

A Multilingual Platform for Building Speech-Enabled Language Courses

Manny Rayner¹, Pierrette Bouillon¹, Nikos Tsourakis¹, Johanna Gerlach¹
Claudia Baur¹, Maria Georgescu¹, Yukie Nakao²

¹ University of Geneva, TIM/ISSCO,
40 bvd du Pont-d'Arve, CH-1211 Geneva 4, Switzerland

² LINA, Nantes University,
2, rue de la Houssinière, BP 92208 44322 Nantes Cedex 03

{Emmanuel.Rayner, Pierrette.Bouillon, Nikolaos.Tsourakis, Johanna.Gerlach}@unige.ch,
Maria.Georgescu@unige.ch, baur5@etu.unige.ch, yukie.nakao@univ-nantes.fr

Abstract

We present CALL-SLT, a generic multilingual speech-enabled Open Source CALL system based on the “translation game” idea of Wang and Seneff, focussing on recent enhancements which allow the instructor to define a structured language course divided up into a set of lessons. Each lesson picks out a subset of the corpus using a combination of semantic and syntactic constraints. We describe how the “structured lesson” framework interacts with the spoken help facilities offered by the system, and outline the initial sets of lessons we have constructed for the English, French and Japanese versions of CALL-SLT.

Index Terms: CALL, speech recognition, translation, interlingua, online help, English, French, Japanese

1. Introduction

As everyone knows, it is almost impossible to achieve any fluency in a foreign language without conversational practice; the most effective way to pick up a language is to spend time in a country where it is spoken, or, failing that, to talk regularly with a native speaker. Unfortunately, neither of these options are available for many language students. It is consequently interesting to explore techniques for creating automatic conversation partners. There is a continuum here with respect to the degree of free variation permitted on the student's part. At one extreme, which represents current commercial mainstream systems like TellMeMore and RosettaStone¹, we have “closed-response” systems [1]. The system indicates to the student precisely what they are supposed to say; the student repeats it, and is then graded on their pronunciation. Though undoubtedly useful, this is less than ideal. As pointed out in [2], effective systems for language learning need to allow the learner to produce large quantities of sentences on their own, something that is, by definition, impossible to realise in a closed response application.

At the other extreme, one can try to build a fully interactive system, which carries out a genuine conversation with the student in some kind of simulated environment; a high-profile example is TLCTS [3]. It is, however, extremely challenging to make a system of this kind adequately robust, and in particular to implement sufficiently powerful language understanding. The system we describe here, CALL-SLT [4], explores an intermediate solution. This has its roots in the “translation game” idea of Wang and Seneff [5], who successfully reused speech

and language technology developed at MIT under other projects [6] to build a speech-enabled game for students who wished to practise Chinese. In their game, the system shows English sentences to the student, who has to respond with a spoken Chinese translation. The system matches the student response against the original prompt, and produces informative feedback. Most of the subjects who participated in the initial study were positive about the system.

CALL-SLT uses a similar basic strategy, though many of the details are different; in particular, we use another approach to speech recognition, present prompts to the student in a more principled way, and include a help system which allows students to share experience across languages. In Section 2, we briefly present more background on the system. The rest of the paper describes new enhancements not described in previous publications, which make it possible to organise the corpus of examples used by the system as a set of structured lessons, each one organised around a specified syntactic or semantic theme. We present the “lesson structure” framework and outline the initial sets of lessons we have built for the English, French and Japanese versions of CALL-SLT.

2. The CALL-SLT System

CALL-SLT is an Open Source speech-based CALL application for beginning to intermediate-level language students who wish to improve their spoken fluency. The system runs on a medium-range Windows laptop; it can also be deployed on a mobile platform, using the client/server architecture described in [7], with performance identical to that of the laptop version. The current version uses a restaurant domain, and supports English, French, Japanese, Swedish, Arabic and German as L2s, with English or French as the L1². Vocabulary varies from around 150 to around 500 words per language, and covers basic situations such as reserving a table, ordering food and drink, asking for the bill, and so on. Table 1 shows typical examples of coverage.

CALL-SLT leverages earlier work on Regulux, a platform for building systems based on grammar-based speech understanding [8] and MedSLT, an interlingua-based speech translation framework [9, 10], to develop a generic CALL platform centered on the “spoken translation game” idea. Our initial experiences, including an extensive test carried out on several hundred Swiss high-school students [4], have demonstrated that the Regulux/MedSLT architecture is a good fit to this type of ap-

¹tellmemore.com; rosettastone.com.

²Versions with Japanese L1 and Greek L2 are in preparation.

English
I would like a mint tea
A tea and a coffee please
Do you have a table for four people
Could I reserve a table for seven thirty
Do you accept credit cards
French
Puis-je avoir une bière (Could I have a beer)
J'aimerais du fromage rape (I would like some grated cheese)
Je voudrais une table dans le coin (I would like a table in the corner)
Est-ce que je pourrais voir le menu (Could I see the menu)
Japanese
Biiru nihai onegai shi masu (I would like two beers)
Madogawa no seki wa arimasu ka (Is there a table by the window)
Betsubetsu ni haraemasu ka (Can we pay separately)
Hachi ji han kara futari no teeburu wo yoyaku shitai no desu ga (I would like to reserve a table for two people for half past eight)

Table 1: Examples of CALL-SLT coverage in the restaurant domain, for English, French and Japanese.

plication. In particular, the grammar-based approach to recognition gives a response profile with accurate recognition on in-grammar utterances and poor or no recognition on out-of-grammar utterances, automatically giving the student feedback on the correctness of their language usage. Also, the platform's rapid development facilities, based on semi-automatic specialisation of general resource grammars, have made it easy to create good speech recognisers for our initial domain (a tourist restaurant scenario), despite the very limited availability of training data. Although the recognisers for the various L2 languages are all built from development corpora of at most a few hundred examples, native speakers typically get per-sentence semantic error rates of under 10%.

Two other differences between CALL-SLT and the MIT system are also worth highlighting. First, one of the main weaknesses of Wang's and Seneff's work is that prompts are in the student's own language (the L1). This has the undesirable effect of tying the language being studied (the L2) too closely to the L1 in the student's mind, and is quite contrary to mainstream theories of language acquisition. Instead of prompting student with sentences in the L1, our system shows them interlingua representations; these are created using semantic grammars based on our previous work on human-readable representations of interlingua [11].

Second, instead of focussing on a single language pair, we think of the problem more broadly as an activity in the multi-lingual language learning community. We structure learning activities so as to encourage students to contribute data both in the L1 and in the L2. Each student's recorded native speaker data is used as a resource to help other students studying that language.

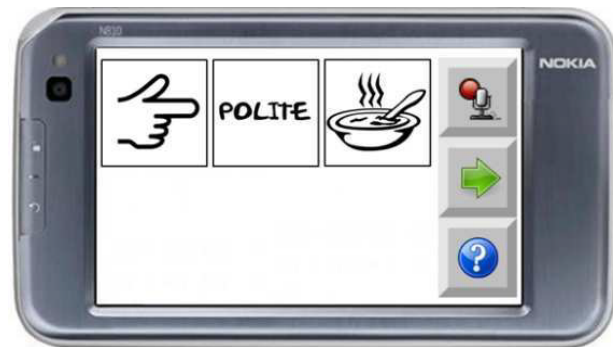


Figure 1: Mobile version of the CALL-SLT system, running on a Nokia tablet and using the graphical interlingua. The pictorial string is the graphical prompt, which here represents “Ask politely for soup”. The buttons on the right are, from top to bottom, “recognise”, “next prompt” and “help”.

In more detail, the game that forms the basis of CALL-SLT is as follows. The system is loaded with a set of possible prompts, created by translating the development corpus into the interlingua. Each turn starts with the student asking for the next prompt. The system responds by showing her a surface representation of the underlying interlingua for the sentence they are supposed to produce in the L2. This representation can either be textual or pictorial. For example, a student whose L1 is French and whose L2 is English might be given the textual prompt

COMMANDER DE_MANIERE_POLIE SOUPE

or the graphical prompt shown in Figure 1. In both cases, an appropriate response would be something like “Could I have the soup?”, “I would like some soup”, or simply “Soup, please”; the grammar supports most of the normal ways to formulate this type of request.

The student decides what she is going to say, presses the “recognise” button, and speaks. The system performs speech recognition using a Nuance 8.5 recognition package³ compiled from a grammar-based language model, translates the result into the interlingua, matches it against the underlying interlingua representation of the prompt, gives the student feedback on the match, and adjusts the level of difficulty up or down. If the match was successful, the student's recorded speech is also saved for future use.

The student may ask for help at any time. The system can give help either in speech or text form. Text help examples are taken from the original corpus, and can also be produced by translating from the interlingua back into the L1; speech help examples are created by recording successful interactions, or by doing TTS on text examples.

3. Constructing language courses in CALL-SLT

A frequent comment from early users of CALL-SLT has been that the system would be more useful if it included structured language exercises. Over the last few months, we have extended

³We are currently porting the system so that it also runs under Nuance 9.0. Despite the similarity of names, Nuance 8.5 and Nuance 9.0 are very different, deriving from separate codebases.

the platform to include a framework to support this type of functionality. The basic idea is simple: we allow the course designer to divide up the set of examples into a number of possibly overlapping subsets, each one defined by a list of syntactic and semantic criteria. We call a grouping of this kind a *lesson*. At runtime, the student sees a menu of available lessons, and is able to navigate between them using a menu. Each lesson is associated with a page of explanation, which covers the topic or topics that it introduces.

Initially, we had planned to define the content of lessons in terms of semantic properties; for example, we might have a lesson that focussed on time expressions, or on numbers. A strategy of this kind can easily be implemented by defining constraints on the interlingual representations associated with the examples. This strategy did indeed seem appropriate for languages like English and Japanese, which have relatively simple syntax. French syntax, however, is considerably more challenging, with concepts like grammatical gender, agreement and verb inflection playing a crucial role; it also turned out to be useful to have the option of including lexical constraints.

Although this complicated the architecture of the system, we decided to allow lesson content to be defined using any Boolean combination of semantic, syntactic and lexical constraints. A lexical constraint is implemented simply as a requirement that a specific sequence of words should be present in the surface string; similarly, a semantic constraint is a requirement that a given, possibly partially instantiated, structure should match some part of the interlingua representation.

The least trivial question was how to permit definition of syntactic constraints. After some experimentation, we decided that we could implement a sufficiently expressive framework by defining a syntactic constraint S to be a partially instantiated syntactic category C_S . We parse each example E using the feature grammar which forms the basis for the recogniser's language model (cf. [8]), and extract all the syntactic categories $C(E)$ from the analysis tree. We then say that E has syntactic property S if there is at least one $C(E)$ which unifies with C_S . In the French version, the framework lets us define constraints like "uses inverted word order", "includes an adjective" or "includes a verb in infinitive form"; for example, the first of these constraints is coded as a match against the syntactic category vp : $[inv=inverted]$. Similarly, in the Japanese version, we can define constraints like "includes a counter" or "uses the plain form of a verb".

At system build time, each help example is analysed to determine the set of lessons it matches, as follows. Recorded spoken help examples are first replaced by their transcripts, and the text form of the example is parsed to produce an analysis tree and a semantic representation. Syntactic categories are then extracted from each parse tree, while the semantic representation is converted into interlingual form. Finally, the set of syntactic and semantic constraints associated with each lesson is matched against the extracted information, and the result is cached.

It is important that the help system only offers help appropriate to the current lesson, since there will in general be multiple help examples for any given prompt. For example, suppose in the French version that the prompt is

ORDER POLITELY SOUP

In an early lesson, where the theme is to learn to ask for things in as simple a way as possible and practise using singular nouns, the help example might be a recording of *Je voudrais la soupe* ("I would like the soup"). In a later lesson, where the theme

```
lesson([lesson_id=singular_nouns,
        constraints=[request=yes,
                     fr_singular_noun=yes,
                     fr_adjective=no,
                     fr_voudrais_only=yes,
                     fr_infinitive=no,
                     fr_time=no,
                     food_and_drink=yes,
                     french_number=no,
                     de=no,
                     du=no],
        help_file='sing_nouns_help.txt']
).

lesson([lesson_id=times,
        constraints=[request=yes,
                     fr_time=yes,
                     fr_inf_libre=yes,
                     person=no,
                     french_loc=no],
        help_file='time_help.txt']
).
```

Figure 2: Two examples of French lesson definitions. The format has been slightly simplified for expositional reasons.

is making requests using questions, the help example might be *Puis-je avoir la soupe?* ("Could I have the soup?"). Similar considerations apply in Japanese. In an early lesson, a help example for something like

ASK-FOR POLITELY TABLE 2 PERSON

will be formed using the basic requesting construction ... *onegai shimasu*, while in a later one it might use the more formal ... *ga arimasu ka* or ... *-tai no desu ga*.

An interesting point arises when the user response is semantically correct, but fails to obey the constraints associated with the current lesson. Continuing the French example immediately above, suppose that the lesson is "simple requests using singular nouns", and the student is given the prompt above. If she replies *Puis-je avoir la soupe?*, her response is semantically correct (it will produce the correct interlingua), but it fails to match the current syntactic constraints. We considered the idea of warning the student in these circumstances, but this strategy seems in practice overly strict; our observation is that it can confuse students, particularly when it is inappropriately applied due to a recognition error. In the current version of the system, students are allowed to respond using any valid syntactic form, irrespective of whether or not it conforms to the theme of the lesson.

3.1. Initial lesson structures

We have implemented initial sets of lessons for English, French and Japanese. As already noted, the very different structures of these three languages mean that the lesson structure differs significantly from language to language. English and Japanese have relatively simple grammars, and this is reflected in an arrangement primarily organised according to semantic/pragmatic considerations. The English version of the system currently has five lessons, as follows:

1. Greetings and politeness (“hello”, “good evening”, “thank you”, etc).
2. Asking for things (“I would like” + a noun phrase)
3. Asking for things using questions (“Could I have/do you have” + a noun phrase)
4. Numbers (“I would like two beers/a table for three people/etc”)
5. Times (“I would like a table for six o’clock/half past six/quarter to seven/nineteen hundred/etc”)

The set of Japanese lessons is similar to the English one. Syntax, however, plays a larger role in French; although we still have lessons based on themes similar to those in the English and Japanese systems, several more are organized explicitly around syntactic issues:

1. Singular nouns. Simple requests involving only singular nouns, e.g. *Je voudrais le lapin* (“I would-like the rabbit”).
2. Plural nouns. Requests involving plural nouns and the future tense, e.g. *Je prendrai les fraises* (“I will-take the strawberries”).
3. Compound nominals. Noun phrases with *de* or *à*, e.g. *Je prendrai les fruits de mer* (“I will-take the seafood”).
4. Adjectives. French adjectives agree in number and gender, e.g. *Je voudrais un steak bien cuit* (“I would-like a-MASC steak-MASC well done-MASC”).
5. Requests using questions. More complex ways to phrase requests, e.g. *Auriez-vous une bière* (“Might you have a beer?”) or *Quels vins recommandez-vous?* (“What wines do you recommend?”)
6. Infinitives. Requests using the infinitive form, e.g. *Puis-je avoir une bière?* (“Can I have-INF a beer?”)

Figure 2 shows examples of two French lesson definitions, for singular nouns and times respectively. The first is intended to be done early in the course, so most non-trivial constructions are blocked; the second is intended to be done near the end, so most constructions are allowed.

4. Summary and further directions

We have presented an overview of the CALL-SLT translation game/conversation partner, focussing on recent work where we have introduced a framework that permits the instructor to divide up material into a structured set of lessons. We have only just begun to experiment with this new functionality, and it is still very much under development. It is, unfortunately, not at all easy to evaluate CALL systems in a fully objective way [12]; the question we really wish to address is whether they help students learn, and this requires an elaborate methodology where a group using the system is contrasted against a similar one that is not doing so. We are currently working on an Internet-enabled version of CALL-SLT, which we expect to have operational by Q3 2010. This should make it much easier to perform evaluation experiments.

It is likely that we will by then also have elaborated the currently very simple lesson structure framework. At a minimum, we will certainly have extended the existing set of lessons, and implemented lesson plans for some of the other CALL-SLT languages, in particular German and Arabic. We are also considering experimenting with some more ambitious ideas; a particularly interesting one is to provide the student with an interface

that allows them to define their own lessons, so that they can practise a specific topic or set of topics, while excluding others. We will report on further progress at the workshop, where we will also be able to demo the system.

5. References

- [1] F. Ehsani and E. Knodt, “Speech technology in computer-aided language learning: Strengths and limitations of a new call paradigm,” in *Language Learning and Technology*, vol. 2(1), pp. 45–60, 1998.
- [2] H. Chen, “Evaluating five speech recognition programs for ESL learners,” in *Papers from the ITMELT 2001 Conference*, 2001.
- [3] W. Johnson, “Serious use of a serious game for language learning,” *Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work*, p. 67, 2007.
- [4] M. Rayner, P. Bouillon, N. Tsourakis, J. Gerlach, M. Georgescu, Y. Nakao, and C. Baur, “A multilingual call game based on speech translation,” in *Proceedings of LREC 2010*, Valetta, Malta, 2010, http://www.issco.unige.ch/pub/lrec2010_callslt.pdf.
- [5] C. Wang and S. Seneff, “Automatic assessment of student translations for foreign language tutoring,” in *Proceedings of NAACL/HLT 2007*, Rochester, NY, 2007.
- [6] D. Goddeau, E. Brill, J. Glass, C. Pao, M. Phillips, J. Polifroni, S. Seneff, and V. Zue, “Galaxy: A human-language interface to on-line travel information,” in *Proceedings 3rd International Conference on Spoken Language Processing (ICSLP)*, Yokohama, Japan, 1994.
- [7] N. Tsourakis, M. Georgescu, P. Bouillon, and M. Rayner, “Building mobile spoken dialogue applications using Regulus,” in *Proceedings of LREC 2008*, Marrakesh, Morocco, 2008.
- [8] M. Rayner, B. Hockey, and P. Bouillon, *Putting Linguistics into Speech Recognition: The Regulus Grammar Compiler*. Chicago: CSLI Press, 2006.
- [9] P. Bouillon, M. Rayner, N. Chatzichrisafis, B. Hockey, M. Santaholma, M. Starlander, Y. Nakao, K. Kanzaki, and H. Isahara, “A generic multi-lingual open source platform for limited-domain medical speech translation,” in *Proceedings of the 10th Conference of the European Association for Machine Translation (EAMT)*, Budapest, Hungary, 2005, pp. 50–58.
- [10] P. Bouillon, G. Flores, M. Georgescu, S. Halimi, B. Hockey, H. Isahara, K. Kanzaki, Y. Nakao, M. Rayner, M. Santaholma, M. Starlander, and N. Tsourakis, “Many-to-many multilingual medical speech translation on a PDA,” in *Proceedings of The Eighth Conference of the Association for Machine Translation in the Americas*, Waikiki, Hawaii, 2008.
- [11] P. Bouillon, S. Halimi, Y. Nakao, K. Kanzaki, H. Isahara, N. Tsourakis, M. Starlander, B. Hockey, and M. Rayner, “Developing non-European translation pairs in a medium-vocabulary medical speech translation system,” in *Proceedings of LREC 2008*, Marrakesh, Morocco, 2008.
- [12] J. Nerbonne, “Natural language processing in computer-assisted language learning,” in *Handbook of Computational Linguistics*, R. Mitkov, Ed. Oxford University Press, 2003, pp. 670–698.