

Traduction de parole État des lieux et perspectives



Hervé Blanchon
GETA-CLIPS
herve.blanchon@imag.fr

Plan de la présentation

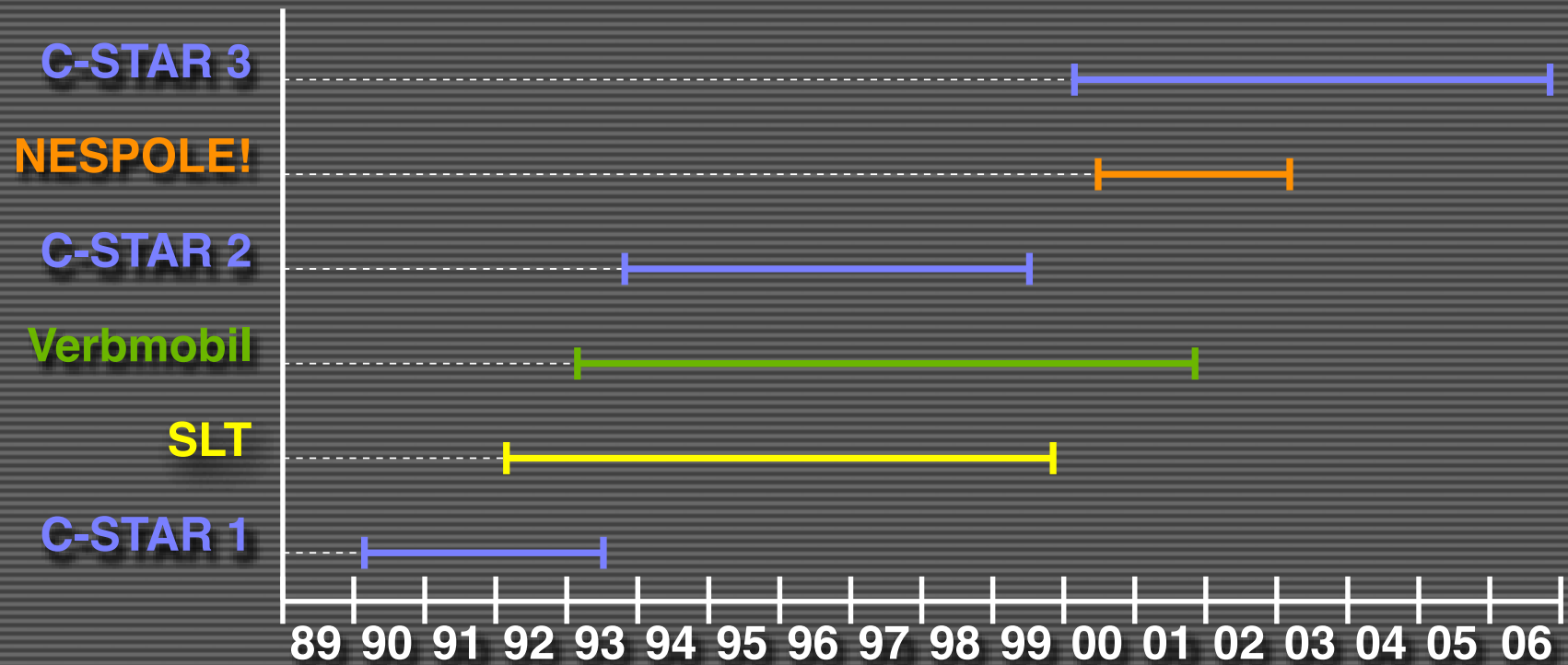
- Introduction
- État de l'art
- Travaux du CLIPS
- Tendances actuelles & Verrous
- Pré-requis pour de développement d'outils
- Bibliographie

La traduction de parole

- Depuis 14 ans
- Convergence
 - Traitement de la parole
 - Traitement de l'écrit

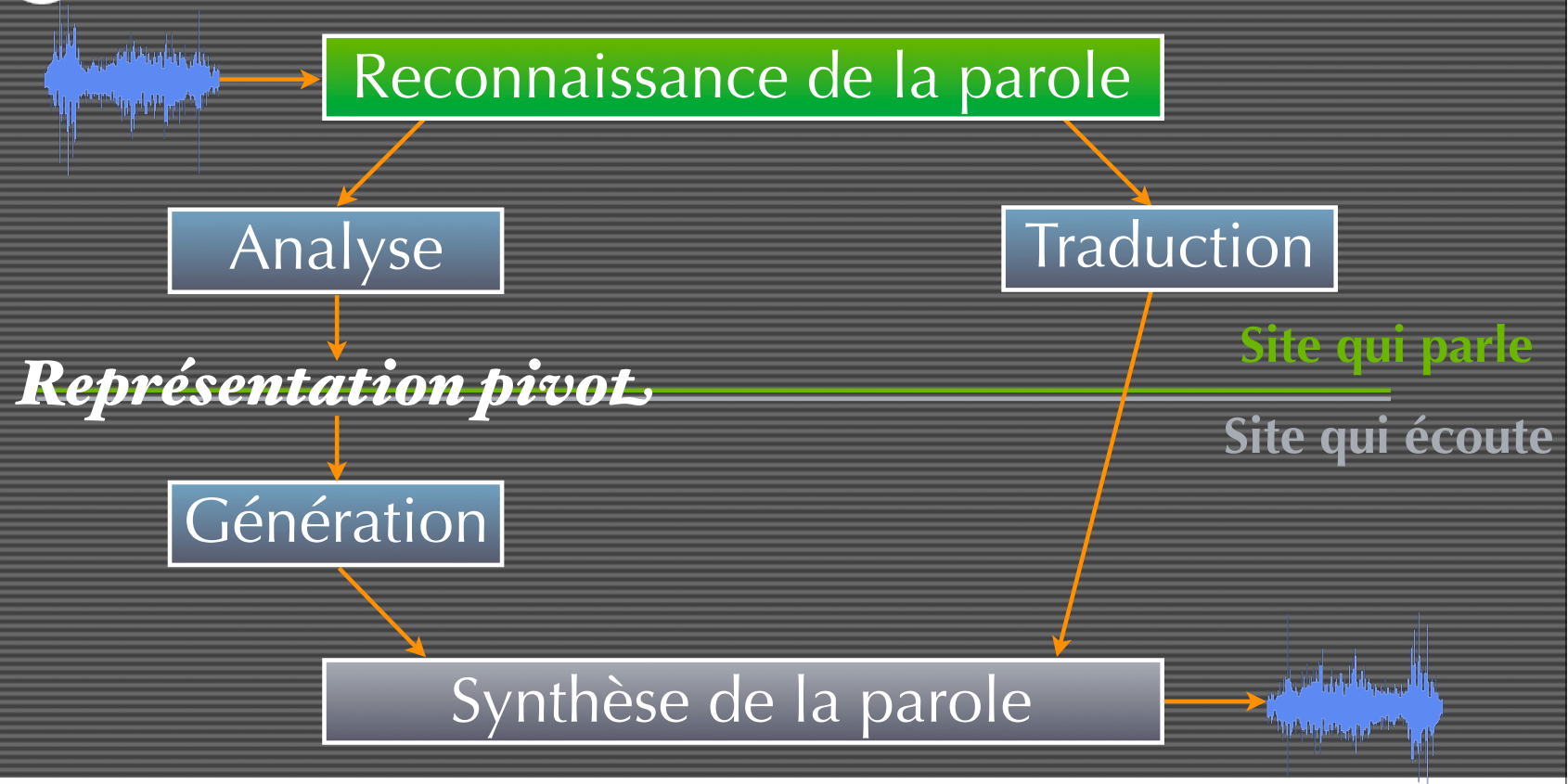
Introduction

○ Bref historique



Introduction

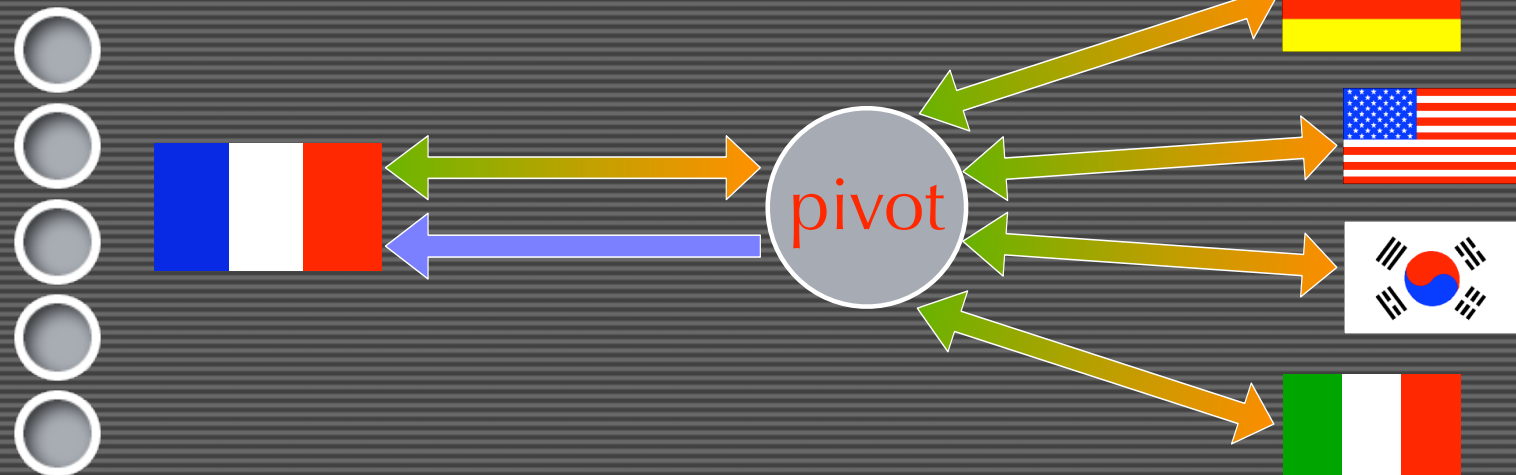
○ Les modèles de traduction



Introduction

○ Avantages de l'architecture par pivot

- ➔ n langues cibles pour le coût d'une
- ➔ rétro-génération en langue source (contrôle)
- ➔ m langues sources pour le coût d'une



○ Inconvénients de l'architecture par pivot

- ➔ Difficulté à faire croître la couverture

Introduction

- Types d'évaluation de la traduction (1/2)

- **Subjectives** — Évaluation humaine

- ➔ Qualité de la langue

- ➔ Transmission du contenu du message

Introduction

- Types d'évaluation de la traduction (2/2)
 - **Objectives** — Évaluation automatique
 - ➔ Comparer la traduction du système à des traductions "étalon"
 - ➔ Scores habituellement utilisés
 - BLEU (BilinLingual Evaluation Understudy)
 - NIST[résoudre des pbs de BLEU]

Introduction

“la qualité de la traduction ne dit pas tout”

- Évaluation de l'**utilisabilité**, **utilité** des systèmes
- Rareté des travaux
- Difficulté de les conduire

État de l'art

○ Les différents projets

● AT&T — IBM

● SLT

● Verbmobil

● C-STAR

● NESPOLE!

État de l'art

	Partenaires	Financement	Langues	Domaine
SLT	SW, CH, GR	Telia Research SRI-I, SICS	FR, EN, SW	Renseignement Aérien (ATIS)
Verbmobil	DE, US (CMU), JP (ATR)	MER allemand	GE, EN, JP	voyage d'affaires ; rendez-vous
C-STAR	US, JP, DE KR, IT, FR, CN	à la charge des partenaires	EN, JP, GE KR, IT, FR, CN	Tourisme
NESPOLE!	IT, FR, DE EN	UE + NSF	IT, FR, GE EN	tourisme ; aide médicale

État de l'art

	Interface	Traduction	∫ Composants	Évaluation
SLT	Microphone	Bimoteur (Surface, QLF)	Boîtes Noires	Subjective
Verbmobil	Téléphone Cellulaire	Multimoteur	utilisation du contexte	Subjective
C-STAR	Visio- Conférence	Pivot ; Vers Lg Cible	Boîtes Noires	Subjective (2) Objective (2-3)
NESPOLE!	Visio- Conférence	Pivot	Boîtes Noires	Subjective

État de l'art

AT&T

- VLVR (Very Large Vocabulary speech Recog.)
 - Broadcast News
- LVCSR (VL Conversational Speech Recog.)
 - Speech-To-Text d'interventions téléphoniques
- TTS (Text-To-Speech)
- Démonstration de TA de parole en 1998
 - Chinois—Japonais

État de l'art

IBM


- MASTOR
 - Multilingual Automatic Speech-to-Speech Technology
 - Anglais—Chinois
- Via Voice Translator
 - Entrée textuelle
 - Traduction synthétisée

MASTOR

Interface

IBM MASTOR System Version 1.50

IBM Multilingual Automatic Speech-to-Speech Translator




Translation | Configuration

LANGUAGES English-to-Chinese

Speech I/O Not Activated

English	i need to see your identification
Chinese	我需要看见你的身份证
Back Translation	(I) (necessary) (saw) (your) (identification/photo id/id)



IBM Corporation 2002

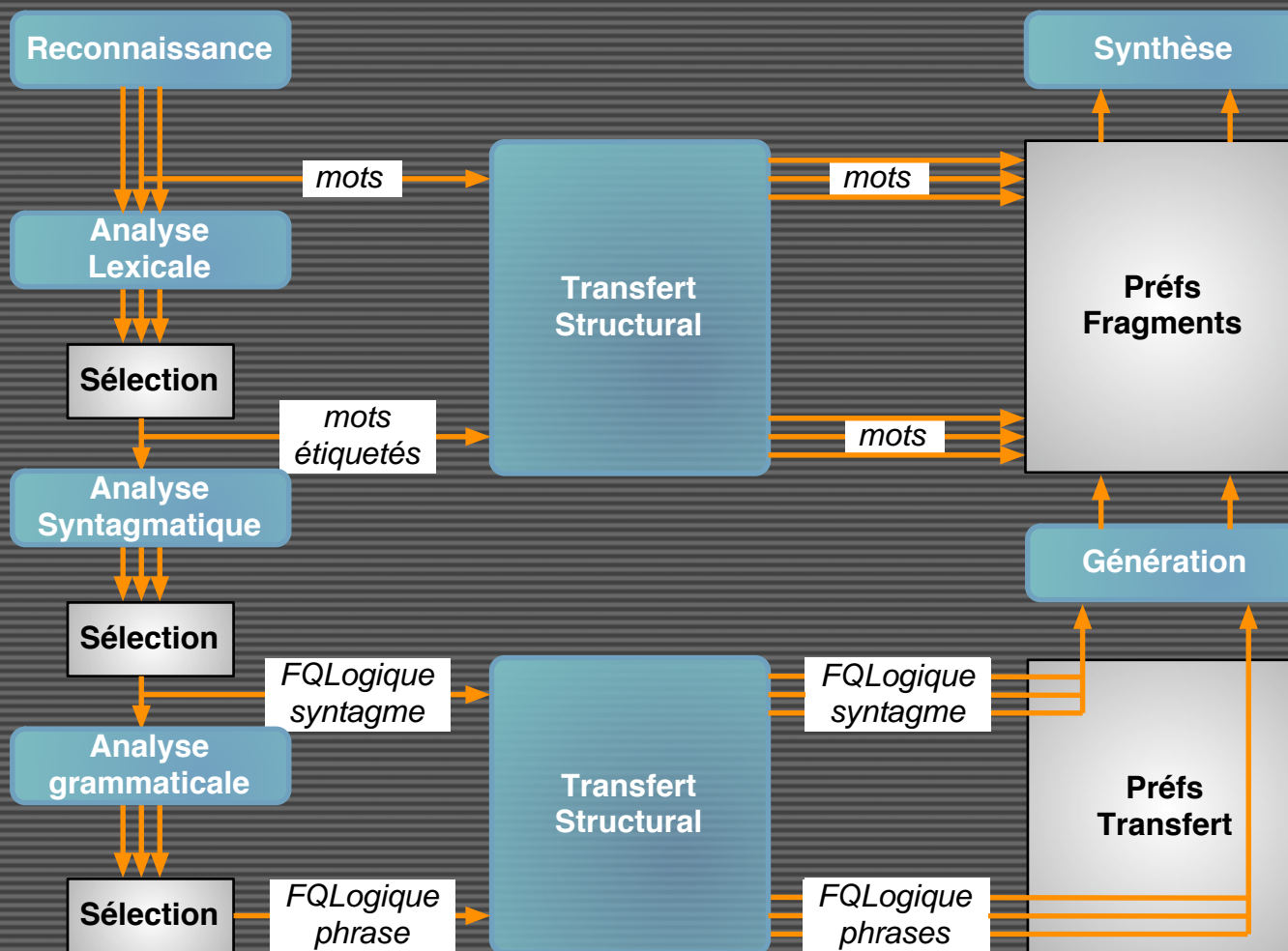
État de l'art

SLT

- Architecture
- Évaluation
- Commentaires

SLT

Architecture



- Évaluation subjective : Parole vers Texte
- Protocole
 - Visualisation transcription
 - Évaluation Hypothèse (acceptable ?)
 - Évaluation de la traduction (7 valeurs)
 - ➔ Utile, Limite, Mauvaise
 - anglais → suédois & anglais → français
 - 400 énoncés non vus en apprentissage
 - (rien sur les évaluateurs)

SLT

Évaluation

	Hyp. OK	Utile	Limite	Mauvaise
En → Sw	80.2%	75.0%	16.0%	9.0%
En → Fr	71.0%	65.3%	7.0%	27.8%

SLT

Commentaires

- Deux niveaux de traduction
 - Mot
 - Forme Quasi Logique
- Bons scores en évaluation
 - Domaine très (extrêmement) restreint

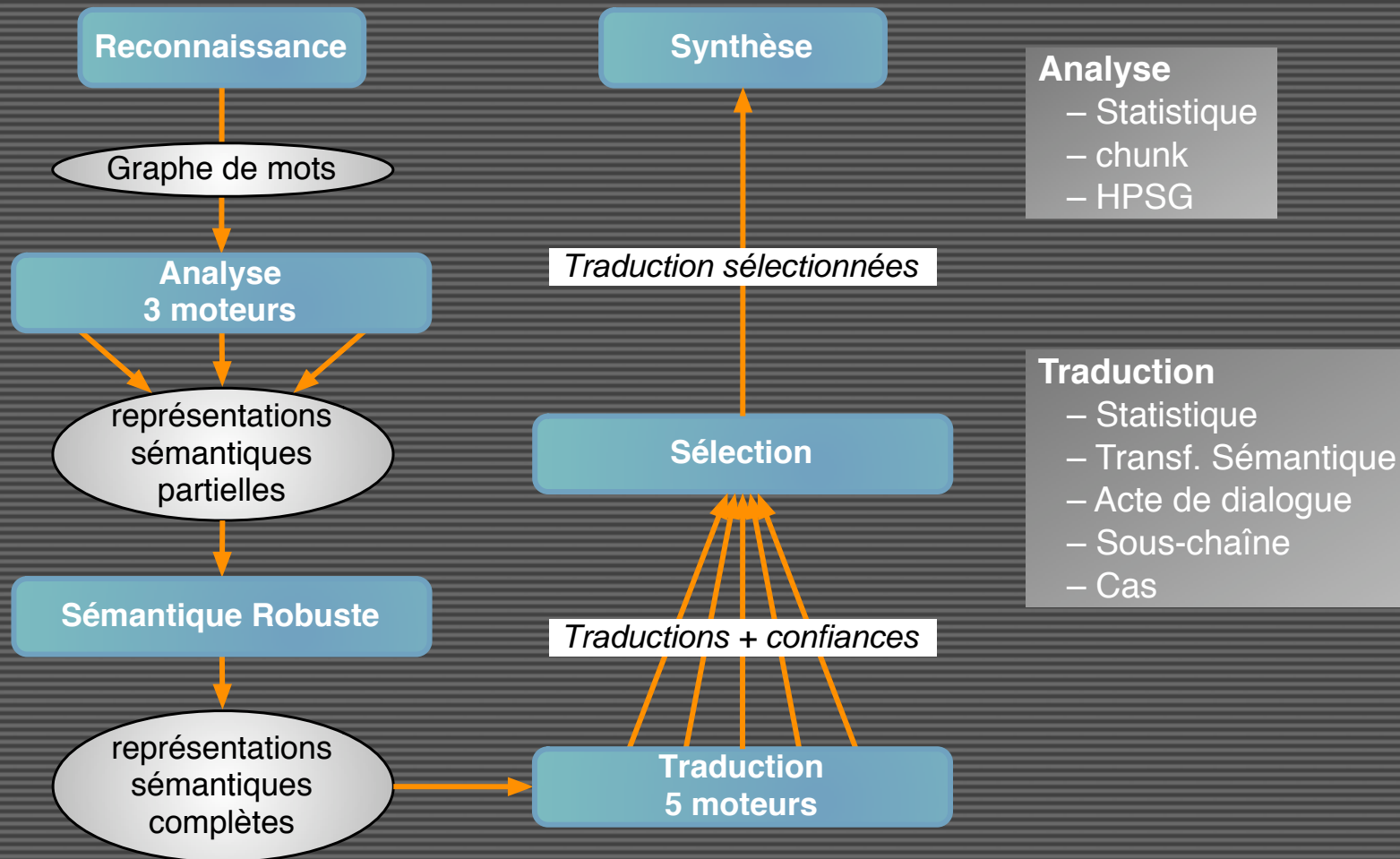
État de l'art

Verbmobil

- Architecture
- Évaluation
- Commentaires

Verbmobil

Architecture



- Évaluation subjective : Parole vers Texte
- Protocole
 - anglais ↔ allemand
 - Un rendez-vous en situation d'usage
 - Qualité Linguistique des tours
 - Évaluation de la traduction (3 valeurs)
 - ➔ Bonne, Moyenne, Mauvaise
 - + Style (humain ou machine)

- 2 campagnes
- Des résultats croisés sur plusieurs critères
 - ➔ Impossible d'avoir un résultat absolu
- 1999
 - 45 dialogues, 1221 tours
- 2000
 - 5069 tours allemand & 4136 tours anglais
 - 65 évaluateurs

- Intégration

- ➔ ≠ analyseur

- ➔ ≠ méthodes de traduction

- Gestion du dialogue

- Évaluation

- ➔ Très complète

- ➔ Résultats multi-critères

État de l'art

C-STAR



- Partenaires
- C-STAR 2
 - Démonstration
 - Architecture
 - Évaluation
 - Commentaires
- C-STAR 3
 - Évaluation (appel à participation)

C-STAR II

Partenaires



CLIPS

Communication in
Natural Language
and Person-System
Interaction



UKA

University
of Karlsruhe



ETRI

Electronics and
Telecommunication
Research Institute



ATR-SLT

Spoken Language Translation
Research Laboratories



CMU

Carnegie Mellon
University



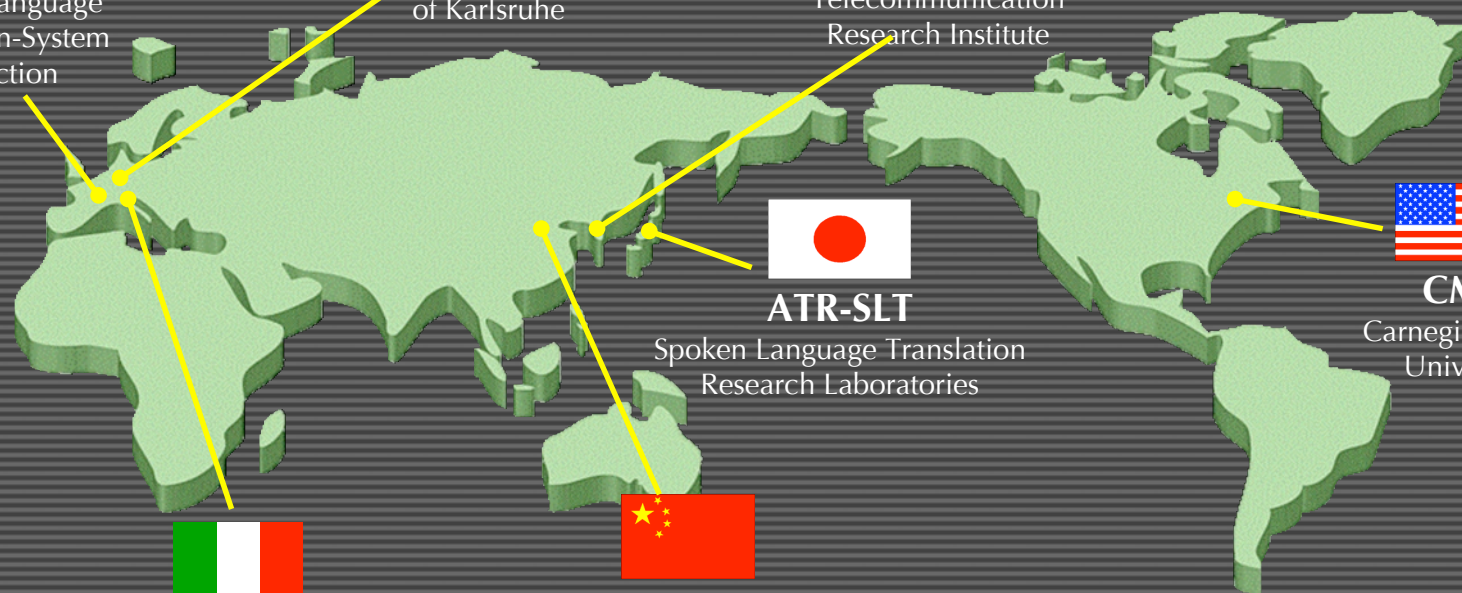
ITC-IRST

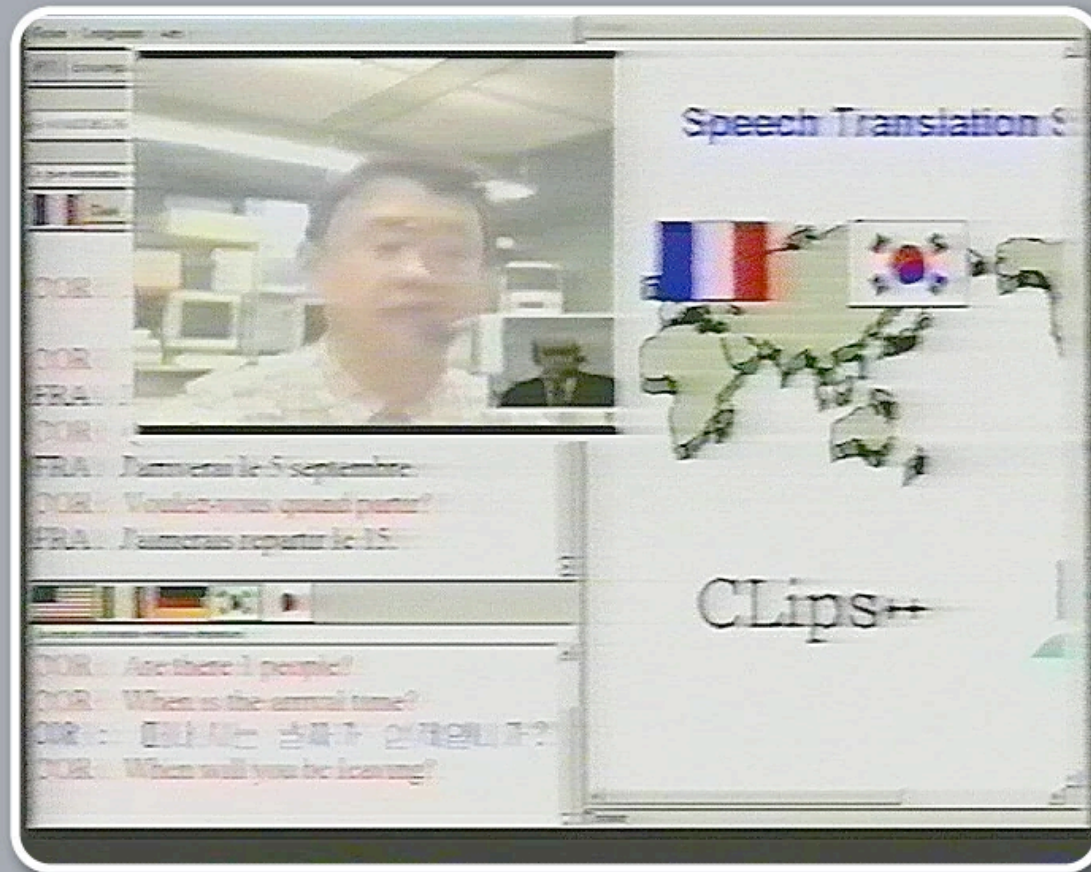
Institute of Scientific
and Technological Research



NLPR

National Laboratory of
Pattern Recognition



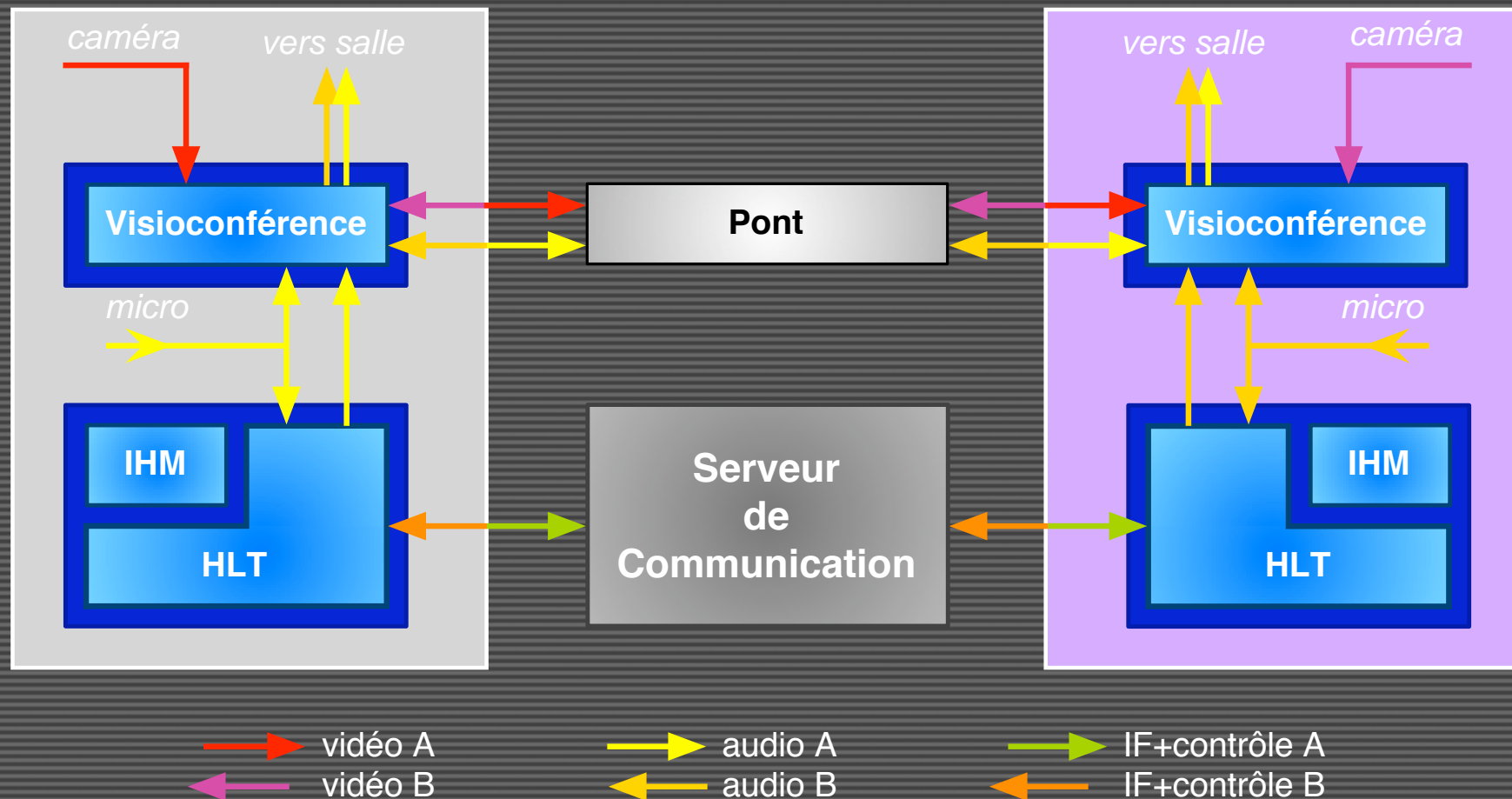


C-STAR 2

Français-Coréen — septembre 1999

C-STAR II

Architecture



- Mesure originale proposée par ATR
- Comparer les performances du système aux performances d'un japonais anglophone dont on connaît le score au TOEIC

- Pas de vraie campagne d'évaluation commune
 - ➔ objectif des démonstrations
- Amélioration incrémentale des systèmes
 - ➔ grâce à des sessions d'entraînement

C-STAR III

- Corpus
- BTEC en langue des partenaires
 - ➔ Traduction + Paraphrases pour évaluation
- Évaluations communes sur BTEC
 - ➔ Internes (membres du consortium)
 - ➔ Ouvertes (appel à participation)

- Corpus used for the evaluation campaign:
 - Basic Travel Expression Corpus (BTEC)
 - Languages: Chinese-English, Japanese-English
 - Domain: tourism-related sentences
 - Media: text in utterance style
 - # of Sentence Pairs: not fixed yet

- Tracks of the Evaluation Campaign:
 - Language Pairs: C-to-E & J-to-E translation
 - Resources Used:
 - ➔ Supplied corpus only (C-to-E, J-to-E)
 - ➔ Supplied corpus + additional linguistic resources
 - ➔ Unrestricted (C-to-E, J-to-E)
- Evaluation Methodology of Translated Results
 - Subjective Evaluation
 - Automatic Evaluation (BLEU, NIST, WER, etc)

État de l'art

NESPOLE!



- Démonstration
- Architecture
- Évaluation
- Commentaire



NESPOLE!

Vidéo de présentation du projet

- Tourisme en situation réelle
 - Un vrai agent de voyage italien
 - Analyse de vraies données (5 classes de scénarii)
 - Collectes en situations correspondantes
- Infrastructure standardisée pour l'intégration des services (*Voice Over IP*)
 - H.323 (NetMeeting, une interface et un tableau blanc)
- 2 plates→formes
 - 2001 : tourisme (vacances d'été et d'hiver)
 - 2002 : tourisme étendu (châteaux, lacs, forfaits)

NESPOLE!

Interface

The screenshot displays a NetMeeting window titled "NetMeeting - Not in a Call" with a menu bar (Call, View, Tools, Help) and a toolbar. A video window shows a smiling woman with a headset. Below the video is the "Nespole Monitor" window, which contains a translation interface. The "System hears:" section shows the text "Hallo I would like to take a trip to Trentino" with "Cancel Translation" and "Send" buttons. The "System understands:" section shows the translated text "Hallo I would like to visit Trentino". An "Audio" section has an "Enable" button and a microphone icon. At the bottom, a "Dialog History Window - Expert mode" shows a list of messages: "IT S.U.: APT informazioni buongiorno", "EN R.S.T:APT information,Hallo", "EN S.H.: Hallo I would like to take a trip to Trentino", "EN S.U.: Hallo I would like to visit Trentino", and "IT R.S.T.: Buongiorno vorrei fare un viaggio in Trentino".

Whiteboard - Aethra Telecomunicazioni s.r.l.
File View Tools ?

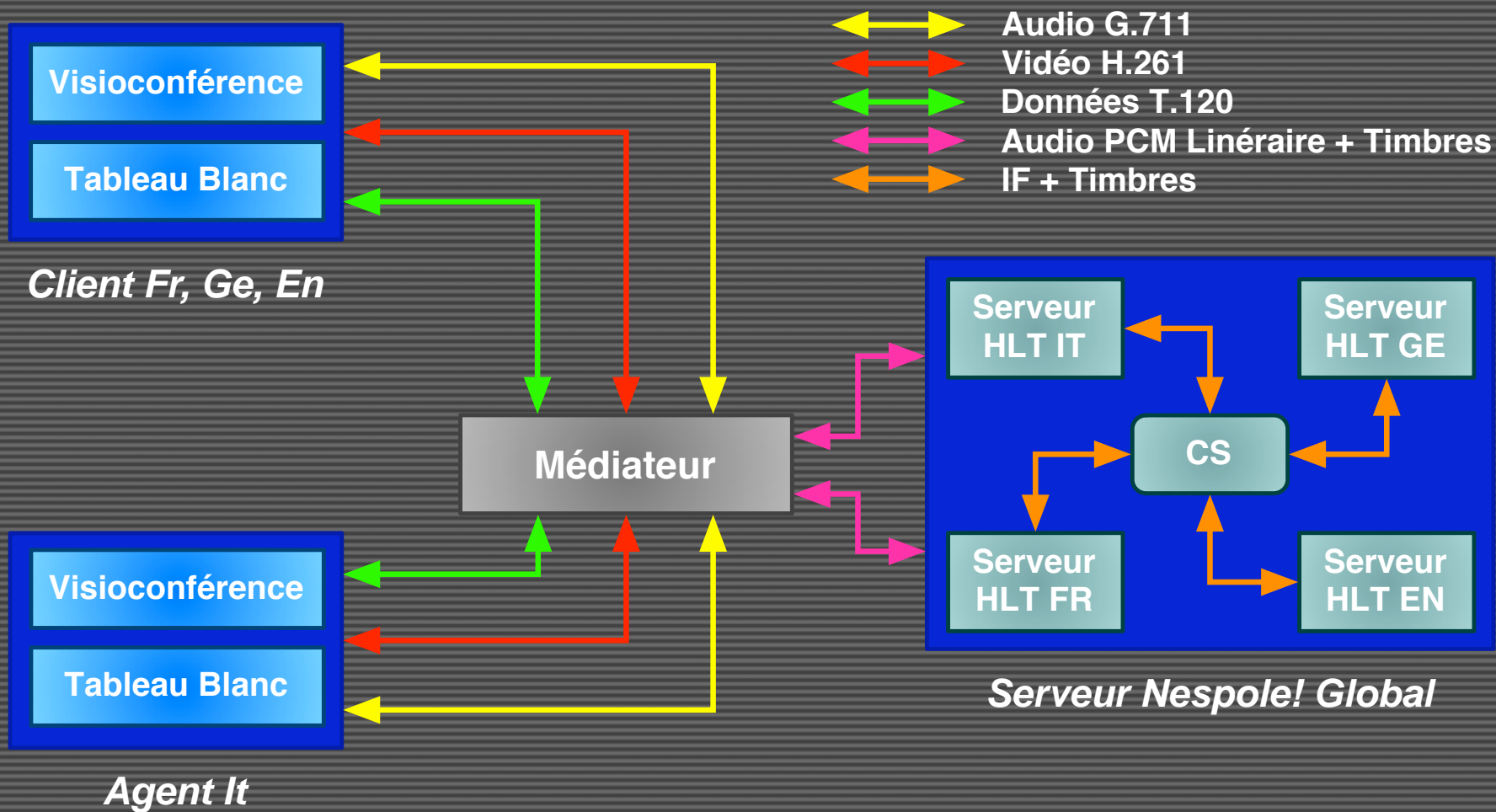
NetMeeting - Not in a Call
Call View Tools Help

Nespole Monitor
System hears:
Hallo I would like to take a trip to Trentino
Cancel Translation Send
System understands:
Hallo I would like to visit Trentino
Audio
Enable
LOCAL REMOTE

Dialog History Window - Expert mode
IT S.U.: APT informazioni buongiorno
EN R.S.T:APT information,Hallo
EN S.H.: Hallo I would like to take a trip to Trentino
EN S.U.: Hallo I would like to visit Trentino
IT R.S.T.: Buongiorno vorrei fare un viaggio in Trentino

NESPOLE!

Architecture



- **Acte de parole** : ce que veut ou fait celui qui parle
 - Suivi de concepts ; évt complété d'arguments
give-information request-information, ...
 - Terminal ; évt complété d'arguments
contradict, dialog-xxx, exclamation, introduce-self, ...
- **Concept** : focus informationnel
 - Suivis de concepts ; évt complétés d'arguments
attitudes (+disposition, +knowledge, ...) ; *actions [prédicats] (+arrival, +reservation, +action, ...)* ; *+availability, +help, +recommendation, ...*
 - Focus concepts (terminaux) ; évt complétés d'arguments
+accommodation, +activity, +information-object, ...

- **Argument** : valeur des variables du discours
 - Top-Level (*Instanciés par : actes de parole et concepts*)
disposition, room-spec, time, location, duration, ...
 - Embedded (*Instanciés par : arguments*)
time-unit, identifiability, quantity, ...
- **Ensemble de valeurs**
 - Les valeurs d'un même type
**times-units* ::= afternoon, day, hour, week, ...*
**objects* ::= *2d-objects*, *acc-boards*, ..., *weather-types*,
account_balance, adaptor, ..., width*

NESPOLE!

IF : Spécification

● Acte de parole

DEF: **greeting**

continuations: (\$bottom\$)

arguments: (<after-rr=> <anti-condition=> <before-rr=> <besides=>
<destination=> <for-whom=> <greeting=> <location=> <manner=>
<origin=> <to-whom=> <while=> <wish=> <with-whom=> <with=>)

comment: comment source — greeting

comment: Both hello and goodbye are now covered by greeting,
with mandatory argument to specify which one.

comment: Ex: "Good morning." = greeting (greeting=good_morning)

comment: Ex: "Goodbye." = greeting (greeting=informal_goodbye)

comment: Ex: "Welcome to Pittsburgh." = greeting (greeting=welcome, location=pittsburgh)

comment: Ex: "Hello from Pittsburgh." = greeting (greeting=informal_hello, origin=pittsburgh)

comment: Ex: "Nice to meet you." = greeting (greeting=first_meeting, to-whom=you)

comment: Ex: "Happy New Year!" = greeting (wish=happy_new_year)

comment: comment source — *cross-sa*

comment: Cross-domain speech act.

NESPOLE!

IF : Spécification

● Concept

DEF: **+accommodation**

continuations: (\$bottom\$)

arguments: (<accommodation-board=> <accommodation-class=>
<accommodation-spec=> <bed-spec=> <class=> <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 -- +accommodation

comment: Ex: "A three star hotel would be fine." = accept+accommodation

comment: Ex: "I want to stay in a hotel." = give-information+disposition+accommodation

comment: Ex: "Does the hotel have a restaurant?" = request-information+contain+accommodation

comment: Ex: "Does the hotel shuttle run every hour?" = request-information+transportation

comment: comment source -- *focus-concepts*

comment: MAIN-PREDICATION

comment: PRED-PARTICIPANT

NESPOLE!

IF : Spécification

● Argument

DEF: **accommodation-spec**

Definition 1 (of 1)

:values

question relative pronoun *accommodations* *hotel-names*
park-names *campsite-names* *shelter-names* *castle-names*

:relations

<accommodation-board=> <accommodation-class=> <called=>
<cultural-style=> <distance=> <duration=> <for-whom=> <for=>
<include=> <location=> <per-unit=> <price=> <specifier=>

:attributes

<identifiability=> <operator=> <polarity=> <quantity=>

: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
+ACCOMMODATION
shared comments from - accommodation-spec=

NESPOLE!

IF : Exemple

- Hypothèse de reconnaissance
 - client : d accord et je voudrais une chambre simple du 10 au 15 septembre
- Simple Dialogue Units (SDUs)
 - SDU1 : d accord ; SDU2 : et je ... septembre
- IFs
 - {c:acknowledge}
 - {c:give-information+disposition+room
(conjunction=discourse, disposition=(desire, who=i),
room-spec=(identifiability=no, single_room),
time=(start-time=(md=10), end-time=(md=15, month=9)))}

État de l'art

- Des systèmes relativement divers
 - Interface
 - Architecture linguicielle
- Des systèmes très proches
 - Taille du domaine à couvrir

Travaux du CLIPS

- C-STAR II
- NESPOLE!

Travaux du CLIPS

C-STAR 2



- Interface Homme Machine
- Intégration des composants
- Reconnaissance vocale du Français
- Analyse Français vers IF

Analyse

- Deux approches
 - Ariane-G5
 - Analyse d'îlots
 - Méthode fondée sur des patrons
 - Mise en œuvre aussi dans NESPOLE!

Travaux du CLIPS

NESPOLE!



- Reconnaissance du français
- Analyse : français vers l'IF
- Génération : IF vers français
- Évaluation

Analyse

Méthode fondée sur des patrons

Automates de reconnaissance de séquences pertinentes

- Étape 1 : Segmentation en SDUs
 - Phrases simples et articulations
- Étape 2 : Détection du type de SDU
 - (1) Focus Concept : thème (activity, accommodation)
 - (2) Acte de Parole Terminal : (acknowledge, exclamation)
 - (3) Unknown
- Étape 3 (1, 3) : Calcul de l'acte de dialogue
 - Relations rhétoriques, acte de parole, attitudes, actions
- Étape 4 : Instanciation des arguments acceptés et complément de l'acte de dialogue

Analyse

Résultat de la reconnaissance automatique de la parole

Découpage en SDUs

SDU₁

SDU₂

SDU_i

SDU_n

Détection d'un acte de parole terminal, focus concepts, unknown

SDU_j Focus_k ou unknown

Construction de l'acte de dialogue

SDU_j acte de parole terminal

Traitement focus concept ou unknown

Rhétoriques

Traitement acte de parole terminal

Construction de l'acte de dialogue

Acte de parole

Construction des arguments

Construction des arguments

Attitudes

Construction de l'acte de dialogue

Complémentation de l'acte de dialogue

Actions

Construction de l'IF

Construction de l'IF

Construction des arguments

Argument A₁

Argument A_j

Argument A_m

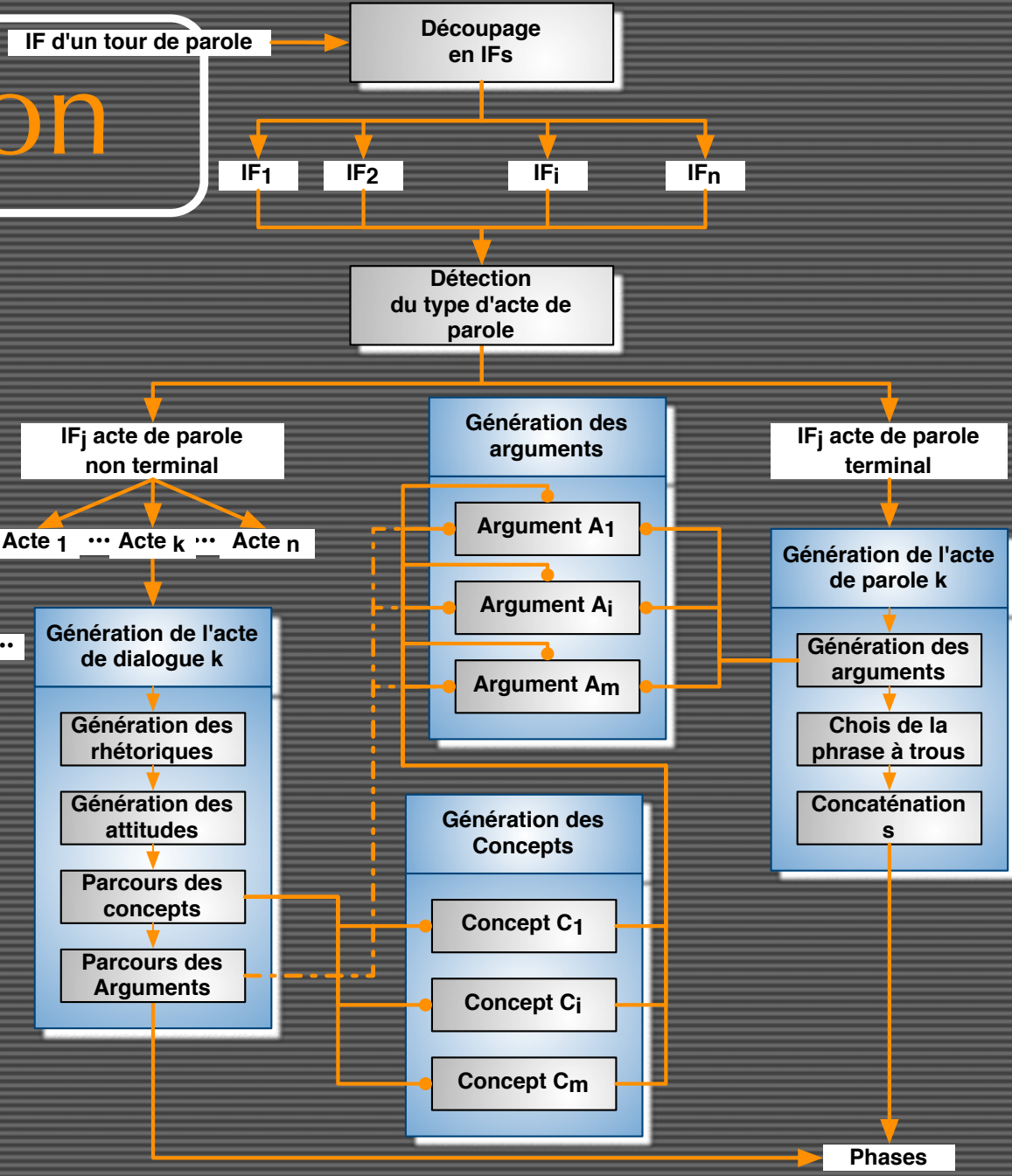
IFs

Génération

Génération concaténative

- Acte de Parole Terminal
 - Une formule figée + quelques arguments indispensables
- Acte de Dialogue complexe
 - Générer les rhétoriques
 - Le DA peut donner un début de phrase figé ou non
 - Générer les attitudes
 - Parcourir l'acte de parole en instanciant au fur et à mesure les différents concepts en utilisant leurs arguments pertinents
 - Générer dans leur ordre d'apparition les arguments non consommés durant la phase précédente

Génération



Évaluation

Exemples Fr → Fr

d'accord oui je le vois aussi	D'accord. Oui. Aussi je vois ce dont nous venons de parler.
oui je vois qu y a un château	Oui. Je vois un château.
d accord et c est possible de visiter le château	D'accord. Et je peux visiter le château.
d accord je en fait je prévois de de venir vers le quinze août	D'accord. ??????? J'ai l'intention d'arriver vers le 15 août.
ben en fait je voudrais savoir parce qu en fait on compte visiter le château de buonconsiglio à trente et je voudrais savoir le quel hôtel serait le plus proche pour s y rendre	Est-ce que vous connaissez ? Nous avons l'intention de visiter Castello Del Buonconsiglio à Trento. Et est-ce que vous connaissez un hôtel proche ?

Évaluation

Exemples Fr → Fr

donc qu'est ce que vous pourriez me conseiller comme hôtel peut être à cavalese

Donc pouvez-vous me recommander un hôtel à Cavalese ?

bonjour madame j aimerais organier une semaine de vacances dans un parc

Bonjour ! J'aimerais un voyage des vacances pour une semaine dans un parc.

et je voudrais aussi une chambre simple à cavalese du quinze au vingt septembre

Et aussi je désirerais une chambre simple à Cavalese à partir du 15 jusqu'au 20 septembre.

oui je vous entends

Oui. Je vous entends.

Évaluation

Protocole

- 2 dialogues de la collecte ShC2 NESPOLE!
 - Non vus pendant la phase de développement
- Données de références (Transcriptions manuelles)
- Hypothèses de reconnaissance (données - 8 kHz, G711)
- 216 tours de parole (Seg. manuelle en 313 SDUs)
- Évaluateurs en école de traduction
 - 3 francophones (→Fr) + 3 italophones (→It)
 - Entraînement + fichier test avant et après la tâche
- 4 scores Very good, Good, Bad, Very Bad
 - Very good + Good = Acceptable
 - Vote majoritaire

Évaluation

Résultats

Reco	56,4%	51%	57,6%	75,3%
Hypo as Paraphrase	66,7%	61,6%	59,5%	76,1%
Monolingue	En→En	All→All	Fr→Fr	It→It
sur Refs/Hypos	68,1/50,4	44/53,4	77,2/57,6	51,1/41,6
Bilingue	En→It	All→It	Fr→It	
sur Refs/Hypos	69,7/50,2	39,7/51,7	76,7/58,1	
	It→En	It→All	It→Fr	
	32,8/29,7	44,6/37,4	37,2/32,5	

- 57,6% de SDUs traduites avec 59,5% atteignables
- Résultats meilleurs que les autres en $Xx \rightarrow Xx$ & $Xx \rightarrow It$
- Résultats comparables en $It \rightarrow Fr$

Travaux du CLIPS

C-STAR 3



- Traduction du corpus BTEC
- Fabriquer rapidement des systèmes
 - ➔ Statistique (\pm tuning)
- Évaluation

Travaux du CLIPS

- Maintenant : Travaux plus exploratoires
 - Intégration reconnaissance-analyse
 - Gestion du dialogue

Tendances actuelles & verrous

- Les **méthodes statistiques** sur des chaînes
 - pas même d'abstraction en classes
 - "lieux", "dates", "adresses"
 - des essais pour monter plus haut avec des structures
- L'**évaluation** en qualité de traduction

Tendances actuelles & verrous

- **Intégration** des composants :
 - De l'approche "boîtes noires"
 - ➔ parole - texte - parole
 - À une véritable "synergie"
 - ➔ En entrée : reconnaissance-analyse
 - ➔ treillis de mots
 - ➔ modèles de langages sémantique
 - ➔ En sortie : génération-synthèse
 - ➔ marques pour la prosodie

Tendances actuelles & verrous

- Gestion du dialogue
 - Aide à la traduction
 - ➔ Liste de référents possibles
 - Aide à la reconnaissance
 - ➔ Thèmes (Modèle de langage dynamique)
 - ➔ Cache de vocabulaire (influence sur le décodage)

Pré-requis pour le développement d'outils

- Cerner un (des) domaine(s) d'application
 - ➔ Des tâches, activités précises
 - ➔ Définir le(s) couple(s) de langues
- Disposer de Corpra
 - ➔ Étude des phénomènes à traiter
 - ➔ Modèles de langages
 - ➔ Vocabulaire

Pré-requis pour le développement d'outils

- Choisir une architecture de traduction
 - ➔ Pivot sémantique lié à la tâche ?
- Mettre en place une architecture générale flexible (plug and play)
 - ➔ Un tableau blanc
 - ➔ dépôt partagé des informations
 - ➔ Fusionner plusieurs niveaux de traitement

Bibliographie

- Wahlster, W. E. (2000). *Verbmobil: Foundation of Speech-to-Speech Translation*. Springer-Verlag. Berlin. 677 p.
- Rayner, M., Carter, D., Bouillon, P., Digalatakis, V., et al. (2000). *The Spoken Language Translator*. Cambridge University Press. Cambridge. 360 p.
- Sugaya, F., Yasuda, K., Takezawa, T. & Yamamoto, S. (2001). *Precise Measurement Method of a Speech Translation System's Capabilities with a Paired Comparison Method between the System and Humans*. Proc. MTSUMMIT VIII. Santiago de Compostela, Spain. 18-22 September, 2001. 6 p.

Bibliographie / NESPOLE!

- 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.
- Lavie, A., Metze, F., Pianesi, F., Burger, S., et al. (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. pp. 6 p.
- Metze, F., Mc Donough, J., Soltau, H., Waibel, A., et al. (2002). *The NESPOLE! Speech-to-Speech Translation System*. Proc. HLT 2002. San Diego, California, USA. March 22-27, 2002. pp. 6 p.
- 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.

Bibliographie / C-STAR 2

- 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. MSC-2000. Kyoto, Japan. October 11-13, 2000. vol. 1/1: pp. 83-87.*
- Boitet, C. & 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.*
- Blanchon, H. & 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.*

Merci!

Questions?