CLIPS++ CONTRIBUTION WITHIN THE C-STAR II CONSORTIUM

Hervé Blanchon

CLIPS-IMAG
BP53
38041 Grenoble Cedex 9, France
herve.blanchon@imag.fr

ABSTRACT

As a late comer partner of the C-STAR II consortium, the CLIPS++ group finally succeeded in building a complete demonstrator for French. The joined effort of our four laboratories (CLIPS, LATL, LAIP, LIRMM) allowed French to be one of the source and target languages demonstrated on July 22nd 1999 during the final demonstration of the six C-STAR partners. The reaction of people attending the demos and the content of the media coverage proved the success of our enterprise. In this paper we will give some information about the CLIPS++ demonstrator and the outcomes of the demonstration itself.

1. INTRODUCTION

The CLIPS++ group joined the C-STAR II consortium as a partner in September 1997. Our group is composed of four research laboratories: CLIPS-IMAG (Communication Langagière et Interaction Personne Système), Université Grenoble I - France, LATL (Laboratoire d'Analyse et de Technologie du Langage), Université de Genève- Suisse, LAIP (Laboratoire d'Analyse Informatique de la Parole), Université de Lausanne – Suisse, and the LIRMM (Laboratoire d'Informatique, de Robotique et de Micro-électronique de Montpellier), Université de Montpellier - France.

The group chose to build its demonstrator based on the interlingua approach. We thus developed the four following modules: a speech recognition module, a French to IF module, an IF to French module, a speech synthesis module. Those modules cooperate between each other and with other partners modules through an integrator.

In the first part of this part of this paper, we will first give a description of the individual components involved in the speech translation process. We will then describe the overall architecture of the CLIPS++ demonstrator. We will then give an account of July 22nd demonstration. The conclusion will be devoted to some global remarks and prospects.

2. COMPONENTS

The CLIPS++ group has been set up and leaded by the CLIPS laboratory according to the know-how of each research team. Joining the consortium as late as we joined it implied to gather some proven technologies so as to be able to fill the gap rapidly with other partners technologies. We also chose to use the pivot approach for the translation process.

2.1 Overall approach

The pivot named IF (Interchange Format) relies on dialogue acts acts, concepts and arguments. Dialogue acts describe the speaker's intention, goal, need (example: `give-information`, `request-information`, `introduce-self`, …). Concepts define the "about what" that is the focus of the dialogue act. Several concepts may appear in one IF (examples: `confirmation`, `reservation`, `features`, `train`, `flight`, `room`, …). Arguments are giving the values of the discourse variables (example: `person-name`, `location`, `price`, `time`, `date`, …). Here is the IF for the sentence "The week of the twelfth we have single and double rooms available" pronounced by an agent: `a:give-information+availability+room(room-type=(single ; double), time=(week, md12))`.

The global architecture for speech translation using the IF approach is thus the following:

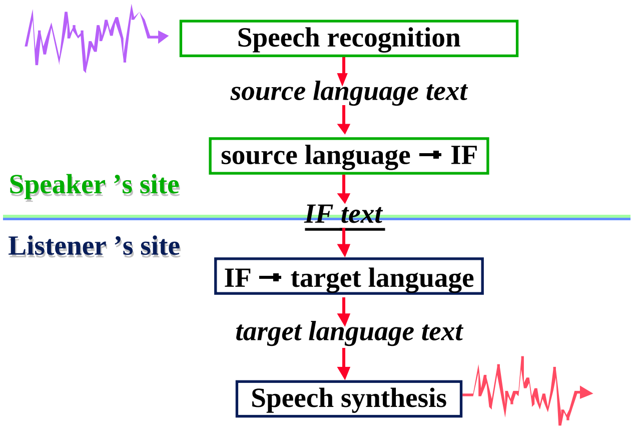


Figure 1: Overall components interaction

The GEOD team of the CLIPS took care of French speech recognition. The GETA team of the CLIPS designed and implemented the French —> IF module. The LATL was in charge of the IF —> French module. Finally, the LAIP developed the text to speech module for French. The LIRMM has been working on French —> IF with a methodology different than the one used at the GETA. This component has not been used in the demonstrator yet.

2.2 Speech recognition

The module is designed for speaker independent continuous speech recognition with a vocabulary specialized for the tourism domain of about 10k words. It is based on a client-server architecture. A speech recognition server is used through "light" clients on the network. It is built with the JANUS III toolbox in collaboration with Carnegie Mellon University.

The module is implemented using both:

- A context independent markovian acoustic model trained on 10 hours of continuous speech (BREF-80 corpus),
- A stochastic language model trained on a 140 million words corpus and optimized for the tourism task.

2.3 French —> IF

This module is developed with Ariane-G5, a generator of machine translation systems that use five specialized languages for linguistic programming, running under VM/ESA/CMS. The module is running on three machines: CLIPS, Grenoble (IBM 9221-130, 3.5 mips); CCSJ, Marseille (IBM 9672-R14, 40 mips); IBM, Montpellier (IBM 9672-RX5, 60 MIPS). For the demonstrations, the last site is used.

The module input is an orthographic transcription of a spoken utterance. The following steps are done one after the other:

- Morphological analysis and lemmatization of the words of the text,
- First access to transfer dictionaries towards IF,
- Syntactical analysis for the recognition of semantically interesting structures: dates, quantity, numbers, prices, …,
- Second access to transfer dictionaries towards IF,
- Syntactical and morphological generation of the resulting IF.

2.4 IF —> French

The IF –> French module was partly developed with GB-Gen, a broad lexical and syntactical coverage syntactic generation tool. This is a deterministic tool based on a generative grammar.

The transformation between an IF and an French text is made in three steps:

- Mapping of an IF into a GB-Gen semantic structure,
- Application of the GB-Gen generation procedures to produce a syntactic structure,
- Application of the GB-Gen morphological rules to produce a text in French.

2.5 Speech synthesis

The LAIPTTS module is a “text to speech” synthesizer based on rules. Synthesis is made in three steps: -text to phoneme mapping, - prosody generation, - signal generation.

Text to phoneme mapping is made through a set of general rules (540) and specialized ones for numbers, abbreviations, fixed expressions, etc. This step is using general (7000 words) and specialized dictionaries (proper nouns). The prosody generation uses psycholinguistic rules. The signal generation uses the Mbrola technique developed at the Mons University.

3. ARCHITECTURE & INTERFACE

3.1 Demonstrators integration

Two kinds of data are exchanged between the systems: video and sound for the visio-conference, data supporting the translation process. The visio-conference is handled by commercial visio-conference systems. According to the number of participant involved in a session, the connections are made point to point or through a MCU (Multi-Cast Unit).

Data exchanged for the translation process itself are reduced to a minimum and are made of strings of characters. What is transferred are the IF structures (mandatory for the translation), and the recognition hypothesis and the generation from the locally produced IFs (for trace purposes). Those exchanges are made through a communication server. The systems that are willing to communicate are connected to this communication server via the telnet protocol.

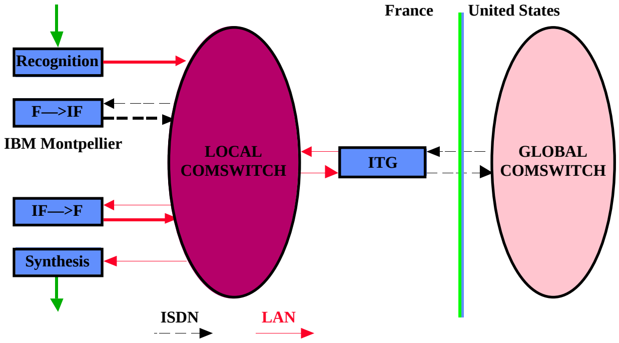


Figure 2: Global architecture of the demonstrator

3.2 Local components integration

All the components of the CLIPS++ demonstrator are servers. Thus none of them is communicating directly with an other one. They are all connected to a Local ComSwitch. We chose this architecture, because it is very convenient for distributed development and versatile.

We do also believe that in the foreseen commercial systems, it is an interesting one because what will only be necessary to provide the customers with are light clients. Those clients are going to interact with very powerful servers whose software will be updated for the benefit of all the users at the same time. This architecture fits also the needs of mobile applications.

3.3 Interface

The whole interface is distributed among two screens: one screen for the visio-conference and one screen for the user interface.

In the client setting the user interface screen is divided in two parts. On the left hand side there is the interface of the speech translation system. On the right hand side is devoted to a web viewer. A picture of this interface is given in Annex 1.

As far as the interface of the speech translation system is concerned it is described in Annex 2. The PTT button is used to control the speech recognition module. Once pushed the button turns orange meaning that the recognizer is ready. When the user starts speaking the button turns green meaning that recording and recognition are going on. When there is no more acoustic signal the recording ends and the button turns gray again. It would thus have been possible to handle continuous recognition without any push to talk. We chose not to produce IFs for the recognition strings that were not matching enough the spoken utterance. That is a way to save time. We have thus implemented a ‘recognition hypothesis validation’ button.

In the agent setting, the user interface screen is also divided in two parts. On the left hand side there is the interface of the speech translation system (the same as the one used in the client setting). The right hand side is shared between a web viewer and a web selector. The web selector allows the travel agent to send web pages to the client. Sent web pages are displayed by the web viewer to give some feedback to the agent.

4. JULY 22nd DEMONSTRATION

4.1 Setting, audience and media coverage

July 22nd CLIPS++ hosting demonstration was held in the lecture hall of the Maison Jean Kuntzmann of the IMAG institute in front of 90 persons. We had successfully invited representative from the Universities (Université Joseph Fourier, Institut d’Informatique et de Mathématiques Appliquées de Grenoble, Institut National Polytechnique de Grenoble, Université Marseille Saint Jérôme), the CNRS, Industries and research centers (IBM, Xerox, Digigram, Azimut, FranceTelecom, Magellan Ingeniérie, CNET, Spacio Guide) and ANVAR.

Several private demonstrations of the whole system or components have also been made for Alcatel, Thomson, Texas Instrument, Schneider, Parrot, CyberStudio, EDF, Soprane, Gany Média.

The media coverage was quite good on the 22nd of July and the week after. There is still a great interest in the project. We have also several publications in September and, may be, a prime time TV presentation.

The picture is the following:

- News agency: AFP, AGRAP, Temps d’m, Agence REA, Editing corporate,
- Press: Le Monde (Interactif), Libération, Le Dauphiné Libéré, Science et vie, Science et Vie junior, Micro-Hebdo, Science et Vie Micro, France Soir
- Television: France 3,
- Radio: Radio France Isère, France Inter, France Info, BFM, Europe 2, RMC, RTL, RCF Isère.

4.2 Shown scenario

July 22nd day was organized as follows. We had an introductory session in the morning with a talk about C-STAR II, its history, the translation techniques and the CLIPS++ demonstrator. Afterward, we offered a buffet to our guests before the actual demonstration at 14 hours.

We actually demonstrated the following scenario:

- A client was entering a virtual tourism agency with branches in the United-States, Korea and Germany, he had first a greeting session with all the travel agents available,
- The client planned next a trip to Taejon with the Korean travel agent: booking a flight from France, asking directions to reach Taejon from Seoul, Booking a hotel, asking for tourist attraction around Taejon, paying with a credit card,
- Then, the client organized a trip to New-York with the American travel agent: booking a flight for France, asking for an hotel and a base-ball game, paying with a credit card,
- Afterwards, the client arranged a trip to Heidelberg: booking a train and a hotel, asking for information about tourist attraction, paying with a credit card.
- For the demo purpose the client finally said thank you to the three travel agents.

We also participated in the ETRI hosting demonstration as a travel agent.

4.3 Outcomes

Although we felt, preparing the demonstration, that the proposed scenario was a little bit too repetitive, the audience did not make any remark on that matter. As late comers it was somehow hard for us to have another one. We carefully tried to use different wordings to express several time the same goal. The opening ceremony was well received and a very nice entry. Some mistakes were also entertaining. The demonstration lasted for about half an hour and the people in the public said that they have not seen the time fly.

The dialogue with the Korean agent has been appreciated a lot (the Hangul script and the almost never heard language), picturing clearly the need for speech translation when there is a need for communication but no common language to support it. In Europe people are less sensitive to that matter as far as German and English are concerned. Most of the examples taken by the media for future application were about French-Korean and French-Japanese dialogues. We tried then to explain the need for that technology even if some communication is possible when the message has not to be distorted or misunderstood.

The audience did not comment also the speed of the overall translation process. It seemed to go at a reasonable pace. The exchange of web pages with live picture was also highly appreciated. It is something that did not seemed very new, except live picture, but people would have been surprised if we had not presented such a thing.

5. CONCLUSION & PROSPECT

We are very enthusiasts on pursuing our work on speech translation. It proved to be a goal that is full of research topics for academic people and, according to the feedback we got, an area for good potential commercial applications.

The acceptance of the NESPOLE! (Negotiating though Spoken Language in E-commerce) project submitted in the User-friendly Information Society chapter of the fifth framework program of the European community will allow us to work with better conditions on that topic.

IBM France assistance allowed us to reach acceptable answer time for our French to IF module written in Ariane-G5. Their interest for our work and their will to advertise it among several communities of potential users (we will have a demo for IBM and those potential users on the 27th of October) is also a great perspective.

6. REFERENCES

[1] Blanchon, H., Boitet Ch., Caelen, J. (1999) La participation francophone au Consortium C-STAR II. A paraître dans la Tribune des Industries de la Langue.

[2] Boitet, Ch. (1998) Problèmes scientifiques intéressants en traduction de parole. Proceedings of TAL+AI/NLP+IA, Moncton, Canada. Septembre 98.

[3] Boitet, Ch. (1997) GETA’s MT methodology and its current development towards networking communication and speech translation in the context of the UNL and CSTAR projects. Proceedings of PACLING-97, Ome, Japan. 2-5 September 97.

[4] Keller, E., & Zellner, B (1998). Motivations for the prosodic predictive chain. Proceedings of ESCA Symposium on Speech Synthesis. Paper 76, pp. 137-141. Jenolan Caves, Australia.

[5] Keller, E. (1997). Simplification of tts architecture vs. Operational quality. Proceedings of EUROSPEECH '97. Paper 735. Rhodes, Greece. September 1997.

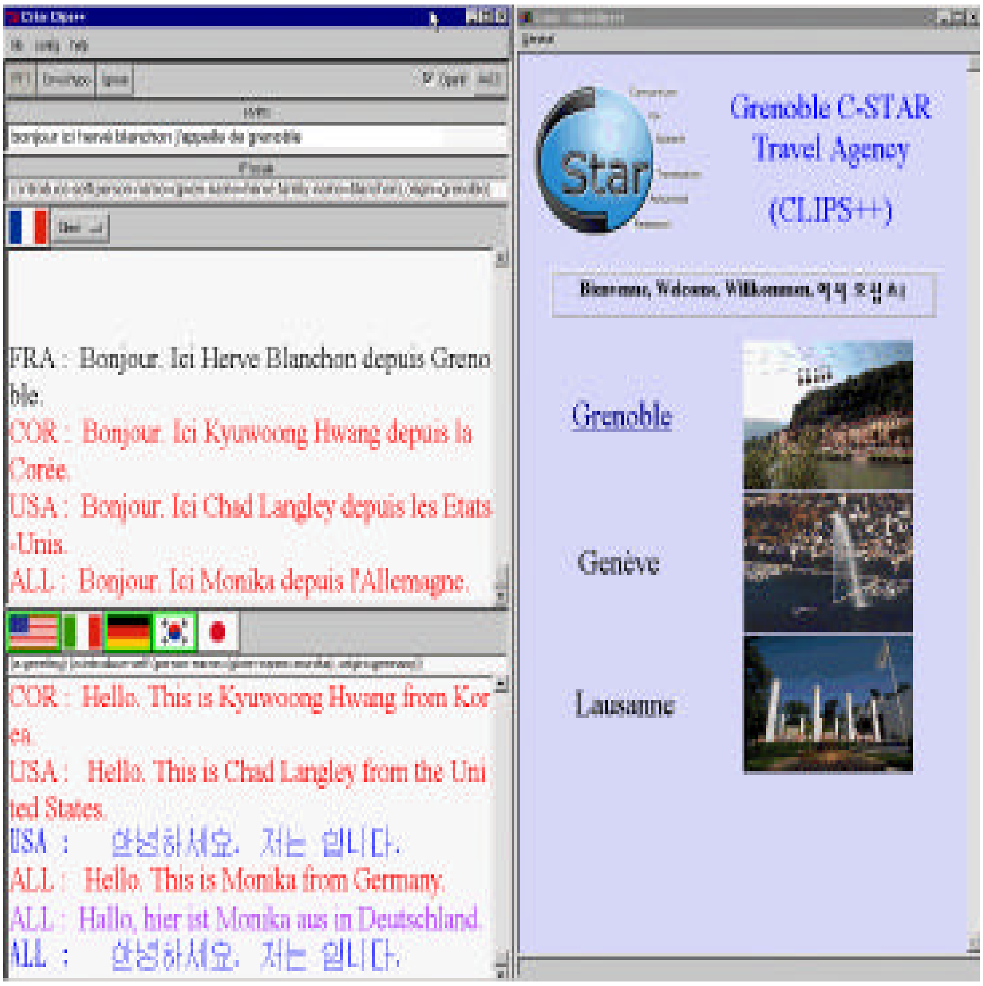
[6] Keller, E., & Werner, S. (1997). Automatic Intonation Extraction and Generation for French. 14th CALICO Annual Symposium. ISBN 1-890127-01-9, West Point. NY. June 1997.

[7] Vaufreydaz D., Akbar M., Rouillard J. (1999) A Network Architecture for Building Application that Use Speech Recognition and/or Synthesis. Proceedings of Eurospeech'99, Budapest, pp 2159-2162

[8] Wehrli, E. & Wehrle T (1998) Overview of GBGen. Proceedings of 9th International Workshop on Natural Language Generation, Niagara-on-the-lake, Canada. August 1998.

[9] Wehrli, E., Wehrle T., Mengon J., Vandeventer A. (1999) Une approche efficace à la génération syntaxique. Le système GBGen. Proceedings GAT'99, Grenoble, France. octobre 1999

ANNEX 1: USER INTERFACE



bonjour ici hervé blanchon j'appelle de grenoble

FRA : Bonjour. Ici Herve Blanchon depuis Grenoble.

COR : Bonjour. Ici Kyuwoong Hwang depuis la Corée.

USA : Bonjour. Ici Chad Langley depuis les Etats-Unis.

ALL : Bonjour. Ici Monika depuis l'Allemagne.

COR : Hello. This is Kyuwoong Hwang from Korea.

USA : Hello. This is Chad Langley from the United States.

USA : 안녕하세요. 저는 입니다.

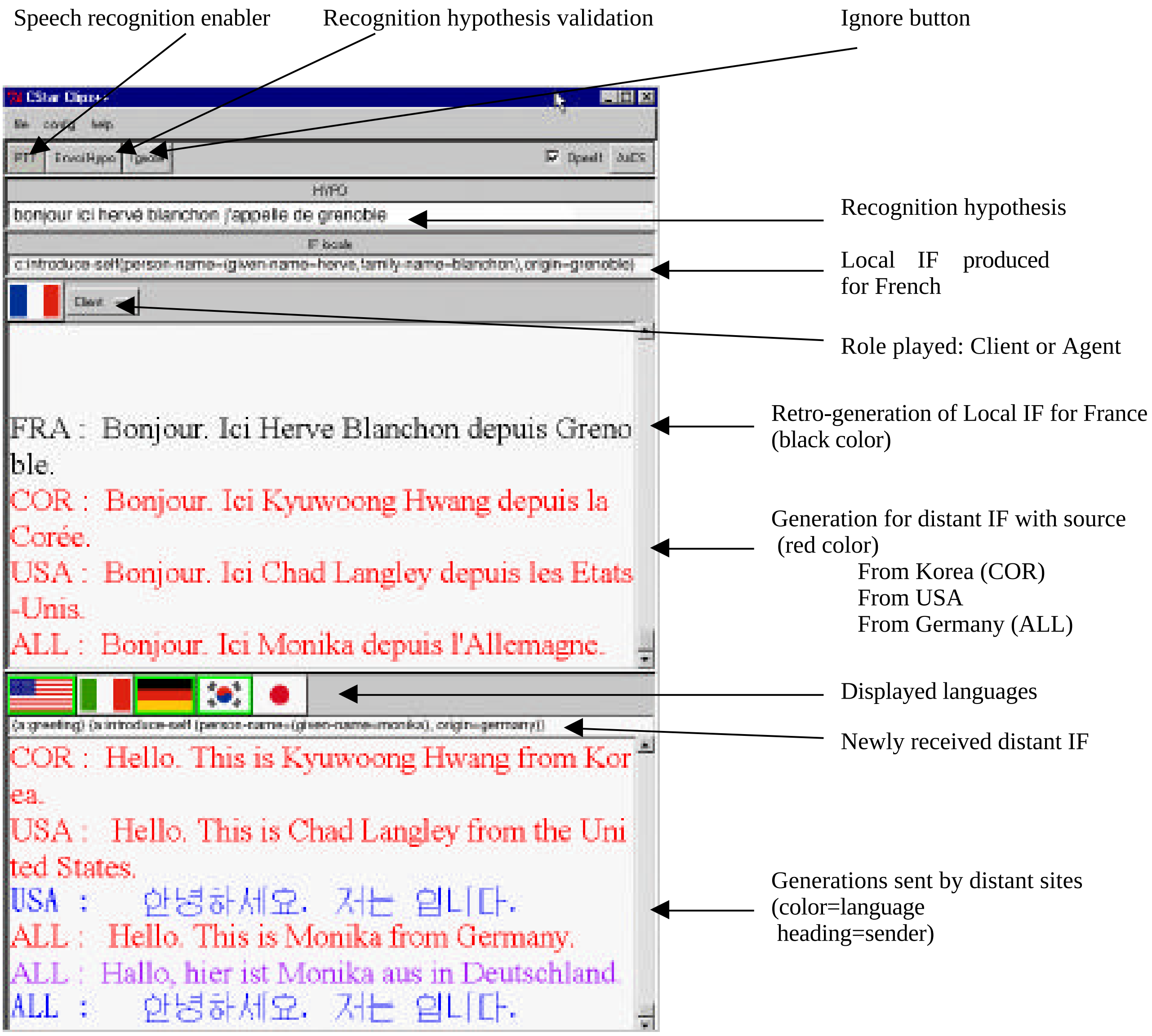
ALL : Hello. This is Monika from Germany.

ALL : Hallo, hier ist Monika aus in Deutschland.

ALL : 안녕하세요. 저는 입니다.

Figure 3: User interface of the demonstrator

ANNEX 2: SPEECH TRANSLATION INTERFACE



bonjour ici hervé blanchon j'appelle de grenoble

c:introduce-self(person-name=(given-name=herve,family-name=blanchon),origin=grenoble)

FRA : Bonjour. Ici Herve Blanchon depuis Grenoble.

COR : Bonjour. Ici Kyuwoong Hwang depuis la Corée.

USA : Bonjour. Ici Chad Langley depuis les Etats-Unis.

ALL : Bonjour. Ici Monika depuis l'Allemagne.

{a:greeting} {a:introduce-self (person-name=(given-name=monika), origin=germany)}

COR : Hello. This is Kyuwoong Hwang from Korea.

USA : Hello. This is Chad Langley from the United States.

USA : 안녕하세요. 저는 입니다.

ALL : Hello. This is Monika from Germany.

ALL : Hallo, hier ist Monika aus in Deutschland.

ALL : 안녕하세요. 저는 입니다.

Speech recognition enabler

Recognition hypothesis validation

Ignore button

Recognition hypothesis

Local IF produced for French

Role played: Client or Agent

Retro-generation of Local IF for France (black color)

Generation for distant IF with source (red color)
From Korea (COR)
From USA
From Germany (ALL)

Displayed languages

Newly received distant IF

Generations sent by distant sites (color=language heading=sender)

Figure 4: Speech translation interface