des systèmes Systran pour la paire japonais-anglais

XII.1.2.2.3. Résultats de l'évaluation objective pour le jeu J_4

Nous espérons bien sûr des résultats meilleurs pour les traductions Systran révisés. Notre intuition a été confirmée.

	BLEU	GMT	NIST	PER	WER
J_4	0.4691	0.7777	9.9189	0.3236	0.3711

Table XII-5 : Résultats de l'évaluation objective du système J_3 Systran japonais-anglais révisé

XII.1.2.2.4. Résultats des évaluations subjectives et objectives compétitives pour la paire chinois-anglais

Neuf systèmes ont participé à la condition chinois-anglais illimitée. Systran **C_3** se classe huitième.

Évaluation subjective

	Fluidité	Adéquation
CE_8	3.7760 1	3.6620 1
CE_3	3.0360 4	2.9960 6
CE_7	2.9340 6	3.2540 3
CE_5	3.7760 1	3.5260 2
CE_9	3.4000 3	2.8000 8
CE_2	2.6480 8	3.1880 4
CE_6	2.9540 5	2.7840 9
C_3	2.5700 9	2.9600 7
CE_4	2.7180 7	3.0820 5

Table XII-6: Résultats de l'évaluation subjective des systèmes japonais-anglais illimités

Évaluation objective

	BLEU	GMT	NIST	PER	WER
CE_8	0.5249 1	0.7482 1	9.5603 1	0.3198 1	0.3795 1
CE_3	0.3505 3	0.6849 2	7.3691 3	0.4428 4	0.5255 3
CE_7	0.2753 5	0.6669 4	7.5002 2	0.4276 3	0.5313 4
CE_5	0.4409 2	0.6720 3	7.2413 4	0.3930 2	0.4570 2
CE_9	0.3113 4	0.5639 8	5.9217 7	0.5310 7	0.5788 6
CE_2	0.2438 6	0.6119 5	6.1354 5	0.4872 5	0.5941 7
CE_6	0.2430 7	0.6023 6	5.4250 8	0.4998 6	0.5735 5
C_3	0.1620 8	0.5845 7	6.0061 6	0.5429 8	0.6582 8
CE_4	0.0798 9	0.3862 9	3.6443 9	0.7650 9	0.8466 9

Table XII-7: Résultats de l'évaluation objective des systèmes chinois-anglais illimités

XII.1.2.2.5. Résultats des évaluations subjectives et objectives compétitives pour la paire japonais-anglais

Quatre systèmes ont participé à la condition japonais-anglais illimitée. Systran **J_3** se classe quatrième.

Évaluation subjective

	Fluidité	Adéquation
JE_1	4.3080 1	4.2080 1
JE_3	4.0360 2	4.0660 2
JE_4	3.6500 3	3.3160 3
J_3	2.4720 4	2.6020 4

Table XII-8: Résultats de l'évaluation subjective des systèmes japonais-anglais illimités

Évaluation objective

	BLEU	GMT	NIST	PER	WER
JE_1	0.6306 1	0.7967 2	10.7201 2	0.2333 1	0.2631 1
JE_3	0.6190 2	0.8243 1	11.2541 1	0.2492 2	0.3056 2
JE_4	0.3970 3	0.6722 3	7.8893 3	0.4202 3	0.4857 3
J_3	0.1320 4	0.5687 4	5.6476 4	0.5978 4	0.7304 4

Table XII-9: Résultats de l'évaluation objective des systèmes japonais-anglais illimités

	BLEU	GMT	NIST	PER	WER
JE_1	0.6306 1	0.7967 2	10.7201 2	0.2333 1	0.2631 1
JE_3	0.6190 2	0.8243 1	11.2541 1	0.2492 2	0.3056 2
J_4	0.4691 3	0.7777 3	9.9189 3	0.3236 3	0.3711 3
JE_4	0.3970 4	0.6722 4	7.8893 4	0.4202 4	0.4857 4
J_3	0.1320 5	0.5687 5	5.6476 5	0.5978 5	0.7304 5

Table XII-10: Résultats de l'évaluation objective des systèmes japonais-anglais illimités avec J_4

Le jeu de traductions **J_4** se classe troisième. Cette position peut sembler basse pour des traductions humaines !

XII.1.2.3. Discussion

XII.1.2.3.1. Résultats pour la paire chinois-anglais

La version Web de la paire chinois-anglais de Systran (**C_1**) a de meilleures performances que la version Premium v5 (**C_2**) (Table XII-3). Cela peut sembler surprenant ! Il faut savoir que la version Premium de cette paire de langue a été figée en mai-juin, alors que la version Web a été mise à jour plus tard et continue d'évoluer. La meilleure version de Systran se classe huitième sur neuf pour cette paire (Table XII-6 and Table XII-7).

XII.1.2.3.2. Résultats pour la paire japonais-anglais

Les systèmes **J_3**, **J_2**, et **J_1** sont classés dans un ordre en accord avec notre intuition que plus un système est finement réglé pour une tâche donnée, meilleurs sont les résultats (Table XII-4). Pour cette paire de langues, le meilleur système Systran se classe dernier (Table XII-8 and Table XII-9).

Pourquoi les traductions révisées (**J-4**, Table XII-10) se classent seulement en troisième position et non en première ? La réponse est extrêmement simple ! Il faut se rappeler que les méthodes d'évaluation objective ne sont pas des **mesures de qualité** en tant que telle, mais des **mesures de similarité**. Ce rappel fait, on peut seulement conclure que les traductions révisées sont moins similaires aux traductions de références que ne le sont les traductions produites par les systèmes **JE_1** et **JE_3**.

XII.1.2.3.3. Accord inter-juges pour les évaluations subjectives

Tous les coefficients Gamma (Table XII-11, troisième colonne) sont supérieurs à 0,6, ce qui indique que les juges sont d'accord sur l'ordre relatif des notes.

Avant de calculer les coefficients Kappa, nous avons vérifié l'hypothèse des effectifs marginaux équilibrés pour chaque paire de langue. Pour chacune des deux évaluations en fluidité, il apparaît que l'un des juges utilise la note 2 dans 46% des cas, alors que les deux autres juges n'ont pas ce comportement. Cela implique que pour ces deux évaluations, il n'est pas possible d'interpréter la valeur du coefficient Kappa global. Pour les deux évaluations en adéquation, nous n'avons pas observé de différence significative pour les effectifs marginaux.

			Gamma	Kappa
chinois-anglais	fluidité	J1 & J2	0.803	0.384
		J2 & J3	0.703	0.197
		J1 & J3	0.712	0.317
		tous		-----
	adéquation	J1 & J2	0.720	0.318
		J2 & J3	0.686	0.305
		J1 & J3	0.656	0.308
		tous		0.309
japonais-anglais	fluidité	J4 & J5	0.745	0.345
		J5 & J6	0.647	0.203
		J4 & J6	0.645	0.272
		tous		-----
	adéquation	J4 & J5	0.724	0.320
		J5 & J6	0.747	0.340
		J4 & J6	0.695	0.298
		tous		0.318

Table XII-11 : Coefficients Gamma et Kappa pour l'accord inter-juges de l'évaluation subjective IWSLT 2002 pour Systran

Les coefficients Kappa deux à deux sont compris entre 0,19 et 0,38. L'accord entre les juges est donc modéré. Pour l'adéquation, les valeurs du coefficient Kappa de chacun des deux groupes d'évaluateurs sont : 0,309 pour la paire chinois-anglais, et 0,318 pour la paire japonais-anglais. L'accord entre les trois juges de chaque groupe est donc modéré.

C'est, à notre connaissance, la première fois que de tels chiffres sont publiés.

XII.1.2.3.4. Analyse des résultats produits par Systran pour la paire japonais-anglais¹¹¹

Tous les tours de parole proposés dans le jeu de test peuvent être considérés comme des transcriptions nettoyées (sans hésitation, ni autocorrection) de tours de parole réels de dialogues dans le domaine du tourisme. Le niveau de langue est plutôt poli.

Quand le tour de parole est un euphémisme, la particule **が** est toujours traduite par "but".

Certains tours de parole sont incompréhensibles sans contexte (**切りますよ。→ "it cuts" ?**).

Lorsque le sujet de la première personne est omis en japonais, il est toujours traduit par "it" (**ここで降ります。→ "It gets off here."**¹¹²).

Le jeu de test contient beaucoup de formes interrogatives. Dans les traductions, le pronom ou l'adverbe interrogatif est toujours placé en fin de traduction. L'ordre standard des mots en anglais n'est donc pas respecté (**オペラ座はどこですか。→ "Is the opera house where?"**¹¹³).

Aucune des expressions orales de la vie quotidienne ne sont dans les dictionnaires Systran (**どういたしまして。→ "How doing."**¹¹⁴, **もしもし。→ "It does."**¹¹⁵, **さようなら。→ "Way if."**¹¹⁶).

¹¹¹ Des exemples de traductions produites par Systran sont donnés en Annexe XII.

¹¹² « Je descends ici. »

¹¹³ « Où est l'opéra ? »

¹¹⁴ « Pas de quoi ! » en réponse à un « Merci ! »

¹¹⁵ « Allo ! » au téléphone.

Les requêtes et les invitations ne sont pas toujours bien traduites (注文したいのです。→ “It is to like to order.”¹¹⁷ 一緒に行きましょう。→ “It will go together.”¹¹⁸).

La politesse en japonais, lorsqu'elle est lexicalisée, est correctement traduite (そのまま切らずにお待ち下さい。→ “Without cutting that way, please wait.”).

Quand la valence du verbe pour deux expressions en japonais et en anglais est différente, la traduction est presque toujours mauvaise (寒気がする。→ “Chill does.”¹¹⁹).

Enfin, l'aspect des prédicats japonais n'est pas rendu correctement en anglais (航空券を家に忘れてしまいました。→ “The air ticket was forgotten in the house.”¹²⁰).

XII.1.3. Conclusion

L'ajout des mots inconnus de Systran, ils étaient en fait peu nombreux, dans des dictionnaires utilisateur, n'est pas suffisant pour que les systèmes puissent atteindre un niveau de performance comparable aux systèmes entraînés spécifiquement pour le corpus BTEC.

Nous avons cependant pu montrer que des traductions humaines « parfaites » obtiennent des scores qui ne sont pas excellents. Cela nous a permis de rappeler à la communauté que les méthodes d'évaluation objectives n'évaluent pas la **qualité** des traductions¹²¹ soumises à évaluation, elles évaluent la **ressemblance** des traductions candidates avec les traductions de référence. Lorsque la ressemblance est parfaite, alors la traduction candidate est parfaite (sous réserve que la traduction de référence le soit !). Lorsque la ressemblance n'est pas parfaite, il nous semble que l'on ne peut rien dire du tout car on ne voit pas les traductions produites (cf. infra).

Les promoteurs des techniques d'évaluation objectives disent que les méthodes objectives leur permettent de suivre l'évolution des performances de leurs systèmes. On peut leur objecter que cette évolution est purement quantitative et non qualitative. En effet, avec les approches statistiques, on peut dire que le système doit globalement converger asymptotiquement vers les traductions de références. Par contre, on ne peut rien dire de l'amélioration « locale » de la qualité des traductions pour le traitement de certains phénomènes linguistiques, si cette notion fait encore sens avec les méthodes statistiques !

XII.2. Problèmes¹²²

XII.2.1. Mesure de la qualité linguistique

Le premier problème est que les chiffres produits par les techniques d'évaluation objective ne sont pas directement liés à la qualité de la traduction. Comme nous l'avons

¹¹⁶ « Au revoir. »

¹¹⁷ « J'aimerais bien commander. »

¹¹⁸ « Allons-y ensemble ! »

¹¹⁹ « J'ai froid. »

¹²⁰ « J'ai oublié mon billet d'avion à la maison. »

¹²¹ Ceci est vraiment un problème lorsque l'on communique sur les résultats avec d'autres communautés, telle celle de la recherche d'informations, qui ont des mesures précises de la qualité d'un système. Bien sûr, en recherche d'information, la nature de la tâche n'est pas la même qu'en TA et il n'y a pas plusieurs réponses possibles à une question pour un jeu de test donné. En TA, « **LA** » **bonne traduction** n'existe pas, on peut seulement dire si une traduction est bonne ou pas.

¹²² Cette section est une version adaptée en français et augmentée de [Blanchon, H., *et al.* 2004a].

montré, beaucoup de travaux ont été conduits en vue de corrélérer des résultats d'évaluations objectives avec des résultats d'évaluations subjectives, mais les résultats peuvent être inconsistants. BLEU est supposé bien corrélé avec des évaluations humaines de qualité de traduction, NIST a été proposé pour donner des résultats meilleurs pour des raisons théoriques, mais BLEU et NIST donnent des résultats contradictoires. Ainsi, si la corrélation avec des jugements humains est une mesure de qualité de ces techniques, alors NIST ne peut être meilleur que BLEU... or la corrélation est trop faible pour vouloir dire quelque chose.

[Culy, C. and Riehemann, S. Z. 2003] ont développé le même type d'arguments en rappelant aussi que les techniques d'évaluation fondées sur les n-grammes proposées jusqu'à présent ne sont pas, strictement, des mesures de qualité de traductions. Elles sont plutôt des mesures de similarité de documents. Leur utilité pour la mesure de la qualité de traductions naît de la supposition qu'une bonne traduction d'un texte est similaire à d'autres bonnes traductions du même document. [Culy, C. and Riehemann, S. Z. 2003] a montré que cette supposition ne tient pas toujours, et qu'elle conduit à des problèmes lorsque l'on utilise ces techniques pour des besoins d'évaluation.

Un autre problème est que, comme rapporté lors du colloque ACL-03, ces techniques donnent souvent de très mauvais scores à des traductions humaines de haute qualité. Ainsi une expérience a été conduite au cours de laquelle chacune parmi quinze paraphrases (humaines) a été évaluée en utilisant les quatorze autres comme références. Les scores n'étaient pas aussi mauvais que ceux de traductions automatiques, mais assez mauvais, bien qu'elles fussent des traductions parfaites. Le problème ici est que les références sont toujours assez clairsemées dans l'espace de toutes les traductions parfaites ou très bonnes.

Une solution, pour résoudre ces problèmes, consisterait à demander à des humains post-éditeurs de réviser, au moindre effort, les traductions produites par le système conduisant à l'une des traductions correctes équivalentes la plus proche. Ainsi, le travail à fournir donnerait une meilleure appréciation de la qualité de la traduction produite par le système en mesurant, en quelque sorte, l'effort humain minimal à consentir pour obtenir une traduction de qualité à partir du résultat produit automatiquement. Avec cette approche, des traductions humaines de qualité ne nécessiteraient aucun travail de post-édition et seraient donc jugées parfaites.

Bien que le processus d'évaluation puisse être automatisé, l'inconvénient de cette solution est qu'elle requière plus de travail humain pour produire non seulement les références, mais pour post-éditer les traductions produites par le système. Mais..., on peut justement fusionner les deux tâches en produisant des références cinq fois moins cher en partant de résultat de TA [Boitet, C. 2004].

Une réponse fréquente à nos objections est que ces techniques sont dans tous les cas utiles pour évaluer les progrès accomplis durant le processus de développement d'un système. Nous voudrions contester cette idée sur la base que le progrès est mesuré de manière circulaire. Ce qui est mesuré, c'est uniquement le progrès selon les techniques d'évaluation utilisées, qui ne mesurent pas la qualité linguistique. D'autre part, mesurer la qualité linguistique est certainement utile pour les développeurs, mais pas forcément pour les utilisateurs. Le but ultime de la construction de systèmes de traduction de parole est de les mettre dans les mains d'utilisateurs, n'est ce pas ?

XII.2.2. Mesure de l'utilisabilité

Une propriété malheureuse de la traduction est que la qualité linguistique n'est pas corrélée avec son utilisabilité pratique. Dès 1972, dans un rapport d'utilisabilité du système russe-français de Systran utilisé à Euratom (Ispra, Italie), le système a obtenu la note de 1 sur

5 quant à la qualité linguistique des traductions produites et une note de 4,5 sur 5 quant à son utilité. Ce phénomène est aussi décrit dans [Church, K. W. and Hovy, E. 1993]. Dans le cas de la parole, il est notoire que les transcriptions de monologues ou dialogues interprétés sont jugées comme des traductions de très mauvaise qualité, bien que l'interprétariat soit très difficile et bien payé... parce que le résultat est très utile.

L'utilisabilité peut uniquement être mesurée sur un système en état de fournir des services. Pendant le développement, nous pouvons tout de même essayer de mesurer l'utilisabilité pour estimer l'utilité future du système en cours de développement. Mais, avec les paradigmes d'évaluation actuels utilisant exclusivement des méthodes objectives,

- l'utilisabilité ne peut être mesurée,
- les données brutes ne sont pas accessibles aux utilisateurs de l'évaluation qui ne peuvent pas se faire leur propre « jugement subjectif » et alors interpréter les scores en fonction de leur perception de la qualité et de l'utilisabilité.

Lors de ACL-03, lors de certains exposés présentant des exemples de traductions avec leurs scores, une part importante de l'assistance n'était pas d'accord avec les jugements de qualité, soit qu'ils fussent faux, soit que le sentiment fût que les traductions, bien que fausses, donnaient l'impression d'être parfaitement adéquates pour que les utilisateurs puissent en comprendre le sens et prendre les mesures appropriées.

Il est alors important que la communauté se concentre aussi sur la question de l'utilisabilité, et essaie de corréliser l'utilisabilité avec les techniques d'évaluation objective et/ou subjective. À notre connaissance, ce problème n'a pas été abordé en traduction de parole sauf dans le projet Verbmobil.

XII.2.3. Comparaison avec les systèmes commerciaux

Il est aussi courant que les systèmes soient évalués compétitivement, en particulier avec des systèmes disponibles dans le commerce ou sur le Web. Le raisonnement est le suivant : si les systèmes en cours de développement sont bien meilleurs, alors les traductions sont linguistiquement meilleures, et seront plus utilisables dans le futur.

On peut d'abord s'interroger sur l'équité de telles comparaisons, car les résultats de l'évaluation objective sont très sensibles à certains paramètres. Les systèmes commerciaux utilisés pour les comparaisons devraient au moins être paramétrés pour fournir les meilleures traductions possibles sur le corpus, ou ne pas être utilisés s'ils sont spécialisés pour une tâche très différente : alors que METEO fait bien mieux que des traducteurs juniors sur les bulletins météorologiques, il sera toujours très mauvais dans le domaine du tourisme !

Le premier paramètre à prendre en compte est le « style » des traductions de références avec lesquelles la sortie du système va être comparée. Par exemple, comment peut-on comparer un système verbeux à un système concis ? Alors, quelles sont les « bonnes » références ? De combien en a-t-on besoin ?

Le second paramètre est le domaine pour lequel le système a été développé. En particulier, comment peut-on comparer un système préparé pour un domaine particulier à un système à large couverture sur un jeu de test tiré de ce domaine particulier ?

Le troisième paramètre est la granularité de la comparaison. Les sorties des systèmes de TA peuvent être comparées à différents niveaux (mots, partie du discours, attributs de désinence). À quel niveau veut-on faire l'évaluation et pourquoi ?

Ensuite, l'espoir de mesurer l'utilisabilité ainsi est vaine. L'utilisabilité dépend crucialement de l'ergonomie d'un système intégré ou d'un service. Par exemple, les systèmes de traduction

statistique fondés sur des chaînes sont extrêmement lents comparés à des systèmes commerciaux fondés sur une programmation procédurale ou sur des ATN semi-déterminés ou des DCG, ..., et le temps de traitement de traitement croît très vite en fonction de la longueur des phrases à traduire. Mais le temps de traitement n'est jamais pris en compte. Un autre facteur important est la flexibilité : le système peut-il de quelque façon apprendre par l'usage, accepter des dictionnaires utilisateur, etc. ?

XII.2.4. Évaluation d'autres aspects

Dans l'étude préparée pour le JEIDA (cf. section II.1.2.2), bien d'autres aspects étaient évalués, comme dans les propositions de EAGLES (cf. section II.2.1) et de FEMTI (cf. section II.2.2).

Nous pensons que deux caractéristiques de l'architecture des systèmes de traduction de parole sont importantes pour évaluer leur potentiel pour le futur : l'utilisation du contexte et la richesse des structures de données manipulées.

Verbmobil est doté de ces deux caractéristiques : le système fait un usage important du contexte dialogique et utilise une structure d'interface très riche entre les modules.

Au contraire, l'architecture d'autres systèmes (C-STAR II et III, NESPOLE) est presque triviale. Chaque tour de parole est traduit de façon isolée, ignorant le contexte. Enfin, l'information transmise d'un module à un autre est minimale, l'architecture se réduisant à une simple intégration de différentes boîtes noires.

Nous voudrions insister sur la nécessité de tendre vers une meilleure intégration des composants comme dans Verbmobil. Nous proposerons des pistes dans la section suivante.

XII.3. Perspectives de recherche¹²³

XII.3.1. En évaluation

Comme nous l'avons déjà dit, et d'autres l'ont dit avant nous, les bonnes évaluations doivent s'intéresser à plusieurs critères et donc utiliser plusieurs méthodes de mesure. Le processus ne peut être complètement automatique, même avec des méthodes objectives.

XII.3.1.1. Évaluer la qualité linguistique potentielle et effective

Mesurer les propriétés d'un système en fonction des choix d'architecture linguistique, — et d'une certaine façon la qualité que l'on peut obtenir de manière asymptotique — ne peut être fait que par des spécialistes, et le coût d'une telle évaluation est limité.

Mesurer réellement la qualité linguistique des traductions candidates produites par un système demande beaucoup d'effort, cet effort peut cependant être partagé. Notre équipe construit à cet effet la plateforme PolyphraZ [Hajlaoui, N. and Boitet, C. 2004], accessible par le web, qui permet l'affichage, la traduction (en utilisant plusieurs outils de traduction), l'édition (par post-édition ou traduction), et l'évaluation de corpus parallèles de « polyphrases » multilingues. Une polyphrase correspond au sens d'un énoncé dans une certaine langue. C'est une ligne de tableau dans laquelle chaque colonne contient un ensemble de propositions homogènes telles que des paraphrases, des résultats de

¹²³ Cette section est une version traduite en français et augmentée de [Blanchon, H., *et al.* 2004a].

systèmes de traductions, des scores en évaluation objective ou subjective, des distances d'édition, etc.

XII.3.1.2. Évaluer l'utilisabilité

Pour mesurer l'utilisabilité, il est sans doute possible de proposer des prototypes sur le Web intégrés dans un service sommaire, mais utilisable, disponible tout le temps et permettant d'utiliser soit de la traduction automatique, soit de la traduction humaine (réalisée par un compère en mode Magicien d'Oz, ou par un traducteur humain identifié comme tel), soit une combinaison. Il serait alors possible d'évaluer l'efficacité de chaque configuration (ST) en fonction du temps qu'il faut aux interlocuteurs pour atteindre leurs objectifs dans la configuration ST comparée à la durée de la tâche avec interprète humain.

$$\text{Efficacite Relative}_{ST} = \frac{\text{Temps}_{\text{Humain}}}{\text{Temps}_{ST}}$$

Équation XII-1 : Efficacité relative d'un système de traduction de dialogue

Pour atteindre cet objectif, il est important de proposer une plateforme très flexible afin que les développeurs puissent essayer une grande variété de configurations.

XII.3.1.3. Ne pas se concentrer seulement sur l'évaluation !

Notre dernière proposition est enfin ... de ne pas se focaliser uniquement sur l'évaluation¹²⁴. Si l'on considère l'histoire de la TA, il semble qu'une somme disproportionnée d'efforts (et d'argent) ait été consacrée à l'évaluation de systèmes ou même de maquettes par rapport aux efforts consentis pour le développement de systèmes. Cette tendance semble encore s'accroître ces dernières années. L'évaluation de l'évaluation est aussi devenu un axe de recherche très dynamique qui produit un grand nombre d'articles¹²⁵.

Pour lutter contre ce phénomène, il nous semble indispensable de revenir à une approche multicritère, et donc de réintroduire un critère d'utilisabilité. Cet objectif peut être atteint en construisant des prototypes de Systèmes de Traduction de Dialogue, en les mettant à disposition sur le Web et en procédant alors à l'évaluation de plusieurs critères.

Cette approche doit aussi permettre aux développeurs d'expérimenter différentes approches, plutôt que de converger, comme c'est maintenant le cas, vers des méthodes qui maximisent les scores obtenus pour telle ou telle méthode d'évaluation¹²⁶. Par exemple, l'usage en « situation réelle » pourrait bien montrer qu'une bonne combinaison entre le traitement automatique, le contrôle de l'utilisateur et la désambiguïsation interactive est plus efficace que les solutions envisagées actuellement.

XII.3.2. En architecture des systèmes

L'amélioration de la qualité des systèmes nécessite, à notre avis, de travailler dans trois directions : l'intégration des composants, la gestion du dialogue et l'utilisation du contexte.

XII.3.2.1. Intégration des composants

L'intégration des composants doit permettre aux systèmes de mieux traiter l'entrée en langue source et de produire une meilleure sortie en langue cible.

¹²⁴ Par exemple, [Langlais, P., *et al.* 2004] rapportent une expérience d'évaluation réalisée en un mois en partant uniquement de 20 000 paires de phrases chinois-anglais, de modules disponibles sur le Web et en écrivant deux cents lignes de code en C.

¹²⁵ Cette section est l'illustration que nous n'avons pas échappé nous-mêmes au phénomène.

¹²⁶ Communication orale de S. Vogel lors de IWSLT 2004.

XII.3.2.1.1. En entrée

En ce qui concerne le traitement de l'entrée, nous envisageons une interaction bidirectionnelle entre le module de reconnaissance de la parole et les modules de traduction. En particulier, le module de reconnaissance de la parole peut produire une structure plus riche qu'une suite de mots orthographiques. Cette structure peut être un treillis de mots complété éventuellement d'informations prosodiques et éventuellement, dans le cas d'une architecture linguistique à pivot, un texte ou un treillis partiellement analysé dans le langage pivot.

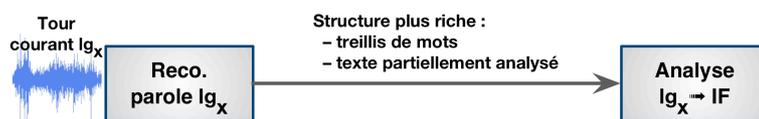


Figure XII-1 : Interaction entre la reconnaissance de la parole et le module d'analyse

Cette proposition semble maintenant prendre corps dans différents travaux [Saleem, S., *et al.* 2004, Vu Minh, Q., *et al.* 2004b, Zhang, R., *et al.* 2004].

Pour aider le module de reconnaissance de la parole d'une **langue x**, les modules d'analyse et de génération qui participent au système peuvent aussi lui transmettre différents types d'information.

Le premier type d'information auquel nous pensons comprend les ensembles de mots suivants :

- les mots utilisés (reconnus) par le locuteur, en **langue x**, lors du tour de parole courant,
- les mots utilisés par le module de génération, en **langue x**, pour produire la régénération du tour de parole courant, en **langue x**, et
- les mots utilisés par le module de génération, en **langue x**, pour produire la traduction du tour de parole précédent, en **langue y**, de l'interlocuteur.

Le poids de ces mots doit être augmenté lors de l'activation suivante du module de reconnaissance de la parole de la **langue x**. En effet, par mimétisme, le locuteur de la **langue x** a tendance réutiliser son propre vocabulaire et le vocabulaire du système.

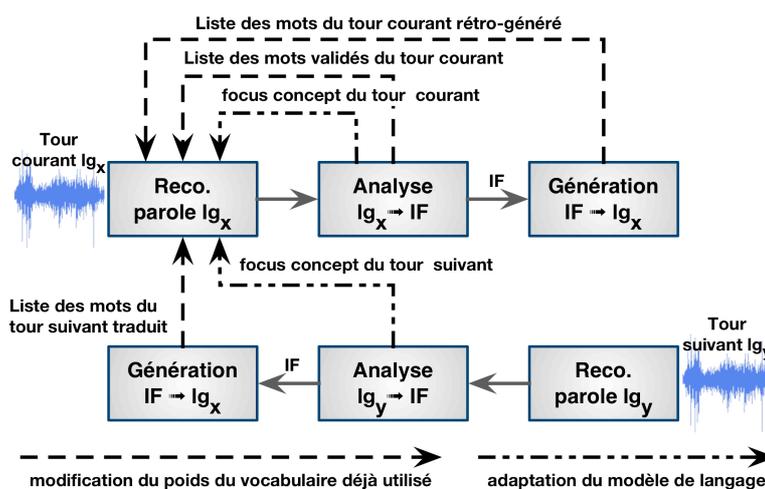


Figure XII-2 : Interaction des modules d'analyse et de génération avec le module de reconnaissance de la parole

Le deuxième type d'information qu'il peut être intéressant de transmettre au module de reconnaissance de la parole de la **langue x** est le ou les domaines du tour de parole en

cours, ainsi que le domaine du tour suivant¹²⁷. Pour le tour courant les informations peuvent être transmises par l'analyseur de la **langue x**, pour le tour suivant les informations peuvent être transmises par le générateur de la **langue y**, ou par l'analyseur de la **langue x**. Cela pourrait permettre au module de reconnaissance de la parole de changer « à la volée » son modèle de langage pour adopter le plus approprié.

XII.3.2.1.2. En sortie

En ce qui concerne la sortie vers la synthèse de la parole, il peut être intéressant de fournir en entrée du module de synthèse une information plus riche qu'une liste de mots. En effet, à partir de cette liste de mots, le module de synthèse commence par faire une analyse afin de détecter les frontières de syntagmes et de propositions, ainsi que la force illocutoire (affirmation, question, exclamation, ...).

Lors de la génération, la force illocutoire est connue ; elle est fournie par l'acte de dialogue. Il est aussi très facile de générer des marques qui signalent les frontières de syntagmes et de propositions. Si le générateur produit ces informations, elles sont sûres, il n'est donc pas nécessaire que le module de synthèse utilise des heuristiques pour faire des choix, par définition, incertains.

XII.3.2.2. Gestion du dialogue

Comme nous l'avons fait remarquer dans le Chapitre IX, le domaine de la traduction de dialogue est conceptuellement très proche du domaine de la communication orale entre personnes et systèmes d'information. La seule différence réside dans le fait que les informations fournies à l'utilisateur proviennent de bases de données dans le cas des systèmes de COP-SI, alors qu'elles proviennent d'un autre humain dans le cas des systèmes de traduction de dialogue.

On peut constater que la communauté des COP-SI continue d'insister énormément sur la gestion du dialogue (voir par exemple [Hardy, H., *et al.* 2003]) alors que notre communauté semble ignorer cet état de fait.

La gestion du dialogue partage certaine caractéristique de ce que nous appelons plus généralement la gestion du contexte.

XII.3.2.3. Utilisation du contexte¹²⁸

Il est généralement admis que la prise en compte du contexte est nécessaire pour atteindre une meilleure « compréhension » et une meilleure traduction d'énoncés en langue naturelle. Cependant, lorsque nous avons cherché des arguments en faveur de cette idée dans les traces de dialogues C-STAR II, nous avons trouvé très peu d'exemples qui étayaient cette idée [Blanchon, H. and Boitet, C. 2000]. Cela est dû au fait que les dialogues avaient été longuement répétés, et que les participants ne produisaient pas vraiment des énoncés spontanés, mais des énoncés dont on était sûr qu'ils permettraient de produire de bonnes traductions. L'IF ne permettait d'ailleurs pas de représenter les phénomènes que nous allons évoquer.

Par exemple, au lieu de répondre « deux » à la question « voulez-vous rester une ou deux semaines ? », la personne qui jouait le rôle du client répondait « deux semaines », bien que cette formulation soit moins naturelle, afin d'éliminer le problème de la résolution de la référence anaphorique. Dans cet exemple, il faut absolument restaurer « jour », « mois », ou

¹²⁷ Une première expérience dans ce sens est décrite dans [Seneff, S., *et al.* 2004] pour le domaine des COP-SI.

¹²⁸ Cette section est une reprise partielle de [Boitet, C., *et al.* 2000].

« semaine » pour faire une traduction correcte en japonais. On ne peut pas traduire « deux » par 二 (ni), mais on doit préciser 二日 (futsuka, deux jours), 二週間 (nishuukan, deux semaines), ou 日課月 (nikkagetsu, deux mois).

Ce problème n'est pas non plus très apparent dans les dialogues monolingues, parce que les participants résolvent la plupart des problèmes sans en avoir conscience. En contrepartie, si l'on étudie des dialogues bilingues spontanés interprétés [Fafiotte, G. and Boitet, C. 1996, Park, Y.-D., *et al.* 1995], ou si l'on restaure la spontanéité dans les traces citées ci-dessus, on se rend compte que la gestion des problèmes relatifs au contexte est essentielle afin d'augmenter la qualité et le naturel de la traduction.

Nous distinguons trois types de contexte : le contexte global, le contexte dialogique et le contexte linguistique.

XII.3.2.3.1. Contexte global

Le contexte global permet de représenter les informations suivantes :

- le type de dialogue : demande de réservation, demande d'information, demande d'assistance, conversation ;
- les caractéristiques des participants, en particulier leurs nom et prénom, leur sexe, leur âge, et le niveau relatif de politesse qu'ils utilisent pour interagir ;
- le rôle des participants (agent/client, médecin/patient, hôte/invité) et leur degré d'accointance (inconnus, amis, ancien étudiant/professeur, ...) ;
- le nom de leur localisation et de leur affiliation, car ils peuvent être personnalisés (« mais Grenoble/le CLIPS vient de me dire... »).

Ce type d'information est aussi nécessaire aux interprètes humains.

En allemand ou en japonais, les noms propres doivent être utilisés lors des salutations. « Bonjour Monsieur » est possible en français, mais on ne peut pas dire „Guten Tag (mein) Herr“ en allemand. Il faut dire, „Guten Tag Herr Müller“. Et si un Japonais dit [Tanaka-san], il faut être capable de rétablir la distinction « Mr Tanaka », « Mme Tanaka », « Melle Tanaka ».

L'âge est aussi important pour choisir en une forme directe ou polie (tu/vous en français) : en France, « tu » sera utilisé pour s'adresser à un enfant jusqu'à onze ans, et vous est obligatoire pour les adultes. En français du Canada, la convention est différente, mais cette information dépend plutôt du contexte linguistique.

Le niveau relatif de politesse est aussi très important dans beaucoup de langues, pas seulement en japonais et en coréen. Contrairement à ce que l'on pense, l'anglais a beaucoup de formes différentes même si il n'a pas la distinction tu/vous :

- Give me a room with bath
- A room with bath, please
- Please give me a room with bath
- Could you give me a room with bath?
- Would you please leave a message for Mr Smith?
- Would you be so kind as to take a message?
- ...

Le rôle des participants et les relations qu'ils entretiennent sont aussi importants pour générer les formes d'apostrophe appropriées : « Docteur, est-ce que c'est sérieux ? », « Je suis désolé, mon ami. », « Mais, Monsieur X, je ne le trouve pas. », ...

XII.3.2.3.2. Contexte dialogique

Produire le bon acte de dialogue et traduire de façon appropriée les connecteurs du dialogue (« oui », « mais », « je comprends », ...) est très important [Tomokiyo, M. 2000].

Si l'on utilise l'approche IF, dans laquelle l'acte de parole est contenu dans la structure, les générateurs possèdent les informations nécessaires.

Avec d'autres approches, les structures linguistiques ne contiennent généralement pas cette information, les générateurs doivent donc recevoir une paire (acte de parole, structure). Avec toutes les approches, le problème principal est, durant l'analyse, de calculer l'acte de parole à partir de l'énoncé et du contexte précédent.

À long terme, des analyseurs combinant de manière fortement couplée des traitements linguistiques et du dialogue devraient voir le jour. Pour l'instant, il semble plus réaliste de développer séparément le « processeur du dialogue » et de transmettre une représentation du contexte à l'analyseur linguistique qui doit être capable d'utiliser les informations contextuelles.

Le contexte du dialogue doit contenir :

- une représentation du passé du dialogue,
- l'étape en cours si elle respecte un script connu, et
- des prédictions sur le futur.

À court terme, beaucoup serait déjà possible si l'analyseur pouvait disposer d'une liste triée d'actes de parole prédits par un modèle de dialogue adéquat. Dans ce cadre, les analyseurs devraient produire leur sortie habituelle (IF ou structure linguistique), mais aussi un acte de parole identifié s'il n'est pas explicite dans la sortie.

XII.3.2.3.3. Contexte linguistique

Les dialogues en langue naturelle contiennent beaucoup d'anaphores et d'ellipses. Cela limite de manière importante la qualité des analyseurs qui manipulent un énoncé sans aucune information sur les énoncés précédents, même si les informations manquantes sont présentes dans un énoncé précédent au sein du même tour de parole.

La partie la plus importante du contexte linguistique semble être une liste des centres possibles, c'est à dire une liste de référents éventuels pour les éléments anaphoriques et les ellipses. Voici un exemple d'exploitation envisageable pour un dialogue français-allemand :

(1a) Nous avons deux chambres, une sur cour avec WC et l'autre sur rue avec douche et WC .

...2 Zimmer,...

(1b) Pour aller à la gare, ne prenez pas la première rue à droite, mais la seconde .

...die erste Straße...

(2) D'accord, je prends la seconde .

Einverstanden, ich werde das/die zweite nehmen.

Lorsque l'on traduit la phrase (2), le genre sera neutre dans le cas où elle succède à la phrase (1a), mais il sera féminin avec la phrase (1b). L'idée est donc d'enrichir le texte à manipuler (une chaîne orthographique en analyse et une IF en génération) avec toutes les informations disponibles.

Le contexte est aussi important pour la désambiguïsation lexicale (exemple : « Je prendrai un express » → "I will take an espresso/an express train"), et pour une sélection lexicale consistante d'un énoncé au suivant.

XII.3.3. Premières idées concrètes et réalisations autour de l'architecture des systèmes

Vu-Min Quang [Vu Minh, Q. 2004] a conduit deux expériences en vue de réaliser une meilleure intégration des composants de reconnaissance de la parole et d'analyse. La première expérience visait à étudier s'il est possible d'améliorer les résultats produits par les modules d'analyse et de génération du français à partir des meilleures hypothèses du module de reconnaissance de la parole, et non plus seulement de la première. La seconde expérience constitue un premier pas vers un modèle de langage sémantique, dérivé de la spécification de l'IF. Ce modèle de langage sémantique permet au module de reconnaissance de la parole de produire une sortie partiellement analysée.

XII.3.3.1. Intégration de la reconnaissance de la parole et de l'analyse

XII.3.3.1.1. Analyse des n meilleures hypothèses

Afin d'évaluer le gain potentiel de l'analyse d'un treillis de mots, le module de reconnaissance fournit les vingt meilleures hypothèses de reconnaissance classées par ordre décroissant de vraisemblance.

Le Table XII-12 donne les 20 meilleures hypothèses pour deux des phrases du corpus.

	oui je le vois	d accord très bien
1	oui je vois	d accord très bien
2	oui je la voir	d accord très bien
3	oui je la vois	d accord c est bien
4	oui je le vois	d accord c est bien
5	oui je la voir	d accord et bien
6	oui je le vois	décor très bien
7	oui je la vois	la corse est bien
8	oui je le vois	d accord très bien que
9	oui je le vois	d accord et bien
10	oui je vois	d accord très bien que
11	oui je le voir	décor très bien
12	puis je la voir	des corps très bien
13	oui je le voir	la corse est bien
14	mise à soir	les corps très bien
15	sujet le voir	le corps très bien
16	puis je vois	décors très bien
17	sujet le voir	La corse à bien
18	mis à soir	la corse et bien
19	puis je le vois	des corps très bien
20	oui je la soir	les corps très bien

Table XII-12 : Liste des vingt hypothèses de reconnaissance sur les énoncés : « oui je le vois » et « d accord très bien »

Données et protocole

Les données utilisées pour cette évaluation sont les mêmes que celles que nous avons utilisées pour la seconde évaluation NESPOLE!. Nous avons travaillé sur des traductions monolingues.

Figure XII-3 : Interface d'évaluation de la traduction de la première hypothèse et de choix d'une meilleure hypothèse

Les vingt meilleures hypothèses sont analysées séquentiellement par le module d'analyse. En reproduisant le protocole de la seconde évaluation NESPOLE!, les évaluateurs notent la traduction de la meilleure hypothèse. Ils doivent aussi dire si l'une des traductions des hypothèses suivantes obtient une note supérieure à la traduction de la meilleure hypothèse.

Pour cela, nous avons développé une interface Web (Figure XII-3).

Nous avons invité 3 juges de nationalité française à faire l'évaluation à l'aide de cette application. Le résultat est enregistré dans une base de données ACCESS pour faire des statistiques.

Résultats

Rappelons qu'il y a au total 216 phrases de référence qui correspondent à 312 SDUs, avec trois évaluateurs nous avons un total de 936 SDU notées.

Nous avons comparé les notes obtenues par la traduction de la première hypothèse et par la meilleure parmi les traductions des vingt hypothèses.

Note	Évaluation de la première hypothèse		Évaluation des vingt hypothèses		acceptabilité
	Nb. De SDU	%	Nb. De SDU	%	
VERY GOOD	456	48,72	480	51,28	acceptable
GOOD	58	6,20	76	8,11	
BAD	66	7,05	86	9,20	non acceptable
VERY BAD	356	38,03	294	31,41	
Total	936	100	936	100	

Table XII-13 : Évaluation de la meilleure hypothèse et des vingt hypothèses

La Table XII-14 montre la distribution du rang de la meilleure hypothèse.

Position des meilleures hypothèses	Nombre	%
1	585	90,14
2	18	2,77
3	5	0,77
4	7	1,08
5	7	1,08
6	2	0,3
7	6	0,92
8	5	0,77
9	1	0,15
10	2	0,3
11	3	0,46
12	2	0,3
13	1	0,15
14	1	0,15
16	1	0,15
17	0	0
18	2	0,3
19	1	0,15
Total	649	100

Table XII-14 : Distribution des meilleures traductions selon le rang de l'hypothèse

Commentaires

La Table XII-13 indique que la première hypothèse permet de produire une traduction acceptable dans 54,82% des cas. Si l'évaluateur choisit l'hypothèse qui permet d'obtenir la meilleure traduction, le taux de traduction acceptable passe à 59,39%, soit un gain de 4,57%. Si l'on examine finement le gain en qualité de traduction, on observe que celle-ci augmente pour 73 SDU (soit 8% du total). Les cas qui nous intéressent plus particulièrement sont ceux pour lesquels la traduction devient acceptable. La Table XII-15 montre l'évolution observée de la qualité de la traduction.

Évolution de la note	Nombre de cas observés	%
GOOD → VERY GOOD	2	3%
BAD → VERY GOOD	7	10%
VERY BAD → VERY GOOD	16	22%
BAD → GOOD	1	1%
VERY BAD → GOOD	19	26%
VERY BAD → BAD (4→3)	28	38%
Total	73	

Table XII-15 : Évolution des notes lorsque la qualité de la traduction s'améliore

La Table XII-14 montre une grande dispersion de l'indice de l'hypothèse qui permet un gain en traduction.

Le gain (4,57%) que nous avons calculé représente l'espérance maximale. Il faudrait pouvoir utiliser un analyseur et un générateur de meilleure qualité, cf. infra, pour mesurer un gain effectif.

XII.3.3.1.2. Production d'une sortie déjà partiellement analysée¹²⁹

Nous avons aussi travaillé dans une autre direction : adapter le module de reconnaissance de la parole afin qu'il produise une chaîne partiellement ou complètement analysée en IF. Nous avons utilisé un modèle de langage utilisant des classes qui correspondent directement à des concepts de l'IF.

Construction d'un modèle de langage sémantique

Un modèle de langage statistique constitue un élément important dans un système de reconnaissance automatique de la parole. Il a pour but de donner la probabilité d'existence d'une chaîne de n'importe quelle suite de trois mots (dans le cas du modèle trigramme). Pendant l'apprentissage du modèle, certains mots peuvent être regroupés en classes, si on considère que chaque mot, à l'intérieur d'une classe, a la même probabilité d'occurrence. Une classe n'est donc qu'une liste de mots qui sont tous équiprobables dans cette classe. Lorsqu'on a défini des classes, on peut alors directement remplacer les mots du corpus d'apprentissage par leur classe avant l'apprentissage du modèle de langage (ML).

Pour introduire des connaissances sémantiques dans le ML, notre idée est la suivante : regrouper tous les mots (par exemple : « bien », « d'accord », « okay », etc.) en des classes correspondant à des entités sémantiques de l'IF ; ces mots correspondent par exemple à la classe « *c:acknowledgment* ». [Brown, P. F., et al. 1992, Kneser, W. and Ney, H. 1993] ont montré l'intérêt de l'utilisation de classes dans diverses tâches de Traitement Automatique de Langue Naturelle. La plupart des méthodes pour constituer automatiquement des classes (donc des ensembles de mots) utilisent des critères statistiques permettant, par exemple, de diminuer la perplexité d'un modèle de langage.

Dans notre cas, notre critère de choix de classes est guidé par la définition du langage pivot et par les concepts les plus utilisés dans l'IF. Notre approche consiste en deux étapes : (1) la sélection des IFs les plus fréquentes à intégrer comme classes dans le nouveau modèle de langage (2) le calcul du modèle de langage proprement dit. Ces étapes, toutes automatiques, sont détaillées dans le paragraphe suivant.

Sélection des classes-IF les plus fréquentes

Utiliser toutes les unités sémantiques présentes dans la définition de l'IF comme classes dans notre modèle de langage conduirait à un modèle inutilisable pour la reconnaissance

¹²⁹ Le texte de cette section est une adaptation de [Vu Minh, Q. 2004, Vu Minh, Q., et al. 2004a].

automatique de la parole. En effet, le nombre de classes doit être limité et surtout, le nombre d'occurrences de mots d'une même classe doit être suffisamment important pour que les probabilités apprises soient correctes.

Nous avons donc choisi de nous limiter aux classes-IF les plus fréquemment rencontrées dans les dialogues du projet NESPOLE! qui correspondent à la tâche que nous voulons traiter. Dans cette étape, nous identifions donc ces IF les plus fréquentes et les regroupons dans des classes. Une classe correspond alors à un ensemble de mots conduisant à une même représentation IF. On trouve par exemple dans ces classes des actes de dialogue tels que *acknowledge*, *affirm*, *negate* La sélection des IFs les plus fréquentes est réalisée automatiquement à partir d'un corpus textuel brut non annoté manuellement en IF.

L'analyseur du second démonstrateur NESPOLE! a été utilisé pour analyser un corpus qui comprend la transcription de 46 dialogues [Burger, S., *et al.* 2001]. Après analyse, nous disposons d'un corpus aligné français-IF. Bien sûr, ce corpus n'est pas parfait car l'analyseur fait éventuellement des erreurs, mais nous supposons que, malgré ces erreurs, la distribution des différentes IF est correctement respectée. Ces données alignées sont alors triées par IF, toutes les SDU correspondant à une même IF sont listées, et nous obtenons enfin les classes « sémantiques » dont certaines sont présentées dans la Table XII-16. Par exemple, la classe la plus fréquente *affirm* contient les variantes représentant le même sens (*affirmer*) en français.

Classes IF	Variantes de SDU	Pourcentage dans un total de 3194 SDU
{c:affirm}	oui, ouais, mouais...	22%
{c:acknowledge}	d'accord, entendu, ok...	19%
{c:exclamation (exclamation=oh)}	Oh, ah, ha....	4%
...

Table XII-16 : Exemples de classes IF obtenues automatiquement

Le nombre des classes sémantiques obtenues de cette façon étant important, nous avons retenu seulement 41 classes en tenant compte de la taille (le nombre de mots ou variantes dans une même classe) et de la fréquence d'apparition dans les dialogues des classes.

Après avoir obtenu la liste des classes sémantiques, nous les utilisons en combinaison avec les données de l'apprentissage du modèle de langage afin de construire notre nouveau modèle « sémantique ». Ce processus est illustré dans la Figure XII-4.

Dans le corpus d'apprentissage du ML, nous remplaçons tous les mots (ou séquences de mots) qui sont éléments de nos nouvelles classes sémantiques, par le nom de la classe. Il en résulte alors un corpus « préparé » qui contient à la fois des mots français et des IF. Ensuite les outils traditionnels de calcul de ML sont utilisés pour obtenir notre nouveau modèle de langage « sémantique ».

Ce nouveau ML a été intégré au système de reconnaissance. La section suivante présente quelques résultats expérimentaux.

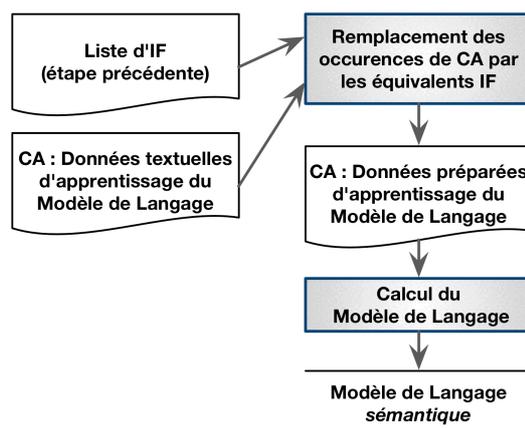


Figure XII-4 : Étapes de calcul du modèle de langage sémantique

Premiers résultats

Les signaux de test sont 216 tours de parole extraits du corpus de dialogues du projet NESPOLE!. La Table XII-17 illustre quelques exemples de ces tours de parole de test. Dans la deuxième colonne sont présentées les hypothèses textuelles obtenues à la sortie du module de reconnaissance utilisant notre modèle de langage sémantique.

Phrases de référence (client)	Sorties du module de reconnaissance avec nouveau ML
oui je vous entends	c:affirm c:dialog-hear(who=i, to-whom=you)
euh je vous entends pas très fort mais c'est correct	euh c:dialog-hear(who=i, to-whom=you) pas très forme ce_qu on est
oh oui c'est bon	c:exclamation (exclamation=wow) c:affirm c:acknowledge
Oui	c:affirm
d'accord	c:acknowledge

Table XII-17 : Exemples d'hypothèses de reconnaissance obtenues avec le modèle de langage sémantique

On peut remarquer que des tours de parole simples sont déjà complètement analysés en IF. D'autres tours de parole plus complexes sont, eux aussi, analysés partiellement ou complètement en IF.

Analyse des résultats

Premièrement, afin de vérifier que ces changements dans le modèle de langage, permettant une analyse partielle en IF, ne dégradent pas la performance intrinsèque du système de reconnaissance initial, nous avons comparé le taux d'erreur du système initial avec celui de notre nouveau système.

Le taux d'erreur de mots (Word Error Rate -WER) du système initial qui utilise des classes construites manuellement est de 31,9% alors que le taux d'erreur du système utilisant le nouveau modèle, après avoir reconstitué les mots français à partir de la classe IF, est 32,9%. Ainsi, nous pouvons constater que le nouveau modèle n'apporte pas une dégradation trop importante de la performance du système initial.

Les 216 tours de parole de test comprennent 915 mots. Parmi ces 915 mots, 35% ont été analysés directement en IF dès la phase de reconnaissance.

Au niveau des tours de parole, 125 tours sur un total de 216 (58%) ont été analysés directement en IF, dès la phase de reconnaissance. Bien sûr, ce sont surtout les tours de parole courts qui sont totalement analysés en IF, mais ce résultat reste encourageant car,

désormais, un volume important du travail du module d'analyse peut être réalisé directement par l'usage d'un module de reconnaissance utilisant un modèle de langage sémantique.

Par ailleurs, sur les 58% de tours directement analysés, 84% sont proprement analysés sans erreur. Les 16% d'erreur restant sur cette partie du corpus de test correspondent essentiellement aux erreurs de reconnaissance faites par le système, avec ou sans modèle sémantique, et qui ne seront jamais rétablies par le module d'analyse.

Conclusion

Avec notre nouveau modèle de langage, le module de reconnaissance peut réaliser directement une partie du travail d'analyse vers la représentation sémantique (IF) : 35% des mots du dialogue test ; 58% des tours de parole du dialogue test. Parmi ces 58% tours analysés directement, 84% sont correctement analysés.

Évidemment, la principale limitation de notre approche est que ce sont majoritairement les tours de parole les plus courts, et donc les plus faciles à analyser, qui sont traités dès la phase de reconnaissance. Par ailleurs, le module d'analyse devra être adapté afin de pouvoir traiter en entrée un mélange de mots français et d'IF. Néanmoins, cette modification reste relativement facile à implémenter.

Cette première expérience est cependant encourageante car les performances du module de reconnaissance ne sont pas dégradées. Pour aller plus loin, il faudrait multiplier les classes et réussir à instancier des « micro-arguments » composés simplement d'un nom et d'une valeur. Le rôle de l'analyseur consisterait alors à construire des arguments complets sur la base des micro-arguments construits par la reconnaissance de la parole.

Nous devons cependant souligner que cette approche de la reconnaissance de la parole empêche de présenter le résultat de la reconnaissance au locuteur. Celui-ci n'aurait plus de retour sur la qualité de la reconnaissance. Afin de contourner ce problème, il faudrait faire tourner en parallèle un module de reconnaissance avec un modèle de langage classique et un module de reconnaissance avec un modèle de langage sémantique.

XII.3.3.2. Nouvelle vision de l'unité de traduction

En communication interpersonnelle orale spontanée spécialisée telle que nous l'avons pratiquée dans le contexte des projets C-STAR et NESPOLE!, nous proposons qu'une unité de traduction soit maintenant composée d'un contexte et d'un tour de parole. Un processeur de dialogue est ajouté pour chaque langue utilisée dans la conversation multipartie.

XII.3.3.2.1. Schéma général

Chaque processeur de dialogue a accès au contexte pour chaque unité de traduction. La Figure XII-5 montre l'architecture générale envisagée pour manipuler le contexte.

Lorsqu'un locuteur parle dans la langue L_x , le module de reconnaissance de L_x produit une ou plusieurs hypothèses sous forme textuelle (①). Le module d'intégration demande au module de gestion du dialogue de L_x de produire une représentation du contexte (**Contexte- $A^1_{L_x}$**) et l'associe avec la sortie du module de reconnaissance afin de produire l'entrée de l'analyseur de L_x (②).

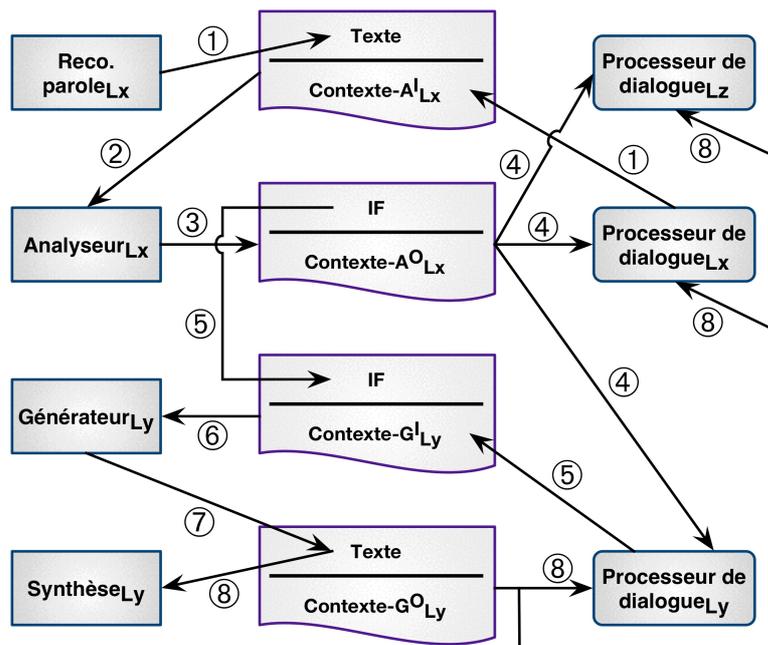


Figure XII-5 : Architecture de gestion du dialogue

L'analyseur de Lx produit alors sa sortie conventionnelle (une IF ou une liste d'IF) du tour de parole du locuteur, ainsi qu'un nouveau contexte (Contexte-A^0_{Lx}) dans lequel l'acte de dialogue calculé remplace la liste des actes de parole prédits ; la liste des centres possibles a aussi été mise à jour en ajoutant les éléments restaurés élidés (③).

Le processeur de dialogue stocke le nouveau contexte (④). Bien sûr, seul le gestionnaire de dialogue de Lx mémorise le contexte linguistique de Lx.

Pour Ly, comme pour toutes les autres langues, le gestionnaire de dialogue de Ly produit le contexte pour la génération dans la langue Ly (Contexte-G^1_{Ly}) en conservant les mêmes contextes global et dialogique que ceux du Contexte-A^0_{Lx} (⑤). Ce nouveau contexte est associé aux IF pour être transmis comme entrée au générateur de Ly (⑥).

Le générateur de Ly produit sa sortie normale pour le tour de parole du locuteur de la langue Lx et un nouveau contexte (Contexte-G^0_{Ly}), dans lequel la liste des centres possibles a été mise à jour (⑦). Le texte produit par le générateur de Ly est alors transmis au module de synthèse de la parole de Ly, pendant que les gestionnaires de dialogue mémorisent le nouveau contexte (ici, seul le gestionnaire de dialogue de Ly mémorise le contexte ⑧).

XII.3.3.2.2. Représentation possible sous forme de texte annoté

Au lieu d'échanger des structures compilées, nous proposons de combiner les représentations du contexte et du tour de parole (sortie de la reconnaissance de la parole, IF ou arbre linguistique) sous la forme d'un texte annoté qui sera l'unité de traduction effectivement soumise au composant de traduction (analyse ou génération).

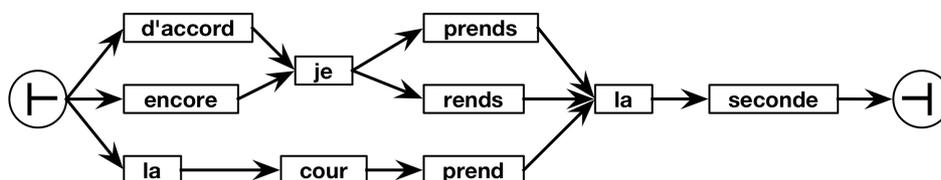
Voici un exemple de représentation possible d'une entrée du module d'analyse pour le tour de parole « d'accord, je prends la seconde. » sous une forme XML qui contient un contexte et un treillis de mots hypothétique produit par le module de reconnaissance de la parole.

```

<ctxt_glob>
  <speaker/>client</speaker>
  <client/>
    <name>Durand</name>
    <firstname>Marguerite</firstname>
    <age>70</age>
    <gender>F</gender>
    <title>Madam</title>
  </client>
  <agent/>
    <name>Biedermeyer</name>
    <firstname>Hans</firstname>
    <age>52</age>
    <gender>M</gender>
    <title>Herr</title>
  </agent>
  <firm>NTG</firm>
  <topic>hotel reservation</topic>
</ctxt_glob>
<ctxt_dial>
  <stage>central episode</stage>
  <past_sp_acts>request-information give-information request-action</past_sp_acts>
  <future_sp_acts>accept reject request-information</future_sp_acts>
</ctxt_dial>
<ctxt_ling> pension-hotel_NF réserver_VT chambre_NF cour_NF réserver_VT pension-
régime_NF prendre_VT rue_NF
</ctxt_ling>
<utterance>
  [[[d'accord|encore] je [prends|rends]]] la cour prend la seconde
</utterance>

```

La représentation graphique du treillis de mots est la suivante :



XII.3.3.3. Esquisse de la gestion du contexte en analyse et en génération

XII.3.3.3.1. Analyse

Dans [Boitet, C., *et al.* 2000], nous décrivons comment une analyse vers l'IF en ARIANE-G5 pourrait utiliser le contexte. Pour ce qui est de l'analyseur implémenté pour NESPOLE!, le contexte pourrait être utilisé dans différentes circonstances.

Choix de la meilleure hypothèse de reconnaissance

Lorsque l'analyseur doit choisir la meilleure parmi les différentes hypothèses produites par le module de reconnaissance de la parole, le contexte du dialogue pourrait être très utile. Imaginons que le tour courant du locuteur L_x commence par une confirmation (acknowledge) et que plusieurs hypothèses soient produites pour le reste du tour de parole. Si le tour de parole du locuteur L_y était de type give-information, alors l'analyseur devrait choisir en priorité l'hypothèse dont les mots étaient aussi présents dans la génération en L_x du tour de parole de L_y .

Résolution des anaphores et des ellipses

En cas d'anaphore et d'ellipse, si le locuteur répond à une question de l'interlocuteur, alors il est presque certain que le référent est le focus concept de la question. L'exemple suivant se répète suivant dans les corpus NESPOLE!

Agent : voyez-vous la carte ?

Client : oui, je la vois.

L'IF correspondant au tour de parole de l'agent est :

```
{a:request-information+view+information-object(who=you, info-object=(map,
identifiability=yes))}
```

Les IF qui correspondent au tour de parole du client sont :

```
{c:affirm} {c:give-information+view(who=i, what=pronoun)}
```

La génération en français donne :

Agent : Est-ce que vous voyez la carte ?

Client : Oui. Je vois ce dont nous venons de parler. (et simplement en italien Si. Vede)

En effet, pour être passe-partout `what=pronoun` est rendu en français par « ce dont nous venons de parler ».

Par contre, si l'analyseur peut savoir que le focus concept du tour précédent est `information-object` et que l'argument qui lui est associé est `info-object=(map, identifiability=yes)`, il pourrait alors produire l'IF suivante :

```
{c:give-information+view+information-object (who=i, info-object=(map,
identifiability=yes))}
```

La génération donnerait alors :

Client : Oui. Je vois la carte.

Résolution interactive des problèmes par le locuteur ou l'interlocuteur

Si l'analyseur n'est pas capable de choisir une interprétation, on peut envisager deux stratégies pour résoudre l'ambiguïté : désambiguïsation par le locuteur ou désambiguïsation par l'interlocuteur.

Dans le cadre de la première approche, une question peut être posée au locuteur du tour de parole en cours de traitement afin de lui permettre de choisir la bonne interprétation. C'est ce que ferait un interprète humain. Dans une étude souvent citée, [Oviatt, S. and Cohen, P. 1991] ont montré que, dans un dialogue bilingue médiatisé par un interprète humain professionnel, jusqu'à 30% des tours de parole servent à la clarification entre le locuteur et l'interprète.

Avec la seconde approche, une interprétation est choisie au hasard et synthétisée pour l'interlocuteur en l'avertissant que d'autres interprétations du tour de parole du locuteur sont disponibles. Si l'interlocuteur n'est pas satisfait de la traduction produite, il peut, oralement, demander à entendre la traduction suivante avant de demander au locuteur de clarifier son tour de parole. Cette idée est déjà mise en pratique dans les systèmes de dictée automatique.

XII.3.3.3.2. Génération

Comme nous l'avons dit précédemment, l'acte de parole de l'énoncé à générer est immédiatement disponible dans le cas de l'approche IF. Sinon, l'acte de parole doit être retrouvé dans le contexte disponible pour l'unité de traduction.

D'autres éléments du contexte peuvent être utilisés pour générer les formes correctes d'adresse, le bon niveau de politesse, et peut-être des attributs auxiliaires qui seront transformés en marques prosodiques utilisées par le module de synthèse de la parole afin de produire une sortie plus naturelle. Il s'agit [Campbell, N. 2004] de l'état de santé, de l'âge, de l'intérêt pour le sujet, de l'implication dans le dialogue, l'état émotif, etc.

Le contexte linguistique peut être utilisé comme une sorte de mémoire pour guider la sélection lexicale. Par exemple, le client peut parler d'une « pension » qui serait codée par `accommodation-spec=(hotel, accommodation-board=half_board)` dans une version des spécifications de l'IF. En répondant, l'agent, pourrait utiliser le mot « hôtel », si dans cette version des spécifications, les deux termes sont associés au même concept. En sachant par le contexte linguistique que `PENSION_NF` est apparu plus récemment que `HOTEL-NF` et par le contexte dialogique que c'est l'agent qui parle, le générateur du français pourrait produire « pension », respectant ainsi la terminologie du client.

Lors des choix lexicaux, le générateur devrait mettre à jour le contexte linguistique en lui ajoutant la liste des mots utilisés et en supprimant les doublons déjà présents dans le contexte passé.

Le générateur finirait par produire un résultat sous la forme d'un texte annoté contenant le contexte mis à jour et le résultat effectif de l'étape de génération linguistique. La génération linguistique pourrait produire plusieurs formes : une forme orthographique normale et une forme avec marques prosodiques à destination du module de synthèse.

Conclusion

La *nouvelle frontière* pour la traduction de parole et la traduction de dialogue est le domaine illimité [Lazzari, G., *et al.* 2004, Waibel, A. 2004]. Si cet objectif peut être considéré comme respectable du point de vue scientifique, c'est un objectif à très long terme, et de plus, nous savons que l'outil universel n'existe pas, ou alors qu'il est très imparfait. Cinquante ans de traduction de l'écrit l'ont bien montré. Nous pensons qu'une autre voie doit être poursuivie en parallèle pour produire des résultats à plus court terme et viser de vrais utilisateurs.

Il nous semble en effet réaliste de concevoir, non pas un système de traduction en domaine illimité, mais un **système multidomaine**. On se rapproche encore de l'architecture des systèmes de communication personne-système informationnels et de l'architecture GALAXY-II du MIT. Nous imaginons de solliciter différents systèmes de traduction spécialisés pour des domaines restreints via un centre de routage. Quand bien même le choix du module de traduction spécialisé au domaine courant ne pourrait être effectué automatiquement, l'utilisateur pourrait lui-même faire ce choix directement, en indiquant au système, au moyen d'un choix dans un menu, le sujet courant de sa conversation.

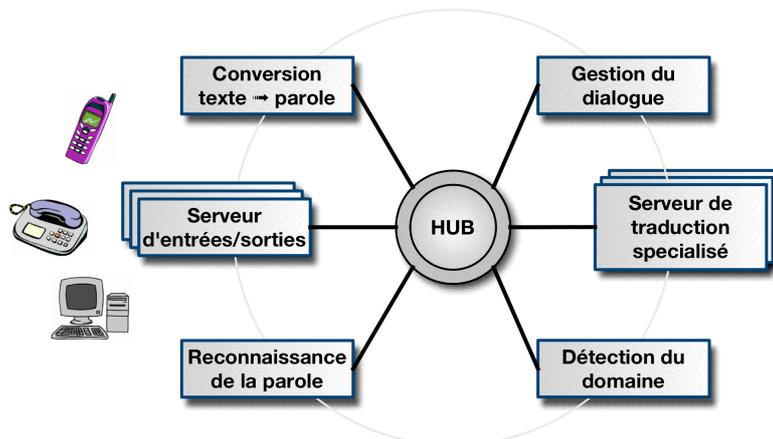


Figure II : Architecture d'un système de traduction de dialogue multidomaine

Pour ce qui est de l'approche à utiliser pour les systèmes de traduction proprement dite, l'approche pivot semble idéale. Quel pivot ? Ce pourrait être l'anglais comme le propose [Waibel, A. 2004]. Il faudrait bien sûr que la représentation pivot ne soit pas un texte en langue anglaise, on serait ramené aux problèmes de l'analyse, mais une représentation plus ou moins profonde d'un énoncé en anglais. Mieux vaut une représentation relativement profonde. Dans ce cas, un tel formalisme

existe déjà, et a été expérimenté sur une quinzaine de langues pour l'écrit, c'est UNL [Tsai, W.-J. 2004]. Pour l'oral, il conviendra sans doute d'y introduire de nouvelles conventions, ou de le relier à des annotations parallèles, pour exprimer les actes de parole, la force illocutoire, l'affect du locuteur, et le contexte en général. Des efforts en ce sens ont commencé, notamment avec la proposition *VoiceUNL* de M.Tomokiyo, S. Hollard et G. Chollet, élaborée dans le cadre des projets LingTour et Normalangue.

Enfin, en ce qui concerne l'utilisabilité et l'étude des usages dans des situations réelles, il faudrait mettre gratuitement des systèmes en version *beta* à la disposition des utilisateurs. Cela se fait déjà en traduction de l'écrit, où les grandes sociétés mettent gratuitement, sur le Web, des systèmes à la disposition des utilisateurs potentiels. Ces services servent de produits d'appel pour les logiciels payants. En traduction de dialogue, les systèmes mis à la disposition du public pourraient aussi être instrumentés pour collecter, à coût nul, de vraies données.

Conclusion

Deux tendances se développent actuellement dans le domaine de la TA, jusqu'à devenir presque hégémoniques :

- l'invasion de l'évaluation et de la méta-évaluation, et
- l'invasion des méthodes statistiques.

Leurs partisans s'entretiennent mutuellement dans une course à la publication de nouveaux résultats et dans la manière dont sont construites les campagnes d'évaluation compétitives.

On peut ainsi faire deux constats :

1. Les couples de « langues surprises » proposés aux compétiteurs ne permettent pas de mettre en œuvre des méthodes de traduction autres que statistiques. En effet, il n'y a pas de système commercial rodé pour ces couples, qui ne représentent pas un marché solvable, et il faut produire des résultats en trois semaines ! Comme il n'y a aucun financement pour le développement, il est également impossible à un groupe de recherche ayant une maquette ou même un prototype pour un tel couple de spécialiser son système au corpus d'entraînement en si peu de temps, même si, au bout du compte, il pourrait obtenir des résultats incomparablement meilleurs¹³⁰.
2. La nature des publications dans les grandes conférences du domaine, ACL et COLING, s'est complètement transformée depuis quatre ans. Nous faisons déjà cette remarque en 2000 [Blanchon, H. 2002]. Il n'est plus question de formalismes de représentation et de résolution de certains phénomènes linguistiques, il n'est question que de déploiement rapide et d'évaluation.

Pour améliorer la qualité, tout en restant dans le cadre de méthodes statistiques, il faut aussi monter vers des niveaux plus abstraits, en utilisant des représentations morphosyntaxiques (mots étiquetés), syntaxiques (syntagmatiques ou dépendancielles), voire sémantiques ; mais il va falloir préparer les données ! Qu'en devient-il alors du « développement » rapide ? Eh bien, il n'est pas plus rapide que dans l'approche par règles, comme le montrent les exemples d'IBM avec MASTOR et de Microsoft avec MSR-MT.

Quelle que soit l'approche envisagée, pour obtenir une qualité « utilisable », il faut disposer de ressources linguistiques « raffinées », et pas seulement « brutes », donc très chères si elles sont élaborées par des spécialistes. C'est pourquoi nous prêchons pour leur construction collaborative mutualisée, qu'il s'agisse de dictionnaires, de corpus annotés, ou de corpus « arborés ».

Nous affirmons que, sans DI, on n'arrivera pas à dépasser un niveau de qualité assez bas, en univers ouvert. On peut alors dire que « la qualité passe par la DI ».

¹³⁰ On pense par exemple aux nombreux prototypes entre une langue indienne et l'anglais, au coréen-anglais du KAIST, etc.

On peut et on doit préciser ce slogan :

- Il faut que le niveau d'interaction soit réglable en fonction de la qualité à atteindre.
- La DI peut « apprendre l'utilisateur » : profils, construction syntaxiques préférées... Le bon choix doit devenir le choix par défaut.

Dans le cadre de NESPOLE!, nous avons d'ailleurs montré que des méthodes « simples »¹³¹ produisent des résultats aussi bons.

En communication interpersonnelle multilingue écrite réfléchie tout terrain, nous avons proposé :

- une architecture adaptée aux exigences de la diffusion de la TAFD,
- une méthode de préparation de sessions de désambiguïsation interactive dont nous avons évalué l'utilisabilité et l'utilité,
- le concept de Documents Auto-Explicatifs.

En communication interpersonnelle multilingue orale spontanée spécialisée, nous avons proposé :

- de mettre les systèmes gratuitement entre les mains des utilisateurs pour enfin mesurer l'utilisabilité,
- d'offrir l'accès à différents systèmes spécialisés à travers une interface commune qui permettrait à un même utilisateur d'utiliser la traduction de dialogue dans différentes circonstances.

Nous avons aussi reconnu que l'approche par pivot sémantique est certainement très bonne dans un contexte fortement multilingue.

¹³¹ On peut se borner à présélectionner deux réponses, la dernière et la plus fréquente. On peut aussi proposer de petits menus pour obtenir l'affect du locuteur (comme les menus d'émojis dans les tchats), etc.

Bibliographie

- [Akiba, Y., *et al.* 2004] Akiba, Y., Federico, M., Kando, N., Nakaiwa, H., Paul, M. and Tsujii, J.-I. (2004) *Overview of the IWSLT04 Evaluation Campaign*. Proc. IWSLT 2004 (ICLSP 2004 Satellite Workshop). Kyoto, Japan. September 30-October 1, 2004. vol. **1/1**: pp. 1-12.
- [Akiba, Y., *et al.* 2001] Akiba, Y., Imamura, K. and Sumita, E. (2001) *Using Multiple Edit Distances to Automatically Rank Machine Translation Output*. Proc. MT Summit VIII. Santiago de Compostela, Spain. 18-22 September, 2001. vol. **1/1**: pp. 15-20.
- [Allauzen, C., *et al.* 2003] Allauzen, C., Mohri, M. and Roark, B. (2003) *The AT&T GRM library: Grammar library*. <http://www.research.att.com/sw/tools/grm/>.
- [ALPAC 1966] ALPAC (1966) *Language and Machine: Computers in Translation and Linguistics*. Rap. Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Science - National Research Council. Washington, D. C. n° 1416. November 1966. 124 p. Disponible à l'adresse: <http://books.nap.edu/books/ARC000005/html/>.
- [Amengual, J. C., *et al.* 1997] Amengual, J. C., Benedi, J. M., Beulen, K., Casacuberta, F., Castaño, A., Castellanos, A., Jiménez, V. M., Llorens, D., Marzal, A., Ney, H., Prat, F., Vidal, E. and Vilar, J. M. (1997) *Speech Translation Based on Automatically Trainable Finite-State Models*. Proc. EUROSPEECH'97. Rhodes, Greece. September 22-25, 1997. vol. **3/5**: pp. 1439-1442.
- [Amengual, J. C., *et al.* 2000] Amengual, J. C., Castaño, A., Castellanos, A., Jiménez, V. M., Llorens, D., Marzal, A., Prat, F. and Vilar, J. M. (2000) *The EUTRANS-I Spoken Language Processing Translation System*. in *Machine Translation*. vol. **15**(1-2): pp. 73-103.
- [Angelini, B., *et al.* 1997] Angelini, B., Cettolo, M., Corazza, A., Falavigna, D. and Lazzari, G. (1997) *Multilingual Person to Person Communication at IRST*. Proc. ICASSP'97. Munich, Germany. April 21-24, 1997. vol. **1/5**: pp. 91-94.
- [Antoine, J.-Y. and Goulian, J. 1999] Antoine, J.-Y. and Goulian, J. (1999) *Le français est-il une langue à ordre variable?* Proc. JILA'99. Nice, France. 24-25 juin 1999. vol. **1/1**: pp. 29-32.
- [Antoine, J.-Y. and Goulian, J. 2001] Antoine, J.-Y. and Goulian, J. (2001) *Word order variations and spoken man-machine dialogue in French: a corpus analysis on the ATIS domain*. Proc. Corpus Linguistics'2001. Lancaster, UK. 30 March - 2 April 2001. vol. **1/1**: pp. 22-29.
- [Babych, B. and Hartley, A. 2004a] Babych, B. and Hartley, A. (2004a) *Extending BLEU MT Evaluation Method with Frequency Weightings*. Proc. ACL 2004. Barcelona, Spain. July 21-26, 2004. 8 p.
- [Babych, B. and Hartley, A. 2004b] Babych, B. and Hartley, A. (2004b) *Modelling legitimate translation variation for automatic evaluation of MT*. Proc. LREC-2004. Lisbon, Portugal. May 26-28, 2004.
- [Baker, K., *et al.* 1994a] Baker, K., Franz, A. and Jordan, P. (1994a) *Coping With Ambiguity in Knowledge-based Natural Language Analysis*. Proc. FLAIRS-94. Pensacola Beach, Florida, USA. May 5-7, 1994. 5 p.

- [Baker, K., *et al.* 1994b] Baker, K., Franz, A., Jordan, P., Mitamura, T. and Nyberg, E. (1994b) *Coping With Ambiguity in a Large-Scale Machine Translation System*. Proc. COLING-94. Kyoto, Japan. August 5-9, 1994. vol. **1/2**: pp. 90-94.
- [Balkan, L. 1992] Balkan, L. (1992) *Translation Tools*. in *Meta*. vol. **37**(3): pp. 408-420.
- [Bangalore, S. and Riccardi, G. 2001] Bangalore, S. and Riccardi, G. (2001) *A Finite-State Approach to Machine Translation*. Proc. NAACL-2001. Pittsburgh, PA, USA. 8 p.
- [Bateman, R. 1985] Bateman, R. (1985) *Introduction to Interactive Translation*. in Lawson, V. (ed.), *Tools for the Trade*. Aslib. London. pp. 193-197.
- [Belkhatir, M. 2001] Belkhatir, M. (2001) *Réalisation d'un client de rédaction en langage Tcl/Tk pour une nouvelle architecture distribuée LIDIA*. Rap. Université Joseph Fourier (Grenoble I). Rapport de TER de licence d'informatique. avril 2001. 45 p.
- [Ben-Ari, D., *et al.* 1988] Ben-Ari, D., Berry, D. M. and Rimon, M. (1988) *Translational Ambiguity Rephrased*. Proc. Second International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language. Carnegie-Mellon University, Pittsburgh. June 12-14, 1988.
- [Bernth, A. and Gdaniec, C. 2001] Bernth, A. and Gdaniec, C. (2001) *MTranslatability*. in *Machine Translation*. vol. **16**(3): pp. 175-218.
- [Berthouzoz, C. and Merlo, P. 1995] Berthouzoz, C. and Merlo, P. (1995) *Statistical Ambiguity Resolution for Principle-based Parsing*. Proc. RANLP'95 (Recent Advances in NLP). Tzigov Chark, Bulgaria. 14-16 September, 1995. vol. **1/1**: pp. 179-186.
- [Berthouzoz, C. and Merlo, P. 2000] Berthouzoz, C. and Merlo, P. (2000) *Statistical ambiguity resolution for grammar-based parsing*. in Nicolov, N. and Mutkov, R. (ed.), *Recent Advances in Natural Language Processing II (Selected papers from RANLP'97)*. John Benjamins Publishing Company. Amsterdam/P. pp. 93-104.
- [Biber, D. 1988] Biber, D. (1988) *Variation across speech and writing*. Cambridge University Press. 299 p.
- [Biber, D. 1989] Biber, D. (1989) *A typology of English texts*. in *Linguistics*. vol. **27**: pp. 3-43.
- [Biber, D. 1995] Biber, D. (1995) *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge University Press. 446 p.
- [Black, E., *et al.* 1993] Black, E., Garside, R. and Leech, G. (1993) *Statistically-Driven Grammars of English: the IBM/Lancaster Approach*. Rodopi. Amsterdam. 248 p.
- [Blanche-Benveniste, C. 2000] Blanche-Benveniste, C. (2000) *Approches de la langue parlée en français*. Ophris. Paris. 164 p.
- [Blanchon, H. 1992a] Blanchon, H. (1992a) *Désambiguïsation interactive en TAO personnelle: poser les bonnes questions*. Proc. Avignon-92, conférence spécialisée: Le Traitement Automatique des Langues Naturelles et ses Applications. Avignon, France. 3-5 juin 1994. vol. **3/4**: pp. 69-80.

- [Blanchon, H. 1992b] Blanchon, H. (1992b) *A Solution to the Problem of Interactive Disambiguation*. Proc. COLING-92. Nantes, France. July 23-28, 1992. vol. 4/4: pp. 1233-1238.
- [Blanchon, H. 1994a] Blanchon, H. (1994a) *LIDIA-1: une première maquette vers la TA interactive "pour tous"*. Thèse. Université Joseph Fourier - Grenoble 1. 1994a. 321 p.
- [Blanchon, H. 1994b] Blanchon, H. (1994b) *Pattern-based approach to interactive disambiguation: first definition and implementation*. Rap. ATR-Interpreting Telecommunications Research Laboratories. Kyoto, Japan. Technical Report. n° TR-IT-0073. Sept. 8, 1994. 91 p.
- [Blanchon, H. 1994c] Blanchon, H. (1994c) *Perspectives en TAFD pour auteur monolingue après une première expérience: la maquette LIDIA-1*. Proc. TALN'94. Marseille, France. 7-8 avril 1994. vol. 1/1: pp. 12-23.
- [Blanchon, H. 1995a] Blanchon, H. (1995a) *Clarification: Towards more User-Friendly Natural Language Human-Computer Interaction*. Proc. Poster session of HCI'95, 6th International Conference on Human-Computer Interaction. Yokohama, Japan. July 9-14, 1995. vol. 1/1: pp. 42-42.
- [Blanchon, H. 1995b] Blanchon, H. (1995b) *An Interactive Disambiguation Module for English Input: an Engine and the Associated Lingware*. Rap. ATR-ITL. Kyoto, Japan. Technical Report. n° TR-IT-0129. August, 1995. 110 p.
- [Blanchon, H. 1995c] Blanchon, H. (1995c) *An Interactive Disambiguation Module for English Natural Language Utterances*. Proc. NLPRS'95. Seoul, Korea. Dec 4-7, 1995. vol. 2/2: pp. 550-555. *Best paper award*.
- [Blanchon, H. 1996] Blanchon, H. (1996) *A Customizable Interactive Disambiguation Methodology and Two Implementations to Disambiguate French and English Input*. Proc. MIDDIM'96. Le Col de Porte, Isère, France. August 12-14, 1996. vol. 1/1: pp. 190-200.
- [Blanchon, H. 1997] Blanchon, H. (1997) *Interactive Disambiguation of Natural Language Input: a Methodology and Two Implementations for French and English*. Proc. IJCAI-97. Nagoya, Japan. August 23-29, 1997. vol. 2/2: pp. 1042-1047.
- [Blanchon, H. 2002] Blanchon, H. (2002) *Area Report: Machine Translation*. Proc. COLING 2002, Post-Conference Workshop: A Roadmap for Computational Linguistics. Taipei, Taiwan. Jeu de 15 transparents. Accessible sur: <http://www.elsnet.org/roadmap-coling2002.html>.
- [Blanchon, H. 2004] Blanchon, H. (2004) *HLT Modules Scalability within the NESPOLE! Project*. Proc. ICSLP 2004. Jeju Island, Korea. October 4-8, 2004. 4 p.
- [Blanchon, H. and Besacier, L. 2004] Blanchon, H. and Besacier, L. (2004) *Traduction de dialogues: mise en perspective des résultats du projet NESPOLE! et pistes pour le domaine*. Proc. TALN 2004. Fès, Maroc. 19-21 avril 2004. vol. 1/1: pp. 55-60.
- [Blanchon, H. and Boitet, C. 2000] Blanchon, H. and Boitet, C. (2000) *Speech Translation for French within the C-STAR II Consortium and Future*

- Perspectives*. Proc. ICSLP 2000. Beijing, China. Oct. 16-20, 2000. vol. 4/4: pp. 412-417.
- [Blanchon, H. and Boitet, C. 2003] Blanchon, H. and Boitet, C. (2003) *Two Steps Towards Self-Explaining Documents*. Proc. Convergence 03. Alexandria, Egypt. December 2-6, 2003. paper 032-324, 6 p.
- [Blanchon, H. and Boitet, C. 2004a] Blanchon, H. and Boitet, C. (2004a) *Deux premières étapes vers les documents auto-explicatifs*. Proc. TALN 2004. Fès, Maroc. 19-21 avril 2004. vol. 1/1: pp. 61-70.
- [Blanchon, H. and Boitet, C. 2004b] Blanchon, H. and Boitet, C. (2004b) *Les Documents Auto-Explicatifs: une voie pour offrir l'accès au sens aux lecteurs*. Proc. CIDE 7: Approches Sémantique du Document Électronique. La Rochelle, France. 22-25 juin 2004. vol. 1/1: pp. 273-290.
- [Blanchon, H., et al. 2004a] Blanchon, H., Boitet, C. and Besacier, L. (2004a) *Spoken Dialogue Translation System Evaluation: Results, New Trends, Problems and Proposals*. Proc. IWSLT 2004. Kyoto, Japan. September 30 - October 1, 2004. vol. 1/1: pp. 95-102.
- [Blanchon, H., et al. 2004b] Blanchon, H., Boitet, C., Brunet-Manquat, F., Tomokio, M., Hamon, A., Hung, V. T. and Bey, Y. (2004b) *Towards Fairer Evaluation of Commercial MT Systems on Basic Travel Expressions Corpora*. Proc. IWSLT 2004 (ICSLP 2004 Satellite Workshop). Kyoto, Japan. September 30-October 1, 2004. vol. 1/1: pp. 21-26.
- [Blanchon, H., et al. 1999] Blanchon, H., Boitet, C. and Caelen, J. (1999) *Participation Francophone au Consortium C-STAR II*. in La tribune des industries de la langue et du multimedia/Linguistic engineering and multimedia tribune. vol. 31-32(August-December 1999): pp. 15-23.
- [Blanchon, H., et al. 2000] Blanchon, H., Boitet, C. and Guilbaud, J.-P. (2000) *Traduction de la parole pour le français: une première étape et quelques perspectives*. Proc. TALN 2000. Lausanne, Suisse. 16-18 octobre 2000. vol. 1/1: pp. 373-382.
- [Blanchon, H. and Fais, L. 1996a] Blanchon, H. and Fais, L. (1996a) *How to ask Users About What they Mean: Two Experiments & Results*. Proc. MIDDIM'96. Le Col de Porte, Isère, France. August 12-14, 1996. vol. 1/1: pp. 238-259.
- [Blanchon, H. and Fais, L. 1996b] Blanchon, H. and Fais, L. (1996b) *Pilot Experiment on the Understandability of Interactive Disambiguation Dialogues*. Rap. ATR-ITL. Kyoto, Japan. Technical report. n° TR-IT-0177. July, 1996. 18 p.
- [Blanchon, H. and Fais, L. 1996c] Blanchon, H. and Fais, L. (1996c) *A Second Experiment on the Understandability of Interactive Disambiguation Dialogues*. Rap. ATR-ITL. Kyoto, Japan. Technical Report. n° TR-IT-0167. April, 1996. 22 p.
- [Blanchon, H. and Fais, L. 1997] Blanchon, H. and Fais, L. (1997) *Asking Users About What They Mean: Two Experiments & Results*. Proc. HCI'97. San Francisco, California. August 24-29, 1997. vol. 2/2: pp. 609-912.
- [Boguslavsky, I., et al. 2000] Boguslavsky, I., Frid, N., Iomdin, L., Kreidlin, L., Sagalova, I. and Sizov, V. (2000) *Creating a Universal Networking Language*

- module within an advanced NLP system*. Proc. COLING 2000. Saarbrücken, Germany. July 31 - August 4, 2000. vol. **1/2**: pp. 83-89.
- [Boitet, C. 1990a] Boitet, C. (1990a) *La TAO à Grenoble en 1990*. in École d'été de LANNION sur le TALN. CNET, juillet 1990. 65 p.
- [Boitet, C. 1990b] Boitet, C. (1990b) *Towards Personal MT: general design, dialogue structure, potential role of speech*. Proc. COLING-90. Helsinki, Finland. August 20-25, 1990. vol. **3/3**: pp. 30-35.
- [Boitet, C. 1993] Boitet, C. (1993) *Multimodal Interactive Disambiguation: first report on the MIDDIM project*. Rap. ATR Interpreting Telecommunications. Technical Report. n° TR-IT-0014. August 1993. 16 p.
- [Boitet, C. 1994] Boitet, C. (1994) *On the design of MIDDIM-DB, a data base of ambiguities and disambiguation methods*. Rap. ATR Interpreting Telecommunications. Technical Report. n° TR-IT-0072. September 1994. 14 p.
- [Boitet, C. 1997] Boitet, C. (1997) *GETA's methodology and its current development towards networking communication and speech translation in the context of the UNL and C-STAR projets*. Proc. PACLING-97. Ome, Japan. 2-5 September, 1997. vol. **1/1**: pp. 23-57.
- [Boitet, C. 2004] Boitet, C. (2004) *Progress report on building the French BTEC and participating in the MT evaluation campaign (CSTAR project)*. Rap. GETA-CLIPS. Internal Report. August 5, 2004. 10 p.
- [Boitet, C. and Blanchon, H. 1994] Boitet, C. and Blanchon, H. (1994) *Promesses et problèmes de la "TAO pour tous" après LIDIA-1, une première maquette*. in Langages, "Le traducteur et l'ordinateur". vol. **116, décembre 1994**: pp. 20-47.
- [Boitet, C. and Blanchon, H. 1995] Boitet, C. and Blanchon, H. (1995) *Multilingual Dialogue-Based MT for monolingual authors: the LIDIA project and a first mockup*. in Machine Translation. vol. **9(2)**: pp. 99-132.
- [Boitet, C., et al. 2000] Boitet, C., Blanchon, H. and Guilbaud, J.-P. (2000) *A way to integrate context processing in the MT component of spoken, task-oriented translation systems*. Proc. MSC-2000. Kyoto, Japan. October 11-13, 2000. vol. **1/1**: pp. 83-87.
- [Boitet, C. and Guilbaud, J.-P. 2000] Boitet, C. and Guilbaud, J.-P. (2000) *Analysis into a Formal Task-Oriented Pivot without Clear Abstract Semantics is Best Handled as "Usual" Translation*. Proc. ICSLP 2000. Beijing, China. Oct. 16-20, 2000. vol. **4/4**: pp. 436-439.
- [Boitet, C. and Nédobejkine, N. 1981] Boitet, C. and Nédobejkine, N. (1981) *Recent Developments in Russian-French Machine Translation at Grenoble*. in Linguistics. vol. **19**: pp. 199-271.
- [Boitet, C., et al. 1995] Boitet, C., Planas, E., Blanchon, H., Blanc, É., Guilbaud, J.-P., Guillaume, P., Lafourcade, M. and Sérasset, G. (1995) *LIDIA-1.2, une maquette de TAO personnelle multicible, utilisant la messagerie usuelle, la désambiguïsation interactive et la rétrotraduction*. Rap. GETA. Rapport final de la convention n° 95 2 93 0104 entre le Ministère de l'industrie et l'Université Joseph Fourier et réalisée par le GETA dans le cadre du Projet Eurêka EUROLANG. septembre 1995. 273 p.

- [Boitet, C. and Tomokio, M. 1995] Boitet, C. and Tomokio, M. (1995) *Ambiguities & ambiguity labelling: towards ambiguity databases*. Proc. RANLP'95 (Recent Advances in NLP). Tzigov Chark, Bulgaria. 14-16 September, 1995. vol. **1/1**: pp. 13-26.
- [Boitet, C. and Tomokiyo, M. 1995] Boitet, C. and Tomokiyo, M. (1995) *Towards ambiguity labelling for the study of interactive disambiguation methods*. Rap. ATR-ITL. Technical Report. n° TR-IT-0112. April 27, 1995. 22 p.
- [Brown, P. F., et al. 1990] Brown, P. F., Cocke, J., Della Pietra, S. A., Jelinek, F., Lafferty, J. D., Mercer, R. L. and Roosin, P. S. (1990) *A statistical approach to machine translation*. in Computational Linguistics. vol. **16**(2): pp. 79-85.
- [Brown, P. F., et al. 1992] Brown, P. F., Della Pietra, V. J., deSouza, P. V., Lai, J. C. and Mercer, R. L. (1992) *Class-Based N-gram Models of Natural Language*. in Computational Linguistics. vol. **18**(4): pp. 467-479.
- [Brown, R. D. 1989] Brown, R. D. (1989) *Augmentation*. in Machine Translation. vol. **4**: pp. 129-147.
- [Brown, R. D. 1998] Brown, R. D. (1998) *Improving Embedded Machine Translation with User Interaction*. Proc. AMTA Workshop on Embedded Machine Translation. Langhorne, Pennsylvania. October 28, 1998. 6 p.
- [Brown, R. D. and Nirenburg, S. 1990] Brown, R. D. and Nirenburg, S. (1990) *Human-Computer Interaction for Semantic Disambiguation*. Proc. COLING-90. Helsinki. August 20-25, 1990. vol. **3/3**: pp. 42-47.
- [Burger, S., et al. 2001] Burger, S., Besacier, L., Metze, F., Morel, C. and Coletti, P. (2001) *The NESPOLE! VoIP Dialog Database*. Proc. Eurospeech. Aalborg, Denmark. September 3-7, 2001. 4 p.
- [Campbell, N. 2004] Campbell, N. (2004) *Development in Corpus-Based Speech Synthesis: Approaching Natural Conversational Speech*. in IEICE Trans. - Special Issue on Corpus-Based Speech Technologies: 8 p. To be published.
- [Cao, W., et al. 2004] Cao, W., Zong, C. and Xu, B. (2004) *Approach to Interchange-Format Based Chinese Generation*. Proc. ICSLP 2004. Jeju Island, Korea. October 4-8, 2004. 4 p.
- [Carbonell, J. G., et al. 1981] Carbonell, J. G., Cullingford, R. E. and Gershman, A. V. (1981) *Steps Toward Knowledge-Based Machine Translation*. in IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. **3**(4): pp. 376-392.
- [Carbonell, J. G. and Tomita, M. 1985] Carbonell, J. G. and Tomita, M. (1985) *New Approaches to Machine Translation*. Proc. TMI-85. Colgate University, Hamilton, New York. 14-16 août 1985. vol. **1/1**: pp. 59-74.
- [Carbonell, J. G. and Tomita, M. 1987] Carbonell, J. G. and Tomita, M. (1987) *Knowledge-based machine translation, the CMU approach*. in Nirenburg, S. (ed.), Machine translation. Cambridge University Press. Cambridge. pp. 68-89.
- [Casacuberta, F., et al. 2004] Casacuberta, F., Ney, H., Och, F. J., Vidal, E., Vilar, J. M., Barrachina, S., García-Varéa, I., Llorens, D., Martínez, C., Molau, S., Nevado, F., Pastor, M., Picó, D., Sanchis, A. and Tillmann, C. (2004) *Some*

- approaches to statistical and finite-state speech-to-speech translation.* in Computer, Speech & Language. vol. **18**(1): pp. 25-47.
- [Casacuberta, F., *et al.* 2002] Casacuberta, F., Vidal, E. and Vilar, J. M. (2002) *Architectures for speech-to-speech translation using finite-state models.* Proc. ACL-02, Workshop on Speech-to-Speech Translation: Algorithms and Systems. University of Pennsylvania, Philadelphia, USA. July 7-12, 2002. pp. 39-34.
- [Cettolo, M., *et al.* 1999] Cettolo, M., Corazza, A., Lazzari, G., Pianesi, F., Pianta, E. and Tovenà, L. M. (1999) *A Speech-to-Speech Translation-based Interface for Tourism.* Proc. ENTER'99. Innsbruck, Austria. January 20-23, 1999. pp. 191-200.
- [Chandioux, J. 1988] Chandioux, J. (1988) *10 ans de METEO (MD).* Proc. Traduction Assistée par Ordinateur. Actes du séminaire international sur la TAO et dossiers complémentaires organisé par l'Observatoire Francophone des Industries de la Langue (OFIL). Paris, France. mars 1988. vol. **1/1**: pp. 169-173.
- [Chandler, B., *et al.* 1987] Chandler, B., Holden, N., Horsfall, H., Pollard, E. and McGee Wood, M. (1987) *N-Tran Final Report. Alvey Project.* Rap. CCL/UMIST, Manchester. n° 87/9. p.
- [Chanod, J.-P. 1993] Chanod, J.-P. (1993) *Problèmes de robustesse en analyse automatique.* Proc. Conférence Informatique et Langue Naturelle. Nantes, France. vol. **1/1**.
- [Church, K. W. and Hovy, E. 1993] Church, K. W. and Hovy, E. (1993) *Good Applications for Crummy Machine Translation.* in Machine Translation. vol. **8**(4): pp. 239-258.
- [Corazza, A. 1999] Corazza, A. (1999) *An Inter-Domain Portable Approach to Interchange Format Construction.* Proc. EUROSPEECH'99. Budapest, Hungary. September 5-9, 1999. vol. **6/6**: pp. 2419-2422.
- [Coughlin, D. 2003] Coughlin, D. (2003) *Correlating Automated and Human Assessments of Machine Translation Quality.* Proc. MT Summit IX. New Orleans, USA. September 23-27, 2003. 8 p.
- [Culy, C. and Riehemann, S. Z. 2003] Culy, C. and Riehemann, S. Z. (2003) *The Limits of N-Gram Translation Evaluation Metrics.* Proc. MT Summit IX. New Orleans, U.S.A. September 23-27, 2003. 8 p.
- [Damerau, F. J. 1964] Damerau, F. J. (1964) *A technique for computer detection and correction of spelling errors.* in Communication of the ACM. vol. **7**(3): pp. 171-176.
- [Davies, K. and al. 1999] Davies, K. and al. (1999) *The IBM conversational telephony system for financial applications.* Proc. EUROSPEECH'99. Budapest, Hungary. September 5-9, 1999. 4 p.
- [Dénos, N. 1993] Dénos, N. (1993) *Typage interactif de fragments de textes en "style d'énoncé" et "genres de textes".* Rap. Institut National Polytechnique de Grenoble (ENSIMAG) & Université Joseph Fourier (UFR IMA). Rapport de DEA. 22 juin 1993. 79 p.
- [Doan-Nguyen, H. 2003] Doan-Nguyen, H. (2003) *A Psycholinguistics-Based Parsing Model and Prototype.* Proc. 7th IASTED International Conference, Artificial

- Intelligence and Soft Computing, ASC 2003. Banff, Alberta, Canada. July 14-16, 2003. vol. **1/1**: pp. 170-175.
- [Doddington, G. 2002] Doddington, G. (2002) *Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics*. Proc. HLT 2002. San Diego, California. March 24-27, 2002. vol. **1/1**: pp. 128-132 (note book proceedings).
- [EAGLES-EWG 1996] EAGLES-EWG (1996) *EAGLES Evaluation of Natural Language Processing Systems*. Rap. Center for Sprogteknologi. Copenhagen, Denmark. Final Report EAG-EWG-PR.2, Project LRE-61-100. October, 1996. 287 p. Disponible à l'adresse: <http://www.issco.unige.ch/projects/ewg96/ewg96.html>.
- [EAGLES-EWG 1999] EAGLES-EWG (1999) *EAGLES Evaluation of Natural Language Processing Systems*. Rap. Center for Sprogteknologi. Copenhagen, Denmark. Final Report EAG-II-EWG-PR.2, Project LRE-61-100. April, 1999. 173 p. Disponible à l'adresse: <http://www.issco.unige.ch/projects/eagles/>.
- [Esteban, J., et al. 2004] Esteban, J., Lorenzo, J., Valderrábanos, A. and Lapalme, G. (2004) *TransType2 - An Innovative Computer-Assisted Translation System*. Proc. ACL 2004. Barcelona, Spain. July 21-26, 2004. 4 p.
- [Fafiotte, G. 2004] Fafiotte, G. (2004) *Interprétariat à distance et collecte de dialogues spontanés bilingues, sur une plate-forme générique multifonctionnelle*. Proc. TALN 2004. Fès, Maroc. 19-21 avril 2004. vol. **1/1**: pp. 173-182.
- [Fafiotte, G. and Boitet, C. 1994] Fafiotte, G. and Boitet, C. (1994) *Report on first EMMI Experiments for the MIDDIM project in the context of Interpreting Telecommunications*. Rap. ATR-ITL. Technical Report. n° TR-IT-0074. September 1994. 11 p.
- [Fafiotte, G. and Boitet, C. 1996] Fafiotte, G. and Boitet, C. (1996) *An Analysis of the First EMMI-Based Experiments on Interactive Disambiguation in the Context of Automated Interpreting Telecommunications*. Proc. MIDDIM'96. Le Col de Porte, Isère, France. August 12-14, 1996. vol. **1/1**: pp. 224-237.
- [Fais, L. and Blanchon, H. 1996] Fais, L. and Blanchon, H. (1996) *Ambiguities in Task-Oriented Dialogues*. Proc. MIDDIM'96. Le col de Porte, Isère, France. August 12-14, 1996. vol. **1/1**: pp. 263-275.
- [Fais, L. and Loken-Kim, K.-H. 1994] Fais, L. and Loken-Kim, K.-H. (1994) *Effects of communicative mode on spontaneous English speech*. Rap. Institute of Electronics, Information and Communication Engineers. Technical Report. n° NLC94-22. Oct. 94. p.
- [Fais, L. and Loken-Kim, K.-H. 1995] Fais, L. and Loken-Kim, K.-H. (1995) *Lexical Accommodation in Human-interpreted and Machine-interpreted Dual Language Interactions*. Proc. ESCA Workshop on Spoken Dialogue Systems. Visgo, Denmark. May 30-June 2, 1995. vol. **1/1**: pp. 69-72.
- [Fais, L., et al. 1995] Fais, L., Loken-Kim, K.-H. and Park, Y.-D. (1995) *Speakers' responses to requests for repetition in a multimedia cooperative dialogue*. Proc. International Conference on Cooperative Multimodal Communication. Eindhoven, The Netherlands. May 24-26, 1995. vol. **1/1**: pp. 129-144.

- [Falaise, A. 2004] Falaise, A. (2004) *Premier pas vers une TA interactive pour le tchat*. Rap. INPG. Grenoble. Rapport de stage de seconde année de master recherche Information, Cognition, Apprentissages, option sciences cognitives; effectué sous la direction de Hervé Blanchon et Christian Boitet. 11 juin 2004. 62 p.
- [Fisher, W. M. and Fiscus, J. G. 1993] Fisher, W. M. and Fiscus, J. G. (1993) *Better alignment procedure for speech recognition evaluation*. Proc. ICASSP-93. Minneapolis, USA. April 27-30; 1993. vol. **2/5**: pp. 59-62.
- [Foster, G., et al. 1997] Foster, G., Isabelle, P. and Plamondon, P. (1997) *Target Text Mediated Interactive Machine Translation*. in Machine Translation. vol. **12**(1-2): pp. 175-194.
- [Frederking, R., et al. 1997] Frederking, R., Rudnicky, A. and Hogan, C. (1997) *Interactive Speech Translation in the DIPLOMAT Project*. Proc. ACL-97, Spoken Language Translation Workshop. Madrid, Spain. 6 p.
- [Frederking, R., et al. 2000] Frederking, R., Rudnicky, A., Hogan, C. and Lenzo, K. (2000) *Interactive Speech Translation in the DIPLOMAT Project*. in Machine Translation. vol. **15**(1-2): pp. 27-42.
- [Frey, V. 1996] Frey, V. (1996) *De la reconnaissance à la découverte de "patrons d'ambiguïtés" en TAFD*. Rap. Université Joseph Fourier (MATIS - Archamps). Rapport de DEA. 24 juin 1996. 67 p.
- [Fuchs, C. 1996] Fuchs, C. (1996) *Les ambiguïtés du français*. Ophris. Paris. 184 p.
- [Gadet, F. 1989] Gadet, F. (1989) *Le français ordinaire*. Armand Collin. Paris. 153 p.
- [Gala Pavia, N. 2003] Gala Pavia, N. (2003) *Un modèle d'analyseur syntaxique robuste fondé sur le modularité et la lexicalisation de ses grammaires*. Thèse. Université Paris XI, Orsay. 2003. 247 p.
- [Gao, Y., et al. 2004] Gao, Y., Zhou, B., Diao, Z., Sorensen, J. and Picheny, M. (2004) *Mars: A Statistical semantic parsing and generation based multilingual automatic translation system*. in Machine Translation. vol. **To Appear**.
- [Ghorayeb, A. 2002] Ghorayeb, A. (2002) *Mémoire de désambiguïstation et documents auto-explicatifs*. Rap. Université Joseph Fourier (Grenoble 1). Rapport de DEA Informatique: Systèmes d'Information. juin 2002. 83 p.
- [Goddeau, D., et al. 1994] Goddeau, D., Brill, E., Glass, J., Pao, C., Philips, M., Polifroni, J., Seneff, S. and Zue, V. (1994) *GALAXY: a Human-Language Interface to On-Line Travel Information*. Proc. ICSLP 94. Yokohama, Japan. September 18-22, 1994. vol. **2/4**: pp. 707-710.
- [Good, R. L. 1988] Good, R. L. (1988) *Automated Lookup: AutoTerm of ALP System*. in Vasconcellos, M. (ed.), *Technology as Translation Strategy*. State University of New York at Binghamton. Binghamton. pp. 87-91.
- [Goodman, K. and Nirenburg, S. 1991] Goodman, K. and Nirenburg, S. ed., (1991) *The KBMT Project: A case study in knowledge-based machine translation*. Morgan Kaufmann. San Mateo, California. 331 p.
- [Gorin, A. L., et al. 1997] Gorin, A. L., Riccardi, G. and Wright, J. H. (1997) *How may I Help You?* in Speech Communication. vol. **23**: pp. 113-127.

- [Goulian, J. 2002] Goulian, J. (2002) *Stratégie d'analyse détaillée pour la compréhension automatique robuste de la parole*. Thèse. Université de Bretagne Sud. 2002. 200 p.
- [Gressard, G. 2001] Gressard, G. (2001) *Une nouvelle architecture distribuée pour LIDIA en Java*. Rap. Université Pierre Mendès France (Grenoble 2), IUT 2, Département informatique. Rapport de stage. Juin 2001. 62 p.
- [Gressard, G. and Delaval, G. 2003] Gressard, G. and Delaval, G. (2003) *Une interface d'accès à un système de traduction automatisée fondée sur le dialogue: cahier des charges et document de spécification externe*. Rap. ENSIMAG. Rapport de projet en Interface Homme-Machine. mars 2003. 47 p.
- [Gu, L. and Gao, Y. 2003] Gu, L. and Gao, Y. (2003) *Improving Statistical natural concept generation in interlingua-based speech-to-speech translation*. Proc. EUROSPEECH 2003. Geneva, Switzerland. September 1-4, 2003. 4 p.
- [Gu, L. and Gao, Y. 2004] Gu, L. and Gao, Y. (2004) *On Feature Selection in Maximum Entropy Approach to Statistical Approach to Statistical Concept-based Speech-to-Speech Translation*. Proc. IWSLT 2004 (ICLSP 2004 Satellite Workshop). Kyoto, Japan. September 30-October 1, 2004. vol. 1/1: pp. 115-121.
- [Habert, B. 2000] Habert, B. (2000) *Des corpus représentatifs: de quoi, pour quoi, comment?* in Bilger, M. (ed.), *Linguistique sur corpus. Étude et réflexions*. Presses Universitaires de Perpignan. Perpignan. pp. 11-58.
- [Hajlaoui, N. and Boitet, C. 2004] Hajlaoui, N. and Boitet, C. (2004) *PolyphraZ: a tool for the quantitative and subjective evaluation of parallel corpora*. Proc. IWSLT 2004 (ICLSP 2004 Satellite Workshop). Kyoto, Japan. September 30-October 1, 2004. vol. 1/1: pp. 123-129.
- [Hardy, H., et al. 2003] Hardy, H., Strzalkowski, T. and Wu, M. (2003) *Dialogue Management for an Automated Multilingual Call Center*. Proc. HLT-NAACL 2003: Workshop on Research Directions in Dialogue Processing. Edmonton, Canada. May 31, 2003. 3 p.
- [Hetherington, L. 2004a] Hetherington, L. (2004a) *The MIT Finite-State Transducer Toolkit for Speech and Language Processing*. Proc. ICSLP 2004. Jeju Island, Korea. October 4-8, 2004. 4 p.
- [Hetherington, L. 2004b] Hetherington, L. (2004b) *The MIT Finite-State Transducer Toolkit for Speech and Language Processing*. <http://www.sls.csail.mit.edu/ilh/fst/>.
- [Hiraoka, S., et al. 1997] Hiraoka, S., Hoshimi, M., Matsui, K. and Junqua, J.-C. (1997) *An Experimental Bidirectional Japanese/English Interpreting Video Phone System Using Internet*. Proc. ICASSP'97. Munich, Germany. April 21-24, 1997. vol. 1/5: pp. 115-118.
- [Hirst, G. 1987] Hirst, G. (1987) *Semantic interpretation and the resolution of ambiguity*. Cambridge University Press. Cambridge. 263 p.
- [Hirst, G. 1988] Hirst, G. (1988) *Resolving Lexical Ambiguity Computationally with Spreading Activation and Polaroid Words*. in Small, S. I., Cottrell, G., W and Tanenhaus, M. K. (ed.), *Lexical Ambiguity Resolution*. Morgan Kaufmann Publishers, Inc. San Mateo, California, USA. pp. 73-107.

- [Hovy, E., *et al.* 2002] Hovy, E., King, M. and Popescu-Belis, A. (2002) *Principles of Context-Based Machine Translation Evaluation*. in *Machine Translation*. vol. 17(1): pp. 43-75.
- [Huang, X. 1990a] Huang, X. (1990a) *Machine Translation in a Monolingual Environment*. Proc. TMI-90. University of Texas at Austin. pp. 33-42.
- [Huang, X. 1990b] Huang, X. (1990b) *A Machine Translation System for the Target Language Inexpert*. Proc. COLING-90. Helsinki. August 20-25, 1990. vol. 3/3: pp. 364-367.
- [Hutchins, J. 2003] Hutchins, J. (2003) *ALPAC: the (in)famous report*. in Nirenburg, S., Somers, H. and Wilks, Y. (ed.), *Readings in machine translation*. The MIT Press. Cambridge, Mass. pp. 131-135.
- [Hutchins, W. J. 1986] Hutchins, W. J. (1986) *Machine Translation: past, present, future*. Ellis Horwood Limited. Chichester, England. 382 p.
- [Iida, H., *et al.* 1996] Iida, H., Sumita, E. and Furuse, O. (1996) *Spoken-Language Translation Method using Examples*. Proc. COLING'96. Copenhagen, Denmark. August 5-9, 1996. vol. 2/2: pp. 1074-1077.
- [Ikeda, T., *et al.* 2002] Ikeda, T., Ando, S.-i., Satoh, K., Okumura, A. and Watanabe, T. (2002) *Automatic Interpretation System Integrating Free-style Sentence Translation and Parallel Text Based Translation*. Proc. ACL-02, Workshop on Speech-to-Speech Translation: Algorithms and Systems. University of Pennsylvania, Philadelphia, USA. July 7-12, 2002. pp. 85-92.
- [Irgadian, F. 2002] Irgadian, F. (2002) *Traduction Interactive Fondée sur le Dialogue*. Rap. Université Stendhal. Grenoble, France. Rapport de stage de Maîtrise en Industries de la Langue. 20 juin 2002. 109 p.
- [ISO/IEC 2001] ISO/IEC (2001) *Software engineering – Product Quality – Part 1*. Rap. International Organisation for Standardization & International Electrotechnical Commission. Geneva, Switzerland. June, 2001. 25 p.
- [Isotani, R., *et al.* 2002] Isotani, R., Yamabana, K., Ando, S.-i., Hanazawa, K., Ishikawa, S.-y., Emori, T., Iso, K.-i., Hattori, H., Okumura, A. and Watanabe, T. (2002) *An Automatic Speech Translation System on PDA for Travel Conversation*. Proc. ICMI-2002. Pittsburgh, PA, USA. October 2002. vol. 1/1: pp. 211-216.
- [JEIDA 1989] JEIDA (1989) *A Japanese view of machine translation in light of the considerations and recommendations reported by ALPAC, U.S.A.* Rap. Japan Electronic Industry Development Association. Tokyo, Japan. July, 1989. 197 p.
- [JEIDA 1992] JEIDA (1992) *JEIDA Methodology and Criteria on Machine Translation Evaluation*. Rap. Japan Electronic Industry Development Association. Tokyo, Japan. November, 1992. 129 p.
- [Johnson, R. and Whitelock, P. 1985] Johnson, R. and Whitelock, P. (1985) *Machine Translation as an Expert Task*. Proc. Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages. Colgate University, Hamilton, New York. 14-16 août 1985. pp. 59-74.

- [Jones, D. and Tsujii, J. 1990] Jones, D. and Tsujii, J. (1990) *Interactive High-quality machine translation for monolinguals*. Proc. Third International Conference on Methodological Issues in Machine Translation of Natural Language. pp. 43-46.
- [Junqua, J.-C. and Haton, J.-P. 1996] Junqua, J.-C. and Haton, J.-P. (1996) *Robustness in Automatic Speech Recognition - Fundamentals and Applications*. Kluwer Academic Publishers. London. 440 p.
- [Jurafsky, D. and Martin, J. H. 2000a] Jurafsky, D. and Martin, J. H. (2000a) *Lexicalized and Probabilistic Parsing*. in *Speech and Language Processing: An Introduction to Natural language Processing, Computational Linguistics, and Speech Recognition*. Prentice-Hall, Inc. Upper Saddle River, New Jersey, USA. pp. 447-476.
- [Jurafsky, D. and Martin, J. H. 2000b] Jurafsky, D. and Martin, J. H. (2000b) *Word Sense Disambiguation and Information Retrieval*. in *Speech and Language Processing: An Introduction to Natural language Processing, Computational Linguistics, and Speech Recognition*. Prentice-Hall, Inc. Upper Saddle River, New Jersey, USA. pp. 631-666.
- [Karaorman, M., et al. 1996] Karaorman, M., Applebaum, T. H., Itoh, T., Endo, M., Ohno, Y., Hoshimi, M., Kamai, T., Matsui, K., Hata, K., Pearson, S. and Janqua, J.-C. (1996) *An Experimental Japanese/English Interpreting Video Phone System*. Proc. ICSLP. Philadelphia, PA. vol. 3: pp. 1676-1679.
- [Kauers, M., et al. 2002] Kauers, M., Vogel, S., Fügen, C. and Waibel, A. (2002) *Interlingua-Based Statistical Machine Translation*. Proc. ICSLP 2002. Denver, USA. 16-20 September, 2002. 4 p.
- [Kay, M. 1973] Kay, M. (1973) *The MIND system*. in Rustin, R. (ed.), *Courant Computer Science Symposium 8: Natural Language Processing*. Algorithmics Press, Inc. New York. pp. 155-188.
- [Kay, M. 1980] Kay, M. (1980) *The Proper Place of Men and Machines in Language Translation*. Rap. Xerox, Palo Alto Research Center. Research Report. n° CSL-80-11. octobre 1980. 20 p.
- [Keller, E. and Werner, S. 1997] Keller, E. and Werner, S. (1997) *Automatic Intonation Extraction and Generation for French*. Proc. 14th CALICO Annual Symposium. West Point, NY, USA. September 1997.
- [Keller, E. and Zellner, B. 1998] Keller, E. and Zellner, B. (1998) *Motivations for the prosodic predictive chain*. Proc. ESCA Symposium on Speech Synthesis. Jenolan Caves, Australia. November 28-29, 1998. vol. 1/1: pp. 137-141.
- [Kikiu, G., et al. 2004] Kikiu, G., Takezawa, T. and Yamamoto, S. (2004) *Multilingual Corpora for Speech-to-Speech Translation Research*. Proc. ICSLP 2004. Jeju Island, Korea. October 4-8, 2004. 4 p.
- [King, M. 1984] King, M. (1984) *When is the next ALPAC report due?* Proc. COLING-84. Stanford University, California. 2-6 July, 1984. vol. 1/1: pp. 352-353.
- [King, M. 1996] King, M. (1996) *Evaluating Natural Language Processing Systems*. in *Communication of the ACM*. vol. 29(1): pp. 73-79.

- [King, M. 2003] King, M. (2003) *Living up to standards*. Proc. EACL 2003 – Workshop on Evaluation Initiatives in Natural Language Processing: are evaluation methods, metrics and ressources reusable? Budapest, Hungary. April 14, 2003. vol. **1/1**: 8 p.
- [Kitamura, M. and Murata, T. 2003] Kitamura, M. and Murata, T. (2003) *Practical Machine Translation System allowing Complex Patterns*. Proc. MT Summit IX. New Orleans, U.S.A. September 23-27, 2003. 8 p.
- [Kneser, W. and Ney, H. 1993] Kneser, W. and Ney, H. (1993) *Improved Clustering Techniques for Class-Based Statistical Language Modeling*. Proc. EUROSPEECH'93. Berlin, Germany. 21-23 September 1993. vol. **2/3**: pp. 973-976.
- [Koo, M.-W., et al. 1995] Koo, M.-W., Sohn, I.-H., Kim, W.-S. and Chang, D.-S. (1995) *KT-STTS: a Speech Translation System for Hotel Reservation and a Continuous Speech Recognition System for Speech Translation*. Proc. EUROSPEECH'95. Madrid, Spain. September 18-21, 1995. vol. **2/3**: pp. 1227-1230.
- [Kurdi, M. Z. 2003] Kurdi, M. Z. (2003) *Contribution à l'analyse du langage oral*. Thèse. Université Joseph Fourier - Grenoble 1. 2003. 379 p.
- [L'haire, S., et al. 2000] L'haire, S., Mengon, J. and Laenzlinger, C. (2000) *Outils génériques de transfert hybride pour la traduction automatique sur Internet*. Proc. TALN 2000. Lausanne, Suisse. 16-18 octobre 2000. vol. **1/1**: pp. 253-262.
- [Laenzlinger, C. and Wehrli, É. 1991] Laenzlinger, C. and Wehrli, É. (1991) *Un analyseur syntaxique pour grammaires lexicales-fonctionnelles*. in T.A. Information. vol. **32**(2): pp. 35-49.
- [Lafourcade, M., et al. 2002] Lafourcade, M., Prince, V. and Schwab, D. (2002) *Vecteurs conceptuels et structuration émergente de terminologies*. in TAL. vol. **43**(1): pp. 43-72.
- [Lafourcade, M. and Standford, E. 1999] Lafourcade, M. and Standford, E. (1999) *Analyse et désambiguïsation lexicale par vecteurs sémantiques*. Proc. TALN'1999. Cargèse, France. 12-17 juillet 1999. vol. **1/1**: pp. 351-356.
- [Langlais, P., et al. 2004] Langlais, P., Carl, M. and Streiter, O. (2004) *Experimenting with Phrase-Based Statistical Translation within the IWSLT 2004 Chinese to English Shared Translation Task*. Proc. IWSLT 2004 (ICLSP 2004 Satellite Workshop). Kyoto, Japan. September 30-October 1, 2004. vol. **1/1**: pp. 31-38.
- [Langlais, P., et al. 2000a] Langlais, P., Foster, G. and Lapalme, G. (2000a) *Unit Completion for a Computer-Aided Translation Typing System*. in Machine Translation. vol. **15**(4): pp. 267-294.
- [Langlais, P., et al. 2001] Langlais, P., Lapalme, G. and Sauvé, S. (2001) *User Interface Aspects of a Translation Typing System*. in Lecture Notes in Computer Science. vol. **2056**: pp. 246-256.
- [Langlais, P., et al. 2000b] Langlais, P., Sauvé, S., Foster, G., Macklovitch, E. and Lapalme, G. (2000b) *Evaluation of TransType, a Computer-aided Translation Typing System*. Proc. LREC. Athens, Greece. 31 May - 2 June 2000. vol. **2/3**: 8 p.

- [Lavie, A., *et al.* 1997] Lavie, A., Waibel, A., Levin, L., Finke, M., Gates, D., Gavalda, M., Zeppenfeld, T. and Zhan, P. (1997) *JANUS-III: Speech-to-Speech Translation in Multiple Languages*. Proc. ICASSP'97. Munich, Germany. April 21-24, 1997. vol. **1/5**: pp. 99-102.
- [Lazzari, G. 1999] Lazzari, G. (1999) *Robust Spoken Translation at ITC-IRST*. Proc. MT Summit VII. Singapore. September 13-17, 1999. pp. 114-120.
- [Lazzari, G., *et al.* 2004] Lazzari, G., Waibel, A. and Zong, C. (2004) *Wordwide Ongoing Activities on Multilingual Speech to Speech Translation*. Proc. ICSLP 2004. Jeju Island, Korea. October 4-8, 2004. 4 p.
- [Leavitt, J., *et al.* 1994] Leavitt, J., Lonsdale, D. and Franz, A. (1994) *A reasoned interlingua for knowledge-based machine translation*. Proc. Canadian Artificial Intelligence Conference. Banff, Alberta, Canada. May 16-20, 1994. 10 p.
- [Lee, S.-J., *et al.* 2001] Lee, S.-J., Lee, Y. and Yang, J.-W. (2001) *Introduction of ETRI automatic translation system for traveler aid*. Proc. ICSP-01. 4 p.
- [Lee, Y., *et al.* 1995] Lee, Y., Kim, Y.-S., Lee, J.-C., Ryoo, J.-H. and Yang, J.-W. (1995) *Korean-Japanese Speech Translation System for Hotel Reservation - Korean Front Desk Side -*. Proc. EUROSPEECH'95. Madrid, Spain. September 18-21, 1995. vol. **2/3**: pp. 1197-1200.
- [Lee, Y.-S. and Roukos, S. 2004] Lee, Y.-S. and Roukos, S. (2004) *IBM Spoken Language Translation System Evaluation*. Proc. IWSLT 2004 (ICLSP 2004 Satellite Workshop). Kyoto, Japan. September 30-October 1, 2004. vol. **1/1**: pp. 39-46.
- [Lehiste, I. 1973] Lehiste, I. (1973) *Phonetic disambiguation of syntactic ambiguity*. in *Glossa*. vol. **7**(2): pp. 107-122.
- [Lehiste, I., *et al.* 1976] Lehiste, I., Olive, J. P. and Streeter, L. A. (1976) *Role of duration in disambiguating syntactically ambiguous sentences*. in *Journal of the Acoustical Society of America*. vol. **60**(5): pp. 1199-1202.
- [Lepage, Y. 2001] Lepage, Y. (2001) *Défense et illustration de l'analogie*. Proc. TALN 2001. Tours, France. 2-5 juillet 2001. vol. **1/1**: pp. 373-378.
- [Lepage, Y. 2003] Lepage, Y. (2003) *De l'analogie rendant compte de la commutation en linguistique*. Mémoire d'habilitation à diriger des recherches. Université Joseph Fourier (Grenoble 1). 2003. 388 p.
- [Lepage, Y. 2004] Lepage, Y. (2004) *Lower and higher estimates of the number of "true analogies" between sentences contained in a large multilingual corpus*. Proc. COLING 2004. Geneva, Switzerland. August 23-27, 2004. vol. **1/2**: pp. 736-742.
- [Leusch, G., *et al.* 2003] Leusch, G., Ueffing, N. and Ney, H. (2003) *A Novel String-to-String Distance Measure with Application to Machine Translation Evaluation*. Proc. MT Summit IX. New Orleans, U.S.A. September 23-27, 2003. 8 p.
- [Levenshtein, V. I. 1966] Levenshtein, V. I. (1966) *Binary codes capable of correcting deletion, insertions and reversals*. in *Soviet Physics Doklady*. vol. **10**(8): pp. 707-710.

- [Levin, L., *et al.* 2000a] Levin, L., Gates, D., Lavie, A., Pianesi, F., Wallace, D., Watanabe, T. and Woszczyna, M. (2000a) *Evaluation of a Practical Interlingua for Task-Oriented Dialogue*. Proc. ANLP/NAACL 2000 Workshop - Applied Interlinguas: Practical Applications of Interlingual Approaches to NLP. Seattle, Washington, USA. April 30, 2000. vol. **1/1**: pp. 18-23.
- [Levin, L., *et al.* 1998] Levin, L., Gates, D., Lavie, A. and Waibel, A. (1998) *An Interlingua Based on Domaine Actions for Machine Translation of Task-Oriented Dialogues*. Proc. ICSLP'98. Sydney, Australia. 30th November - 4th December 1998. vol. **4/7**: pp. 1155-1158.
- [Levin, L., *et al.* 2003] Levin, L., Gates, D., Wallace, D., Peterson, K., Pianta, E. and Mana, N. (2003) *The NESPOLE! Interchange Format*. Rap. ITC-irst, UKA, CMU, UJF. Nespole! (IST-1999-1156) Public Report. n° D13. February 24, 2003. 62 p.
- [Levin, L., *et al.* 2000b] Levin, L., Lavie, A., Woszczyna, M., Gates, D., Gavaldà, M., Koll, D. and Waibel, A. (2000b) *The JANUS III Translation System: Speech-to-Speech Translation in Multiple Domains*. in Machine Translation. vol. **15**(1-2): pp. 3-25.
- [Lin, C.-Y. 2004a] Lin, C.-Y. (2004a) *Looking for a few Good Metrics: ROUGE and its Evaluation*. Proc. NTCIR Workshop. Tokyo, Japan. June 2-4, 2004. vol. **1/1**: 8 p.
- [Lin, C.-Y. 2004b] Lin, C.-Y. (2004b) *ROUGE: A Package for Automatic Evaluation of Summaries*. Proc. ACL 2004. Barcelona, Spain. July 21-26, 2004. 8 p.
- [Lin, C.-Y. and Och, F. J. 2004a] Lin, C.-Y. and Och, F. J. (2004a) *Automatic Evaluation of Machine Translation Quality Using Longest Common Subsequence and Skip-Bigram Statistics*. Proc. ACL 2004. Barcelona, Spain. July 21-26, 2004. 8 p.
- [Lin, C.-Y. and Och, F. J. 2004b] Lin, C.-Y. and Och, F. J. (2004b) *ORANGE: a Method for Evaluation Automatic Evaluation Metrics for Machine Translation*. Proc. COLING 2004. Geneva, Switzerland. August 23-27, 2004. vol. **1/2**: pp. 501-507.
- [Lin, D. 1994] Lin, D. (1994) *PRINCIPAR: an efficient, broad coverage, principle-based parser*. Proc. COLING-94. Kyoto, Japan. August 5-9, 1994. vol. **1/2**: pp. 482-488.
- [Loffler-Laurian, A.-M. 1983] Loffler-Laurian, A.-M. (1983) *Pour une typologie des erreurs dans la traduction automatique*. in Multilingua. vol. **2**(2): pp. 65-78.
- [Loken-Kim, K.-H., *et al.* 1993] Loken-Kim, K.-H., Yato, F., Kurihara, K., Fais, L. and Furukawa, R. (1993) *EMMI-ATR environment for multi-modal interactions*. Rap. ATR-ITL. Technical Report. n° TR-IT-0018. Sept 30, 1993. 28 p.
- [Lowrence, R. and Wagner, C.-K. 1975] Lowrence, R. and Wagner, C.-K. (1975) *An Extension of the String-to-String Correction Problem*. in Journal of the ACM. vol. **22**(2): pp. 177-183.
- [Mana, N., *et al.* 2003] Mana, N., Burger, S., Cattoni, R., Besacier, L., Maclaren, V., Mc Donough, J. and Metze, F. (2003) *The Nespole! VoIP Corpora in Tourism*

- and Medical Domains*. Proc. EUROSPEECH 2003. Geneva, Switzerland. September 1-4, 2003. 4 p.
- [Maruyama, H., *et al.* 1990] Maruyama, H., Watanabe, H. and Ogino, S. (1990) *An Interactive Japanese Parser for Machine Translation*. Proc. COLING-90. Helsinki. August 20-25, 1990. vol. **2/3**: pp. 257-262.
- [Maruyama, N., *et al.* 1988] Maruyama, N., Morohashi, M., Umeda, S. and Sumita, E. (1988) *A Japanese sentence analyser*. in IBM Journal of Research and Development. vol. **32**(2): pp. 238-250.
- [Matsui, K., *et al.* 2001] Matsui, K., Wakita, Y., Konuma, T., Mizutani, K., Endo, M. and Murata, M. (2001) *An Experimental Multilingual Speech Translation System*. in Lecture Notes in Computer Science. vol. **2116**: pp. 248-251.
- [Mazenot, S. 2002] Mazenot, S. (2002) *Projet NESPOLE! – IF vers français*. Rap. CLIPS-GETA. Rapport interne projet NESPOLE! 28 février 2002. 38 p.
- [McCord, M. 1989a] McCord, M. (1989a) *Design of LMT: A Prolog-Based Machine Translation System*. in Computational Linguistics. vol. **15**(1): pp. 33-52.
- [McCord, M. 1989b] McCord, M. (1989b) *LMT*. Proc. MT SUMMIT II. Munich. August 16-18, 1989. vol. **1/1**: 111-116.
- [McTear, M. F. 2002] McTear, M. F. (2002) *Spoken Dialogue Technology: Enabling the Conversational User Interface*. in ACM Computing Survey. vol. **34**(1): 90-169.
- [Melby, A. K. 1978] Melby, A. K. (1978) *Design and Implementation of a Computer-Assisted Translation System*. Proc. COLING-78. Bergen, Norway. August 14-18, 1978. vol. **1/1**.
- [Melby, A. K. 1981] Melby, A. K. (1981) *Translators and Machines - Can they cooperate?* in META. vol. **26**(1): 23-34.
- [Melby, A. K. 1982] Melby, A. K. (1982) *Multi-Level Translation Aids in a Distributed System*. Proc. COLING-82. Prague. 5-10 juillet 1982. vol. **1/1**: 215-220.
- [Melby, A. K. 1987] Melby, A. K. (1987) *On human-machine interaction in translation*. in Nirenburg, S. (ed.), *Machine Translation*. Cambridge University Press. Cambridge. 145-154.
- [Melby, A. K. 1991] Melby, A. K. (1991) *Pour le traducteur: un poste de travail à trois niveaux d'assistance*. Proc. L'environnement traductionnel; La station de travail du traducteur de l'an 2001. 25-27 avril 1991. vol. **1/1**: 151-153.
- [Melby, A. K., *et al.* 1980] Melby, A. K., Smith, M. R. and Peterson, J. (1980) *ITS: An Interactive Translation System*. Proc. COLING-80. Tokyo. 30 septembre-4 octobre 1980. vol. **1/1**: 424-429.
- [Merlo, P. 1995] Merlo, P. (1995) *Modularity and information content classes in principle-based parsing*. in Computational Linguistics. vol. **21**(4): 515-541.
- [Mitamura, T. 1991] Mitamura, T. (1991) *An Efficient Interlingua Translation system for Multi-lingual Document Production*. Proc. Machine Translation Summit III. Washington, D.C. 1-4 juillet 1991. vol. **1/1**: 55-61.

- [Mitamura, T., *et al.* 2003a] Mitamura, T., Baker, K., Nyberg, E. and Svoboda, D. (2003a) *Diagnostics for Interactive Controlled Language Checking*. Proc. EAMT/CLAW 2003. Dublin City University, Ireland. May 15-17, 2003. pp. 8 p.
- [Mitamura, T., *et al.* 2003b] Mitamura, T., Baker, K., Svoboda, D. and Nyberg, E. (2003b) *Source Language Diagnostics for MT*. Proc. MT Summit IX. New Orleans, U.S.A. September 23-27, 2003. pp. 7 p.
- [Mitamura, T., *et al.* 1999] Mitamura, T., Nyberg, E., Torrejon, E. and Igo, R. (1999) *Multiple Strategies for Automatic Disambiguation in Technical Translation*. Proc. TMI-99. University College, Chester, England. August 23-25, 1999. vol. **1/1**: 218-227.
- [Mohri, M. 1997] Mohri, M. (1997) *Finite-state transducers in language and speech processing*. in Computational Linguistics. vol. **23**(2): 269-312.
- [Mohri, M., *et al.* 2003] Mohri, M., Pereira, F. C. N. and Riley, M. D. (2003) *The AT&T FSM library: Finite-state machine library*. <http://www.research.att.com/sw/tools/fsm/>.
- [Morel, M.-A. 2004] Morel, M.-A. (2004) *Intonation, Regard et Genres dans le dialogue à bâtons rompus*. in Langages. vol. **153 (Les genres de la parole)**: 15-27.
- [Morimoto, T., *et al.* 1992] Morimoto, T., Suzuki, M., Takezawa, T., Kikui, G. i., Nagata, M. and Tomokio, M. (1992) *A Spoken Language Translation System: SL-TRANS2*. Proc. COLING-92. Nantes, France. 23-28 juillet 1992. vol. **3/4**: 1048-1052.
- [Morimoto, T., *et al.* 1993] Morimoto, T., Takezawa, T., Yato, F., Sagayama, S., Tashiro, T., Nagata, N. and Kurematsu, A. (1993) *ATR's Speech Translation System: Asura*. Proc. EUROSPEECH'93. Berlin, Germany. September 21-23, 1993. vol. **2/3**: 1291-1294.
- [Morishima, S. and Nakamura, S. 2002] Morishima, S. and Nakamura, S. (2002) *Multi-modal Translation System and its Evaluation*. Proc. ICMI-2002. Pittsburgh, PA, USA. October 2002. vol. **1/1**: 223-228.
- [Nagao, M. and Tsujii, J.-I. 1986] Nagao, M. and Tsujii, J.-I. (1986) *The Transfert Phase of the MU Machine Translation System*. Proc. COLING-86. Bonn. 25-29 août 1986. vol. **1/1**: 97-103.
- [Ney, H. 2004] Ney, H. (2004) *Statistical Machine Translation and its Challenges*. Proc. ICSLP 2004. Jeju Island, Korea. October 4-8, 2004. pp. 4 p.
- [Nießen, S., *et al.* 2000] Nießen, S., Och, F. J., Leusch, G. and Ney, H. (2000) *An Evaluation Tool for Machine Translation: Fast Evaluation for MT Research*. Proc. LREC 2000. Athens, Greece. 31 May - 2 June 2000. vol. **1/3**: 39-45.
- [Nirenburg, S. 1989] Nirenburg, S. (1989) *Knowledge-based Machine Translation*. in Machine Translation. vol. **4**: 5-24.
- [Nirenburg, S. and al. 1989] Nirenburg, S. and al. (1989) *KBMT-89 Project Report*. Rap. Center for Machine Translation, Carnegie Mellon University, Pittsburg. avril 1989. 286 p.

- [Norvig, P. 1992] Norvig, P. (1992) *Paradigms of Artificial Intelligence Programming: Case Studies in Common Lisp*. Morgan Kaufmann Publishers. San Mateo, California. 945 p.
- [Nyberg, E. and Mitamura, T. 1992] Nyberg, E. and Mitamura, T. (1992) *The KANT System: Fast, Accurate, High-Quality Translation in Practical Domains*. Proc. COLING-92. Nantes, France. July 23-28, 1992. vol. **3/4**: pp. 1069-1073.
- [Nyberg, E. and Mitamura, T. 2000] Nyberg, E. and Mitamura, T. (2000) *The KANTOO Machine Translation Environment*. Proc. AMTA-2000. Cuernavaca, Mexico. October 10-14, 2000. 5 p.
- [O'Shaughnessy, D. 1989] O'Shaughnessy, D. (1989) *Specifying accent marks in French text for teletext and speech synthesis*. in International Journal of Man-Machine Studies. vol. **31**(4): pp. 405-414.
- [Och, F. J. and Ney, H. 2002] Och, F. J. and Ney, H. (2002) *Discriminative Training and Maximum Entropy Models for Statistical Machine Translation*. Proc. ACL-02. University of Pennsylvania, Philadelphia, USA. July 7-12, 2002. vol. **1/1**: pp. 395-302.
- [Oviatt, S. and Cohen, P. 1991] Oviatt, S. and Cohen, P. (1991) *Discourse structure and performance efficiency in interactive and non interactive spoken modalities*. in Computer, Speech & Language. vol. **5**(4): pp. 297-326.
- [Pankowicz, Z. 1966] Pankowicz, Z. (1966) *Commentary on ALPAC Report*. Rap. RADC, Griffiss Air Force Base. Rome, New York, USA. Personal Memorandum. December 1966. 160 p.
- [Papagaaij, B. C. 1986] Papagaaij, B. C. (1986) *Word Expert Semantics, an Interlingual Knowledge-Based Approach*. Floris Publications. Dordrecht, Holland. 240 p.
- [Papineni, K., et al. 2002] Papineni, K., Roukos, S., Ward, T. and Zhu, V. (2002) *BLEU: a Method for Automatic Evaluation of Machine Translation*. Proc. ACL-02. Philadelphia, USA. July 7-12, 2002. vol. **1/1**: pp. 311-318.
- [Park, J., et al. 1999] Park, J., Hwang, K., Choi, U.-C., Hosaka, J., Yi, S. H. and Yang, J.-W. (1999) *Spoken Language Translation System: Development and Demonstration*. Proc. ICSP-99. Seoul, Korea. August 18-20, 1999. vol. **2/2**: pp. 535-538.
- [Park, Y.-D., et al. 1995] Park, Y.-D., Loken-Kim, K.-H., Mizunashi, S. and Fais, L. (1995) *Transcription of the Collected Dialogues in a Telephone and Multimedia/Multimodal WOZ Experiment*. Rap. ATR-ITL. Kyoto, Japan. Technical Report. n° TR-IT-0090. Feb., 1995. 123 p.
- [Pastor, M., et al. 2001] Pastor, M., Sanchis, A., Casacuberta, F. and Vidal, E. (2001) *EUTRANS: a Speech-to-Speech Translator Prototype*. Proc. Eurospeech. Aalborg, Denmark. September 3-7, 2001. 4 p.
- [Pianta, E. and Toven, L. M. 1999] Pianta, E. and Toven, L. M. (1999) *XIG: Generating from Interchange Format using Mixed Representation*. Proc. AI*IA'99. Bologna, Italy. September 14-17, 1999. pp. 14-17.
- [Pierozak, I. 2003] Pierozak, I. (2003) *Le "français tchaté", un objet à géométrie variable?* in Langage et Société. vol. **104**: pp. 123-144.

- [Planas, É. 1994] Planas, É. (1994) *TELA, Structures et algorithmes pour la Traduction Fondée sur la Mémoire*. Thèse. Université Joseph Fourier - Grenoble 1. 1994. 376 p.
- [Ponsard, C. 2002] Ponsard, C. (2002) *Module d'analyse par pivot sémantique dans le cadre du projet NESPOLE!* Rap. Université Stendhal (Grenoble III). Rapport de stage de Maîtrise Industrie De la Langue, Orientation TALEP. 14 juin 2002. 50 p.
- [Precoda, K., et al. 2004] Precoda, K., Franco, H., Dost, A., Frandsen, M., Fry, J., Kathol, A., Richey, C., Riehemann, S., Vergyri, D. and Zheng, J. (2004) *Limited-Domain Speech-to-Speech Translation between English and Pashto*. Proc. HLT-NAACL 2004. Boston, Massachusetts, USA. May 2-7, 2004. vol. **1/1**: 4 p.
- [Prudhomme, É. 1998] Prudhomme, É. (1998) *Création d'une application client-serveur pour le logiciel de TAFD LIDIA*. Rap. CLIPS-GETA. Grenoble. Rapport de Stage d'IUT deuxième année. 15 juin 1998. 54 p.
- [Quint, V. and Vatton, I. 1994] Quint, V. and Vatton, I. (1994) *Making structured documents active*. in *Electronic Publishing Origination, Dissemination, and Design*. vol. **7(2)**: pp. 55-74.
- [Rayner, M. and Bouillon, P. 2002a] Rayner, M. and Bouillon, P. (2002a) *A Flexible Speech to Speech Phrasebook Translator*. Proc. ACL-02, Workshop on Speech-to-Speech Translation: Algorithms and Systems. University of Pennsylvania, Philadelphia, USA. July 7-12, 2002. pp. 69-76.
- [Rayner, M. and Bouillon, P. 2002b] Rayner, M. and Bouillon, P. (2002b) *A phrasebook Style Medical Speech Translator*. Proc. ACL-02. University of Pennsylvania, Philadelphia, USA. July 7-12, 2002. 2 p.
- [Rayner, M., et al. 2003] Rayner, M., Bouillon, P., Van Dalsem, V., Isahara, H., Kanzaki, K. and Hockey, B. A. (2003) *A Limited-Domain English to Japanese Medical Speech Translator Built Using REGULUS 2*. Proc. ACL-2003, Companion Volume. Sapporo, Japan. July 7-12, 2003. vol. **1/1**: pp. 137-1470.
- [Rayner, M., et al. 1993] Rayner, M., Bretan, I., Carter, D., Collins, M., Digalakis, V., Gambäck, B., Kaja, J., Karlgren, J., Lyberg, B., Pulman, S., Price, P. and Samuelsson, C. (1993) *Spoken Language Translation with mid-90's Technology: a Case Study*. Proc. EUROSPEECH'93. Berlin, Germany. 21-23 September 1993. vol. **2/3**: pp. 1299-1302.
- [Rayner, M., et al. 2000] Rayner, M., Carter, D., Bouillon, P., Digalatakis, V. and Virén, M. E. (2000) *The Spoken Language Translator*. Cambridge University Press. Cambridge. 360 p.
- [Richardson, S. and Braden-Harder, L. 1993] Richardson, S. and Braden-Harder, L. (1993) *The Experience of Developing a Large-Scale Natural Language Processing System: Critique*. in Jensen, K., Heidorn, G. E. and Richardson, S. D. (ed.), *Natural Language Processing: The PLNLP Approach*. Kluwer Academic Publishers. London. pp. 77-90.

- [Rimon, M., *et al.* 1991] Rimon, M., McCord, M. C., Schwall, U. and Martínez, P. (1991) *Advances in Machine Translation Research in IBM*. Proc. Machine Translation Summit III. Washington, D.C. 1-4 juillet 1991. vol. **1/1**: pp. 11-18.
- [Romier, N. 2004] Romier, N. (2004) *Intégration des services de Traduction Automatisée Fondée sur le Dialogue de LIDIA dans un éditeur XML*. Rap. CLIPS-GETA. Grenoble. Rapport de Stage d'IUT année spéciale. 5 septembre 2004. 74 p.
- [Rossato, S., *et al.* 2002a] Rossato, S., Blanchon, H. and Besacier, L. (2002a) *Speech-to-Speech Translation System Evaluation: Results for French for the NESPOLE! Project First Showcase*. Proc. ICSLP 2002. Denver, USA. 16-20 September, 2002. 4 p.
- [Rossato, S., *et al.* 2002b] Rossato, S., Blanchon, H. and Besacier, L. (2002b) *Évaluation du premier démonstrateur de traduction de parole dans le cadre du projet NESPOLE!* Proc. TALN 2002, Atelier "couplage de l'écrit avec l'oral". Nancy, France. 24-27 juin 2002. vol. **2/2**: pp. 149-161.
- [Sadler, V. 1989] Sadler, V. (1989) *Working with analogical semantics: Disambiguation techniques in DLT*. Floris Publications. Dordrecht, Holland. 256 p.
- [Saito, H. and Tomita, M. 1986] Saito, H. and Tomita, M. (1986) *On Automatic Composition of Stereotypic Documents in Foreign Languages*. Rap. CMU. Rapport de recherche. n° CMU-CS-86-107. 1986. p.
- [Saleem, S., *et al.* 2004] Saleem, S., Jou, S.-C., Vogel, S. and Schultz, T. (2004) *Using Word Lattice Information for a Tighter Coupling in Speech Translation Systems*. Proc. ICSLP 2004. Jeju Island, Korea. October 4-8, 2004. 4 p.
- [Sauvageot, A. 1972] Sauvageot, A. (1972) *Analyse du français parlé*. Hachette. Paris. 184 p.
- [Schank, R. and Abelson, R. 1977] Schank, R. and Abelson, R. (1977) *Scripts, Plans, Goals and Understanding*. Laurence Erlbaum Associates. Hillsdale, NJ, USA. 266 p.
- [Schonek, E. 2003] Schonek, E. (2003) *Présentation de documents auto-explicatifs*. Rap. Université Joseph Fourier (Grenoble I). Rapport de TER de maîtrise d'informatique. mai 2003. 23 p.
- [Schubert, K. 1986] Schubert, K. (1986) *Linguistic and Extra-Linguistic Knowledge. A catalog of language-related rules and their computational application in machine translation*. vol. **1(3)**: pp. 125-152.
- [Schubert, K. 1988] Schubert, K. (1988) *The Architecture of DLT - Interlingual or Double Direct?* Proc. New Directions in Machine Translation. Budapest. 18-19 août 1988. pp. 131-143.
- [Schultz, T., *et al.* 2004] Schultz, T., Alexander, D., Black, A. W., Pererson, K., Suebvisai, S. and Waibel, A. (2004) *A Thai Speech Translation System for Medical Dialogs*. Proc. HLT-NAACL 2004. Boston, Massachusetts, USA. May 2-7, 2004. vol. **1/1**: pp. 34-35.

- [Seligman, M. 1996] Seligman, M. (1996) *Interactive MT and Speech Translation via the Internet*. Proc. MIDDIM'96. Le col de Porte, Isère, France. August 12-14, 1996. vol. **1/1**: pp. 309-317.
- [Seligman, M. 1997] Seligman, M. (1997) *Interactive Real-Time Translation via the Internet*. Proc. AAAI Spring Symposium on Natural Language Processing for the World Wide Web. Stanford University. March 24-26, 1997. 7 p.
- [Seligman, M. 2000] Seligman, M. (2000) *Nine Issues in Speech Translation*. in Machine Translation. vol. **15**(1-2): pp. 149-185.
- [Seligman, M., et al. 1998] Seligman, M., Flanagan, M. and Toole, S. (1998) *Dictated Input for Broad-Coverage Speech Translation*. Proc. AMTA Workshop on Embedded Machine Translation. Langhorne, Pennsylvania. October 28, 1998. 6 p.
- [Seneff, S., et al. 1998] Seneff, S., Hurley, E., Lau, R., Pao, C., Schmid, P. and Zue, V. (1998) *Galaxy-II A Reference Architecture of Conversational System Development*. Proc. ICSLP. Sidney, Australia. vol. **3/4**: pp. 931-934.
- [Seneff, S., et al. 1999] Seneff, S., Lau, R. and Polifroni, J. (1999) *Organization, Communication and Control in the Galaxy II Conversational System*. Proc. Eurospeech-99. Budapest, Hungary. September 5-9, 1999. vol. **3/6**: pp. 1271-1274.
- [Seneff, S., et al. 2004] Seneff, S., Wang, C., Hetherington, L. and Chung, G. (2004) *A Dynamic Vocabulary Spoken Dialogue Interface*. Proc. ICSLP 2004. Jeju Island, Korea. October 4-8, 2004. 4 p.
- [Sibley, J. 1988] Sibley, J. (1988) *Le système ALPS*. Proc. Traduction Assistée par Ordinateur — Perspectives technologiques, industrielles et économiques envisageables à l'horizon 1990 — l'offre, la demande, les marchés et les évolutions en cours. Paris. mars 1988. pp. 95-104.
- [Sigurdson, R. and Greatex, R. 1987] Sigurdson, R. and Greatex, R. (1987) *MT of on-line searches in Japanese Data Bases*. Rap. RPI, Lund University. 124 p.
- [Slocum, J. 1987] Slocum, J. (1987) *METAL: The LRC Machine Translation System*. in King, M. (ed.), Machine Translation Today. Edinburgh University Press. pp. 319-350.
- [Slocum, J., et al. 1985] Slocum, J., Bennett, W. S., Whiffin, L. and Norcross, E. (1985) *An Evaluation of METAL: the LRC Machine Translation System*. Proc. EACL-85. Geneva, Switzerland. March 23-27, 1985. vol. **1/1**: pp. 62-69.
- [Somers, H. and Jones, D. 1992] Somers, H. and Jones, D. (1992) *La génération de textes multilingues par un utilisateur monolingue*. in Meta. vol. **37**(4): pp. 647-656.
- [Somers, H. L., et al. 1990] Somers, H. L., Tsujii, J.-I. and Jones, D. (1990) *Machine Translation without a source text*. Proc. COLING-90. Helsinki. August 20-25, 1990. vol. **3/3**: pp. 271-276.
- [Soricut, R. and Brill, E. 2004] Soricut, R. and Brill, E. (2004) *A unified Framework for Automatic Evaluation using N-Gram Co-Occurrence Statistics*. Proc. ACL 2004. Barcelona, Spain. 8 p.

- [Streeter, L. A. 1978] Streeter, L. A. (1978) *Acoustic determinants of phrase boundary perception*. vol. **60**(6): pp. 1582-1592.
- [Su, K.-Y., et al. 1992] Su, K.-Y., Wu, M.-W. and Chang, J.-S. (1992) *A New Quantitative Quality Measure for Machine Translation System*. Proc. COLING-92. Nantes, France. 23-28 août 1992. vol. **2/4**: pp. 433-439.
- [Sugaya, F., et al. 1999] Sugaya, F., Takezawa, T., Yokoo, A. and Yamamoto, S. (1999) *End-to-End Evaluation in ATR-MATRIX Speech Translation System between English and Japanese*. Proc. EUROSPEECH'99. Budapest, Hungary. September 5-9, 1999. vol. **6/6**: pp. 2431-2434.
- [Sugaya, F., et al. 2000] Sugaya, F., Takezawa, T., Yokoo, A. and Yamamoto, S. (2000) *Evaluation of ATR-MATRIX Speech Translation System with Pair Comparison Method Between Human and System*. Proc. ICSLP 2000. Beijing, China. October 16-20, 2000. vol. **3/4**: pp. 1105-1108.
- [Sugaya, F., et al. 2001] Sugaya, F., Yasuda, K., Takezawa, T. and Yamamoto, S. (2001) *Precise Measurement Method of a Speech Translation System's Capabilities with a Paired Comparison Method between the System and Humans*. Proc. MT Summit VIII. Santiago de Compostela, Spain. 18-22 September, 2001. vol. **1/1**: pp. 345-350.
- [Suhm, B., et al. 1994] Suhm, B., Levin, L., Coccaro, J., Carbonell, K., Horiguchi, R., Isotani, A., Lavie, A., Mayfield, C. P., Rosé, C., Van Ess-Dykema, C. and Waibel, A. (1994) *Speech-language integration in a multi-lingual speech translation system*. Proc. AAAI Workshop on speech and language processing. Seattle, Washington, USA. July 31 - August 1, 1994. 6 p.
- [Sumita, E., et al. 2004] Sumita, E., Akiba, Y., Doi, T., Finch, A., Imamura, K., Okuma, H., Paul, M., Shimohata, M. and Watanabe, T. (2004) *EBMT, SMT, Hybrid and More: ATR Spoken Language Translation System*. Proc. IWSLT 2004 (ICLSP 2004 Satellite Workshop). Kyoto, Japan. September 30-October 1, 2004. vol. **1/1**: pp. 13-20.
- [Sumita, E., et al. 1999] Sumita, E., Yamada, S., Yamamoto, K., Paul, M., Kashioka, H., Ishikawa, K. and Shirai, S. (1999) *Solutions to Problems Inherent in Spoken-Language Translation: The ATR-MATRIX Approach*. Proc. MT Summit VII. Singapore. September 13-17, 1999. vol. **1/1**: pp. 229-235.
- [Suzuki, M., et al. 1995] Suzuki, M., Inoue, N., Yato, F., Takeda, K. and Yamamoto, S. (1995) *A Prototype of a Japanese-Korean Realtime Speech Translation System*. Proc. EUROSPEECH'95. Madrid, Spain. September 18-21, 1995. vol. **3/3**: pp. 1951-1954.
- [Takezawa, T., et al. 1998] Takezawa, T., Morimoto, T., Sagisaka, Y., Campbell, N., Iida, H., Sugaya, F., Yokoo, A. and Yamamoto, S. (1998) *A Japanese-to-English speech translation system: ATR-MATRIX*. Proc. ICSLP. Sydney, Australia. Nov. 30-Dec. 4, 1998. pp. 2779-2782.
- [Takezawa, T., et al. 2002] Takezawa, T., Sumita, E., Sugaya, F., Yamamoto, H. and Yamamoto, S. (2002) *Towards a Broad-coverage Bilingual Corpus for Speech Translation of Travel Conversation in the Real World*. Proc. LREC-2002. Las Palmas, Spain. May 29-31, 2002. vol. **1/3**: pp. 147-152.

- [Tenny, M. D. 1985] Tenny, M. D. (1985) *Machine translation, machine-aided translation and machine-imposed translation*. in Lawson, V. (ed.), *Tools for the Trade*. Aslib. London. pp. 105-113.
- [Tillmann, C. 2003] Tillmann, C. (2003) *A projection extension algorithm for statistical machine translation*. Proc. EMNLP 2003. Sapporo, Japan. July 11-12, 2003. vol. **1/1**: pp. 1-8.
- [Tillmann, C., et al. 2000] Tillmann, C., Vogel, S., Ney, H. and Sawaf, H. (2000) *Statistical Translation of Text and Speech: First Results with the RWTH System*. in *Machine Translation*. vol. **15**(1-2): pp. 43-73.
- [Tomita, M. 1984] Tomita, M. (1984) *Disambiguating Grammatically Ambiguous Sentences by Asking*. Proc. COLING-84. Stanford University, California. 2-6 juillet 1984. vol. **1/1**: pp. 476-480.
- [Tomita, M. 1985] Tomita, M. (1985) *Feasibility Study of Personal/Interactive Machine Translation System*. Proc. TMI-85. 14-16 août 1985. vol. **1/1**: pp. 289-297.
- [Tomita, M. 1986] Tomita, M. (1986) *Sentence disambiguation by asking*. in *Computers and Translation*. vol. **1**(1): pp. 39-51.
- [Tomita, M. 1987] Tomita, M. (1987) *An efficient Augmented Context-Free Parsing Algorithm*. in *Computational Linguistics*. vol. **13**(1-2): pp. 31-46.
- [Tomita, M., et al. 1993] Tomita, M., Shirai, M., Tsutsumi, M., Matsumura, J. and Yoshikawa, Y. (1993) *Evaluation of MT Systems by TOEFL*. Proc. TMI-93. Kyoto, Japan. July 14-16, 1993. vol. **1/1**: pp. 252-259.
- [Tomokiyo, M. 2000] Tomokiyo, M. (2000) *Discursive analysis of task-oriented spoken dialogues in Japanese, French and English dialogues – Analyse discursive de dialogues oraux finalisés en français, japonais et anglais: élaboration et validation par analyse de corpus réels*. Thèse. Paris VII. 2000. 275 p.
- [Tomás, J., et al. 2003] Tomás, J., Mas, J. À. and Casacuberta, F. (2003) *A Quantitative Method for Machine Translation Evaluation*. Proc. EACL 2003 – Workshop on Evaluation Initiatives in Natural Language Processing: are evaluation methods, metrics and resources reusable? Budapest, Hungary. April 14, 2003. vol. **1/1**: 8 p.
- [Tovena, L. M. 1998] Tovena, L. M. (1998) *A user's point of view in the interlingua IF*. Rap. ITC-IRST. Trento, Italy. technical Report. n° 9809-01. September 1998. 23 p.
- [Tsai, W.-J. 2004] Tsai, W.-J. (2004) *La coédition Langue-UNL pour partager la révision entre langues d'un document multilingue*. Thèse. Université Joseph Fourier (Grenoble 1). 2004. 314 p.
- [Tsujii, J.-I. 1986] Tsujii, J.-I. (1986) *Future Directions of Machine Translation*. Proc. COLING-86. Bonn. 25-29 août 1986. vol. **1/1**: pp. 655-668.
- [Tsutsumi, T. 1986] Tsutsumi, T. (1986) *A prototype English-Japanese machine translation system for translating IBM computer manuals*. Proc. COLING-86. Bonn. 25-29 août 1986. vol. **1/1**: pp. 646-648.

- [Tsutsumi, T., *et al.* 1993] Tsutsumi, T., Watanabe, H., Maruyama, H., Uramoto, N., Morohashi, M., Takeda, K. and Nasukawa, T. (1993) *Example-Based Approach to Machine Translation*. Proc. Premières journées franco-japonaises sur la traduction assistée par ordinateur. Ambassade de France au Japon, Tokyo, Japon. 15-16 mars 1993. vol. **1/1**: pp. 161-169.
- [Valderrábanos, A., *et al.* 2003] Valderrábanos, A., Esteban, J. and Iraola, L. (2003) *TransType2 - A New Paradigm for Translation Automation*. Proc. MT Summit IX. New Orleans, USA. September 23-27, 2003. 4 p.
- [Van Slype, G. 1979] Van Slype, G. (1979) *Critical Study of Methods for Evaluating the Quality of Machine Translation*. Rap. Bureau Marcel van Dijk. Bruxelles, Belgique. Prepared for the Commission for the European Communities. Final Report. n° BR 19142. 30 novembre 1979. 187 p. Disponible à l'adresse: <http://issco-www.unige.ch/projects/isle/van-slype.pdf>.
- [Vasconcellos, M. and León, M. 1988] Vasconcellos, M. and León, M. (1988) *SPANAM and ENGSPAN: Machine Translation at the Pan American Health Organization*. in Slocum, J. (ed.), *Machine Translation Systems*. Cambridge University Press. pp. 187-236.
- [Vaufreydaz, D., *et al.* 1999] Vaufreydaz, D., Akbar, M. and Rouillard, J. (1999) *A Network Architecture for Building Applications that Use Speech Recognition and/or Synthesis*. Proc. EUROSPEECH'99. Budapest, Hungary. September 5-9, 1999. vol. **5/6**: pp. 2159-2162.
- [Vauquois, B. 1975] Vauquois, B. (1975) *Some problems of optimization in multilingual automatic translation*. Proc. First International Conference on the Application of Mathematical Models and Computers in Linguistics. Varna, Bulgaria. May, 1975. 7 p.
- [Vauquois, B. 1988] Vauquois, B. (1988) *BERNARD VAUQUOIS et la TAO, vingt-cinq ans de Traduction Automatique, ANALECTES. BERNARD VAUQUOIS and MT, twenty-five years of MT*. Ch. Boitet (ed), Ass. Champollion & GETA, Grenoble. 700 p.
- [Vauquois, B. and Boitet, C. 1985] Vauquois, B. and Boitet, C. (1985) *Automated Translation at Grenoble University*. in *Computational Linguistics*. vol. **11**(1): pp. 28-36.
- [Vernant, D. 2003] Vernant, D. (2003) *Communication interpersonnelle et communication personne-système*. in Miège, B. (ed.), *Communication personnes systèmes informationnels*. Lavoisier. Paris, France. pp. 73-102.
- [Veronis, J. and Ide, N. M. 1990] Veronis, J. and Ide, N. M. (1990) *Word Sense Disambiguation with Very Large Neural Networks Extracted from Machine Readable Dictionaries*. Proc. COLING-90. Helsinki. August 20-25, 1990. vol. **2/3**: pp. 389-394.
- [Vogel, S., *et al.* 2004] Vogel, S., Hewavitharana, S., Kolss, M. and Waibel, A. (2004) *The ISL Statistical Translation System for Spoken Language Translation*. Proc. IWSLT 2004. Kyoto, Japan. September 30 - October 1, 2004. vol. **1/1**: pp. 65-72.

- [Vogel, S., *et al.* 2003] Vogel, S., Zhang, Y., Huang, F., Tribble, A., Venugopal, A., Zhao, B. and Waibel, A. (2003) *The CMU Statistical Machine Translation System*. Proc. MT Summit IX. New Orleans, USA. September 23-27, 2003. 8 p.
- [Vu Minh, Q. 2004] Vu Minh, Q. (2004) *Expériences en vue d'une meilleure intégration des modules de reconnaissance automatique de la parole et d'analyse dans un système de traduction de dialogues oraux finalisés*. Rap. Université Joseph Fourier (Grenoble 1). Rapport de DEA. juin 2004. 60 p.
- [Vu Minh, Q., *et al.* 2004a] Vu Minh, Q., Besacier, L., Blanchon, H. and Bigi, B. (2004a) *Modèle de langage sémantique pour la reconnaissance automatique de parole dans un contexte de traduction*. Proc. TALN 2004. Fès, Maroc. 19-21 avril 2004. vol. **1/1**: pp. 477-482.
- [Vu Minh, Q., *et al.* 2004b] Vu Minh, Q., Besacier, L., Castelli, E., Bigi, B. and Blanchon, H. (2004b) *Interchange format-based language model for automatic speech recognition in speech-to-speech translation*. Proc. RIVF'04. *To be published in a special issue of Studia Informatica Universalis [Suger Editor]*. February 2-5, 2004. vol. **1/1**: pp. 47-50.
- [Wagner, C.-K. and Fisher, M.-J. 1974] Wagner, C.-K. and Fisher, M.-J. (1974) *The String-to-String Correction Problem*. in Journal of the ACM. vol. **21**(1): pp. 168-173.
- [Wahlster, W. 2000] Wahlster, W. *ed.*, (2000) *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer-Verlag. Berlin. 677 p.
- [Waibel, A. 1996] Waibel, A. (1996) *Interactive Translation of Conversational Speech*. in IEEE Computer. vol. **29**(7): pp. 41-48.
- [Waibel, A. 2004] Waibel, A. (2004) *Speech Translation: Past Present and Future*. Proc. ICSLP 2004. Jeju Island, Korea. October 4-8, 2004. 4 p.
- [Waibel, A., *et al.* 2003a] Waibel, A., Badran, A., Black, A., Frederking, R., Gates, D., Lavie, A., Levin, L., Lenzo, K., Mayfield Tomokiyo, L., Reichert, J., Schultz, T., Wallace, D., Woszczyna, M. and Zhang, J. (2003a) *Speechlator: two-way speech-to-speech translation on a consumer PDA*. Proc. Eurospeech 2003. Geneva, Switzerland. September 1-4, 2003. 4 p.
- [Waibel, A., *et al.* 2003b] Waibel, A., Badran, A., Black, A. W., Frederking, R., Gates, D., Lavie, A., Levin, L., Lenzo, K., Mayfield Tomokiyo, L., Reichert, J., Schultz, T., Wallace, D., Woszczyna, M. and Zhang, J. (2003b) *Speechlator: Two-Way Speech-to-Speech Translation in Your Hand*. Proc. HLT-NAACL 2003. Edmonton, Canada. May 27-June 1, 2003. 2 p.
- [Watanabe, T., *et al.* 2000] Watanabe, T., Okumura, A., Sakai, S., Yamabana, K., Doi, S. and Hanazawa, K. (2000) *An Automatic Interpretation System for Travel Conversation*. Proc. ICSLP 2000. Beijing, China. Oct. 16-20, 2000. vol. **4/4**: pp. 444-447.
- [Weaver, A. 1988] Weaver, A. (1988) *Two Aspects of Interactive Machine Translation*. in Vasconcellos, M. (ed.), *Technology as Translation Strategy*. State University of New York at Binghamton. Binghamton. pp. 116-123.
- [Wehrli, É. 1990] Wehrli, É. (1990) *STS: An Experimental Sentence Translation System*. Proc. COLING-90. Helsinki. August 20-25, 1990. vol. **1/3**: pp. 76-78.

- [Wehrli, É. 1991] Wehrli, É. (1991) *Pour une approche interactive au problème de la traduction automatique*. Proc. L'environnement traductionnel; La station de travail du traducteur de l'an 2001. 25-27 avril 1991. vol. 1/1: pp. 59-68.
- [Wehrli, É. 1992] Wehrli, É. (1992) *The IPS System*. Proc. COLING-92. 23-28 juillet 1992. vol. 3/4: pp. 870-874.
- [Wehrli, É. 1993] Wehrli, É. (1993) *Vers un système de traduction interactif*. in Bouillon, P. and Clas, A. (ed.), *La traductique*. Les Presses de l'Université de Montréal, AUPELF/UREF. pp. 423-432.
- [Wehrli, É. 1994] Wehrli, É. (1994) *Traduction Interactive: problèmes et solutions*. in Bouillon, P. and Clas, A. (ed.), *TA-TAO: recherches de pointe et applications immédiates*. Les presses de l'Université de Montréal, AUPELF/UREF. pp. 333-342.
- [Wehrli, É. and Ramluckun, M. 1993] Wehrli, É. and Ramluckun, M. (1993) *ITS-2: an interactive personal translation system*. Proc. EACL-93. Utrecht, The Netherlands. April 21-23, 1993. 1 p.
- [Wehrli, É. and Wehrle, T. 1998] Wehrli, É. and Wehrle, T. (1998) *Overview of GBGen*. Proc. 9th International Workshop on Natural Language Generation. Niagara-on-the-lake, Canada. August 1998.
- [Wehrli, É., et al. 1999] Wehrli, É., Wehrle, T., Mengon, J. and Vandeventer, A. (1999) *Une approche efficace à la génération syntaxique. Le système GBGen*. Proc. GAT'99. Grenoble, France. Octobre 1999.
- [White, J. S., et al. 1994] White, J. S., O'Connell, T. and O'Mara, F. E. (1994) *The ARPA MT Evaluation Methodologies: Evolution, Lessons and Further Approaches*. Proc. Technology Partnerships for Crossing the Language Barrier (the First Conference of the Association for Machine Translation in the Americas). Columbia, Maryland, USA. October 5-8, 1994.
- [Whitelock, P. J., et al. 1986] Whitelock, P. J., Wood, M. M., Chandler, B. J., Holden, N. and Horsfall, H. J. (1986) *Strategies for Interactive Machine translation: the experience and implications of the UMIST Japanese project*. Proc. COLING-86. Bonn. 25-29 août 1986. vol. 1/1: pp. 25-29.
- [Witkam, A. P. M. 1983] Witkam, A. P. M. (1983) *Distributed Language Translation – Feasibility Study of a Multilingual Facility for Videotex Information Networks*. BSO. Utrecht, The Netherlands. 340 p.
- [Witkam, T. 1988] Witkam, T. (1988) *DLT - An Industrial R&D Project for Multilingual MT*. Proc. COLING-88. Budapest. 22-27 août 1988. vol. 2/2: pp. 576-579.
- [Wood, M. M. 1989] Wood, M. M. (1989) *Japanese for speakers of English: The UMIST/Sheffield Machine Translation Project*. in Peckham, J. (ed.), *Recent Developments and Applications of Natural Language Processing*. Kogan Page Limited. London. pp. 56-64.
- [Wood, M. M. G. and Chandler, B. 1988] Wood, M. M. G. and Chandler, B. (1988) *Machine Translation For Monolinguals*. Proc. COLING-88. Budapest. 22-27 août 1988. vol. 2/2: pp. 760-763.

- [Woszczyna, M., *et al.* 1993] Woszczyna, M., Coccaro, N., Eisele, A., Lavie, A., McNair, A., Polzin, T., Rogina, I., Rose, C. P., Sloboda, T., Tomita, M., Tsutsumi, M., Aoki-Waibel, N., Waibel, A. and Ward, W. (1993) *Recent Advances in JANUS: a Speech Translation System*. Proc. EUROSPEECH'93. Berlin, Germany. September 21-23, 1993. vol. **2/3**: pp. 1295-1298.
- [Yamabana, K., *et al.* 2000] Yamabana, K., Ando, S.-i. and Mimura, K. (2000) *Lexicalized Tree Automata-Based Grammars for Translating Conversational Texts*. Proc. COLING 2000. Saarbrücken, Germany. July 31 - August 4, 2000. vol. **2/2**: pp. 926-932.
- [Yamabana, K., *et al.* 1997] Yamabana, K., Muraki, K., Kamei, S., Satoh, K., Doi, S. and Tamura, S. (1997) *An interactive Translation Support Facility for Non-Professional Users*. Proc. 5th Conference on Applied Natural Language Processing. Washington DC. 31 March - 3 April 1997. pp. 324-331.
- [Yamamoto, K. 2002] Yamamoto, K. (2002) *Machine Translation by Interaction between Paraphraser and Transfer*. Proc. COLING. Taipei, Taiwan. 24 August - 1 September, 2002. vol. **2/2**: pp. 1107-1113.
- [Yang, J.-W. and Park, J. 1997] Yang, J.-W. and Park, J. (1997) *An Experiment on Korean-to-English and Korean-to-Japanese Spoken Language Translation*. Proc. ICASSP'97. Munich, Germany. April 21-24, 1997. vol. **1/5**: pp. 87-90.
- [Zaki, A. B. and Noor, H. 1991] Zaki, A. B. and Noor, H. (1991) *Malay Official Letters Translation*. Proc. International Conference on Current Issues in Computational Linguistics. Universiti Sains Malaysia, Penang, Malaisie. 10-14 juin 1991. vol. **1/1**: pp. 413-420.
- [Zhang, R., *et al.* 2004] Zhang, R., Kikui, G., Yamamoto, H., Soong, F., Watanabe, T., Sumita, E. and Lo, W.-K. (2004) *Improved Spoken Language Translation Using N-best Speech Recognition Hypotheses*. Proc. ICSLP 2004. Jeju Island, Korea. October 4-8, 2004. 4 p.
- [Zhou, B., *et al.* 2004] Zhou, B., Déchelotte, D. and Gao, Y. (2004) *Two-way Speech-to-Speech Translation on Handheld Devices*. Proc. ICSLP 2004. Jeju Island, Korea. October 4-8, 2004. 4 p.
- [Zhou, B., *et al.* 2003] Zhou, B., Gao, Y., Sorensen, J. and Gao, Y. (2003) *A Hand-held speech-to-speech translation system*. Proc. ASRU 2003. Virgin Island, USA. November 30 - December 3, 2003. 6 p.
- [Zirkle, A. T. 1990] Zirkle, A. T. (1990) *The role of computer-aided translation in translation services*. in Mayorcas, P. (ed.), *Translation and the computer* 10. Aslib. London. pp. 11-17.
- [Zong, C., *et al.* 2000a] Zong, C., Huang, T. and Xu, B. (2000a) *An Improved Template-Based Approach to Spoken Language Translation*. Proc. ICSLP 2000. Beijing, China. Oct. 16-20, 2000. vol. **4/4**: pp. 440-443.
- [Zong, C., *et al.* 2000b] Zong, C., Huang, T. and Xu, B. (2000b) *Japanese-to-Chinese Spoken Language Translation Based on the Simple Expression*. Proc. ICSLP 2000. Beijing, China. Oct. 16-20, 2000. vol. **4/4**: pp. 418-421.
- [Zong, C., *et al.* 2002] Zong, C., Xu, B. and Huang, T. (2002) *Interactive Chinese-to-English Speech Translation Based on Dialogue Management*. Proc.

ACL-02, Workshop on Speech-to-Speech Translation: Algorithms and Systems. University of Pennsylvania, Philadelphia, USA. July 7-12, 2002. pp. 61-68.

[Zuo, Y., *et al.* 2004] Zuo, Y., Zhou, Y. and Zong, C. (2004) *Multi-Engine Based Chinese-to-English Translation System*. Proc. IWSLT 2004. Kyoto, Japan. September 30 - October 1, 2004. vol. **1/1**: pp. 73-77.

Bibliographie personnelle

Articles

Revue d'audience internationale avec comité de rédaction

Blanchon, H., Boitet, C. & Besacier L. (2004) *Spoken Dialogue Translation Systems: Selected Evaluation Results, New Evaluation Trends, Problems and Proposals to Go Further*. in IEEE Transactions on Speech and Audio Processing:15 p. Submitted.

Blanchon, H., Boitet, C. & Caelen, J. (1999) *Participation Francophone au Consortium C-STAR II*. in La tribune des industries de la langue et du multimedia/Linguistic engineering and multimedia tribune. vol. **31-32** (August-December 1999) : pp. 15-23.

Boitet, C. & Blanchon, H. (1995) Multilingual Dialogue-Based MT for monolingual authors: the LIDIA project and a first mockup. in Machine Translation. vol. **9(2)** : pp 99-132.

Blanchon, H. (1995) *Interagir pour traduire : la TAO personnelle pour rédacteur monolingue*. in La Tribune des Industries de la Langues. vol. **17-18-19** : pp. 28-34.

Boitet, C. & Blanchon, H. (1994) *Après la maquette LIDIA-1, quel avenir pour la "TAO personnelle" pour auteur monolingue?* in Langages, "L'ordinateur et l'ordinateur". Vol. **116**, décembre 1994 : pp. 20-47.

Communications

Manifestations d'audience internationale avec comité de sélection

Blanchon, H. (2004) *HLT Modules Scalability within the NESPOLE! Projet*. Proc. ICSLP 2004. Jeju Island, Korea. October 4-8, 2004. 4 p.

Blanchon, H., (2002) *A Pattern-Based Analyzer for French in the Context of Spoken Language Translation: First Prototype and Evaluation*. Proc. COLING. Taipei, Taiwan. 24 August - 1 September, 2002. vol. 1/2: pp. 92-98.

Blanchon, H. (1997) *Interactive Disambguation of Natural Language Input: a Methodology and Two Implementations for French and English*. Proc. IJCAI-97. Nagoya, Japan. August 23-29, 1997. vol. 2/2 : pp. 1042-1047.

Blanchon, H. (1995) *Clarification: Towards more User-Friendly Natural Language Human-Computer Interaction*. Proc. Poster session of HCI'95, 6th International Conference on Human-Computer Interaction. Yokohama, Japan. July 9-14, 1995 . vol. 1/1 : pp. 42-42.

Blanchon, H. (1994) *Perspectives of DBMT for monolingual authors on the basis of LIDIA-1, an implemented mock-up*. Proc. COLING-94. Kyoto, Japon. Aug. 5-9, 1994. vol. 1/2 : pp. 115-119.

- Blanchon, H.** (1992a) *A Solution to the Problem of Interactive Disambiguation*. Proc. COLING-92. Nantes, France. 23-28 juillet 1992. vol. 4/4 : pp. 1233-1238.
- Blanchon, H.** (1992b) *Désambiguïsation interactive en TAO personnelle : poser les bonnes questions*. Proc. Avignon-92, conférence spécialisée : Le Traitement Automatique des Langues Naturelles et ses Applications. Avignon, France. 3-5 juin 1992. vol. 3/4 : pp. 69-80.
- Blanchon, H.** (1990) *LIDIA-1 : Un prototype de TAO personnelle pour rédacteur unilingue*. Proc. Avignon-90, conférence spécialisée : Le Traitement Automatique des Langues Naturelles et ses Applications. Avignon, France. 28 mai-1 juin 1990 : pp. 51-60.
- Blanchon, H. & Besacier, L.** (2004) *Traduction de dialogues : mise en perspectives des résultats du projet NESPOLE! et pistes pour le domaine*. Proc. TALN 2004. Fès, Maroc. 19-21 avril 2004. vol. 1/1: pp. 55-60.
- Blanchon, H. & Boitet, C.** (2004) *Deux premières étapes vers les documents auto-explicatifs*. Proc. TALN 2004. Fès, Maroc. 19-21 avril 2004. vol. Internal Report.: pp. 61-70.
- Blanchon, H. & Boitet, C.** (2003) *Two Steps Towards Self-Explaining Documents*. Proc. Convergence 03. Alexandria, Egypt. December 2-6, 2003. paper 032-324, 6 p.
- Blanchon, H. & Boitet, C.** (2000) *Speech Translation for French within the C-STAR II Consortium and Future Perspectives*. Proc. ICSLP 2000. Beijing, China. 16-20 October, 2000. vol. 4/4 : pp. 412-417.
- Blanchon, H., Boitet, C. & Besacier, L.** (2004) *Spoken Dialogue Translation System Evaluation: Results, New Trends, Problems and Proposals*. Proc. Kyoto, Japan. September 30 - October 1, 2004. vol. 1/1: pp. 95-102.
- Blanchon, H., Boitet, C., Brunet-Manquat, F., Tomokio, M., Hamon, A., Hung, V. T. & Bey, Y.** (2004) *Towards Fairer Evaluation of Commercial MT Systems on Basic Travel Expressions Corpora*. Proc. IWSLT 2004 (ICLSP 2004 Satellite Workshop) Kyoto, Japan. September 30-October 1, 2004. vol. 1/1: pp. 21-26.
- Blanchon, H., Boitet, C. & Guilbaud J.-P.** (2000) *Traduction de la parole pour le français : une première étape et quelques perspectives*. Proc. TALN 2000. Lausanne, Suisse. 16-18 octobre 2000. vol. 1/1 : pp. 373-382.
- Blanchon, H. & Fais, L.** (1997) *Asking Users About What They Mean: Two Experiments & Results*. Proc. HCI'97. San Francisco, California. August 24-29, 1997. vol. 2/2 : pp.609-912.
- Blanchon, H., Loken-Kim, K-H, Morimoto, T.** (1995) *An Interactive Disambiguation Module for English Natural Language Utterances*. Proc. NLPRS'95. Seoul, Korea. Dec 4-7, 1995. vol. 2/2 : pp. 550-555. Best paper award.
- Besacier, L., Blanchon, H., Fouquet, Y., Guilbaud, J.-P., Helme, S., Mazenot, S., Moraru, D. and Vaufraydaz, D.** (2001) *Speech Translation for*

French in the NESPOLE! European Project. Proc. Eurospeech. Aalborg, Denmark. September 3-7, 2001. vol. 2/4: pp. 1291-1294.

Boitet, C. & Blanchon, H. (1993) *Dialogue-based MT for monolingual authors and the LIDIA project*. Proc. NLPRS'93. Fukuoka, Japon. Dec. 6-7, 1993. vol. 1/1 : pp. 208-222.

Boitet, C., Blanchon, H, Guilbaud J.-P. (2000) *A way to integrate contexte processing in the MT component of spoken, task-oriented translation systems*. Proc. MSC2000. Kyoto, Japon. October 11-13, 2000. vol. 1/1 : pp. 83-87.

Cattoni, R., Lazzari, G., Mana, N., Pianesi, F., Pianta, E., Burger, S., Gates, D., Lavie, A., Levin, L., Langley, C., Pererson, K., Schultz, T., Waibel, A., Wallace, D., Metze, F., Mc Donough, J., Soltau, H., Besacier, L., Blanchon, H., Vaufreydaz, D., Costantini, E. and Taddei, L. (2002) *Not only Translation Quality: Evaluation the NESPOLE! Speech-to-Speech Translation System along other viewpoints*. Proc. ACL-02. University of Pennsylvania, Philadelphia, USA. July 7-12, 2002. 8 p.

Lavie, A., Metze, F., Pianesi, F., Burger, S., Gates, D., Levin, L., Langley, C., Pererson, K., Schultz, T., Waibel, A., Wallace, D., Mc Donough, J., Soltau, H., Cattoni, R., Lazzari, G., Mana, N., Pianta, E., Costantini, E., Besacier, L., Blanchon, H., Vaufreydaz, D. & Taddei, L., (2002) *Enhancing the Usability and Performance of NESPOLE! – a Real-World Speech-to-Speech Translation System*. Proc. HLT 2002. San Diego, California, USA. March 22-27, 2002. 6 p.

Metze, F., Mc Donough, J., Soltau, H., Waibel, A., Lavie, A., Burger, S., Langley, C., Levin, L., Schultz, T., Pianesi, F., Cattoni, R., Lazzari, G., Mana, N., Pianta, E., Besacier, L., Blanchon, H., Vaufreydaz, D. & Taddei, L. (2002) *The NESPOLE! Speech-to-Speech Translation System*. Proc. HLT 2002. San Diego, California, USA. March 22-27, 2002. 6 p.

Rossato, S., Blanchon, H. & Besacier, L. (2002) *Speech-to-Speech Translation System Evaluation: Results for French for the NESPOLE! Project First Showcase*. Proc. ICSLP. Denver, USA. 16-20 September, 2002. 4p.

Rossato, S., Blanchon, H. & Besacier, L. (2002) *Évaluation du premier démonstateur de traduction de parole dans le cadre du projet NESPOLE!* Proc. TALN 2002, Atelier Couplage de l'écrit avec l'oral. Nancy, France. 24-27 juin 2002. vol. 2/2: pp. 149-161.

Vu Minh, Q., Besacier, L. & Blanchon H. (2004) *Modèle de langage sémantique pour la reconnaissance automatique de parole dans un contexte de traduction*. Proc. TALN 2004. Fès, Maroc. 19-21 avril 2004. vol. 1/1: pp. 477-482.

Vu Minh, Q., Besacier, L., & Blanchon H. (2004) *Interchange format-based language model for automatic speech recognition in speech-to-speech translation*. Proc. RIVF'04 (Recherche Informatique Vietnam-Francophonie) To be published in a special issue of Studia Informatica Universalis [Sugar Editor]. February 2-5, 2004. vol. 1/1: pp. 47-50.

Manifestations d'audience nationale avec comité de sélection

Blanchon, H. & Boitet, C. (2004) *Les Documents Auto-Explicatifs : une voie pour offrir l'accès au sens aux lecteurs*. Proc. CIDE 7 : Approches Sémantique du Document Électronique. La Rochelle, France. 22-25 juin 2004. vol. 1/1 : pp. 273-290.

Blanchon, H. (1994) *Perspectives en TAFD pour auteur monolingue après une première expérience : la maquette LIDIA-1*. Proc. TALN'94. Marseille, France. 7-8 avril 1994. vol. 1/1 : 12p.

Blanchon, H., Fais, L., Loken-Kim, K. H. and Morimoto, T. (1995) *A Pattern-Based Approach for Interactive Clarification of Natural Language Utterances*. in Bulletin of the Information Processing Society of Japan (95-NL-107) vol. 95(52) : pp. 11-18.

† **Blanchon, H. & Loken-Kim, K. H.** (1994) *Towards More Robust, Fault-Tolerant and User-Friendly Software Integrating Natural Language Processing Components*. in Bulletin of the Information Processing Society of Japan. vol. 94(109) : pp. 17-24. (aussi sous ‡)

Manifestations sans comité de sélection, communications invitées

Blanchon, H., (2002) *Area Report: Machine Translation*. Proc. COLING 2002, Post-Conference Workshop: A Roadmap for Computational Linguistics. Taipei, Taiwan. Jeu de 15 transparent. Accessible sur : <http://www.elsnet.org/roadmap-coling2002.html>.

Blanchon, H., Besacier, L., Vaufreydaz, D. and Rossato, S. (2002) *French Contribution to Showcase One of the NESPOLE! Project: Components and Evaluation*. Proc. ESSLLI. Trento, Italy. 5-16 August, 2002. Jeu de 24 transparents

Blanchon, H. (2002) *Traduction de dialogues oraux et finalisés en réseau (jeu de 19 transparents) Présentation invitée*. Proc. Colloque annuel du RNRT. Grenoble, France. 7-8 février 2002. Téléchargeable sur <http://www.telecom.gouv.fr/rnrt/coll/coll2002pr.htm>.

Blanchon, H. (1996) *A Customizable Interactive Disambiguation Methodology and Two Implementations to Disambiguate French and English Input*. Proc. MIDDIM'96. Le col de porte, Isère, France. 12-14 Août 1996. vol. 1/1 : pp. 190-200.

Blanchon, H. (1991) *Problèmes de désambiguïsation interactive en TAO Personnelle*. Proc. L'environnement traductionnel ; La station de travail du traducteur de l'an 2001. Mons, Belgique. 25-27 avril 1991. vol. 1/1 : pp. 31-48.

Blanchon, H. & Fais, L. (1996) *How to ask Users About What they Mean: Two Experiments & Results*. Proc. MIDDIM'96. Le col de porte, Isère, France. 12-14 Août 1996. vol. 1/1 : pp. 238-259.

Fais, L. & Blanchon, H. (1996) *Ambiguities in Task-oriented Dialogues*. Proc. MIDDIM'96. Le col de porte, Isère, France. 12-14 Août 1996. vol. 1/1 : pp. 263-275.

Rapports de Recherche

Blanchon, H. (1995) *An Interactive Disambiguation Module for English Input: an Engine and the Associated Lingware*. Rap. ATR-ITL. Technical Report. n° TR-IT-0129. August, 1995. 110 p.

Blanchon, H. (1994) *Pattern-based approach to interactive disambiguation: first definition and implementation*. Rap. ATR-Interpreting Telecommunications Research Laboratories. Technical Report. n° TR-IT-0073. Sept. 8, 1994. 91 p.

Blanchon, H. (1991) *Le réseau LIDIA*. Rap. GETA, IMAG. Rapport interne. juillet 1991. 7 p.

Boitet, C. & Blanchon, H. (1993) *La Traduction Automatisée Fondée sur le dialogue pour auteur monolingue et le projet LIDIA*. Rap. IMAG. Rapport de Recherche. n° RR-918-I. mai 1993. 19 p.

Blanchon, H. & Fais, L. (1996a) *Pilot Experiment on the Understandability of Interactive Disambiguation Dialogues*. Rap. ATR-ITL. Technical report. n° TR-IT-0177. July, 1996. 18 p.

Blanchon, H. & Fais, L. (1996b) *A Second Experiment on the Understandability of Interactive Disambiguation Dialogues*. Rap. ATR-ITL. Technical Report. n° TR-IT-0167. April, 1996. 22 p.

‡ **Blanchon, H. & Loken-Kim, K. H.** (1994) *Towards More Robust, Fault-Tolerant and User-Friendly Software Integrating Natural Language Processing Components*. Rap. Institute of Electronics, Information and Communication Engineering. Technical Report. n° SP94-65, NLC94-34. December, 1994. 8 p. (aussi sous †)

Rapports (intermédiaires) de contrats de recherche

Blanchon, H. (1993) *Report on a stay at ATR*. Rap. GETA & ATR-ITL. Internal Report, MIDDIM Project. July 1993. 30 p.

Blanchon, H., Boitet, C., Guilbaud, J.-P., Nédobejkine, N., Blanc, É., Guillaume, p. & Quézel-Ambrunaz, M., (1995) *Étude et maquettage de procédés de désambiguïsation interactive pour la TAO personnelle*. Rap. GETA. Rapport final de la convention n° 94 2 93 0136 entre le Ministère de l'industrie et l'Université Joseph Fourier et réalisée par le GETA dans le cadre du Projet Eurêka EUROLANG. février 1995. 296 p.

Boitet, C., Planas, E., Blanchon, H., Blanc, É., Guilbaud, J.-P., Guillaume, p., Lafourcade, M. and Sérasset, G., (1995) *LIDIA-1.2, une maquette de TAO personnelle multicible, utilisant la messagerie usuelle, la désambiguïsation interactive et la rétrotraduction*. Rap. GETA. Rapport final de la convention n° 95 2 93 0104 entre le Ministère de l'industrie et l'Université Joseph Fourier et réalisée par le GETA dans le cadre du Projet Eurêka EUROLANG. septembre 1995. 243 p.

Lavie A., Cattoni, A. & Blanchon, H. (2003) *The NESPOLE! Speech-to-Speech Translation Systems*. Rap. Nespole! (IST-1999-1156) Internal Report. n° D15. 26 February 2003. 19 p.

ITC-irst, UKA, CMU, UJF, AETHRA, APT (2003) *Evaluation of the NESPOLE! Showcase-2 System*. Rap. Nespole! (IST-1999-1156) Public Document. n° D18. 26 February 2003. 7 p.

ITC-irst, UKA, CMU, UJF, AETHRA, APT (2002) *Requirements for second showcase*. Rap. Nespole! (IST-1999-1156) Internal Report. n° D11. 10 July 2002. 18 p.

ITC-irst, UKA, CMU, UJF, AETHRA, APT (2002) *Evaluation of the NESPOLE! Showcase-1 System*. Rap. Nespole! (IST-1999-1156) Public Document. n° D9. 15 June 2002. 14 p.

ITC-irst, UKA, CMU, UJF, AETHRA, APT (2001) *First Showcase documentation*. Rap. Nespole! (IST-1999-1156) Public Document. n° D8. 10 November 2001. 27 p.

TC-irst, UKA, CMU, UJF, AETHRA, APT (2001) *Multimedial Space for First Showcase*. Rap. Nespole! (IST-1999-1156) Internal Report. n° D7. 25 January 2001. 7 p.

ITC-irst, UKA, CMU, UJF, AETHRA, APT (2000) *Requirements for 1st showcase*. Rap. Nespole! (IST-1999-1156) Internal Report. n° D1. 17 July 2000. 27 p.

ITC-irst, UKA, CMU, UJF, AETHRA, APT (2000) *Assessment and evaluation Plan*. Rap. Nespole! (IST-1999-1156) Internal Report. n° D21. 4 April 2000. 10 p.

Animation scientifique (jeux de transparents + notes ou actes de séminaires)

Film documentaire et presse

Santoni, J-F. (1999) *"Recherche transfert... passionnément"*. Produit par Mille Images en partenariat avec le CEA, le CNRS et l'INRIA. La démonstration finale de C-STAR II des coréens en dialogue avec la France y est présentée, ainsi qu'une courte interview.

Divers interview presse écrite et radio. Sont ou ont été sur le web:

Traduction simultanée orale assistée par ordinateur dans **Le Monde**

Trois questions à Hervé Blanchon dans **Le Monde**

L'ordinateur a appris la traduction simultanée dans **Libération**

Séminaires pour l'industrie

Blanchon, H. (2004) *Traduction de parole : état des lieux et perspectives*. Séminaire pour la société Prosodie. Paris. 2 février 2004. Jeu de transparents KeyNote. 58 p.

Blanchon, H. (2003) Traduction de parole : état des lieux et perspectives. Séminaire sur la Traduction de France Telecom. Paris, France. 15 décembre 2003. Jeu de transparents KeyNote. 70 p.

Blanchon, H. (2002) *Traduction automatique – une brève introduction & travaux du GETA. Jeu de transparents et notes.* Séminaire du club ECRIN sur la traduction automatique. Solaize (Rhône) 31 janvier 2002. 80 p.

Blanchon, H. (1993) *Traduction automatique – aspects et réalités. Jeu de transparents et notes.* Séminaire de l'Institut d'Expertise et Prospective de l'École Normale Supérieure : Nouvelles techniques linguistiques et informatiques en documentique intelligente. Paris. 29 avril 1993. 53 p.

Séminaires de troisième cycle

Blanchon, H. (1999) *La traduction automatique : une brève introduction et présentation des travaux du GETA. Jeu de transparents et notes.* Séminaire MIM2 (Magistère d'Informatique de l'École Normale Supérieure de Lyon) Les Deux Alpes. 13-17 décembre 1999 : 71 p.

Blanchon, H. (1996a) Désambiguïsation interactive pour monolingues naïfs : l'approche du projet LIDIA et les expériences du projet MIDDIM (jeu de transparents) Séminaire Alpin en Informatique et Linguistique sur la traduction automatique. Archamps. 7 novembre 1996 : 40 p.

Blanchon, H. (1996b) *Revue des systèmes de TA interactive anciens et actuels (jeu de transparents)* Séminaire Alpin en Informatique et Linguistique sur la traduction automatique. Archamps. 7 novembre 1996 : 40 p.

Blanchon, H. (1994) *De la traduction interactive pour spécialiste à la TADF "pour tous".* Cours de DESS "traduction spécialisée et production de textes multilingues". Université Stendhal (Grenoble 3), Département de Langues Étrangères Appliquées. France. 19 avril 1994, 70 p.

Blanchon, H. (1993) *La traduction automatisée fondée sur le dialogue et la maquette LIDIA-1.* Séminaire Alpin en Informatique et Linguistique sur la traduction automatique. Archamps. 16 décembre 1993. 68 p.

Workshops et conférences

Blanchon, H. (1999a) *Le projet C-STAR II (jeu de transparents)* Journée des projets IMAG. Grenoble, France. 21 octobre 1999 : 31 p.

Blanchon, H. (1999b) *CLIPS++ Contribution within the C-STAR II Consortium.* Proc. C-STAR Workshop. Schwetzingen (Germany) September 22nd-24th, 1999 : 6 p.

Blanchon, H. (1990a) *TAO personnelle et promotion des langues nationales : projet LIDIA du GETA. Jeu de transparents.* Les industries de la langue : perspectives des années 1990. Montréal, Canada. 22-24 novembre 1990. 20p.

Blanchon, H. (1990b) *Ambiguity resolution and paraphrase selection*. Proc. DBMT-90, Post-COLING seminar on Dialogue-Based MT. Blanc É & Boitet C. eds. Le Sappey, France. 26-28 août 1990, vol. 1/1 : pp. 38-41 & 238-241.

Besacier, L. Blanchon, H. (2001) *Traduction de parole pour le français dans le projet Européen NESPOLE! (jeu de transparents)* Journée de l'ATALA "Vers un meilleur couplage du traitement de la parole et du traitement de l'écrit". Paris, France. 9 juin 2001. 48 p.

Rapport de DEA

Blanchon, H. (1990) *HyperAriane : architecture de la partie "micro" d'un système de TAO personnelle*. Rap. GETA, IMAG, UJF&INPG. Rapport de DEA d'informatique. juin 1993. 63 p.

Thèse

Blanchon, H. (1994) *LIDIA-1 : une première maquette vers la TA interactive "pour tous"*. Thèse. Université Joseph Fourier - Grenoble 1. 21 janvier 1994. 321 p.

Logiciels

Blanchon, H. (2003) *Générateur IF vers français*, Vers. 2.0 (janvier 2003) TCL code, 31918 lines. Second NESPOLE! Showcase Delivrable. GETA-CLIPS (IMAG) 2003.

Blanchon, H. (2003) *Analyseur français vers IF*, Vers. 2.0 (janvier 2003) TCL code, 16787 lines. Second NESPOLE! ShowCase Delivrable. GETA-CLIPS (IMAG) 2003.

Blanchon, H. (2001) *Générateur IF vers français*, Vers. 1.0 (novembre 2001) TCL code, 6793 lines. First NESPOLE! Showcase Delivrable. GETA-CLIPS (IMAG) 2001.

Blanchon, H. (2001) *Analyseur français vers IF*, Vers. 1.0 (novembre 2001) TCL code, 9753 lines. First NESPOLE! ShowCase Delivrable. GETA-CLIPS (IMAG) 2001.

Blanchon, H. (1999) *Module français vers IF en TCL/TK*. Maquette de démonstration. CLIPS-IMAG, Grenoble, France. Multi plate-forme.

Blanchon, H. (1995) *Disambiguation engine v2 & English lingware v1*. Maquette de démonstration. ATR-ITL, Kyoto, Japan. Macintosh Common Lisp.

Blanchon, H., Blanc, É., Gaschler, J., Guilbaud, J.-P., Guillaume, p., Lafourcade, M., Nédobekine, N., Quézel-Ambrunaz, M. & Sérasset, G. (1994) *LIDIA-1.0 (première maquette d'un système de TAFD)* Maquette de démonstration. GETA-IMAG, France. Macintosh et IBM.

Blanchon, H., Guilbaud, J. p. & Nédobekine, N. (1992) *LIDIA : the disambiguation process — le processus de désambiguisation*. Pile HyperCard

présentée à l'exposition COLING-92, Nantes, 23-28 juillet 1992. GETA-IMAG, juillet 1992. 160 p.

CLIPS++ (1999) *Démonstateur C-STAR II du groupe CLIPS++*. Semi-prototype de démonstration. CLIPS-IMAG, Grenoble & LATL, Université de Genève & LAIP, Université de Lausanne.

Annexes

Annexe I

Ambiguïtés dans les corpus de ATR

Nous présentons ici la liste des ambiguïtés que nous avons trouvées dans les corpus que nous avons étudiés.

La première colonne de chaque tableau indique le type d'ambiguïté, la seconde indique la situation (TL : téléphone, MM : multimodal), la troisième colonne indique le rôle du locuteur (A : Agent, C : Client, I : Interprète, W : Compère)

I.1. Conversations monolingues (C1)

CO	TL	A	...you can just [uh] tell him that you're trying to get to thi International Conference Center and it should be about a twenty minute ride
D	MM	A	...you've come in on thi second platform so we're right about here
D	MM	A	...you can pay for it right on the bus
D	TL	C	oh hello I'm calling for information
D	TL	C	{/breath intake/} hi I just arrived in Kyoto
D	TL	A	you'll {want} to leave by exit number six
D, P	TL	A	...that will be east if you just walk right out of the main doors of the sta{tion}
P	MM	A	yes bus five should leave every half hour
P	MM	C	right OK (so) and it's one hundred yen including the change
P	MM	A	and change subways to thi Keihen-Kyotsu Line
P	MM	A	and then I'll show you the blow up of thi station
P	MM	A	so let me go to thi station and
P	MM	A	so you can [ah] catch the subway
P	MM	A	you're going to take the [n] subway north
P	MM	A	you're at Kyoto station and you can travel a number of different ways
P	MM	C	OK and how long will the ride be
P	MM	A	...there you can pick up a taxi
P	MM	C	...do they make change on the bus
P	MM	C	{{[umm] I} am on the first floor and that's about all I can tell
P	TL	C	{{(how mu}ch) how much is the bus
P	TL	C	How many stories
P	TL	A	thi [uh] taxi costs ten thousand yen
P, PH	MM	A	OK let me call up a map here and this will help you OK you're at Kyoto station I will draw up [th] map of Kyoto station
PH	MM	A	{{[z]} do you [ah] wanna go over that again or
PH	MM	A	well it's difficult to get out of Kyoto station
PH	TL	C	OK thats fine I thi{nk} I got it +straight+
S	TL	A	..that is where the main ticket office is
S	TL	C	/breath/ yes I'm trying to find my way to thi International Conference Center

S	TL	A	{theres} only one platform and the trains are only going in one direction
S	TL	A	{[uhuhn]} yes [ah] this is an English speaking agent (English law professor)
S, SV	MM	C	...I'm trying to figure out how to get to thi International Conference Center...where [um] the conference is I believe
SC	MM	A	Good morning conference office
SC	MM	A	/ls/ the bus ride is about five hundred yen
SC	MM	A	OK the quickest route would really be taking a taxi
SC	TL	A	{line} and you can catch that at the second level floor platform
SC	TL	C	{OK} (and that so that) I take that [um] one stop
SC,SC	TL	A	Let me get those maps out OK [ah] you can travel by subway bus or taxi
SV	MM	A	OK Kyoto station is located in the south part of the map right here
SV	MM	A	OK [ah] so you can cross the street from the station
SV	MM	A	...going to go directly across the street to the middle of the bus station
SV	MM	A	...we can [ah] show you where you are in relation to transportation
SV	MM	C	[um] yes I'm calling about thi [ah] International Symposium on Interpreting Telecommunications right now
SV	MM	A	and thi taxi stand is located to the north of the train station [ah] to thi [ah] east of the bus stop
SV	MM	C	OK [umm] where can I catch a taxi {from Kyoto Station}
SV	TL	C	I think subway sounds like the best way to me
SV	TL	A	that's where you can pick up a taxi as well
SV	TL	A	{/breath/} OK let me pull up my maps to help you here
SV	TL	C	...{it} says here on my flyer
SV, P	MM	A	OK I'm going to [ah] tell you roughly how you're gonna go

I.2. Conversations bilingues interprétées (C2)

CO	MM	I	So as I said there's a swimmin pool and also the gym that you can go and train
CO	MM	C	[ah] (is) Is that like an international hotel they're going to have like regular western beds and regular western toilets and stuff
CO	MM	I	All right then can I ask you to please type in your name and the telephone number
CO	TL	C	OK if possible I'd like a room with a large double or queen size bed
CO	TL	C	OK but is there like a restaurant or something in the hotel or nearby where I'll be able to eat
CO	TL	I	K [um] so if you standing at outside a gate you should be able to see the sign saying exit whatever maybe the number or the name
CO	TL	C	is there a train or bus or something from here
CO	TL	I	Well there are many temples and castles like kinkaguchi temple or nijogo castle
CO,	MM	I	K could I ask you to type in your full name and your telephone number either at your office or home
CO,S,SV	MM	C	...am I going to be eating like a <i>Japanese style breakfast</i> with fish and stuff
D	MM	C	Can you spell that for me
D	TL	C	{!;(he);!} Hello is this reservations for hotels
D, SV	TL	C	Yeah [um] my name is [ah] Tex and I'm stuck down here at Kyoto Station
D, SV	TL	C	OK so I go straight up the stairs in front of me [uh]
D, SV	TL	I	Yes [ah] go out from the gate ten where you are
D,SV	MM	I	If we could [ah] we like to be [ah] making such arrangement for you also
S	MM	I	They should have Japanese style rooms
S	MM	I	July the seventh through fourteenth for two in one twin room
S	TL	I	{C1;You;C1} should tell him the name of the temple which is nanzenji
S	TL	C	is it possible to get a room in a hotel close to the conference center
S	TL	I	K fifteen thousand per night including breakfast
S	TL	I	[ah] Five minutes [ah] by taxi from the [ah] conference center
S	TL	I	[ah] Then a double bed room will be fine

S	TL	C	OK so this is like at the bottom of the stairs there's a gate on the left
S	TL	C	Yeah that's right I'd like to do some sightseeing the day after the conference before I leave
S	TL	I	[ah] You have to go to number five bus stop number five
S,SV	MM	C	I'm calling from the Kyoto Station and I'm looking for a hotel reservation near the International Conference Center
SC	MM	C	[um] I'll go by bus and train
SV	MM	I	...you should be able to enjoy jogging around there
SV	MM	I	You will see many taxis waiting there so you can get on the taxi there
SV	MM	C	OK yeah I can see the top of the stairs there
SV	MM	I	All right then I'll make sure you will have your reservation at the Royal Hotel
SV	MM	I	Do you know where you can get on a taxi at the Kyoto Station
SV	MM	I	{C1;But;C1} [ah] the Royal Hotel will be more convenient for the transportation because you can get on the Keihan Line from the hotel
SV	TL	C	K [um] if you are staying in a room with one large bed
SV	TL	I	Well it has a very good environment and we're sure that you'll [um] enjoy some sightseeing around there too
SV	TL	I	And when you come to the end of the overpass [ah] you see the stairs to go down on your right hand side
SV	TL	I	{C;And;C} I'm sure that you'll be able to find some good Japanese restaurants also
SV	TL	I	[ah] Could you exit out from [ah] the shinkansen
SV	TL	I	That's right yes that's where you get on the bus
SV, D	MM	C	OK let me give you my number here then [ah]
SV, S	TL	C	...I'd [ah] like to make a reservation at a hotel nearby if possible

I.3. Conversations bilingues traduites M0z (C3)

D	TL	C	OK, [um] should I make a [um] deposit with my credit card?
S	MM	C	/breath/ OK, do I wait for bus number three at the sightseeing bus stop
S	MM	W	Yes, OK adult single room
S, CO	MM	C	I would like a small Japanese style hotel near shops and restaurants please
SC	MM	C	Can I do it all over again?!
SC	TL	C	[ah] Taxi
SC	TL	W	You can go in about twenty minutes
SC	TL	W	Will you come by bus or taxi ?
SC	TL	C	OK, [ah] that's all right .
SV	MM	C	Can you please draw the directions on your map ?
SV	MM	C	Can you see my location now ?
SV	MM	C	OK, can you book me a room for three nights , starting tonight?
SV	MM	C	OK, [ahum] what do I tell them, where do I tell the taxi to go?
SV	MM	C	/ls/ Yes, I see the map, so you want me to follow thi arrow into (thi) this taxi stop ?
SV	MM	C	I'm standing at thi mark near thi Kintetsu Line
SV	MM	C	[ah] How much does a taxi cost to thi Conference Center ?
SV	TL	C	...so how long would it take from the station here ? (cf p 74)
SV	TL	W	Yes, please ride bus five to Sanjo Keihan Station
SV,S	MM	C	/ls/ So, also I need information about a hotel for the weekend

I.4. Effet de la répétition sur l'ambiguïté

La première colonne du tableau décrit l'effet de la répétition sur l'ambiguïté : elle peut disparaître, apparaître, ou bien demeurer.

SC→∅	TL	C: Is that M Y A K O, Myako WOZ: Please repeat C: How do you spell Miyako?>
∅→ PH→PH	TL	C: [ah] When [ah] can I check in? (it's [ah]) It's morning now WOZ: Please repeat C: When can I check in to the hotel, what time? WOZ: Please repeat C: What time can I check in to the hotel?
SC→∅	TL	C: One hyphen, four zero two hyphen, seven seven two hyphen, six three nine eight WOZ: Please repeat C: One, four o six, four zero, pardon me, I'll begin again. One, four zero six,
PH→PH	TL	C: All right, I would like to check in to the hotel on October twenty-fifth and I would like to check out of the hotel on October twenty-seventh WOZ: Please repeat C: I would like to check in to the hotel on October twenty-fifth and I would like to check out of the hotel on October twenty-seventh
SV→ SV→ ∅→∅	MM	C: OK, [um] what do I tell them, where do I tell the taxi to go? WOZ: Please repeat C: OK, when I take the taxi, where do I tell the taxi to go? ![oh] that one, OK! WOZ: Please repeat C: /laugh/ [hum] /ls/ OK, I see which taxi stand. [um] (what) Which building do I say to the taxi driver? WOZ: Please repeat C: [um] /ls/ When I'm in the taxi, (how) [ah] where does the taxi go to?
SV→SV, SV→∅	MM	C: OK, can you book me a room for three nights, starting tonight? WOZ: Please repeat C: OK, I need a room for three nights. Can you book?
PH→∅	TL	C: That's correct. Does it come with meals, /ls/ how many meals? WOZ: Please repeat C: /ls/ That's correct. How many meals a day are included in the price?
P→ ∅→∅	MM	C: But that says Keage. Is Keage Keitsu? Are they the same? WOZ: Please repeat C: [um] I don't see Keitsu on the window. Where is Keitsu? (can you sh) WOZ: Please repeat C: [um] Where is the Keitsu Line? Can you show me on the window?
∅→D	MM	C: OK, great. Thank you. [um] I want to rent a hotel room. [ah] Can you help me get a hotel room? WOZ: Please repeat C: [ah] Can you help me with a hotel reservation?
S, S→ ∅→S	MM	C: [um] Actually (I would) I would like a traditional Japanese hotel , a traditional Japanese inn. Is there one available? WOZ: Please repeat C: OK. [um] I would like to stay in a Japanese inn if possible WOZ: Please repeat C: [um] I want to stay in a traditional Japanese hotel
CO→ ∅→∅	MM	C: [um] (is it OK if I) Can I arrive and leave my luggage (in the) in the lobby? WOZ: Please repeat C: [ah] Can I drop off my luggage so I don't have to carry it around all day? WOZ: Please repeat C: [ah] Can I bring my suitcase before three o'clock?

SC→∅	TL	C: [ah] (is it thi) Is it a straight walk or should I take a taxi or bus ? WOZ: Please repeat C: (what's the) What's the best way to get there?
∅→SV	MM	C: I'm at thi Kintetsu Line. I'm putting a mark where I'm standing WOZ: Please repeat C: I'm standing at thi mark near the Kintetsu Line
SV→∅	MM	C: /ls/ Yes, I see the map, so you want me to follow the arrow into (thi) this taxi stop ? WOZ: Please repeat C: I can follow this line to the taxi stop ?
SV→ S, SV	MM	C: {l;l} am at Kyoto Station and I need to get some directions to a conference in Kyoto . Can you help me? WOZ: Please repeat C: Hello, /ls/ I want to find out how to get to thi International Conference Center in Kyoto
PH→PH	MM	C: OK, so I will check out October twenty-seventh (a) at ten a.m. WOZ: Please repeat C: I will check out October twenty-seventh at ten a.m.
SV→SV	MM	C: So I have a reservation October twenty-fifth and twenty-sixth at the Miyako Hotel . Is that correct? WOZ: Please repeat C: I have a reservation October twenty-fifth and twenty-sixth at Miyako Hotel . Is that correct?
P→P	TL	C: Walk twenty meters, turn right at the gas station, and then I will see the Conference Center? WOZ: Please repeat C: After I turn right , where do I go?
D→∅	MM	C: [um] OK, (l) I also need to [um] book a hotel room. Can you please advise me on a hotel room? WOZ: Please repeat C: Can you please advise me on a hotel room reservation?
SV→ D→SV	MM	C: [um] /ls/ OK, that sounds good, (l) I see where it is. [uh] Can you please make a reservation for me ? WOZ: Please speak slowly C: Yeah, (l'd like to like) I'd like to make a reservation there then WOZ: Please repeat C: I would like to make a reservation for that hotel
D→∅, SV→SV	TL	C: [ah]OK, [um] [um] (kay) (l I'd, l'd l'd also) I also need to book a hotel room. Can you please [ah] advise me on a hotel room near the Conference Center ? WOZ: Please repeat C: [ah] (l need a) I need to book a hotel room near the Conference Center . Can you help me? WOZ: Yes, there is thi Kyoto Royal Hotel
SV→SV	TL	C: [um] OK, (can) Can you please make a reservation for me for five nights please WOZ: Please repeat C: Can you please make a reservation for me for five nights ?
SV→∅	TL	C: [um] [well] [ah] /ls/ I'd like to go there now, so how long would it take from the station here ? WOZ: Please repeat C: [ah] How long would it take to get there from the station?
S→S	MM	C: So, [um] what is thi address of thi International Conference Center ? WOZ: Please repeat C: What is thi address of the International Conference Center ?
S→S	MM	C: Actually, I would like to stay in a very traditional Japanese style hotel . Can you recommend one? WOZ: Please repeat

		C: I would like to stay in a traditional Japanese hotel
S→∅	TL	C: Yes, so one single room from October [um] twenty-fourth through the twenty-eighth WOZ: Please repeat C: OK, [um] yes, a single room for one person
SV→∅	MM	C: OK, yeah, that's good. Yeah, please [ah] book me a room for four nights WOZ: Please speak slowly C: [ah] Yeah, that's good
PH→∅	MM	C: [ah] Yeah I'd like to check out October thirtieth WOZ: Please speak slowly C: [ah] Until the thirtieth
SV→∅	TL	C: (e) Excuse me, change lines to where? WOZ: Please repeat C: [ah] (do where) Where do you want me to change, at Keihan Station I get off the bus and then what do I do?
SV→ ∅→∅	TL	C: So (I) I would like to [um] spend [ah] maybe one hundred dollars a night {I2;if possible;I2} WOZ: {C2;Please speak;C2} slowly C: [ah] Yes, I'm looking for a place, maybe [ah] about one hundred dollars a night WOZ: Please speak slowly C: [ah] Yes, I would like a hotel that costs about one hundred dollars a night
SV→ SV→SV	TL	C: OK, [ah] very good. [ah] Can you please book me a room for tonight, [ah] let's see, (tomo) tonight, [ah] and for the next three nights? WOZ: Please repeat C: [ah] Yes, (I would like to book a room in the Kyo) I would like you to book me a room in thi Kyoto Hotel for tonight through the {I;night of;I} WOZ: {C;Please speak;C} slowly C: OK, [ah] yes, please book a room from October twenty-fifth to October twenty-ninth
PH→∅	TL	C: [ah] (I) I think I'll pass on breakfast WOZ: Please repeat C: [um] [ah] (no) No thank you, (I) I don't want breakfast
SV→SV	MM	C: I would like to find out how I can get to the Conference Center from Kyoto Station please WOZ: Please repeat C: I would like some information about how to get to the Conference Center from Kyoto Station please

Annexe II

Données pour l'évaluation pilote en compréhensibilité des questions

II.1. Texte lu par les sujets

I will be giving you information about traveling to the conference center and about various sightseeing spots in Kyoto. First, I will tell you about travel options and later I'll tell you about sightseeing. So let me **tell you what to do about traveling now**. **SVE** If you will be coming by Shinkansen, please be sure you have all your luggage with you and check your seat number before you leave the train. Then **exit out from the Shinkansen**. **PHH** Once you do that, you will walk down the platform toward the exits. On your left will be the offices where you can buy tickets for the special express trains, the bullet train, or tourist charter trains. Walk past that **train ticket vendor**. **SE** Then look for your exit. The exits have either names, such as South Central, or numbers, such as Exit 3. They are clearly marked above the exit door. So you should be able to see the sign saying the **exit with the number, or the name**. **COH** Follow the signs for Exit 14. At this point, you are on the lowest level of the station. Please take the escalators up. As you go up, you will reach one level that is built like a set of bleachers, in steps, and then two platforms that are level. You can **catch a taxi at the second level platform**. **SCH** When you arrive at this level, there will be additional signs for Exit 14. The exit you want to take is the South Central, right next to Exit 14. So, please **exit by Exit 14**, through South Central. **DH** Once you are at this platform, please continue through the main doors of the station which should be directly in front of you. At this point, you must be careful of your luggage. If you have a large suitcase, **it may be difficult to get out of Kyoto Station**. **PHH** If you have trouble getting through the door with any large luggage, there is a special luggage exit, one door to your left. You should have no trouble with your luggage if you just **take that one door to your left**. **SCE** You will have to go over a drainage conduit cover embedded in the floor of the station just past the exit. It is a little bumpy, and your luggage wheels may get caught, so you may want to be sure to **go over it carefully**. **PHH** Now, once you are out of the station, you have several options for getting to the conference center. You can take a subway, but it is a bit complicated. Most of our customers prefer to taxi to the conference center. I will assume that you would also prefer to taxi, though **you could take a bus or taxi**. **SCH** Once you have left Kyoto station by Exit 14, you should see the taxi stand directly in front of you. The taxis wait here in line to take each customer in the order that they come out from the station. **You will see many taxis waiting**. **SVE** Walk to the front of the line and you will see a sign that says "boarding area." This is the area **where you get on the taxi**. **PHE**. There are a number of different kinds of taxis and they will be different colors. The red ones are taxis from the conference center. Just tell the driver of a red cab where you want to go, and the trip should take you five minutes **by taxi from the conference center**. **SH** Be sure to take a red cab. Once taxis get to the conference center, all other taxis must turn left and stop in a lot that is a little far from the center entrance. However, the red taxis don't have to turn left, they come **right**, to the front of the center. **PH** Because these are special conference center taxis, the conference center will pay part of the cost of the taxi. Please inquire at the conference information desk to find out the cost of the taxi to you and **the cost of the taxi to the conference center**. **DH** The conference information desk can also help you with information about your return. Please consult them to help you make reservations to **take the taxi back**. **PHE**

The area around the conference center is rich with cultural sights. There are numerous shrines and temples and the famous castle, Nijo Castle. The conference information desk can also help you with transportation to **these temples or castles like Nijo**. **COH** Most conference goers will not leave immediately after the conference. If you are free on the day after the conference, perhaps you would like to join the sightseeing tour arranged especially for conference participants at that time. Please notify the desk if you would like to take part in this **tour the day after the conference before you leave**. **SE** Or you might like to arrange your own

sightseeing. Although many taxi drivers in Japan speak English, not all of them do. Therefore, you may require help with communicating with your taxi driver. Please just ask anyone at the conference information desk. They know what to say and can help you with **where you should tell the taxi to go. DE**

If you do not wish, for some reason, to come to the center by taxi, please refer to the conference packet you received when you sent in your pre-registration. In it you will find a brochure. Consult it if you want to **look at** another method of travel. **PH** There are trains from the eastern terminal of the subway, or buses from Kyoto Station itself which can get you to the conference center. Please read the brochure concerning taking **a train or bus from the Station. COH**

I represent a travel service company. We specialize in helping visitors to Kyoto make arrangements for their stay here. We understand that you are coming to Kyoto for a conference. And **we'd also like to make such arrangements for you. DE** We have a number of agents, both human and mechanical. Our mechanical agents, or robots, can only display information for you graphically. Among our speaking agents, we have some who are French, some who are German and some who are English. I am an **English speaking agent. SCH**

To begin with, you have several options for making your arrangements. You can make a reservation through a foreign travel agency, a Japanese travel agency or by calling the hotel itself. However, out of all these options, **the most efficient one will be calling us to make your arrangements. SCE** Of course you may be calling some other places to help you as well. We would like to know who else you are calling after you call us so that we can know who our competitors are. For that reason, we'd like to know **where you are calling, from now. SVH** The information I will give you may be rather complex. We want to be sure that you fully understand it. So, please feel free to ask any questions you might have so that you can be sure to **get it down correctly. PE** Now I'd like to give you some information about making reservations for your hotel. At most Japanese hotels, you can make a reservation for a night including breakfast or for a night not including breakfast. The price is thirteen thousand yen for a night not including breakfast and **fifteen thousand for a night including breakfast. SH** Of course, hotels have rooms in both Japanese and Western styles. You can reserve a Japanese room with either a Japanese-style or a Western-style bath. However, the western-style rooms only have western-style baths. If you reserve a western-style room, **it will have a regular western bed and bath. COE** There are deluxe Japanese rooms available as well. These look out on their own private gardens. Often these gardens have ponds with decorative carp. You could be **eating your Japanese style breakfast with fish** outside your window. **SVH** We have a number of sizes of rooms to choose from also. You can arrange to have a suite with one bedroom or a suite with two bedrooms. Just let us know if you'd like a single bedroom suite or a **double bedroom. SH** Some hotels have extensive recreational equipment. The Miyako, for instance, has an Olympic size swimming pool where many guests swim laps for exercise, and a gym complete with weights and exercise bicycles. You might enjoy staying at such a hotel where **there is a swimming pool and also a gym where you can train. COE**

Once you have chosen the hotel you'd like to stay in, we will handle your reservation, but we require a deposit. The deposit must be in yen, and it *must* be made by postal money order or by bank transfer. No cash or credit cards, please. Although you cannot make the deposit with a credit card, you can pay the final bill with a credit card. If you would like to pay the final bill with your credit card, we will need to see your credit card before you register. For that reason, please let us have the **deposit with your credit card. DH** Because we are particular about our clients, we assume our clients are particular, too. For that reason, you will have an opportunity to check out the hotel and be sure that it is satisfactory before you register. Please let us know when you would like to check out your hotel. Most guests plan to **check out** the day before they register. **PH**

Or, if you would rather make your own reservation, you can do that, too. Right next to the conference center where your conference is being held is a Tourist Information and Hotel Reservation Office with a number of **stories. PE** They can help you make reservations for hotels anywhere in the city. So it is very easy for you to **make a reservation for a hotel near the conference center. SVH**

II.2. Libellé des questions de désambiguïisation

Human	Machine
let me tell you what to do about traveling now let me tell you now about traveling let me tell you about your forthcoming travel	let me tell you what to do about traveling now now, let me tell you what to do about traveling let me tell you what to do about (traveling now)
exit out from the Shinkansen to out of the Shinkansen area	exit out from the Shinkansen to leave from the Shinkansen

to exit the Shinkansen itself	to leave the shinkansen
that train ticket vendor someone selling train ticket someone selling tickets to buy a train	that train ticket vendor that vendor for train ticket that ticket vendor for train
the exit with the number, or the name the name the exit with the name	the exit with the number, or the name (the exit with the number) or (the name) (the exit with the number) or (the exit with the name)
level the platform is on the second level the platform itself is level	level a platform that is (second level) a (second platform) that is level
exit by Exit 14, through South Central get out near Exit 14 get out through Exit 14	exit by Exit 14, through South Central exit next to Exit 14 exit through Exit 14
to get out of Kyoto Station to get yourself out of Kyoto Station to take something else out of Kyoto Station	to get out of Kyoto Station to leave Kyoto Station to get (out of Kyoto Station)
that go through that door to your left take that through the door to your left	that take (that one door) to your left take (that) one door to your left
go over it carefully look it over and consider it carefully walk or move above and over carefully	go over it carefully to examine it carefully to go (over it) carefully
taxi you could take a bus or a taxi you could take a bus or you could taxi	taxi you could take a taxi you could taxi
you will see many taxis waiting while you are waiting, you will see many taxis you will see many taxis that are waiting	you will see many taxis waiting waiting, you will see many taxis you will see many (taxis waiting)
get on the taxi actually climb into the taxi arrive at somewhere on the taxi	get on the taxi board the taxi get (on the taxi)
by taxi from the conference center the trip is five minutes from the conference center the taxi is the conference center's property	by taxi from the conference center from the conference center, five minutes by taxi by (taxi from the conference center); five minutes
right on the right side, not on the left straight there; no detour	right opposite to the left directly
the cost of the taxi to the conference center the taxi is going to the conference center The conference center will pay for the taxi	the cost of the taxi to the conference center the cost of the taxi towards the conference center the cost of the taxi in case of the conference center
take the taxi back take the taxi and return it take the taxi and go back	take the taxi back return the taxi take the taxi (back)
these temples or castles like Nijo these temples these temples like Nijo	these temples or castles like Nijo (these temples) or (these castles like Nijo) (these temples like Nijo) and (these castles like Nijo)
the day after the conference before you leave it the day before you leave it the day after the conference which is right before you leave	the day after the conference before you leave (the day before you leave) after the conference the day after (the conference before you leave)
where you should tell the taxi to go where you should speak where the taxi should go	where you should tell the taxi to go at what place you should tell the taxi to go to what place you should tell the taxi to go
look at actually looking with your eyes think about or think over	look at to see to consider
a train or bus from the station the bus only is from the station the train and the bus are from the station	a train or bus from the station (a train) or (a bus from the station) (an train from the station) or (a bus from the station)

<p>arrangements for you the arrangements are for you the arrangement are made on your behalf, possibly by someone else</p>	<p>arrangements for you arrangements appropriate to you arrangements instead of you</p>
<p>speaking someone who speaks English someone who is English and who is speaking</p>	<p>speaking an agent that is (English speaking) A (speaking agent) that is English</p>
<p>calling the most efficient option will be to call us to make the arrangement the most efficient person will be calling us to make the arrangement</p>	<p>calling the most efficient one (will call) us to make the arrangement the most efficient one (will be) calling us to make the arrangement</p>
<p>where are you calling from now where are you calling from at this point of time from now, where are you calling to</p>	<p>where are you calling from now now, where are you calling from from now, where are you calling</p>
<p>get down to make sure something is put down to get a complete understanding</p>	<p>get down to lower to understand</p>
<p>fifteen thousand for a night including breakfast fifteen thousand for a night which happens to also include breakfast fifteen thousand for a particular kind of night which includes breakfast</p>	<p>fifteen thousand for a night including breakfast including breakfast, fifteen thousand for a night for (a night including breakfast), fifteen thousand</p>
<p>a western bed and bath a bath a western bath</p>	<p>a western bed and bath (a western bed) and (a bath) (a western bed) and (a western bath)</p>
<p>You could be eating you Japanese style breakfast with fish outside your window your Japanese breakfast includes fish the fish are outside the window</p>	<p>You could be eating you Japanese style breakfast with fish outside your window outside your window, you could be eating you Japanese style breakfast with fish with fish outside your window, you could be eating you Japanese style breakfast</p>
<p>a double bed-room two bedrooms a room with a double bed</p>	<p>a double bed-room a double (bedroom) a (double bed) room</p>
<p>a swimming pool and also a gym where you can train a swimming pool a swimming pool where you can train</p>	<p>a swimming pool and also a gym where you can train (a swimming pool) and (a gym where you can train) (a swimming pool where you can train) and (a gym where you can train)</p>
<p>the deposit with your credit card use your credit card for the deposit let us have your deposit and your credit card</p>	<p>the deposit with your credit card the deposit using your credit card the deposit and your credit card</p>
<p>check out to settle the bill and leave the hotel to look into or investigate something</p>	<p>check out to pay and leave the hotel to verify</p>
<p>story like fairy tales or legends different levels or floors</p>	<p>story narrative floor</p>
<p>it is very easy to make a reservation for a hotel near the conference center the reservation takes place near the conference center the hotel is near the conference center</p>	<p>it is very easy to make a reservation for a hotel near the conference center near the conference center, it is very easy to make a reservation for a hotel it is very easy to make a reservation for (a hotel near the conference center)</p>

Annexe III

Données pour la seconde évaluation en compréhension des questions

III.1. Texte lu par les sujets

Bill had to go to the library to get a book. But he didn't want to have to pay **the cost of the taxi to the library**. **D** So he decided to drive himself. However, he wasn't sure how to get there, so he called his friend Tom. Tom gave him the directions over the telephone and told Bill to **take them down** **PH** carefully. Tom told Bill to **go by the highway** **D**, and then get off at the Charles Street exit. But Bill told Tom that he hated taking the highway; the cars and the trucks drove so quickly and made so much noise. He especially didn't like all the **noisy cars and trucks** **CO**. So Tom gave him different directions, using a much longer route so that he could avoid the highway. The directions were so complicated that they had to **go over them** **PH** twice before Bill understood. They **talked about an hour** **D** before they were finished.

The next day, Bill drove to the library. He found the sign Tom had told him about--the sign **in front of the gas station with the red roof**. **S** He looked to the left, as Tom had told him to, and there he saw the library building with a number of **stories** **p**. On the third floor, he found the book that he wanted, but the librarian was nowhere to be seen. Bill decided to go home, but he **met the librarian as he was leaving** **SV**. He signed out the book and took it home with him. The trip took him all day, and he was exhausted when he got home. He called Tom and **told him he had had a terrible time that day** **SV**.

There was going to be a big Medieval Arts festival that night in the nearby city of Newton, and Susan really wanted to go. She didn't **have a date** **P** to go to the festival, though, so she wondered what to do. Finally she thought she might call up her **Old English professor** **S** and see if he wanted to go with her. He hadn't heard about the festival and told her that he would **look into it** **PH**. He called back an hour later to say that he had talked to a friend about it and that it sounded quite interesting. He heard that everything there would be medieval: medieval entertainment and **medieval food and drink** **CO**. He agreed to pick her up at 8:00, so she started getting ready.

She searched through her closet for something appropriate to wear. She found a cape, but she also wanted a long dress that would match it. Finally she found a dress that would **go with** **P** the cape. They both were made from velvet and the cape had a feather collar. She put on **the dress and the cape with the feather collar** **CO** and waited for her professor. He arrived with a **bottle of wine from France** **SV**, so they had a glass of wine before they left.

The festival was easy to find--the area it was in was brightly lit and the music could be heard from a long way off. It was a warm night, so she **deposited her cape with the coat check clerk** **D** and she and the professor went to see what there was to eat. The food stalls were lined up along the river under a string of white lights, so they headed for the **bank** **p**. There they found a wide variety of foods available. It was early, so some of the stalls weren't quite open, but they found one that advertised a special French style of mutton. The professor admitted that he was quite fond of **the French school of cooking** **S** and so they went to see if they could get some mutton stew. The **lamb did look ready to eat** **D**, so they ordered a dish to share between the two of them. The stew also contained chestnuts and cranberries, and was delicious. They enjoyed walking along the river **eating their mutton stew with cranberries**. **SV**

They were still hungry, so they continued to wander among the food stalls. Other people were also strolling along the river. Some were eating cakes and **some were eating apples SC**. But soon they got tired, so they walked back to the entrance, picked up her cape, and went home.

It is London, 1943. People's lives are disrupted daily by the air raids. But they still try to carry on as if everything were normal. Joe was doing research in using thin film for information storage before the war; now he uses his knowledge to help the war effort. He works as a **thin film technologist S** in a research university.

Every day, he gets up, **washes his face and dresses SC** as if life were usual. He tries to help out around the house as much as he can. Food is hard to get, but potatoes are always available. Luckily, he **likes potatoes and cooks SC** a lot so that his wife can have a break. He and his wife work hard to keep their house clean after each bombing run that affects their neighborhood. They **wash the floor and dust. SC**

Joe's wife also helps the war effort. She goes every day to a nearby school which has been turned into a shelter for the homeless. There she takes care of **children and aged people who don't have a home. CO**

Joe's son Georgie plays "guard;" he takes his Dad's binoculars and watches out his bedroom window to see if any planes come flying over. Once he saw a **plane with the binoculars SV**. His Daddy told him that **flying planes can be dangerous SC**, so he ran to tell his parents. They **shut the dog up P** in the basement so that he wouldn't run out into the yard, and they all went down there themselves, too. Georgie asked why they had to go into the basement, but Joe didn't want to scare him, so he decided to just brush it off and not to **go into it PH** too much. "**Take it from me PH**, Georgie," said his father, "this is the safest place to be."

The family tries to do something nice each weekend. Joe's wife makes a picnic lunch with whatever is available at the time--any kind of **bread and cheese that they can get CO**. They are a little afraid to go too far away from a shelter, so they try to find a picnic **place near a house with a basement S** where they can go if there is any trouble. They lead as happy a life as they can under the circumstances.

III.2. Libellé des questions de désambiguïisation

Human	Machine
let me tell you what to do about traveling now let me tell you now about traveling let me tell you about your forthcoming travel	let me tell you what to do about traveling now now, let me tell you what to do about traveling let me tell you what to do about (traveling now)
exit out from the Shinkansen to out of the Shinkansen area to exit the Shinkansen itself	exit out from the Shinkansen to leave from the Shinkansen to leave the shinkansen
that train ticket vendor someone selling train ticket someone selling tickets to buy a train	that train ticket vendor that vendor for train ticket that ticket vendor for train
the exit with the number, or the name the name the exit with the name	the exit with the number, or the name (the exit with the number) or (the name) (the exit with the number) or (the exit with the name)
level the platform is on the second level the platform itself is level	level a platform that is (second level) a (second platform) that is level
exit by Exit 14, through South Central get out near Exit 14 get out through Exit 14	exit by Exit 14, through South Central exit next to Exit 14 exit through Exit 14
to get out of Kyoto Station to get yourself out of Kyoto Station to take something else out of Kyoto Station	to get out of Kyoto Station to leave Kyoto Station to get (out of Kyoto Station)
that go through that door to your left take that through the door to your left	that take (that one door) to your left take (that) one door to your left
go over it carefully look it over and consider it carefully walk or move above and over carefully	go over it carefully to examine it carefully to go (over it) carefully
taxi you could take a bus or a taxi you could take a bus or you could taxi	taxi you could take a taxi you could taxi

you will see many taxis waiting while you are waiting, you will see many taxis you will see many taxis that are waiting	you will see many taxis waiting waiting, you will see many taxis you will see many (taxis waiting)
get on the taxi actually climb into the taxi arrive at somewhere on the taxi	get on the taxi board the taxi get (on the taxi)
by taxi from the conference center the trip is five minutes from the conference center the taxi is the conference center's property	by taxi from the conference center from the conference center, five minutes by taxi by (taxi from the conference center); five minutes
right on the right side, not on the left straight there; no detour	right opposite to the left directly
the cost of the taxi to the conference center the taxi is going to the conference center The conference center will pay for the taxi	the cost of the taxi to the conference center the cost of the taxi towards the conference center the cost of the taxi in case of the conference center
take the taxi back take the taxi and return it take the taxi and go back	take the taxi back return the taxi take the taxi (back)
these temples or castles like Nijo these temples these temples like Nijo	these temples or castles like Nijo (these temples) or (these castles like Nijo) (these temples like Nijo) and (these castles like Nijo)
the day after the conference before you leave it the day before you leave it the day after the conference which is right before you leave	the day after the conference before you leave (the day before you leave) after the conference the day after (the conference before you leave)
where you should tell the taxi to go where you should speak where the taxi should go	where you should tell the taxi to go at what place you should tell the taxi to go to what place you should tell the taxi to go
look at actually looking with your eyes think about or think over	look at to see to consider
a train or bus from the station the bus only is from the station the train and the bus are from the station	a train or bus from the station (a train) or (a bus from the station) (an train from the station) or (a bus from the station)
arrangements for you the arrangements are for you the arrangement are made on your behalf, possibly by someone else	arrangements for you arrangements appropriate to you arrangements instead of you
speaking someone who speaks English someone who is English and who is speaking	speaking an agent that is (English speaking) A (speaking agent) that is English
calling the most efficient option will be to call us to make the arrangement the most efficient person will be calling us to make the arrangement	calling the most efficient one (will call) us to make the arrangement the most efficient one (will be) calling us to make the arrangement
where are you calling from now where are you calling from at this point of time from now, where are you calling to	where are you calling from now now, where are you calling from from now, where are you calling
get down to make sure something is put down to get a complete understanding	get down to lower to understand
fifteen thousand for a night including breakfast fifteen thousand for a night which happens to also include breakfast fifteen thousand for a particular kind of night which includes breakfast	fifteen thousand for a night including breakfast including breakfast, fifteen thousand for a night for (a night including breakfast), fifteen thousand
a western bed and bath a bath a western bath	a western bed and bath (a western bed) and (a bath) (a western bed) and (a western bath)

<p>You could be eating you Japanese style breakfast with fish outside your window your Japanese breakfast includes fish</p> <p>the fish are outside the window</p>	<p>You could be eating you Japanese style breakfast with fish outside your window outside your window, you could be eating you Japanese style breakfast with fish with fish outside your window, you could be eating you Japanese style breakfast</p>
<p>a double bed-room two bedrooms a room with a double bed</p>	<p>a double bed-room a double (bedroom) a (double bed) room</p>
<p>a swimming pool and also a gym where you can train a swimming pool a swimming pool where you can train</p>	<p>a swimming pool and also a gym where you can train (a swimming pool) and (a gym where you can train) (a swimming pool where you can train) and (a gym where you can train)</p>
<p>the deposit with your credit card use your credit card for the deposit let us have your deposit and your credit card</p>	<p>the deposit with your credit card the deposit using your credit card the deposit and your credit card</p>
<p>check out to settle the bill and leave the hotel to look into or investigate something</p>	<p>check out to pay and leave the hotel to verify</p>
<p>story like fairy tales or legends different levels or floors</p>	<p>story narrative floor</p>
<p>it is very easy to make a reservation for a hotel near the conference center the reservation takes place near the conference center the hotel is near the conference center</p>	<p>it is very easy to make a reservation for a hotel near the conference center near the conference center, it is very easy to make a reservation for a hotel it is very easy to make a reservation for (a hotel near the conference center)</p>

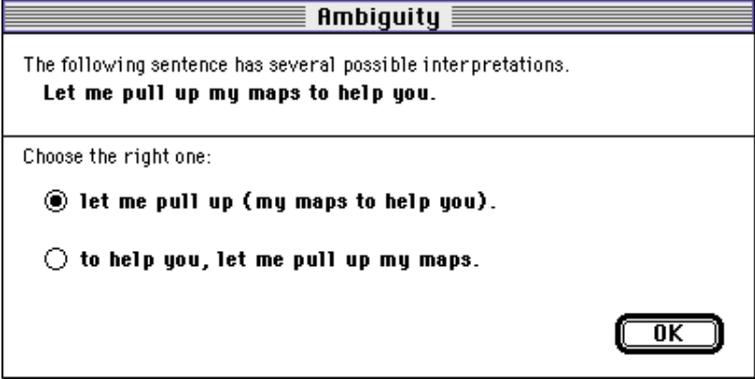
III.3. Réponses aux questions

Le tableau donne le nombre de bonnes réponses à chaque question en fonction du type des questions.

Question	Classe	Nombre de bonnes réponses			
		TH	SH	TM	SM
taxi	D	15	13	10	13
takedown	PH	15	15	14	14
goby	D	15	15	15	15
noisy	C	15	15	15	14
goover	PH	15	13	15	15
talkabout	D	15	14	14	14
sign	S	15	15	14	15
stories	P	15	15	15	15
leaving	SV	15	15	15	15
told	SV	14	8	15	14
date	P	15	15	15	15
professor	S	12	12	10	11
lookinto	PH	15	15	15	15
medieval	CO	15	15	15	15
gowith	P	15	15	14	15
collar	CO	15	15	15	14
wine	SV	11	15	15	14
deposit	D	15	15	15	15
bank	P	15	15	15	15
school	S	15	15	7	2
lamb	D	15	15	15	15
cranberries	SV	15	15	13	13
apples	SC	15	15	4	0
technologist	S	15	15	15	15
face	SC	15	15	15	15
potatoes	SC	15	15	15	15
floor	SC	15	15	15	15
children	CO	7	13	13	13
binoculars	SV	15	15	15	15
flying	SC	15	14	14	15
shutup	P	15	15	14	15
gointo	PH	15	15	15	14
takefrom	PH	15	15	15	15
bread	CO	14	15	14	15
picnic	S	14	15	15	15

Ambiguïtés prises en compte dans le module de désambiguïsation de l'anglais

IV.1. Attachement prépositionnel avec une proposition subordonnée



The following sentence has several possible interpretations.
Let me pull up my maps to help you.

Choose the right one:

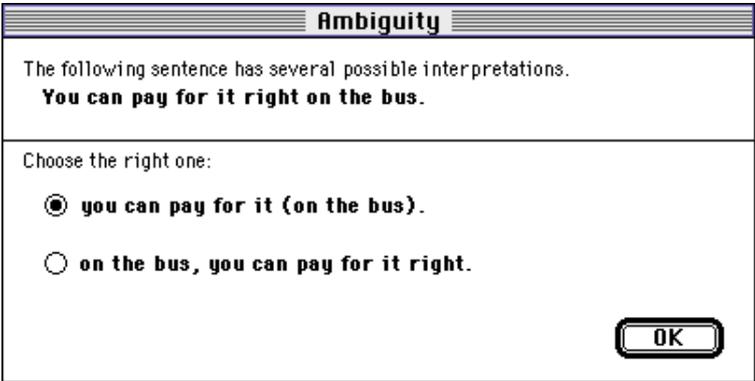
let me pull up (my maps to help you).

to help you, let me pull up my maps.

OK

Figure IV-1 : Dialogue de résolution d'une ambiguïté d'attachement prépositionnel avec une proposition subordonnée

IV.2. Attachement prépositionnel ou adverbial avec la proposition principale



The following sentence has several possible interpretations.
You can pay for it right on the bus.

Choose the right one:

you can pay for it (on the bus).

on the bus, you can pay for it right.

OK

Figure IV-2 : Dialogue de résolution d'une ambiguïté d'attachement prépositionnel avec la proposition principale

Ambiguity
<p>The following sentence has several possible interpretations. It says that here on my flyer.</p>
<p>Choose the right one:</p> <p><input checked="" type="radio"/> it says that (on my flyer).</p> <p><input type="radio"/> on my flyer, it says that here.</p>
<input type="button" value="OK"/>

Figure IV-3 : Dialogue de résolution d'une ambiguïté d'attachement prépositionnel avec la proposition principale

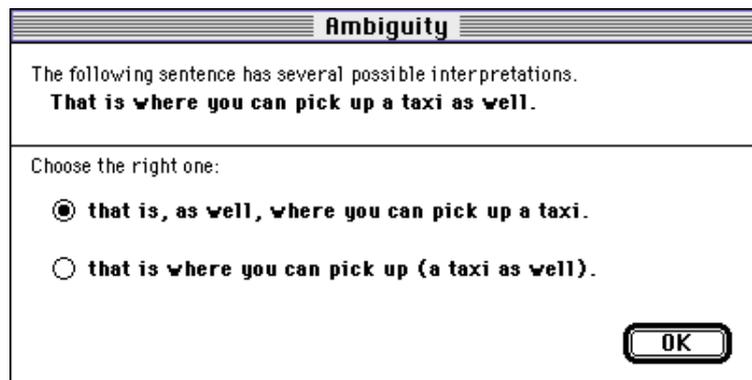
Ambiguity
<p>The following sentence has several possible interpretations. Where can I catch a taxi from Kyoto station?</p>
<p>Choose the right one:</p> <p><input checked="" type="radio"/> where can I catch (a taxi from Kyoto Station).</p> <p><input type="radio"/> from Kyoto Station, where can I catch a taxi.</p>
<input type="button" value="OK"/>

Figure IV-4 : Dialogue de résolution d'une ambiguïté d'attachement prépositionnel avec la proposition principale

Ambiguity
<p>The following sentence has several possible interpretations. Go across the street to the north of the station.</p>
<p>Choose the right one:</p> <p><input checked="" type="radio"/> go (across the street to the north).</p> <p><input type="radio"/> to the north, go across the street.</p>
<input type="button" value="OK"/>

Figure IV-5 : Dialogue de résolution d'une ambiguïté d'attachement prépositionnel avec la proposition principale

IV.3. Attachement adverbial avec une proposition relative



Ambiguity

The following sentence has several possible interpretations.
That is where you can pick up a taxi as well.

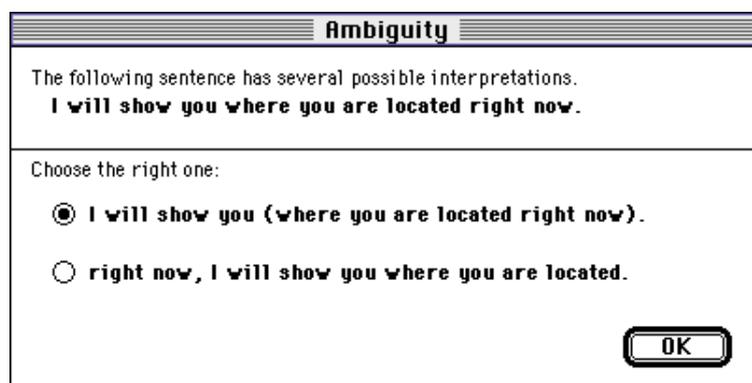
Choose the right one:

that is, as well, where you can pick up a taxi.

that is where you can pick up (a taxi as well).

OK

Figure IV-6 : Dialogue de résolution d'une ambiguïté d'attachement adverbial avec une proposition relative



Ambiguity

The following sentence has several possible interpretations.
I will show you where you are located right now.

Choose the right one:

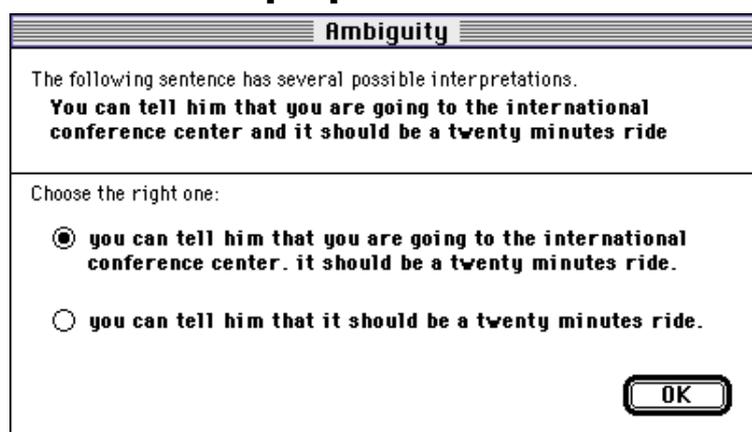
I will show you (where you are located right now).

right now, I will show you where you are located.

OK

Figure IV-7 : Dialogue de résolution d'une ambiguïté d'attachement adverbial avec une proposition relative

IV.4. Coordination de propositions subordonnées



Ambiguity

The following sentence has several possible interpretations.
You can tell him that you are going to the international conference center and it should be a twenty minutes ride

Choose the right one:

you can tell him that you are going to the international conference center. it should be a twenty minutes ride.

you can tell him that it should be a twenty minutes ride.

OK

Figure IV-8 : Dialogue de résolution d'une ambiguïté de coordination de propositions subordonnées

IV.5. Attachement de groupe nominal

Ambiguity
The following sentence has several possible interpretations. You are going to the international conference center.
Choose the right one:
<input checked="" type="radio"/> the (international center) for conference
<input type="radio"/> the center for (international conference)
<input type="button" value="OK"/>

Figure IV-9 : Dialogue de résolution d'une ambiguïté d'attachement de groupe nominal

Ambiguity
The following sentence has several possible interpretations. I want the symposium on interpreting telecommunication at the international conference center.
Choose the right one:
<input checked="" type="radio"/> the symposium (on interpreting telecommunications at the international conference center)
<input type="radio"/> at the international conference center, the symposium on interpreting telecommunications
<input type="button" value="OK"/>

Figure IV-10 : Dialogue de résolution d'une ambiguïté d'attachement de groupe nominal

Ambiguïtés prises en compte dans le module de désambiguïstation du français

V.1 Ambiguïté lexicale

The dialog box has a title bar with the word "ambiguïté". The main text reads: "Il y a plusieurs sens acceptables pour le mot :
capitaine

Below this, it says "Choisissez le bon." followed by three radio button options:

- Officier qui commande une compagnie d'infanterie, un escadron de cavalerie, une batterie d'artillerie.**
- Officier qui commande un navire de commerce**
- Chef d'une équipe sportive**

An "OK" button is located in the bottom right corner.

Figure V-1 : Dialogue de résolution d'une ambiguïté de polysémie

V.2 Ambiguïté de classe morphosyntaxique

V.2.1 Coordination verbale

The dialog box has a title bar with the word "ambiguïté". The main text reads: "Il y a plusieurs interprétations pour la phrase :
Il atteint la grange et la ferme.

Below this, it says "Choisissez la bonne." followed by two radio button options:

- Il atteint la ferme.**
- Il ferme la grange.**

An "OK" button is located in the bottom right corner.

Figure V-2 : Dialogue de résolution d'une ambiguïté de coordination verbale

V.2.2. Ambiguïté de classe simple

ambiguïté	
Il y a plusieurs interprétations pour la phrase : Un savant compromis a été arrêté hier.	
Choisissez la bonne.	
<input checked="" type="radio"/>	Un savant (adjectif) compromis (nom) a été arrêté hier.
<input type="radio"/>	Un savant (nom) compromis (adjectif) a été arrêté hier.
<input type="button" value="OK"/>	

Figure V-3 : Dialogue de résolution d'une ambiguïté de classe simple – exemple 1

ambiguïté	
Il y a plusieurs interprétations pour la phrase : Devant cette somme, il hésite.	
Choisissez la bonne.	
<input checked="" type="radio"/>	Devant (verbe devoir : je dois) cette somme, il hésite.
<input type="radio"/>	Devant (préposition : en face de) cette somme, il hésite.
<input type="button" value="OK"/>	

Figure V-4 : Dialogue de résolution d'une ambiguïté de classe simple – exemple 2

V.3. Ambiguïté de géométrie

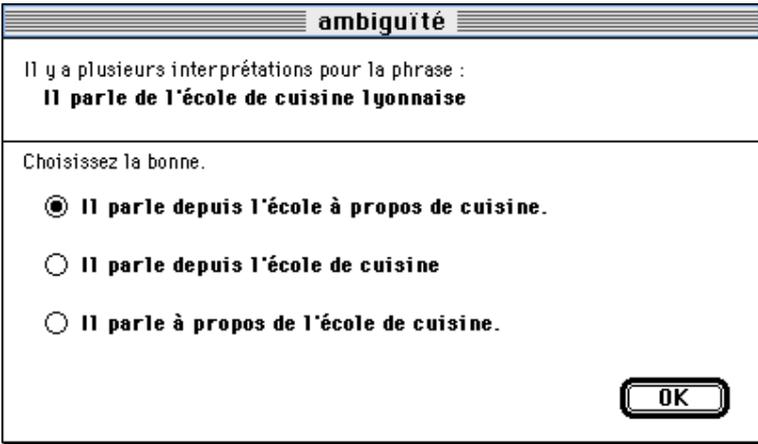
V.3.1. Structure argumentaire du verbe

Type 1

ambiguïté	
Il y a plusieurs interprétations pour la phrase : Il parle depuis l'école de cuisine.	
Choisissez la bonne.	
<input checked="" type="radio"/>	Il parle depuis l'école à propos de cuisine
<input type="radio"/>	Il parle depuis (l'école de cuisine)
<input type="button" value="OK"/>	

Figure V-5 : Dialogue de résolution d'une ambiguïté de structure argumentaire du verbe – type 1

Type 2



ambiguïté

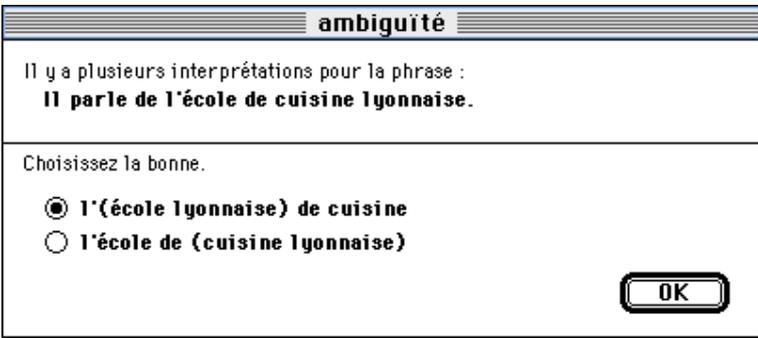
Il y a plusieurs interprétations pour la phrase :
Il parle de l'école de cuisine lyonnaise

Choisissez la bonne.

- Il parle depuis l'école à propos de cuisine.**
- Il parle depuis l'école de cuisine**
- Il parle à propos de l'école de cuisine.**

OK

Figure V-6 : Dialogue de résolution d'une ambiguïté de structure argumentaire du verbe – type 2



ambiguïté

Il y a plusieurs interprétations pour la phrase :
Il parle de l'école de cuisine lyonnaise.

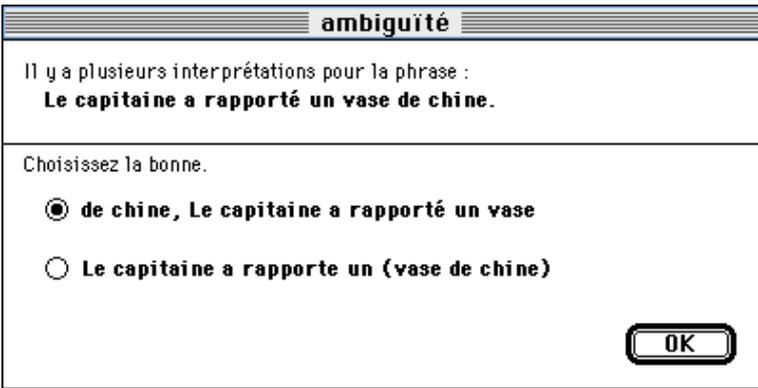
Choisissez la bonne.

- l'(école lyonnaise) de cuisine**
- l'école de (cuisine lyonnaise)**

OK

Figure V-7 : Dialogue de résolution d'une ambiguïté de subordination

Type 3, type 4, type 5



ambiguïté

Il y a plusieurs interprétations pour la phrase :
Le capitaine a rapporté un vase de chine.

Choisissez la bonne.

- de chine, Le capitaine a rapporté un vase**
- Le capitaine a rapporte un (vase de chine)**

OK

Figure V-8 : Dialogue de résolution d'une ambiguïté de structure argumentaire du verbe – type 3

ambiguïté
<p>Il y a plusieurs interprétations pour la phrase :</p> <p>Marie voit l'homme dans le parc avec un télescope.</p>
<p>Choisissez la bonne.</p> <p><input checked="" type="radio"/> dans le parc, Marie voit l'homme</p> <p><input type="radio"/> Marie voit l'(homme dans le parc)</p>
<input type="button" value="OK"/>

Figure V-9 : Dialogue de résolution d'une ambiguïté de structure argumentaire du verbe – type 3

ambiguïté
<p>Il y a plusieurs interprétations pour la phrase :</p> <p>Marie voit l'homme dans le parc avec un télescope.</p>
<p>Choisissez la bonne.</p> <p><input checked="" type="radio"/> avec un télescope, dans le parc, Marie voit l'homme</p> <p><input type="radio"/> (dans le parc avec un télescope), Marie voit l'homme</p>
<input type="button" value="OK"/>

Figure V-10 : Dialogue de résolution d'une ambiguïté de structure argumentaire du verbe – type 4

ambiguïté
<p>Il y a plusieurs interprétations pour la phrase :</p> <p>Marie voit l'homme dans le parc avec un télescope.</p>
<p>Choisissez la bonne.</p> <p><input checked="" type="radio"/> avec un télescope, Marie voit l'(homme dans le parc)</p> <p><input type="radio"/> Marie voit l'((homme avec un télescope) dans le parc)</p> <p><input type="radio"/> Marie voit l'(homme (dans le parc avec un télescope))</p>
<input type="button" value="OK"/>

Figure V-11 : Dialogue de résolution d'une ambiguïté de structure argumentaire du verbe – type 5

Type 6

ambiguïté

Il y a plusieurs interprétations pour la phrase :
Le magistrat juge les enfants coupables.

Choisissez la bonne.

Le magistrat juge les (enfants coupables).

Le magistrat juge coupables les enfants.

OK

Figure V-12 : Dialogue de résolution d'une ambiguïté de structure argumentaire du verbe – type 6

V.3.2. Coordination nominale

Type 1

ambiguïté

Il y a plusieurs interprétations pour la phrase :
On étudie l'évolution de la bourse et des investissements.

Choisissez la bonne.

On étudie des investissements

On étudie l'évolution des investissements

OK

Figure V-13 : Dialogue de résolution d'une ambiguïté de coordination nominale – type 1

Type 2

ambiguïté

Il y a plusieurs interprétations pour la phrase :
On étudie l'évolution de la structure du réseau et des investissements.

Choisissez la bonne.

On étudie des investissements

On étudie l'évolution des investissements

On étudie l'évolution de la structure des investissements

OK

Figure V-14 : Dialogue de résolution d'une ambiguïté de coordination nominale – type 2

Type 3

ambiguïté	
Il y a plusieurs interprétations pour la phrase : L'évolution de la structure du réseau et des investissements est étudiée.	
Choisissez la bonne.	
<input checked="" type="radio"/>	L'évolution des investissements
<input type="radio"/>	L'évolution de la structure des investissements
<input type="button" value="OK"/>	

Figure V-15 : Dialogue de résolution d'une ambiguïté de coordination nominale – type 3

V.3.3. Coordination adjectivale

Type 1

ambiguïté	
Il y a plusieurs interprétations pour la phrase : Il prend des cahiers et des classeurs noirs.	
Choisissez la bonne.	
<input checked="" type="radio"/>	(des cahiers noirs) et (des classeurs noirs)
<input type="radio"/>	(des cahiers) et (les classeurs noirs)
<input type="button" value="OK"/>	

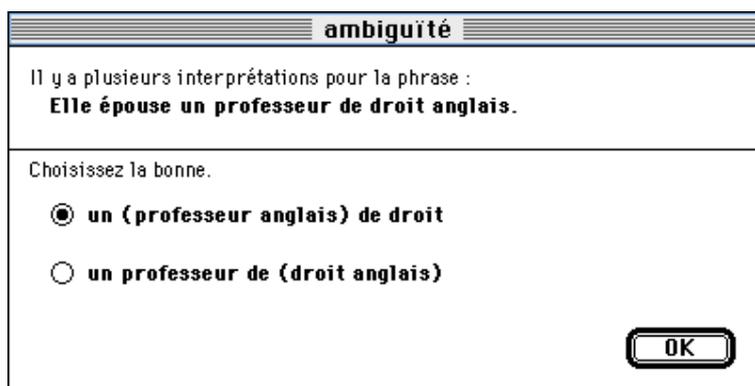
Figure V-16 : Dialogue de résolution d'une ambiguïté de coordination adjectivale – type 1

Type 2

ambiguïté	
Il y a plusieurs interprétations pour la phrase : Il prend des cahiers, des stylos et des classeurs noirs.	
Choisissez la bonne.	
<input checked="" type="radio"/>	(des cahiers noirs) et (des stylos noirs) et (des classeurs noirs)
<input type="radio"/>	(des cahiers) et (des stylos noirs) et (des classeurs noirs)
<input type="radio"/>	(des cahiers) et (des stylos) et (des classeurs noirs)
<input type="button" value="OK"/>	

Figure V-17 : Dialogue de résolution d'une ambiguïté de coordination adjectivale – type 2

V.3.4. Subordination



ambiguïté

Il y a plusieurs interprétations pour la phrase :
Elle épouse un professeur de droit anglais.

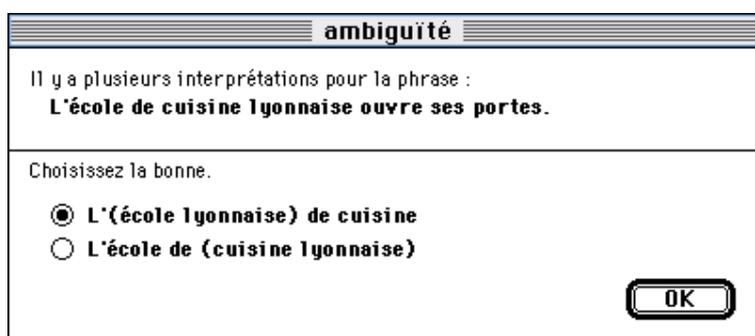
Choisissez la bonne.

un (professeur anglais) de droit

un professeur de (droit anglais)

OK

Figure V-18 : Dialogue de résolution d'une ambiguïté de subordination



ambiguïté

Il y a plusieurs interprétations pour la phrase :
L'école de cuisine lyonnaise ouvre ses portes.

Choisissez la bonne.

L'(école lyonnaise) de cuisine

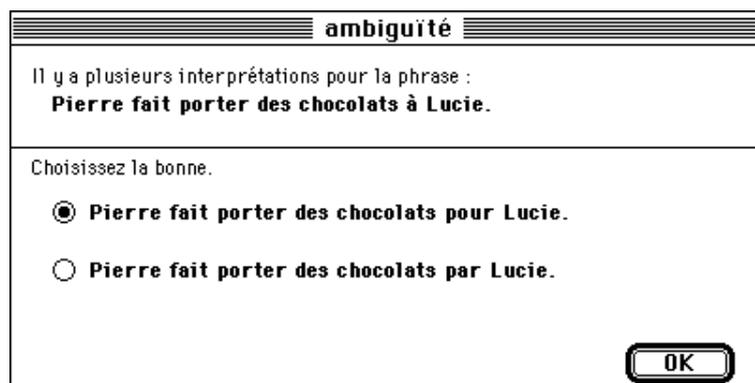
L'école de (cuisine lyonnaise)

OK

Figure V-19 : Dialogue de résolution d'une ambiguïté de subordination

V.4. Ambiguïté de décoration

V.4.1. Étiquetage logico-sémantique



ambiguïté

Il y a plusieurs interprétations pour la phrase :
Pierre fait porter des chocolats à Lucie.

Choisissez la bonne.

Pierre fait porter des chocolats pour Lucie.

Pierre fait porter des chocolats par Lucie.

OK

Figure V-20 : Dialogue de résolution d'une ambiguïté d'étiquetage logico-sémantique

V.4.2. Étiquetage syntaxique

The dialog box has a title bar with the word "ambiguïté" in the center. The main content area is divided into two sections. The top section contains the text: "Il y a plusieurs interprétations pour la phrase :
Je vous parle de la tour Eiffel." The bottom section contains the text: "Choisissez la bonne." followed by two radio button options: " **Je vous parle depuis la tour Eiffel.**" and " **Je vous parle à propos de la tour Eiffel.**" In the bottom right corner, there is an "OK" button.

Figure V-21 : Dialogue de résolution d'une ambiguïté d'étiquetage syntaxique

The dialog box has a title bar with the word "ambiguïté" in the center. The main content area is divided into two sections. The top section contains the text: "Il y a plusieurs interprétations pour la phrase :
Je vote pour ma mère." The bottom section contains the text: "Choisissez la bonne." followed by two radio button options: " **je vote en faveur de ma mère.**" and " **je vote en prenant la place de ma mère.**" In the bottom right corner, there is an "OK" button.

Figure V-22 : Dialogue de résolution d'une ambiguïté d'étiquetage syntaxique

V.4.3. Ordre des arguments d'un verbe transitif direct

The dialog box has a title bar with the word "ambiguïté" in the center. The main content area is divided into two sections. The top section contains the text: "Il y a plusieurs interprétations pour la phrase :
Quel auteur cite ce conférencier ?" The bottom section contains the text: "Choisissez la bonne." followed by two radio button options: " **Un auteur cite un conférencier.**" and " **Un conférencier cite un auteur.**" In the bottom right corner, there is an "OK" button.

Figure V-18 : Dialogue de résolution d'une ambiguïté sur l'ordre des arguments d'un verbe transitif direct

Annexe VI

Extraits de la spécification de l'IF de C-STAR II

VI.1. Actes de dialogue

(*top* (accept acknowledge acknowledge-action affirm affirm-action apologize closing delay-action end-action give-information greeting greeting-welcome greeting-nice-meet greeting-request greeting-response-bad greeting-response-good introduce-self introduce-topic negate negate-action not-understand offer offer-information please-wait reject request-action request-affirmation request-delay-action request-information request-introduce-self request-neg-affirmation request-repeat offer-repeat request-suggestion request-verification return-from-delay suggest thank give-certainty request-certainty request-knowledge verify welcome x-exclamation) nil)

(give-information (x-reception x-addition x-completion connection inform time time-difference weather display budget possibility personal-data write search change temporal availability location address features contain price exchange-rate price-difference price-fluctuation payment spelling numeral telephone-number fax-number expiration-date preference pick-up drop-off purchase order click goto view reservation cancellation confirmation exchange send transfer) (<send-by>))

(request-action (x-reception x-addition x-completion connection inform display write payment change help pick-up drop-off purchase order click goto view reservation cancellation exchange send confirmation transfer) nil)

VI.2. Concepts

(reservation (payment price features contain temporal location availability expiration-date personal-data numeral) nil)

(features (restaurant meal admission seat stay minimum-stay party hotel meeting transportation taxi room flight train trip attraction tour event activity arrival departure currency guide x-car-rental webpage) (<time> <frequency> <location> <duration> <include> <who> <for-whom> <price> <purpose> <distance> <origin> <destination>))

```
(room nil (<hotel-name> <hotel-type> <hotel-style> <room-view> <room-type> <room-size>  
<contain> <bed-type> <room-location> <room-number> <room-name> <occupancy> <for-whom>  
<include>))
```

VI.3. Valeurs

```
(room-type= room luxury modest matrimoniale twin double single family suite junior_suite  
senior_suite bedroom meeting conference smoking non-smoking western_style japanese_style  
*general-modifier*)
```

```
(bed-type= bed twin double king queen *general-modifier* *size-modifier*)
```

```
(*general-modifier* additional best better different good clean typical quiet beautiful famous interesting  
popular fun new other *special-modifier*)
```

```
(*size-modifier* big bigger biggest small smaller smallest medium *special-modifier*)
```

Annexe VII

Extraits de la spécification de l'IF de NESPOLE!

VII.1. Actes de dialogue

DEF: \$top\$

continuations: (accept acknowledge affirm apologize apology-response contradict defer
descriptive dialog-hear dialog-indicate-understand
dialog-instruct-no dialog-instruct-yes dialog-present
dialog-problem dialog-proceed dialog-ready dialog-request-hear
dialog-request-indicate-yes-or-no dialog-request-present
dialog-request-problem dialog-request-proceed
dialog-request-ready dialog-request-scenario dialog-request-see
dialog-request-start dialog-request-stop
dialog-request-testing-ok dialog-request-understand
dialog-scenario dialog-see dialog-set-audio-level dialog-start
dialog-stop dialog-testing dialog-testing-ok dialog-understand
dialog-use-mt dialog-wait ditto do-not-worry empty
end-discussion end-topic end-wait exclamation give-information
greeting greeting-request greeting-response introduce-self
introduce-topic negate negate- no-tag offer please-wait promise
reject request-accept request-acknowledge request-action
request-information request-introduce-self request-reject
request-suggestion request-verification request-verification-
resume-topic spoken-pause suggest suggest-action thank
thank-response verify verify-)

arguments: none

comment: comment source -- \$top\$

comment: Start here when creating a legal DA for an IF representation

comment: comment source -- \$top\$

comment: Start here when creating a legal DA for an IF representation

DEF: give-information

continuations: ((+negation +accommodation) (+negation +action) (+negation +activity) (+negation +address) (+negation +admission) (+negation +airport) (+negation +arrival) (+negation +attraction) (+negation +availability) (+negation +balance) (+negation +begin-operation) (+negation +body-stat) (+negation +calculate) (+negation +cancellation) (+negation +change) (+negation +checkin) (+negation +checkout) (+negation +click) (+negation +completion) (+negation +concept) (+negation +connection) (+negation +contain) (+negation +currency) (+negation +current-time) (+negation +decrease) (+negation +delay) (+negation +departure) (+negation +deposit) (+negation +directions) (+negation +display) (+negation +disposition) (+negation +drop-off) (+negation +end-operation) (+negation +equipment) (+negation +event) (+negation +eventuality) (+negation +evidentiality) (+negation +exclusion) (+negation +exclusion-from) (+negation +existence) (+negation +exit) (+negation +experience) (+negation +expiration-date) (+negation +explain) (+negation +facility) (+negation +fax-number) (+negation +feasibility) (+negation +feature) (+negation +general-action) (+negation +goto) (+negation +health-status) (+negation +hear) (+negation +help) (+negation +improve) (+negation +in-operation) (+negation +inclusion) (+negation +inclusion-in) (+negation +increase) (+negation +indicate) (+negation +inform) (+negation +information-object) (+negation +knowledge) (+negation +learn) (+negation +learn-about) (+negation +make-image) (+negation +meal) (+negation +measure) (+negation +medical-procedure) (+negation +meeting) (+negation +numeral) (+negation +object) (+negation +obligation) (+negation +obtain) (+negation +occurrence) (+negation +order) (+negation +out-of-operation) (+negation +package) (+negation +party) (+negation +payment) (+negation +person) (+negation +personal-data) (+negation +pick-up) (+negation +plan) (+negation +price) (+negation +price-difference) (+negation +price-fluctuation) (+negation +proceed) (+negation +providing) (+negation +purchase) (+negation +read) (+negation +reception) (+negation +recommendation) (+negation +rent) (+negation +repeat) (+negation +reservation) (+negation +restaurant) (+negation +room) (+negation +route) (+negation +search) (+negation +seat) (+negation +send) (+negation +service) (+negation +skill) (+negation +spelling) (+negation +status) (+negation +stay) (+negation +telephone-number) (+negation +time-difference) (+negation +tour) (+negation +transportation) (+negation +trip) (+negation +understand) (+negation +vehicle) (+negation +view) (+negation +weather) (+negation +worsen) (+negation +write) +accommodation +action +activity +address +admission +airport +arrival +attraction +availability +balance +begin-operation +body-stat +calculate +cancellation +change +checkin +checkout +click +completion +concept +connection +contain +currency +current-time +decrease +delay +departure +deposit +directions +display +disposition +drop-off +end-operation +equipment +event +eventuality +evidentiality +exclusion +exclusion-from +existence +exit +experience +expiration-date +explain +facility +fax-number +feasibility +feature +general-action +goto +health-status +hear +help +improve +in-operation +inclusion +inclusion-in +increase +indicate +inform +information-object +knowledge +learn +learn-about +make-image +meal +measure +medical-procedure +meeting +numeral +object +obligation +obtain +occurrence +order +out-of-operation +package +party +payment +person +personal-data +pick-up +plan +price +price-difference +price-fluctuation +proceed +providing +purchase +read +reception +recommendation +rent +repeat +reservation +restaurant +room +route +search +seat +send +service +skill +spelling +status +stay +telephone-number +time-difference +tour +transportation +trip +understand +vehicle +view +weather +worsen +write)

arguments: (<after-rr=> <anti-condition=> <before-rr=> <besides=> <cause=>

<co-occurrence=> <concessive=> <condition=> <conjunction=>
 <contrastive=> <dependency=> <destination=> <digression=>
 <disjunction=> <duration=> <e-time=> <end-of=> <end-of=>
 <exception=> <factuality=> <final=> <first=> <focalizer=> <focus=>
 <for-whom=> <fourth=> <frequency=> <location=> <manner=>
 <middle-of=> <middle-of=> <origin=> <polarity=> <property=>
 <purpose=> <related-to=> <rhematic=> <second=> <start-of=>
 <start-of=> <substitution=> <third=> <time=> <to-whom=> <while=>
 <with-whom=> <with=>)

comment: comment source -- give-information

comment: Any statement.

comment: Ex: "I am doing that." = give-information+action

comment: Compare to: "I will do that." = promise+action

comment: comment source -- *basic-info-sa*
 comment: give-information and request-information cannot appear without a continuation concept.

VII.2. Concepts

DEF: +disposition

continuations: (\$bottom\$ +accommodation +action +activity +admission +airport +arrival
 +attraction +budget +cancellation +change +checkin +checkout +click
 +completion +concept +connection +contain +contained-in +currency
 +delay +departure +deposit +directions +display +drop-off
 +equipment +event +exclusion +exclusion-from +exit +explain
 +facility +feasibility +feature +general-action +goto +hear +help
 +inclusion +inclusion-in +indicate +inform +information-object
 +knowledge +learn +learn-about +meal +meeting +numeral +object
 +obtain +order +package +payment +person +pick-up +plan +price
 +proceed +purchase +read +recommendation +rent +repeat +reservation
 +restaurant +room +search +seat +send +service +skill +stay +tour
 +transportation +trip +vehicle +view +write)

arguments: (<disposition=> <manner=>)

comment: comment source -- +disposition

comment: See IF spec for arguments for possible values, including:

comment: { acceptance bother desire dislike enthusiasm expectation indifferent intention interest
 like need preference question ready try }.

comment: Ex: "How do you feel about skiing?" = request-information+disposition+activity
 (disposition=question, activity-spec=skiing)

comment: May depend on whether the speaker is client or agent.

comment: Ex: "I'm looking for a winter package starting December twelfth."

comment: c:give-information+disposition+package vs. a:give-information+search+package

comment: ATTITUDE

comment: MAIN-PREDICATION

DEF: +room

continuations: (\$bottom\$)

arguments: (<bed-spec=> <destination=> <distance=> <duration=> <for-whom=>
 <frequency=> <location=> <meal-spec=> <occupancy=> <origin=>
 <room-spec=> <scenic-view=> <speed=> <telephone-number=> <time=>
 <via=> <with-whom=> <with=>)

comment: comment source -- *focus-concepts*

comment: MAIN-PREDICATION

comment: PRED-PARTICIPANT

VII.3. Arguments

DEF: disposition=

Definition 1 (of 1)

:values

disposition-types

:relations

<duration=> <frequency=> <phase=> <time=>
<who=>

:attributes

<compared-to=> <degree=> <e-time=> <exactness=>
<factuality=> <focalizer=> <intensity=> <operator=>
<phase=> <polarity=>

:comments

shared comments from argument class -- @all
shared comments from argument class -- @all-mod
shared comments from argument class -- @pred
+DISPOSITION
need: refers to "lacking something" ("bisogno" (Italian)),
"I need some water"
interest: "I'm looking for / I'm interested in a winter package"
like: "I like/love skiing"
question: "how do you feel about ..."
"how do you feel about mushrooms?"
a:request-information+disposition+food
(disposition=(question, who=you), food-spec=mushrooms)
expectation = wait for / expect something
bother: "what's bothering you?" to express annoyance or
irritation, not for Showcase 1
shared comments from - disposition=

DEF: room-spec=

Definition 1 (of 1)

:values

question relative pronoun *rooms*
room-names

:relations

<accompanied-by=> <bed-spec=> <besides=> <called=>
<consist-of=> <contain=> <contained-in=> <contrastive=>
<cultural-style=> <dependency=> <destination=> <distance=>
<duration=> <exclude=> <excluded-from=> <extension=>
<for-whom=> <for=> <frequency=> <include=>
<included-in=> <location=> <object-number=> <origin=>
<per-unit=> <price-style=> <price=> <provider=>

<purpose=> <related-to=> <smoking=> <specifier=>
 <time=> <whose=> <with-whom=>

:attributes

<age=> <audio-level=> <color=> <compass-point=>
 <focalizer=> <grouping=> <identifiability=> <language=>
 <manner=> <max-min=> <modifier=> <nationality=>
 <noise-level=> <normality=> <object-ref=> <operator=>
 <order-ref-ext=> <order-ref-int=> <plus-minus=> <polarity=>
 <portion=> <quantity=> <relative-region=> <sex=>
 <shape=> <side=> <size=> <temperature=>
 <type=>

:comments

shared comments from argument class -- @all
 shared comments from argument class -- @qval
 shared comments from argument class -- @spec
 a double room with a double bed ==> room-spec=(double, bed-spec=(double,id=no))
 Some buildings/campgrounds have special names for rooms/sites.
 the Gertrude Stein room ==> (room-spec=name-gertrude_stein_room)
 the Lincoln Bedroom ==> (room-spec=name-lincoln_bedroom)
 shared comments from - room-spec=

DEF: bed-spec=

Definition 1 (of 1)

:values

question relative pronoun *beds*

:relations

<accompanied-by=> <besides=> <called=> <consist-of=>
 <contain=> <contained-in=> <contrastive=> <cultural-style=>
 <dependency=> <destination=> <distance=> <duration=>
 <exclude=> <excluded-from=> <extension=> <for-whom=>
 <for=> <frequency=> <include=> <included-in=>
 <location=> <origin=> <per-unit=> <price=>
 <provider=> <purpose=> <related-to=> <specifier=>
 <time=> <whose=> <with-whom=>

:attributes

<age=> <audio-level=> <color=> <compass-point=>
 <focalizer=> <grouping=> <identifiability=> <language=>
 <manner=> <max-min=> <modifier=> <nationality=>
 <noise-level=> <normality=> <object-ref=> <operator=>
 <order-ref-ext=> <order-ref-int=> <plus-minus=> <polarity=>
 <portion=> <quantity=> <relative-region=> <sex=>
 <shape=> <side=> <size=> <temperature=>
 <type=>

:comments

shared comments from argument class -- @all
 shared comments from argument class -- @qval
 shared comments from argument class -- @spec
 shared comments from - bed-spec=

DEF: location=**Definition 1 (of 4)****:values**

question relative pronoun *accommodations*
 rental-property-names *hotel-names* *park-names* *campsite-names*
 shelter-names *castle-names*

:relations

<2d-position=> <3d-position=> <above=> <accommodation-board=>
 <accommodation-class=> <accompanied-by=> <across=> <address=>
 <around=> <behind=> <besides=> <between=>
 <called=> <consist-of=> <contain=> <contained-in=>
 <contrastive=> <covering=> <cultural-style=> <dependency=>
 <destination=> <distance=> <duration=> <exclude=>
 <excluded-from=> <extension=> <family-name=> <for-whom=>
 <for=> <frequency=> <given-name=> <in-front-of=>
 <include=> <included-in=> <inside=> <intersection=>
 <location=> <next-to=> <object-number=> <object-topic=>
 <origin=> <path=> <per-unit=> <person-title=>
 <price-style=> <price=> <provider=> <purpose=>
 <related-to=> <secondary-name=> <smoking=> <specifier=>
 <through=> <time=> <under=> <whose=>
 <with-whom=>

:attributes

<age=> <audio-level=> <color=> <compass-point=>
 <exactness=> <focalizer=> <grouping=> <identifiability=>
 <language=> <manner=> <max-min=> <modifier=>
 <nationality=> <noise-level=> <normality=> <object-ref=>
 <operator=> <order-ref-ext=> <order-ref-int=> <plus-minus=>
 <polarity=> <portion=> <quantity=> <relative-region=>
 <sex=> <shape=> <side=> <size=>
 <temperature=> <type=>

:comments

shared comments from argument class -- @all
 shared comments from argument class -- @qual
 shared comments from argument class -- @spec
 shared comments from argument class -- @accommodation-spec
 need to get special sub args of accommodation for coordination
 shared comments from - location=

DEF: location=**Definition 2 (of 4)****:values**

question relative pronoun *package-types*
 package-names

:relations

<2d-position=> <3d-position=> <above=> <accompanied-by=>

<across=> <address=> <around=> <behind=>
 <besides=> <between=> <called=> <consist-of=>
 <contain=> <contained-in=> <contrastive=> <covering=>
 <dependency=> <destination=> <distance=> <duration=>
 <exclude=> <excluded-from=> <extension=> <family-name=>
 <for-whom=> <for=> <frequency=> <given-name=>
 <in-front-of=> <include=> <included-in=> <inside=>
 <intersection=> <location=> <next-to=> <object-number=>
 <object-topic=> <origin=> <package-type=> <path=>
 <per-unit=> <person-title=> <price=> <provider=>
 <purpose=> <related-to=> <secondary-name=> <smoking=>
 <specifier=> <through=> <time=> <under=>
 <whose=> <with-whom=>

:attributes

<age=> <audio-level=> <color=> <compass-point=>
 <exactness=> <focalizer=> <grouping=> <identifiability=>
 <language=> <manner=> <max-min=> <modifier=>
 <nationality=> <noise-level=> <normality=> <object-ref=>
 <operator=> <order-ref-ext=> <order-ref-int=> <plus-minus=>
 <polarity=> <portion=> <quantity=> <relative-region=>
 <sex=> <shape=> <side=> <size=>
 <temperature=> <type=>

:comments

shared comments from argument class -- @all
 shared comments from argument class -- @qval
 shared comments from argument class -- @spec
 shared comments from argument class -- @package-spec
 need to get special sub args of package for coordination
 shared comments from - location=

DEF: location=**Definition 3 (of 4)****:values**

question relative pronoun *vehicles*

:relations

<2d-position=> <3d-position=> <above=> <accompanied-by=>
 <across=> <address=> <around=> <behind=>
 <besides=> <between=> <called=> <consist-of=>
 <contain=> <contained-in=> <contrastive=> <covering=>
 <dependency=> <destination=> <distance=> <duration=>
 <exclude=> <excluded-from=> <extension=> <family-name=>
 <for-whom=> <for=> <frequency=> <given-name=>
 <in-front-of=> <include=> <included-in=> <inside=>
 <intersection=> <location=> <next-to=> <object-number=>
 <object-topic=> <origin=> <path=> <per-unit=>
 <person-title=> <price=> <provider=> <purpose=>
 <related-to=> <rental-agent=> <secondary-name=> <selling-agent=>
 <smoking=> <specifier=> <through=> <time=>
 <under=> <whose=> <with-whom=>

:attributes

<age=> <audio-level=> <color=> <compass-point=>
 <exactness=> <focalizer=> <grouping=> <identifiability=>
 <language=> <manner=> <max-min=> <modifier=>
 <nationality=> <noise-level=> <normality=> <object-ref=>
 <operator=> <order-ref-ext=> <order-ref-int=> <plus-minus=>
 <polarity=> <portion=> <quantity=> <relative-region=>
 <sex=> <shape=> <side=> <size=>
 <temperature=> <type=> <vehicle-make=> <vehicle-model=>
 <vehicle-shape=> <vehicle-size=>

:comments

shared comments from argument class -- @all
 shared comments from argument class -- @qval
 shared comments from argument class -- @spec
 shared comments from argument class -- @vehicle-spec
 need to get special sub args of vehicle for coordination
 shared comments from - location=

DEF: location=**Definition 4 (of 4)****:values**

question relative pronoun *objects-and-locations*
 people *rental-property-names* *travel-agency-names* *hotel-names*
 park-names *shelter-names* *castle-names* *campsite-names*
 attraction-names *building-names* *location-names* *restaurant-names*
 room-names *street-names* *building-names* *path-names*

:relations

<2d-position=> <3d-position=> <above=> <accompanied-by=>
 <across=> <address=> <around=> <behind=>
 <besides=> <between=> <called=> <consist-of=>
 <contain=> <contained-in=> <contrastive=> <covering=>
 <dependency=> <destination=> <distance=> <duration=>
 <exclude=> <excluded-from=> <extension=> <family-name=>
 <for-whom=> <for=> <frequency=> <given-name=>
 <in-front-of=> <include=> <included-in=> <inside=>
 <intersection=> <location=> <next-to=> <object-number=>
 <object-topic=> <origin=> <path=> <per-unit=>
 <person-title=> <price=> <provider=> <purpose=>
 <related-to=> <secondary-name=> <smoking=> <specifier=>
 <through=> <time=> <under=> <whose=>
 <with-whom=>

:attributes

<age=> <audio-level=> <color=> <compass-point=>
 <exactness=> <focalizer=> <grouping=> <identifiability=>
 <language=> <manner=> <max-min=> <modifier=>
 <nationality=> <noise-level=> <normality=> <object-ref=>
 <operator=> <order-ref-ext=> <order-ref-int=> <plus-minus=>
 <polarity=> <portion=> <quantity=> <relative-region=>
 <sex=> <shape=> <side=> <size=>
 <temperature=> <type=>

:comments

shared comments from argument class -- @all
shared comments from argument class -- @qval
shared comments from argument class -- @spec
... right next to it
 (... location=(exactness=exact, next-to=pronoun))
<near=> and <far=> removed
(See distance= for new representation)
general underspecified location
shared comments from - location=

VII.4. Valeurs

DEF: *disposition-types*

acceptance anticipation bother complaint concern
desire dislike enthusiasm expectation hope
indifferent intention interest like need
preference ready tolerance try

DEF: *rooms*

bedroom camping_lot double_room family_room junior_suite
matrimoniale room royal_suite senior_suite single_room
site sleeping_compartment suite tatami_room tent
triple_room twin_room vehicle-cabin

DEF: *beds*

bed couchette double_bed king_bed queen_bed
single_bed sleeping_bag tatami_bed twin_bed waterbed

Annexe VIII

Annotation en IF de tours de parole d'un client

(fn-1100001)

f110fg_16.1 olang FRE lang FRE Prv CLIPS " oui c'est bon "
f110fg_16.1 IF Prv CLIPS c:acknowledge+object

(fn-1100002)

f110fg_16.2.1 olang FRE lang FRE Prv CLIPS " oui "
f110fg_16.2.1 IF Prv CLIPS c:affirm

f110fg_16.2.2 olang FRE lang FRE Prv CLIPS " je vous entends très bien "
f110fg_16.2.2 IF Prv CLIPS c:dialog-hear (manner=(good, intensity=intense))

(fn-1100004)

f110fg_16.4 olang FRE lang FRE Prv CLIPS " d'accord "
f110fg_16.4 IF Prv CLIPS c:acknowledge

(fn-1100006)

f110fg_16.6 olang FRE lang FRE Prv CLIPS " bon je commence ou vous commencez "
f110fg_16.6 IF Prv CLIPS c:dialog-request-start (who=(operator=disjunct, [i, you]))

(fn-1100008)

f110fg_16.8 olang FRE lang FRE Prv CLIPS " bonjour je "
f110fg_16.8 IF Prv CLIPS c:greeting (greeting=good_morning)

(fn-1100009)

f110fg_16.9 olang FRE lang FRE Prv CLIPS " oui "
f110fg_16.9 IF Prv CLIPS c:affirm

(fn-1100013)

f110fg_16.13 olang FRE lang FRE Prv CLIPS " bonj "
f110fg_16.13 IF Prv CLIPS c:descriptive

(fn-1100014)

f110fg_16.14.1 olang FRE lang FRE Prv CLIPS " oui "
f110fg_16.14.1 IF Prv CLIPS c:affirm

f110fg_16.14.2 olang FRE lang FRE Prv CLIPS " je suis organisée "
f110fg_16.14.2 IF Prv CLIPS c:give-information+feature (who=(modifier=organized, i))

f110fg_16.14.3 olang FRE lang FRE Prv CLIPS " je suis intéressée pour partir en vacances dans le Trente cet été et ce serait pour deux personnes en fait "

f110fg_16.14.3 IF Prv CLIPS c:give-information+disposition+trip (disposition=(who=i, interest), visit-spec=vacation, location=name-trento, time=(season=summer, identifiability=non-distant), for-whom=(person, quantity=2))

(fn-1100016)
f110fg_16.16 olang FRE lang FRE Prv CLIPS " du quinze au dix-huit août "
f110fg_16.16 IF Prv CLIPS c:give-information+concept (time=(start-time=(md=15), end-time=(md=18), month=8))

(fn-1100018)
f110fg_16.18.1 olang FRE lang FRE Prv CLIPS " et où exactement "
f110fg_16.18.1 IF Prv CLIPS a:request-information+concept (conjunction=discourse, location=question)

f110fg_16.18.2 olang FRE lang FRE Prv CLIPS " vous avez une idée plus précise "
f110fg_16.18.2 IF Prv CLIPS a:give-information+feature (who=you, feature=(object-spec=(idea, degree=more, precise))

(fn-1100019)
f110fg_16.19 olang FRE lang FRE Prv CLIPS " ben on aimerait bien être "
f110fg_16.19 IF Prv CLIPS c:give-information+disposition (disposition=(who=we, like))

(fn-1100020)
f110fg_16.20.1 olang FRE lang FRE Prv CLIPS " on aimerait bien être proche d'un lac "
f110fg_16.20.1 IF Prv CLIPS c:give-information+disposition (disposition=(who=we, like), location=(distance=(relative-location=(lake, identifiability=no), relative-distance=close)))

f110fg_16.20.2 olang FRE lang FRE Prv CLIPS " si c'est possible "
f110fg_16.20.2 IF Prv CLIPS c:give-information+feasibility (feasibility=feasible, condition=discourse)

(fn-1100022)
f110fg_16.22 olang FRE lang FRE Prv CLIPS " oui "
f110fg_16.22 IF Prv CLIPS c:affirm

(fn-1100024)
f110fg_16.24.1 olang FRE lang FRE Prv CLIPS " deux personnes "
f110fg_16.24.1 IF Prv CLIPS c:give-information+concept (person-spec=(person, quantity=2))

f110fg_16.24.2 olang FRE lang FRE Prv CLIPS " oui "
f110fg_16.24.2 IF Prv CLIPS c:affirm

(fn-1100026)
f110fg_16.26 olang FRE lang FRE Prv CLIPS " ouais "
f110fg_16.26 IF Prv CLIPS c:affirm

(fn-1100027)
f110fg_16.27 olang FRE lang FRE Prv CLIPS " du "
f110fg_16.27 IF Prv CLIPS no-tag

(fn-1100028)
f110fg_16.28 olang FRE lang FRE Prv CLIPS " quinze au dix-huit "
f110fg_16.28 IF Prv CLIPS c:give-information+concept (time=(start-time=(md=15), end-time=(md=18))

(fn-1100031)
f110fg_16.31.1 olang FRE lang FRE Prv CLIPS " plutôt en hôtel en fait avec une pension complète "
f110fg_16.31.1 IF Prv CLIPS c:give-information+disposition+concept (disposition=preference, accommodation-spec=hotel, accommodation-board=full_board)

f110fg_16.31.2 olang FRE lang FRE Prv CLIPS " si c'est possible "
f110fg_16.31.2 IF Prv CLIPS c:give-information+feasibility (feasibility=feasible, condition=discourse)

(fn-1100036)
f110fg_16.36 olang FRE lang FRE Prv CLIPS " oui "
f110fg_16.36 IF Prv CLIPS c:affirm

(fn-1100038)

f110fg_16.38 olang FRE lang FRE Prv CLIPS " ben c'est bien parce que "

f110fg_16.38 IF Prv CLIPS c:give-information+feature (feature=(object-spec=good))

(fn-1100039)

f110fg_16.39.1 olang FRE lang FRE Prv CLIPS " oui "

f110fg_16.39.1 IF Prv CLIPS c:affirm

f110fg_16.39.2 olang FRE lang FRE Prv CLIPS " non "

f110fg_16.39.2 IF Prv CLIPS c:negate

f110fg_16.39.3 olang FRE lang FRE Prv CLIPS " ça va quatre nuits "

f110fg_16.39.3 IF Prv CLIPS c:acknowledge+object (object-spec=(night, quantity=4))

(fn-1100041)

f110fg_16.41 olang FRE lang FRE Prv CLIPS " oui "

f110fg_16.41 IF Prv CLIPS c:affirm

(fn-1100043)

f110fg_16.43 olang FRE lang FRE Prv CLIPS " deux cent quarante-sept euros "

f110fg_16.43 IF Prv CLIPS c:give-information+concept (price=(currency=euro, quantity=247))

(fn-1100046)

f110fg_16.46 olang FRE lang FRE Prv CLIPS " d'accord "

f110fg_16.46 IF Prv CLIPS c:acknowledge

Annexe IX

Extraits du code du premier analyseur NESPOLE!

IX.1. Instanciation des arguments

Par exemple, pour l'argument `room-spec`, une fonction `RoomSpec2If` cherche à localiser des séquences de détermination, de quantité et de valeur (parmi la liste des types de pièces), détermination et quantité étant éventuellement optionnelles. Lorsqu'une telle séquence est reconnue, un argument `room-spec` est construit.

```
proc RoomSpec2If {inString} {
  if {[regexp "(?:une|un) ($fifservdico::frenchroomspec)(?:x|s)?" $inString !Match !RoomSpec]!=0} {
    append the_result "room-spec=(identifiability=no, " [fifservdico::RoomSpec2If $!RoomSpec] ")"
  } elseif {[regexp "(?:des|plusieurs) ($fifservdico::frenchroomspec)(?:x|s)?" $inString !Match
!RoomSpec]!=0} {
    append the_result "room-spec=(identifiability=no, quantity=plural, " [fifservdico::RoomSpec2If $!RoomSpec]
)"
  } elseif {[regexp "(?:mon|ma|le|la|l') ?($fifservdico::frenchroomspec)(?:x|s)?" $inString !Match
!RoomSpec]!=0} {
    append the_result "room-spec=(identifiability=yes, " [fifservdico::RoomSpec2If $!RoomSpec] ")"
  } elseif {[regexp "(?:les) ($fifservdico::frenchroomspec)(?:x|s)?" $inString !Match !RoomSpec]!=0} {
    append the_result "room-spec=(identifiability=yes, quantity=plural, " [fifservdico::RoomSpec2If $!RoomSpec]
)"
  } elseif {[regexp "(?:mes) ($fifservdico::frenchroomspec)(?:x|s)?" $inString !Match !RoomSpec]!=0} {
    append the_result "room-spec=(identifiability=yes, quantity=plural, " [fifservdico::RoomSpec2If $!RoomSpec]
)"
  } elseif {[regexp "(?:les) ([0\-\9\+]) ($fifservdico::frenchroomspec)(?:x|s)?" $inString !Match !Quantity
!RoomSpec]!=0} {
    append the_result "room-spec=(identifiability=yes, quantity=" $!Quantity ", " [fifservdico::RoomSpec2If
$!RoomSpec] ")"
  } elseif {[regexp "([0\-\9\+]) ($fifservdico::frenchroomspec)(?:x|s)?" $inString !Match !Quantity
!RoomSpec]!=0} {
    append the_result "room-spec=(identifiability=no, quantity=" $!Quantity ", " [fifservdico::RoomSpec2If
$!RoomSpec] ")"
  } elseif {[regexp "($fifservdico::frenchroomspec)(?:x|s)?" $inString !Match !RoomSpec]!=0} {
    append the_result "room-spec=" [fifservdico::RoomSpec2If $!RoomSpec]
  } else {
    return ""
  }
}
```

IX.2. Construction de l'acte de dialogue

Nous donnons ci-dessous un exemple complet pour la fonction **Room2lf** du module d'analyse du premier démonstrateur.

```
proc Room2lf {inWho inString} {
```

Recherche et construction de la seule attitude traitée, disposition.
Recherche et construction des arguments.

```
set theDisposition [fifservices::Disposition2lf $inString]
set theAccommodation [fifservices::AccomSpec2lf $inString]
set theRoom [fifservices::RoomSpec2lf $inString]
set theLocation [fifservices::LocationButPlace2lf $inString]
set theDistance [fifservices::Distance2lf $inString]
set theDuration [fifservices::Duration2lf $inString]
set theTime [fifservices::Time2lf $inString]
set thePrice [fifservices::Price2lf $inString]
set theWithWhom [fifservices::WithWhom2lf $inString]
set theForWhom [fifservices::ForWhom2lf $inString]
set theFacility [fifservices::AccomRoomFacility2lf $inString]
set theService [fifservices::Service2lf $inString]

if {$theLocation!=""} {set theLocation [concat "," $theLocation]}
if {$theTime!=""} {set theTime [concat "," $theTime]}
if {$thePrice!=""} {set thePrice [concat "," $thePrice]}
if {$theAccommodation!=""} {set theAccommodation [concat "," $theAccommodation]}
if {$theDistance!=""} {set theDistance [concat "," $theDistance]}
if {$theDuration!=""} {set theDuration [concat "," $theDuration]}
if {$theForWhom!=""} {set theForWhom [concat "," $theForWhom]}
if {$theWithWhom!=""} {set theWithWhom [concat "," $theWithWhom]}
if {$theFacility!=""} {set theFacility [concat "," $theFacility]}
if {$theService!=""} {set theService [concat "," $theService]}
```

Recherche de la présence de syntagmes signalant les concepts possibles
de l'acte de dialogue reservation, suggestion, question, negation, subject,
existence, availability.

```
set Reservation [regexp "($fifservdico::frenchreservation)" $inString IMatch]
set Suggestion [regexp "($fifservdico::frenchsuggestion)" $inString IMatch]
set Question [regexp "($fifservdico::frenchquestion)" $inString IMatch]
set Negation [fifservdico::PNegate $inString]
set Subject [regexp "($fifservdico::frenchpronoun)" $inString IMatch IPronoun]
#availability prime sur existence
set Existence [regexp "($fifservdico::frenchexistence)" $inString IMatch]
set Availability [regexp "($fifservdico::frenchavailability)" $inString IMatch]

if {$Subject!=0 && [fifservdico::NormalizePronoun2lf $IPronoun]!="he"} {
  append theSubject " , who=" [fifservdico::NormalizePronoun2lf $IPronoun]
} else {
  set theSubject ""
}
```

Construction de la partie arguments de l'IF.

```
#argument value construction
if {$theDisposition!=""} {
  append theArguments $theDisposition " , " $theRoom $theFacility $theService $theAccommodation $thePrice
  $theLocation $theDuration $theDistance $theTime $theForWhom $theWithWhom
} else {
  append theArguments $theRoom $theFacility $theService $theAccommodation $thePrice $theLocation
  $theDuration $theDistance $theTime $theSubject $theForWhom $theWithWhom
}
```

Construction de l'acte de dialogue en fonction des arguments subject, question, negation, suggestion.

```
#Da construction
if {$Subject!=0} {
  if {$Question!=0} {
    if {$Negation!=0} {
      if {$Suggestion!=0} {
        #sujet question non-negation suggestion
        set theDA1 "negate-request-suggestion"
      } else {
        #sujet question non-negation non-suggestion
        set theDA1 "negate-request-information"}
    } else {
      if {$Suggestion!=0} {
        #sujet question negation suggestion
        set theDA1 "request-suggestion"
      } else {
        #sujet question negation non-suggestion
        set theDA1 "request-information"}}
    } else {
      if {$Negation!=0} {
        if {$Suggestion!=0} {
          #sujet non-question non-negation suggestion
          set theDA1 "negate-suggest"
        } else {
          #sujet non-question non-negation non-suggestion
          set theDA1 "negate-give-information"}
        } else {
          if {$Suggestion!=0} {
            #sujet non-question negation suggestion
            set theDA1 "suggest"
          } else {
            #sujet non-question negation non-suggestion
            set theDA1 "give-information"}}}
    } else {
      if {$Question!=0} {
        if {$Negation!=0} {
          if {$Suggestion!=0} {
            #sujet question non-negation suggestion
            set theDA1 "negate-request-suggestion"
          } else {
            #sujet question non-negation non-suggestion
            set theDA1 "negate-request-information"}
          } else {
            if {$Suggestion!=0} {
              #sujet question negation suggestion
              set theDA1 "request-suggestion"
            } else {
              #sujet question negation non-suggestion
              set theDA1 "request-information"}}
          } else {
            set theDA1 "give-information+concept"
          }
        }
      }
    }
  }
}
```

Construction de la première partie des concepts de l'acte de dialogue, avec l'argument disposition.

```
if {$theDisposition!=""} {
  set theDA2 "+disposition"
} else {
  set theDA2 ""
}
```

```
}

```

Construction de la seconde partie des concepts avec les arguments
availability, reservation, existence, price.

```
if {$Availability!=0} {
  set theDA3 "+availability"
} elseif {$Reservation!=0} {
  set theDA3 "+reservation"
} elseif {$Existence!=0} {
  set theDA3 "+existence"
} elseif {$thePrice!=""} {
  set theDA3 "+price"
} else {
  set theDA3 ""
}

```

Construction du prédicat principal room.

```
set theDA4 "+room"

```

Construction des prédicats secondaires facility, service.

```
if {$theFacility!=""} {
  set theDA5 "+facility"
} else {
  set theDA5 ""
}

if {$theService!=""} {
  set theDA6 "+service"
} else {
  set theDA6 ""
}

```

Construction de l'acte de dialogue complet.

```
if {$theDA1=="give-information+concept"} {
  append theresult {" $inWho $theDA1 " (" $theArguments ")"}
} else {
  append theresult {" $inWho $theDA1 $theDA2 $theDA3 $theDA4 $theDA5 $theDA6 " (" $theArguments ")"}
}
}

```

Annexe X

Extraits de code des modules du second démonstrateur NESPOLE!

X.1. Analyse

X.1.1. Instanciation des arguments

La fonction RoomSpec2lf est maintenant la suivante :

```
proc RoomSpec2lf {inString} {  
  global partialIF
```

Expressions régulières de reconnaissance de l'argument room-spec :

```
if {[regexp " (?:(une|un) ($fifservdico::frenchroom)(?:x|s)? " $inString IMatch IRoomSpec]!=0} {  
  append the_result "room-spec=(identifiability=no, " [fifservdico::Room2lf $IRoomSpec] ")"  
} elseif {[regexp " (?:(des|plusieurs) ($fifservdico::frenchroom)(?:x|s)? " $inString IMatch IRoomSpec]!=0} {  
  append the_result "room-spec=(identifiability=no, quantity=plural, " [fifservdico::Room2lf $IRoomSpec] ")"  
} elseif {[regexp " (?:(mon|ma|le|la|l) ?($fifservdico::frenchroom)(?:x|s)? " $inString IMatch IRoomSpec]!=0} {  
  append the_result "room-spec=(identifiability=yes, " [fifservdico::Room2lf $IRoomSpec] ")"  
} elseif {[regexp " (?:(les) ($fifservdico::frenchroom)(?:x|s)? " $inString IMatch IRoomSpec]!=0} {  
  append the_result "room-spec=(identifiability=yes, quantity=plural, " [fifservdico::Room2lf $IRoomSpec] ")"  
} elseif {[regexp " (?:(mes) ($fifservdico::frenchroom)(?:x|s)? " $inString IMatch IRoomSpec]!=0} {  
  append the_result "room-spec=(identifiability=yes, quantity=plural, " [fifservdico::Room2lf $IRoomSpec] ")"  
} elseif {[regexp " (?:(les) ([0-9]+) ($fifservdico::frenchroom)(?:x|s)? " $inString IMatch IQuantity IRoomSpec]!=0}  
{  
  append the_result "room-spec=(identifiability=yes, quantity=" $IQuantity ", " [fifservdico::Room2lf  
$IRoomSpec] ")"  
} elseif {[regexp " ([0-9]+) ($fifservdico::frenchroom)(?:x|s)? " $inString IMatch IQuantity IRoomSpec]!=0} {  
  append the_result "room-spec=(identifiability=no, quantity=" $IQuantity ", " [fifservdico::Room2lf $IRoomSpec]  
")"  
} elseif {[regexp " ($fifservdico::frenchroom)(?:x|s)? " $inString IMatch IRoomSpec]!=0} {  
  append the_result "room-spec=" [fifservdico::Room2lf $IRoomSpec]  
} else {  
  set the_result ""  
}
```

Rattachement d'un argument de type location situé juste après dans le
texte d'entrée.

```
#trying to find the location of the accommodation  
if {$the_result==""} {  
  return ""  
} else {  
  regexsub $IMatch $partialIF(CleanedString) " ArgumentX " partialIF(CleanedString)  
  set theLocation [fifservices::ArgumentLocation2lf $partialIF(CleanedString)]
```

```

if {$theLocation!=""} {
  regsub "\\)" $the_result " , $theLocation)" the_result
  return $the_result
} else {
  return $the_result
}
}
}
}

```

X.1.2. Complément de l'acte de dialogue

Voici l'exemple de la fonction Room2If dans sa seconde version de l'analyseur.

```

proc Room2If {inWho inString} {
  global partialF
  global isConcept

```

Construction du préfixe de l'acte de dialogue.

```
ffdaconstruct::ConstructDA $inString
```

Construction des arguments privilégiés, et des arguments qui vont permettre d'augmenter le préfixe de l'acte de dialogue.

```

set theDistance [ffservices::Distance2If $partialF(CleanedString)]
set theRoom [ffservices::RoomSpec2If $partialF(CleanedString)]
set theAccommodation [ffservices::AccomSpec2If $partialF(CleanedString)]
set theLocation [ffservices::Location2If $partialF(CleanedString)]
#set theLocation [ffservices::LocationButPlace2If $partialF(CleanedString)]
set theDuration [ffservices::Duration2If $partialF(CleanedString)]
set theTime [ffservices::Time2If $partialF(CleanedString)]
set thePrice [ffservices::Price2If $partialF(CleanedString)]
set theWithWhom [ffservices::WithWhom2If $partialF(CleanedString)]
set theForWhom [ffservices::ForWhom2If $partialF(CleanedString)]

if {$theLocation!=""} {set theLocation [concat " , " $theLocation]}
if {$theTime!=""} {set theTime [concat " , " $theTime]}
if {$thePrice!=""} {set thePrice [concat " , " $thePrice]}
if {$theAccommodation!=""} {set theAccommodation [concat " , " $theAccommodation]}
if {$theDistance!=""} {set theDistance [concat " , " $theDistance]}
if {$theDuration!=""} {set theDuration [concat " , " $theDuration]}
if {$theForWhom!=""} {set theForWhom [concat " , " $theForWhom]}
if {$theWithWhom!=""} {set theWithWhom [concat " , " $theWithWhom]}

if {[regexp {\+indicate} $partialF(DA)!=0} {
  set theIndicator [ffservices::Indicator2If $partialF(CleanedString)]
  if {$theIndicator!=""} {set theIndicator [concat " , " $theIndicator]}
} else {
  set theIndicator ""
}

if {$partialF(Arguments)!=""} {
  set partialF(Arguments) "$partialF(Arguments),
$theRoom$theAccommodation$thePrice$theLocation$theDuration$theDistance$theTime$theForWhom$theWith
Whom$theIndicator"
} else {
  set partialF(Arguments)
"$theRoom$theAccommodation$thePrice$theLocation$theDuration$theDistance$theTime$theForWhom$theWith
Whom$theIndicator"
}
}

```

```

regsub -all "^,(,)+" $partialF(Arguments) "" partialF(Arguments)
regsub -all ",(,)+" $partialF(Arguments) ", " partialF(Arguments)
regsub -all ",(,)*$" $partialF(Arguments) "" partialF(Arguments)

set isFeature [fifservices::FeaturePredicate $partialF(Arguments)]

```

Complément de l'acte de dialogue.

```

if {$isConcept==0} {
  if {$isFeature==0} {
    if {$thePrice!=""} {
      if {[regexp {\+price} $partialF(DA)]!=0} {
        set partialF(DA) "$partialF(DA)+room"
      } else {
        set partialF(DA) "$partialF(DA)+price+room"
      }
    } else {
      set partialF(DA) "$partialF(DA)+room"
    }
  } else {
    set partialF(DA) "$partialF(DA)+feature+room"
  }
}
}

```

Construction de l'IF.

```

if {$partialF(Arguments)!=""} {
  return "{$inWho$partialF(DA)$partialF(Arguments)}"
} else {
  return "{$inWho$partialF(DA)}"
}
}

```

X.2. Génération

X.2.1. Génération d'un acte de parole terminal

Pour la première IF, nous sommes intéressés par les arguments `greeting=` et `communication-mode=`. En effet pour une salutation au téléphone nous disons « Allô » en français.

```

proc Greeting2Fr {inWho inPrefix inNegate} {
  #greeting, communication-mode=phone, to-whom
  set greetingArgValue [ifsubarguments::GetArgValue "greeting="]
  set communicationmodeArgValue [ifsubarguments::GetArgValue "communication-mode="]

  switch $inPrefix {
    "request-verification-" {
      if {$communicationmodeArgValue=="phone"} {
        return "Allô, c'est ça ?"
      } elseif {$greetingArgValue!=""} {
        return "[ifservdico::CapitalizeFirstLetterString [ifservdico::If2Greeting $greetingArgValue]], n'est-ce pas ?"
      } else {
        return "Bonjour, c'est ça ?"
      }
    }
    "verify-" {
      if {$communicationmodeArgValue=="phone"} {
        return "En effet, allô."
      } elseif {$greetingArgValue!=""} {
        return "En effet, [ifservdico::If2Greeting $greetingArgValue]"
      }
    }
  }
}

```


Annexe XI

Fichier complété par un juge lors de la seconde évaluation NESPOLE!

Nous présentons ci-dessous un extrait de fichier d'évaluation d'un juge français évaluant la qualité **des traductions en français** produites à partir de **références en italien**. L'analyseur italien-IF (IRST) et le générateur IF-français (CLIPS) sont ceux du second démonstrateur.

		*	very			very	*	no
		*	good	good	bad	bad	*	lang
		*	*****	*****	*****	*****	*	*****
TURN	1	*	*****	*****	*****	*****	*	*****
1	APT del trentino	1	x				1	
2	buongiorno	2	x				2	
TRANS	Ici une agence d'informations du Trentin. Bonjour !	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	2	*	*****	*****	*****	*****	*	*****
1	buongiorno	1	x				1	
2	desidera qualche informazione	2	x				2	
TRANS	Bonjour ! Est-ce que vous désirez ?	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	3	*	*****	*****	*****	*****	*	*****
1	sì	1	x				1	
TRANS	Oui.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	4	*	*****	*****	*****	*****	*	*****
1	ho capito	1	x				1	
2	sa dirmi una zona in particolare del trentino o posso suggerirle qualcosa	2		x			2	
TRANS	Je comprends. Est-ce que vous connaissez la région, le Trentino ?	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	5	*	*****	*****	*****	*****	*	*****
1	le consiglieri predazzo non so se ha mai sentito trentino dell' est guardi le faccio vedere la piantina	1		x			1	
TRANS	Je vous recommande Predazzo. Je montre le Trentino.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	6	*	*****	*****	*****	*****	*	*****

1	la piantina di predazzo conosce la zona o vuole che le faccia vedere una piantina	1			X		1	
TRANS	Predazzo. Est-ce que vous désirez la région ?	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	7	*	*****	*****	*****	*****	*	*****
1	prima magari le faccio vedere in che zona del trentino si trova	1	X				1	
TRANS	Je montre plus tôt. Il s'agit de la région du Trentino.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	8	*	*****	*****	*****	*****	*	*****
1	ecco riesce a vedere l' immagine lei da dove chiama scusi	1		X			1	
TRANS	Ce qui est possible, c'est appeler où.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	9	*	*****	*****	*****	*****	*	*****
1	da milano	1	X				1	
2	vede c'è il lago di garda predazzo adesso glielo segno	2	X				2	
TRANS	Vous voyez depuis Milano. Lago Di Garda Predazzo. J'indique maintenant.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	10	*	*****	*****	*****	*****	*	*****
1	con un cerchio blu	1		X			1	
TRANS	J'indique avec le cercle en bleu.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	11	*	*****	*****	*****	*****	*	*****
1	ecco qui il trentino orientale in valle di fiemme	1			X		1	
TRANS	Ici le Trentino. Bonjour !	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	12	*	*****	*****	*****	*****	*	*****
1	lo vede ecco era interessa	1				X	1	
TRANS	Est-ce que je vois ?	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	13	*	*****	*****	*****	*****	*	*****
1	era interessato a qualche attività in particolare non so lei scia oppure ha una famiglia	1		X			1	
TRANS	Vous êtes intéressé par quelques activités une famille.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	14	*	*****	*****	*****	*****	*	*****
1	oppure ha una famiglia	1			X		1	
TRANS	Est-ce qu'avec une famille ?	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	15	*	*****	*****	*****	*****	*	*****
1	il percorso più adatto adesso le faccio vedere lei viene da milano passerà verona	1			X		1	
TRANS	. Je montre maintenant depuis Milano. Verona.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	16	*	*****	*****	*****	*****	*	*****
1	ecco verona che è qui più o meno vede il	1		X			1	

	cerchio blu sempre riesce							
TRANS	J'indique avec le cercle en bleu Verona. Toujours.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	17	*	*****	*****	*****	*****	*	*****
1	lei seguirà poi per l' autostrada del brennero vede dov' e' rovereto trento	1				x	1	
TRANS	Pas de quoi ! Vous voyez Brennero, Trento.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	18	*	*****	*****	*****	*****	*	*****
1	poi la strada migliore è proseguire verso bolzano e uscire ad ora il casello di egna ora	1				x	1	
TRANS	. Il s'agit de ça maintenant.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	19	*	*****	*****	*****	*****	*	*****
1	poi da ora esce dall' autostrada e c'è una strada molto buona si arriva in circa una mezz' ora a predazzo sono più o meno da ora una trentina di chilometri	1		x			1	
TRANS	à partir de maintenant. Bon maintenant, à partir de maintenant à Predazzo 30 kilomètres.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	20	*	*****	*****	*****	*****	*	*****
1	la strada è molto scorrevole facilmente percorribile sì	1				x	1	
TRANS	Il s'agit de. Oui.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	21	*	*****	*****	*****	*****	*	*****
1	guardi ho un pacchetto in un hotel	1		x			1	
TRANS	Je vous recommande un forfait dans un hôtel.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	22	*	*****	*****	*****	*****	*	*****
1	quanto voleva spendere aveva qualche idea non so un hotel 3 stelle poteva andare bene	1	x				1	
TRANS	Est-ce que vous désirez un hôtel trois étoiles ?	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	23	*	*****	*****	*****	*****	*	*****
1	c'è una buona offerta in un hotel 3 stelle 7 notti a predazzo	1	x				1	
2	e l' offerta è valida per dall' 1 dicembre al 202 dicembre in un hotel 3 stelle per 3603 euro a persona	2	x				2	
TRANS	Il y a un bon forfait pour les 7 nuits à l'hôtel, Predazzo. Le forfait à 3603 euros par personne à partir du 1 décembre, en décembre le 202 à l'hôtel.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	24	*	*****	*****	*****	*****	*	*****
1	ecco sì questo è sì vediamo un attimo se trovo una sistemazione in appartamento	1	x				1	
TRANS	Oui. Il s'agit de. Oui. Un instant, merci ! Vous dites un logement à l'appartement.	*	*****	*****	*****	*****	*	*****

		*	*****	*****	*****	*****	*	*****
TURN	25	*	*****	*****	*****	*****	*	*****
1	predazzo ecco qui in appartamento un attimo che guardo i prezzi	1			x		1	
TRANS	Predazzo. Ici. Bonjour ! Un instant, merci ! Vous dites un prix.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	26	*	*****	*****	*****	*****	*	*****
1	ecco in appartamento c'è una sistemazione sempre nello stesso periodo a un prezzo leggermente inferiore di 200 euro a persona	1	x				1	
2	ecco un albergo un appartamento sempre sono appartamenti ristrutturati da poco	2				x	2	
3	perciò sono degli appartamenti belli	3	x				3	
4	per passare una settimana tranquilla	4	x				4	
TRANS	à l'appartement un logement toujours ça un prix 200 euros par personne. Il s'agit de toujours. Un appartement joli pour la semaine.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	27	*	*****	*****	*****	*****	*	*****
1	ecco le faccio vedere magari se vuole no le facevo vedere la cartina di predazzo mi dica se aveva	1			x		1	
TRANS	Est-ce que vous désirez ? Non. Est-ce que je vois la carte à propos de Predazzo ? Est-ce que je peux vous aider ? Vous entendez, n'est ce pas ?	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	28	*	*****	*****	*****	*****	*	*****
1	sì	1	x				1	
TRANS	Oui.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	29	*	*****	*****	*****	*****	*	*****
1	sì	1	x				1	
2	ecco nel pacchetto degli alberghi che le ho suggerito è incluso 1 skipass latemar	2	x				2	
TRANS	Oui. Un forfait un hôtel. C'est inclus un forfait de ski. Latemar.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	30	*	*****	*****	*****	*****	*	*****
1	per sì proprio lo skipass settimanale per la zona di predazzo lusia non so se ha presente se è mai stato son tutte delle zone le faccio vedere perfetto	1		x			1	
TRANS	Merci d'avoir attendu. Oui. Le forfait de ski jusqu'à la région de Predazzo la région. Je montre. D'accord.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	31	*	*****	*****	*****	*****	*	*****
1	ecco questo questa è val di fiemme ecco perfetto vede la cartina le faccio vedere	1	x				1	
TRANS	. Il s'agit de Val Di Fiemme. D'accord. Est-ce que je vois la carte ? Je montre.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	32	*	*****	*****	*****	*****	*	*****

1	in rosso predazzo così ha il rilievo 3d ecco riesce a vedere ecco poi la zona sci_ sciistica c'è il lusia	1		X			1	
TRANS	Est-ce que vous pouvez voir Predazzo ? Il y a la région.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	33	*	*****	*****	*****	*****	*	*****
1	sulla destra ecco qui ci sono piste	1		X			1	
TRANS	Il y a une piste de ski ici.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	34	*	*****	*****	*****	*****	*	*****
1	e non so lei è 1 sciatore provetto principiante	1				X	1	
TRANS	Vous dites le 1. N'est-ce pas ?	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	35	*	*****	*****	*****	*****	*	*****
1	medio lusia potrebbe andar bene ci son delle piste abbastanza lunghe e di media difficoltà	1				X	1	
TRANS	Est-ce que vous acceptez ? C'est inclus dans une piste de ski. .	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	36	*	*****	*****	*****	*****	*	*****
1	per questa zona e per questa zona qui se vede lo ski center del latemar ecco vede qui a sinistra	1			X		1	
TRANS	Vous voyez Latemar jusqu'à une région, une région ici. Vous voyez ici, à gauche.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	37	*	*****	*****	*****	*****	*	*****
1	per entrambe queste sono piste un pochettino più difficili sono	1		X			1	
TRANS	Une piste de ski dur.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	38	*	*****	*****	*****	*****	*	*****
1	facilmente raggiungibili proprio una decina di chilometri sia la zona del lusia sia quella del latemar ecco	1		X			1	
TRANS	10 kilomètres. La région, Latemar.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****
TURN	39	*	*****	*****	*****	*****	*	*****
1	sì	1	X				1	
2	poi ci sono incluse nel pacchetto 4 lezioni di sci e 2 lezioni pattinaggio	2	X				2	
TRANS	Oui. Il y a un forfait avec 4 leçons du ski. Le 2 du patin.	*	*****	*****	*****	*****	*	*****
		*	*****	*****	*****	*****	*	*****

Annexe XII

Traductions produites pour la paire japonais-anglais (IWSLT 2004)

XII.1. Résultats Systran annotés

Source BTEC ¹³²	Sortie SYSTRAN	Traduction Révisée	Commentaire
お勘定をおねがいします。	We request calculation.	Give me the check, please.	Bad lexical choice for 勘定.
この街の見どころを教えてください。	Please teach the places of interest of this town.	Tell me please the places of interest of this town.	Bad lexical choice for 教える.
あの角にありますよ。	There is that angle.	There is one at the corner.	Bad lexical choice for 角. Object missing.
どれが私のですか。	Either one is my?	Which one is mine?	Question: good word order. Wrong lexical choice for どれ.
入場料はいくらですか。	Is admission fee how much?	How much is the admission fee?	Question: wrong word order.
オペラ座はどこですか。	Is the opera house where?	Where is the opera house?	Question: wrong word order.
日本語は話せますか。	You can speak Japanese?	Can you speak Japanese?	Question: wrong word order. Acceptable.
どういたしまして。	How doing.	You are welcome.	Spoken fixed formula.
もしもし。	It does.	Hello.	Spoken fixed formula.
それにします。	It makes that.	I'll take that.	Spoken fixed formula.
ここで降ろしてください。	Please lower here.	Drop me off here, please.	Subject: missing. Bad lexical choice for 降ろ.
ここで降ります。	It gets off here.	I'll get off here.	Subject: missing.

¹³² Le texte japonais est donné tel qu'il est dans BTEC, le japonais est une langue sans séparateur, il ne devrait pas y avoir d'« espaces ».

			Good lexical choice for 降る.
一緒に行きましょう	It will go together.	Let's go together.	Subject: missing.
切符を二枚持っています。	It has two tickets.	I have two tickets.	Subject: missing.
医者を呼んでください。	Please call the doctor.	Please call the doctor.	Subject: missing. Very good translation.
お医者さんに行った方がいいかもよ。	Whether the person where does to the doctor is good.	It's better to go to see a doctor.	Very bad translation.
薬が効きましたか。	Was the medicine effective?	Was the medicine effective?	Good translation. Impersonal sentence.
初心者でも大丈夫ですか。	It is all right even with the beginner?	It is all right even with the beginner?	Very good translation. Impersonal sentence.
パスポート、トラベラーズチェック、航空券。	Passport, traveler's check and air ticket.	Passport, traveler's check and air ticket.	Very good translation. Noun phrases.
観光ツアーはありますか。	Is there a sightseeing tour?	Is there a sightseeing tour?	Very good translation: Impersonal sentence.

XII.2. Résultats Systran bruts (extrait)

Source BTEC	Sortie SYSTRAN	Traduction Révisée
日本大使館に連絡したいのですが。	It is to like to communicate to the Japanese embassy but.	I'd like to take contact with the Japanese Embassy.
両替所はどこですか。	Is the exchange place where?	Where is located the money-exchange office?
そのまま切らずにお待ち下さい。	Without cutting that way, please wait.	Hold the line, please.
市内までいくらですか。	It is how much to the city?	How much is it for the city?
あの角にありますよ。	There is that angle.	There is one at the corner.
切りますよ。	It cuts.	It cuts.
寒気がする。	Chill does.	I feel a chill.
ここで降りしてください。	Please lower here.	Let me off here, please.
ここで降ります。	It gets off here.	I'll get off here.
観光ツアーはありますか。	Is there a sightseeing tour?	Is there a sightseeing tour?
青色の大型のスーツケースです。	It is the blue large-sized suitcase.	It is the blue large-sized suitcase.
入場料はいくらですか。	Is admission fee how much?	How much is the admission fee?
一緒に行きましょう。	It will go together.	Let's go together.
航空券を家に忘れてしまいました。	The air ticket was forgotten in the house.	I have forgotten my air ticket at home
オペラ座はどこですか。	Is the opera house where?	Where is the opera house?
このフィルムの現像と焼き付けをお願いしますか	We can ask developing and baking this film?	Please develop and print this film.

。		
背中 の マッサージ は いか が だす か 。	Is massage of the back how?	How about a massage of your back?
二つ 目 の 角 に あり ます 。	There are second angles.	There is one at the second corner.
海 側 の 部 屋 に 替 え て くだ さ い 。	Please change into the room of sea side.	Please change into the room facing to the sea side.
お 勘 定 を お ね が い し ます 。	We request calculation.	Give me the check, please.
玉 ネギ を お 願 い し ます 。	We request the ball green onion.	Onions, please.
搭 乗 開 始 時 刻 は 何 時 で す か 。	Is embarcation start time when?	What's the boarding time?
二 十 五 秒 ぐ ら い で す 。	It is about 25 second.	It is about 25 second.
こ の 小 包 を 日 本 へ 送 り た い の で す が 。	It is to like to send this parcel to Japan but.	I'd like to send this parcel to Japan.
ど う い た し ま し て 。	How doing.	You are welcome.
時 計 が 一 日 に つ き 十 分 遅 れ ます 。	The clock ten minutes it is late concerning the first.	The watch loses ten minutes a day.
す み ま せ ん 、 わ か り ま せ ん 。	It is not completed, it is not understood.	Sorry, but I don't understand.
助 け て 。	Helping.	Help!
地 元 料 理 み た い な も の は あ り ます か 。	Are there any local cooking like things?	Is there some local cooking?
番 号 通 話 で か け た い の で す が 。	It is station call to like to go out, but.	Station call, please.
フ ラ ッ シ ュ を 使 っ て も い い で す か 。	It is possible to use the flash, is?	Is it possible to use a flash here?
こ の 街 の 見 どころ を 教 え て くだ さ い 。	Please teach the places of interest of this town.	Tell me, please the places of interest of this town.
初 心 者 で も 大 丈 夫 で す か 。	It is all right even with the beginner?	It is all right even with the beginner?
お 魚 を 頼 ん だ の で す が 。	It is to ask the fish but.	I've ordered the fish dish.
も う ひ と ま わ り 、 大 き い サ イ ズ は 、 あ り ます か 。	Already, is there a size which one it turns, is large?	Do you have a size larger one?
ミ ュ ー ジ カ ル を 観 た い の で す が 。	The musical it is view to want, but.	I'd like to see a musical.
軽 く 焼 い て くだ さ い 。	Please burn lightly.	Please bake it a little.
私 の 席 ま で 案 内 し て くだ さ い 。	Please guide to my seat.	Please guide me to my seat.
ハ イ ド パ ー ク ま で 。	To hydro park.	To hydro park.
全 粒 パ ン を 下 さ い 。	All the grain pans.	Azyme bread, please.
朝 食 込 み で す か 。	It is breakfast included?	Is breakfast included?
そ の 便 は 席 が 二 つ 空 い て い ます か 。	As for that flight has the seat been less crowded two?	Are there two vacant seats for the flight?
わ た し よ 。	Transferring.	It's me!
私 に 合 う サ イ ズ の も の を 見 せ て い た だ け ま せ ん か 。	You cannot show those of the size which is agreeable to me?	Could you show me something for my size, please?
ど れ が 私 の で す か 。	Either one is my?	Which one is mine?
診 断 証 明 書 を い た だ け ます か 。	It can receive the diagnostic certificate?	Could you give me the diagnostic certificate?
は い 。 ト ー ス ト コ コ ナ ッ ツ と	It is. We ask with the toast	Yes, the toast and coconut and

パイナップルをお願いします。	coconut and the pineapple.	the pineapple, please
お金がかかりますか。	Is the money required?	Is it expensive?
このステーキをもう少し焼いてください。	Please grill more this steak.	Please grill more this steak.
桜の季節だね。	Don't you think? it is season of the cherry tree.	It's the season of cherry blossom.
いま、どの辺を通過していますか。	Now, which side is passed?	Where are we going through at the moment?
折り返し電話を下さいと伝えて下さい。	That, please convey the lapel telephone.	Tell him to call me back, please
パスポート、トラベラーズチェック、航空券。	Passport, traveler's check and air ticket.	Passport, traveler's check and air ticket.
メインコースは何になさいますか。	You do main course to something?	What do you have for the main courses?
ここに書いてください。	Please write here.	Write here, please.
常備薬は飲んでいますか。	You drink the regular medicine?	Have you taken your household medicine?
本当にごめんなさい。	The excuse me do truly.	I am awfully sorry.
おつりが間違っています。	Hanging by mistake is.	You gave me the wrong change.
何ですか。	It is what?	What is it?
もしもし。	It does.	Hello
すみません。あのビルは何ですか。	It is not completed. Is that building what?	Excuse me, what is that building?
では、一枚私をとってくれますか。	So, am one I taken?	Then, could you take my picture?
注文したいのです。	It is to like to order	We want to order, please.
日本語を話す人はいますか。	As for the person who speaks Japanese it is?	Is there anyone who speaks Japanese?
彼の部屋番号は何番ですか。	Is that room number no turn?	What is his room number?
人気のあるツアーを教えてください。	Please teach the tour which has popularity.	Please tell me popular tours.
わかりました。	It was understood.	I understand. All right.
オペラグラスは借りられますか。	The opera glass is borrowed?	Can I borrow opera glasses?
基礎化粧品をセットでほしいのですが。	The fundamental cosmetics it is to want being set, but.	I want a set of basic cosmetics, please.
どのブランドがお好みですか。	Some brand likes?	Which brand do you prefer?
それはどこで手に入るのですか。	Is that to enter into the hand somewhere?	Where one can get it?
モーグルをしたいのです。	It is to like to do the mogul.	I want to play the mogul.
食べ方を教えてください。	Please teach the method of eating.	How to eat this, please.
毎食後一日三回服用してください。	Every one day three time after the food please take.	Take three times this after a meal.
領収書はあります。	There is territory acquisition.	I have the receipt.
海が見える部屋がいいです。	The room where the sea is visible is good, is.	I prefer the room facing the sea.

チェックアウトしますので、 ポーターを呼んでください。	Because check out it does, please call the porter.	I'm checking out, please call me a porter.
乗っていきませんか。	It doesn't keep riding?	Can I give you a lift in my car?
ボレーの練習をしましょう 。	You will practise the volley.	Let's practice the volley.
あなたに来てもらえて うれしい。	Being able to have coming to you, it is delightful.	I'm happy to have met you.
ワインはありますか。	Is there wine?	Is there wine?
五ドル三十二セントになり ます。	It reaches five dollar 32 cent.	It reaches five dollar thirty two cents.
二泊したいのですが。	It is two Tomari to like to do, but.	I want to stay two nights.
さようなら。	Way if.	goodby.
おもしろい映画をやってるん だ。	Doing the funny movie, the <nfw>[ru]</nfw> it is.	An interesting film is now showing.
窓を開けてもいいですか。	It is possible to open the window, is?	May I open the window?
ありがとう。	Thank you.	Thank you.
ウエスト旅行社でございます 。	It is the west travelling corporation.	It is the west travelling corporation.
これらのレンズはセールに なっているんですか。	These lenses have become sale, it is?	Are these lenses sale items?
不賛成です。	It is disapproval.	I don't agree.
病気になりました。	It became the sickness.	I'm sick.
われわれにしてほしくない事 はありますか。	Are there times when we do not want making we?	Is there anything that you would not like us to do?

Résumé

Ce travail concerne les systèmes de communication interpersonnelle multilingue médiatisée par la machine, dans deux situations : la communication écrite réfléchie tout terrain et la communication orale spontanée spécialisée. Nous en présentons une vue unifiée en détaillant différentes caractéristiques, dont en particulier les obstacles à la qualité. Nous présentons aussi les mesures de la qualité utilisées actuellement, et montrons que, dans les faits, elles ne donnent pas d'évaluations fiables de la qualité linguistique des traductions, ni de l'utilisabilité et de l'utilité de futurs systèmes opérationnels.

Pour augmenter la qualité en traduction de l'écrit sans restreindre trop le domaine, nous proposons que l'utilisateur aide le système à **lutter contre l'ambiguïté**, obstacle majeur à la qualité, au moyen de dialogues de désambiguïsation interactive (DI) construits automatiquement par un système de règles avec patrons associés aux classes d'ambiguïtés à résoudre. Nous avons évalué expérimentalement l'utilisabilité et l'utilité de la méthode proposée. Enfin, la DI nous a permis de proposer et de maquetter le concept de Document Auto-Explicatif, qui devrait révolutionner la manière dont un lecteur peut accéder au sens d'un document.

Pour l'oral, l'utilisateur peut aider le système en adaptant sa façon de parler et en corrigeant les résultats de reconnaissance de la parole. Pour **lutter contre le bruit**, nous utilisons pour l'analyse des transducteurs finis implémentant des grammaires locales. Nous avons utilisé une méthode subjective pour évaluer expérimentalement la qualité des traductions produites en utilisant les modules d'analyse et de génération du français que nous avons implémentés dans les projets CSTAR-II et NESPOLE!. Nous avons aussi participé à une campagne d'évaluation compétitive destinée à comparer des méthodes objectives et subjectives d'évaluation.

Ce travail permet de conclure que, pour progresser en communication **écrite réfléchie tout terrain**, il faut de la DI, et que, pour progresser en communication **orale spontanée spécialisée**, il faut utiliser un pivot sémantique que l'on peut spécialiser pour des domaines donnés, et pas un pivot orienté par le domaine. Enfin, il faut évaluer les systèmes en utilisation, et pour cela mettre en service les systèmes développés pour des domaines restreints, sans attendre d'hypothétiques systèmes en domaine illimité.

Mots Clés

Traduction Automatique ; Traduction Interactive Fondée sur le Dialogue ; Désambiguïsation Interactive ; Documents Auto-Explicatifs ; Traduction de Dialogues Oraux Finalisés ; Évaluation et mesure de qualité des systèmes de Traduction Automatique ; Traitement de l'ambiguïté ; Traitement du bruit