A Multilingual CALL Game Based on Speech Translation

Manny Rayner[1], Pierrette Bouillon[1], Nikos Tsourakis[1], Johanna Gerlach[1]
Maria Georgescul[1], Yukie Nakao[2], Claudia Baur[1]

(1) University of Geneva, TIM/ETI,
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}@unige.ch
{Johanna.Gerlach,Maria.Georgescul}@unige.ch
yukie.nakao@univ-nantes.fr, baur5@etu.unige.ch

Abstract

We describe a multilingual Open Source CALL game, CALL-SLT, which reuses speech translation technology developed using the Regulus platform to create an automatic conversation partner that allows intermediate-level language students to improve their fluency. We contrast CALL-SLT with Wang's and Seneff's "translation game" system, in particular focussing on three issues. First, we argue that the grammar-based recognition architecture offered by Regulus is more suitable for this type of application; second, that it is preferable to prompt the student in a language-neutral form, rather than in the L1; and third, that we can profitably record successful interactions by native speakers and store them to be reused as online help for students. The current system, which will be demoed at the conference, supports four L2s (English, French, Japanese and Swedish) and two L1s (English and French). We conclude by describing an evaluation exercise, where a version of CALL-SLT configured for English L2 and French L1 was used by several hundred high school students. About half of the subjects reported positive impressions of the system.

1. Introduction and background

As the world becomes smaller, an ever growing number of people find that they need to be able to speak foreign languages. It is a cliché that the best way to gain proficiency in a language is to spend time in a country where that language is spoken, and that the second best way is intensive one-on-one tuition with a native speaker. Unfortunately, neither of these options is feasible for the majority of language students. Classroom courses and self-study using books and audio resources are far cheaper, but for most people yield disappointing results. Although it is often possible to gain a reasonable working knowledge of grammar and vocabulary, it seems in general that fluency only comes from extended practical use of spoken language in interactive dialogue situations.

Over the last 20 years, spoken language technology has become vastly more accessible, and it is natural that CALL has become increasingly interested in developing applications which allow students to use the machine as a conversational partner (Nerbonne, 2003). When the desired input from the student is tightly constrained (what (Ehsani and Knodt, 1998) refer to as "closed response design"), it is already possible to build systems that can usefully address certain types of CALL tasks. A standout instance is CMU's reading tutor LISTEN (Mostow et al., 1994), which has been shown in controlled studies to improve children's reading skills significantly.

Despite successes like LISTEN, it is doubtful that closed response design systems can address the central issues in second language learning. In another overview paper, (Chen, 2001) suggests that effective systems for language learning need to prioritise allowing the learner to produce large quantities of sentences on their own. Since this is, by definition, almost impossible to realise in a closed response application, the next obvious thing to try is some kind of free-format spoken dialogue game.

As noted in (Ehsani and Knodt, 1998), much work has been invested in developing methods for building interactive spoken dialogue systems for command-and-control and knowledge-base query, and people have used this technology to construct CALL applications; one convincing example is TLCTS (Johnson, 2007). Systems of this kind are, however, expensive to build. The key problem is that dialogue games require detailed modelling of language meaning at a level deep enough to support valid system responses, a challenging task even in relatively trivial domains. For example, in a flight booking domain, formulating a correct response to a request like "Find me flights from Geneva to London" might involve translating abstract representations of queries into a form suitable for querying Travelocity; asking clarification questions to determine the required date, time of departure and airline; determining that there are several airports in London, and asking if one of them is preferred; and so on.

Since dialogue games as such are difficult to realise, it is natural to seek other kinds of language games which make it possible to practise fluency, but which are simpler to implement. A promising idea in this area is "translation games" (Wang and Seneff, 2007). In general, translation requires a shallower understanding of language than dialogue in order to reach a similar level of performance; for example translating the sentence "Find me flights from Geneva to London" does not involve knowing any details of actual flights, but only the relevant grammar and vocabulary. Wang and Seneff successfully reused speech and language technology developed at MIT under other projects (Goddeau et al., 1994) to build a speech-enabled game for students who wished to practice Chinese. In their game, the

system shows English sentences to the student, who has to respond with a spoken Chinese translation. Most of the subjects who participated in the initial study were positive about the system.
Our system, CALL-SLT, is inspired by Wang and Seneff's work, and further extends their ideas. Section 2. gives an overview of CALL-SLT, and Section 3. describes the processing in more detail. Section 4. described a recent exercise, where a version of CALL-SLT was used by several hundred Swiss high-school students. The final section concludes.

2. The CALL-SLT system

CALL-SLT is an Open Source speech-based CALL application for 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 (Tsourakis et al., 2008), with performance identical to that of the laptop version. The current version uses a restaurant domain, and supports English, French, Japanese and Swedish as L2s, with English or French as the L1. 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 Regulus, a platform for building systems based on grammar-based speech understanding (Rayner et al., 2006b) and MedSLT, an interlingua-based speech translation framework (Bouillon et al., 2005; Bouillon et al., 2008a), to develop a generic CALL platform centered on the "spoken translation game" idea. Our experience so far suggests that the Regulus/MedSLT architecture is a good fit to this type of application. 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.
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 sentences in the L1, our system prompts students using interlingua representations; these are created using semantic grammars based on our previous work on human-readable representations of interlingua (Bouillon et al., 2008b).
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

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 rapé (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)
Terasuseki wa arimasu ka (Is there a table outside)
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)
Swedish
Kunde jag få ett glas (Could I have a glass)
Har ni en vinlista (Do you have a wine list)
Vad är dagens (What is the dish of the day)
Jag skulle vilja betala med euro (I would like to pay in Euros)

Table 1: Examples of CALL-SLT coverage in the restaurant domain, in the four system languages.

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. We will elaborate on this in Section 3.4. below.
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 them 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

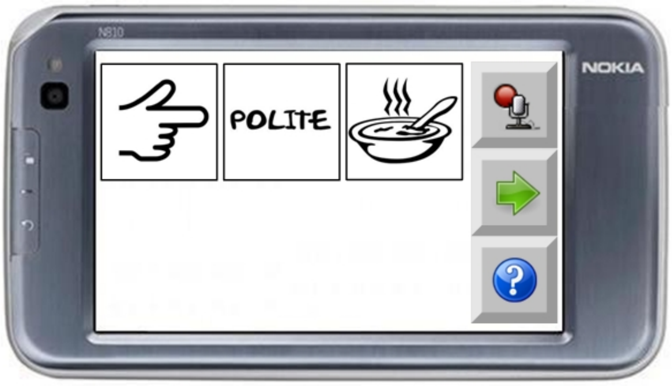

Figure 1: Mobile version of the CALL-SLT system, running on a Nokia tablet and using the graphical interlingua. The pictorial string is the graphical prompt, which here represents "Ask politely for soup". The buttons on the right are, from top to bottom, "recognise", "next prompt" and "help".

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 "recognize" button, and speaks. The system performs speech recognition using a Nuance 8.5 recognition package 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.

In the following section, we describe in more detail the processing involved in the above.

3. Underlying processing

3.1. Grammars and recognisers

As already mentioned, the grammar-based Regulus platform appears very suitable for this type of application. We have been able to build good recognisers for our four L2s (English, French, Japanese and Swedish), using small corpora that contain between 150 and 400 utterances. In each case, we use the normal Regulus method. We start with a general resource grammar, written in a unification-based feature-grammar notation. For English, the grammar is the one described in Chapter 9 of (Rayner et al., 2006b). For French, we use the shared French/Spanish/Catalan grammar of (Bouillon et al., 2007b). An early version of the Japanese resource grammar is sketched in (Rayner et al., 2005); we handle Swedish using a version of the English grammar generalized to cover both English and Swedish. The resource grammar for each language is combined with a domain-specific content-word lexicon. For example, the English content-word lexicon for our initial restaurant domain includes verbs like "reserve" and "order", nouns like "table", "fork" and "hamburger", and adjectives like "large" and "non-smoking". Function words are taken from the existing resource function-word lexica.

For each language, the relevant grammar, together with the new content-word lexicon, is used to parse the domain corpus, giving rise to a set of analysis trees. These trees are then processed by the Explanation Based Learning algorithm, as explained in Chapter 10 of (Rayner et al., 2006b), to produce a specialised feature-grammar tuned to the domain. Next, the specialised grammar is compiled into a CFG grammar in the proprietary Nuance GSL format[1], using the methods of Chapter 9 of (Rayner et al., 2006b). Finally, the Nuance grammar is compiled, using proprietary Nuance utilities, into a recognition package which can be run on the Nuance platform. As part of this process, the domain corpus is used a second time, to perform probabilistic tuning of the recognition grammar. This tuning has a large effect on performance (Rayner et al., 2006a).

The recognition packages produced by this process take speech as input, and yield output in the form of N-best hypothesis lists. Each recognition hypothesis consists of both a word string and a semantic representation. Semantic representations are encoded in Almost Flat Functional (AFF) notion (Rayner et al., 2008). For example, the AFF representation of "Can we have a non-smoking table?" produced by the current grammar is the unordered list

```
[null=[utterance_type,ynq],
 null=[modal,can],
 agent=[pronoun,we],
 null=[action,have],
 null=[voice,active]
 object=[property,non_smoking],
 object=[place_to_sit,table]]
```

where each element consists of a semantic primitive associated with a (possibly null) functional tag. As argued

[1]The initial version of CALL-SLT uses the Nuance 8.5 recogniser. We are currently porting it to allow use of the Nuance 9 platform as well; this involves producing recognition grammars in GrXML format.

in (Rayner et al., 2008), this gives an intelligent compromise between the flat feature/value lists favoured in commercial spoken dialogue systems, and the nested predicate/argument structure typical of linguistically motivated frameworks like LFG and HPSG. Processing is on lists, which are generally much easier to manipulate than nested structures. The functional tags, in practice, appear to retain enough grammatical information to avoid ambiguity. Thus, here, the *object* tag encodes the fact that the adjective *non_smoking* attaches to the noun *table*.

3.2. Interlingua

The L2 semantic form produced by recognition is translated into a corresponding interlingua form, also represented in AFF. Translation on AFF forms is performed using the interpreter described in (Rayner et al., 2008), which applies two types of rules. These, respectively, perform conditional rewriting of lists of tagged elements to lists of tagged elements, and conditional rewriting of tags to tags.

Continuing the example, the interlingua form corresponding to "Can we have a non-smoking table?" is

```
[null=[utterance_type,request],
 null=[politeness,polite],
 arg2=[seating,table],
 arg2=[table_type,non_smoking]]
```

As can be seen, the interlingua form is considerably simpler than the original L2 form, mirroring the fact that the interlingua is designed to abstract away from surface syntax. The space of well-formed interlingua representations is defined by another Regulus grammar; a representation is well-formed if and only if it is possible to generate an analysis tree from it (Bouillon et al., 2008b).

In addition to checking well-formedness, the interlingua grammar also performs a second function, namely associating each interlingua representation with a surface string; this follows naturally from the fact that checking well-formedness is carried out by performing generation from the interlingua representation. The interlingua grammar is parametrized so that it can be compiled in different versions, corresponding to different lexical choices for the surface string. In the current system, we have three different versions, for English-based, French-based and graphical realisations of the interlingua. Returning to the running example, the English version of the generated string is

```
ASK-FOR POLITELY
TABLE IN-NON-SMOKING-AREA
```

The French one is

```
DEMANDER DE_MANIERE_POLIE
TABLE NON-FUMEUR
```

Finally, the graphical version is a list of four JPEG files, which respectively illustrate the concepts "ask for", "politely", "table" and "smoking". Table 2 shows English- and French-based textual interlingua versions of the English examples from Table 1.

The interlingua grammar also performs a third function, acting as an additional component of the language model; by rescoring the N-best hypothesis list to discard alternatives which produce ill-formed interlingua, we can further tune the system in the direction of preferring well-formed grammatical input.

E	I would like a mint tea
EI	ASK-FOR POLITELY MINT-TEA
FI	DEMANDER DE_MANIERE_POLIE THÉ_À_LA_MENTHE
E	A tea and a coffee please
EI	ORDER POLITELY TEA AND COFFEE
FI	COMMANDER DE_MANIERE_POLIE THÉ ET CAFÉ
E	Do you have a table for four people
EI	ASK-FOR POLITELY TABLE 4 PERSON
FI	DEMANDER DE_MANIERE_POLIE TABLE 4 PERSONNE(S)
E	Could I reserve a table for seven thirty
EI	BOOK POLITELY TABLE 19 : 30
FI	RÉSERVER DE_MANIERE_POLIE TABLE 19 H 30
E	Do you accept credit cards
EI	ASK-FOR POLITELY PAY CREDIT-CARD
FI	DEMANDER DE_MANIERE_POLIE PAYER AVEC_CARTE_DE_CRÉDIT

Table 2: Examples of English- and French-based textual interlingua. 'E' = original English; 'EI' = English-based interlingua; 'FI' = French-based interlingua.

3.3. The top-level

Figure 2 presents a screen-shot of the laptop version of the system, showing the state after a successful interaction in the Japanese edition. The student was first given the textual interlingua prompt (top pane)

```
ASK-FOR POLITELY TABLE 1 PERSON
```

They pressed the "recognise" icon (top right), and spoke in Japanese. The smiling sun on the left indicates a correct match.

The recognition result is shown in the second pane down, and is presented in three aligned versions: native script (kana/kanji); Roman script (romaji); and English-language glosses. These multiple versions of the recognition result are produced as follows. For Japanese, Nuance requires that the recognition grammar be encoded using romaji. In order to produce a kana/kanji recognition result as well, we write the grammar and lexicon using macros which expand lexical items out to either romaji or kana/kanji, and compile two aligned parallel versions of the grammar, one for each script. We then reconstruct the kana/kanji of the recognition string by parsing it using the romaji grammar, substituting the corresponding kana/kanji elements from the parallel grammar at the leaves of the parse tree, and then reading off the fringe.

Glosses are produced in the same way. Note in passing that the "gloss" is quite different from the "textual interlingua".

The "textual interlingua" is a human-readable textual representation of the interlingua. The "gloss", in contrast, is a version of the L1, with the original words replaced one-to-one by English equivalents.
The bottom pane shows help results, which are obtained by clicking on the "help" icon (the life-belt: third down on the right). The menu under "Aide" gives three Japanese sentences, all of which have meanings equivalent to "I would like a table for one person". By clicking on an item, the student can hear the corresponding recorded version. The next section describes how these help resources are created.

3.4. Help

Help examples, as already indicated, are in both text and speech forms, which if possible are linked together. Text help for a prompt is generated in two ways. First, each prompt is associated with the set of L1 corpus sentences which are mapped to it by the L1 → Interlingua translation rules. Second, at build time, the system also applies a set of Interlingua → L1 translation rules, to produce a canonical L1 text realisation of each interlingual form. A check is performed to ensure that each canonical text realisation can also be parsed, and translated back into the Interlingua form it came from.
Speech help is also created in two ways. The first method is to generate it by applying TTS to examples of text help. The second is to use recorded human speech; as already noted, the recorded speech file resulting from every successful match is stored automatically. The recorded file is associated with several pieces of information: the prompt used to produce it, the speaker ID, and a text version, which initially is the recognition result. Since the examples are only saved for successful matches, the recognition result is guaranteed to exist.
It is thus possible to create speech help from text help, using TTS, and text help from speech help, using recognition. Both approaches have advantages and disadvantages. The TTS route is easy to implement, but teaching students to imitate artificially produced speech has obvious drawbacks; TTS pronunciation and prosody are never perfect, and are sometimes quite unnatural. In the other direction, recognition results, even for correctly matched examples, are not themselves always correct, since the recognition grammar has some slack in it. For example, the student might have said "I would like **a** pizza", while the system recognised "I would like pizza". If the original recognition results are not cleaned up, the result can again be confusing for the student. A system tool allows developers to edit the transcriptions quickly, using the normally reliable assumption that a transcription which is already the same as a stored text example is correct.
We have experimented with various help strategies. There does indeed seem to be a considerable advantage in using recorded examples. Since these have all resulted in successful matches, the student can feel confident that sufficiently precise imitation of the speaker will mean that they will themselves get a successful match. This gives them a simple way to improve their fluency; they listen to a recorded example, try to imitate it, and see if the system accepts their pronunciation. If it doesn't, they keep on trying until they succeed.
It also appears important to connect speech and text help. While they are trying to acquire the necessary fluency, students want to be able to see a textual representation of a correct example, so that they can if necessary read it aloud. The system consequently prefers to show help examples which can be presented in both forms.

Figure 3: Two students using CALL-SLT during the Geneva University Open Week, November 2009.

4. Using CALL-SLT in practice

The L1-French/L2-English version of the system was tested by about six hundred Swiss high school students at the end of November 2009, as part of the Geneva University Student Week. Most of the subjects were aged 15 to 17, and had a wide range of levels of ability in English, ranging from "beginner" to "fluent, speak English regularly at home". Each student had a half-hour slot. They were first given a five-minute PowerPoint presentation, in which an instructor showed them how to operate the system, after which they were encouraged to experiment with it freely. At the end of the session, they completed a brief questionnaire, which we based on the one reported in (Wang and Seneff, 2007). The results are summarised in Table 3.
Our main conclusions are the following. Encouragingly, a good proportion of the subjects enjoyed playing with the system, and felt that it was teaching them something useful (Questions 3, 4, 5 and 7). This confirmed our intuitive assessments after watching them; it was clear that many people were having a good time (Figure 3), and we could often hear them making favourable comments to the new students coming in. It also agreed with our own experiences of using versions of the game in languages we didn't know well. It was reassuring to see that the prompts were easy to understand (Question 10), and that the help system did what it was supposed to (Question 11). As we had expected, the system appeared to be most popular with beginner/intermediate students; many of the advanced students found it rather too easy.
The only clear-cut negative aspect was in the recognition, which appeared to be too challenging (Question 13). Our

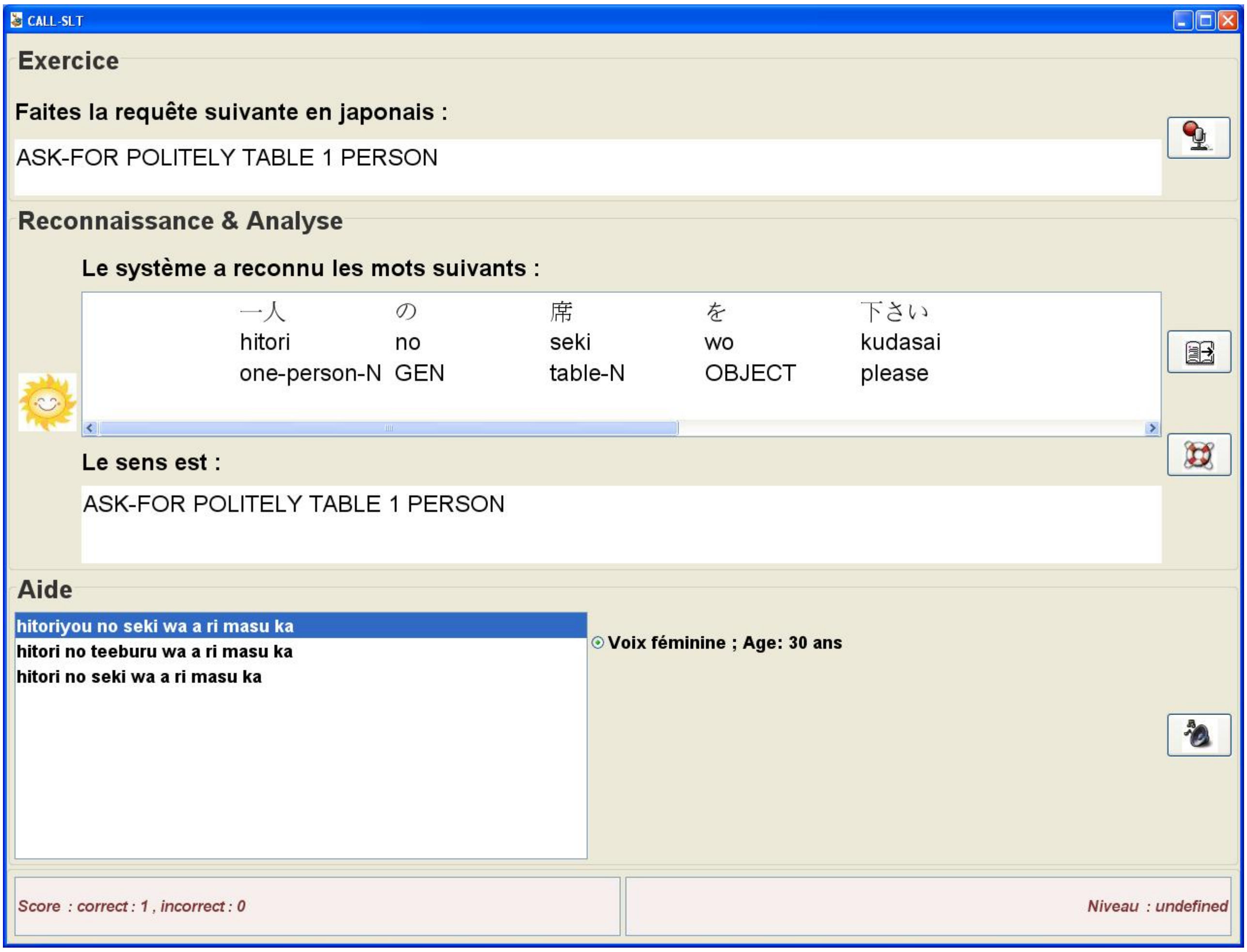

Figure 2: Screenshot of the Japanese version of CALL-SLT, running on a laptop. The prompt is being shown using the English-flavour textual version of the interlingua.

impression was that there were two main reasons for this. First, there were many students who were unable to get anything to recognise for the first few turns. This is common for people who are using a speech-enabled system for the first time; it is often due to difficulties with the push-to-talk button, initial uncertainty about finding an appropriate speaking rate and volume, and similar issues. In every case we examined, the end of the session showed much better recognition than the beginning.

With hindsight, the second reason for the poor recognition was the bad strategy used for choosing the vocabulary and the set of examples. A couple of developers complained that we were missing words for items of food and drink that they considered important. We responded to this by adding several hundred such items, but the result was that the set of examples was dominated by requests like “Ask politely for a chicken tikka masala” or “Ask politely for a J2O”. Many of these items were types of food and drink common in England, but which were unknown to most of the Swiss students. They had trouble pronouncing them properly, and the unnecessarily large vocabulary introduced unexpected phonetic confusions. We have since then reorganized the vocabulary and the set of examples accordingly, to focus more on grammatical issues.

5. Summary and future directions

We have presented an overview of our experiences to date with CALL-SLT, an interactive speech-enabled CALL game for beginner/intermediate level students. The system is based on Wang’s and Seneff’s “translation game” idea, and implemented for four languages using the Regulus toolkit. We reported on an initial informal test, where CALL-SLT was used by about six hundred Swiss high-school students. About half of the group clearly liked it, and felt that it could be tangibly useful to them.

The CALL-SLT project, which began in August 2009, is still quite new. We feel that it has got off to a good start, and are generally optimistic about its prospects. During the next few months, we plan to concentrate on the following topics:

- Incorporation into a structured CALL environment. Our initial plan is to break up the set of examples into a number of (possibly overlapping) groups, each one in effect a “lesson” focussing on a specific topic like “asking for things”, “masculine and feminine” or “telling the time”. Lessons will include an extended help mechanism, so that students can at a minimum access written help on the lesson’s topic, or, if stuck

		++	+	=	–	– –
Beginner (15% of total)						
1	The game was too difficult	2%	9%	45%	36%	8%
2	The game was too easy	3%	6%	38%	50%	3%
3	I enjoyed learning English with the game	23%	37%	26%	10%	4%
4	The game helped me improve my English	1%	33%	42%	16%	8%
5	I would use this game again if I could	11%	27%	44%	14%	4%
6	I would use a similar game which covered another topic	11%	27%	44%	9%	9%
7	I would recommend the game to my friends	14%	35%	36%	10%	5%
8	I would prefer a game which focussed only on vocabulary	1%	10%	32%	50%	7%
9	I would prefer a game which was more interactive	3%	32%	41%	23%	1%
10	It was easy to understand what I was meant to say	13%	33%	36%	12%	6%
11	The help system helped me find what I was meant to say	50%	27%	17%	5%	1%
12	The help system helped me find what I had mispronounced	20%	23%	31%	18%	8%
13	It was easy to get the machine to understand me	1%	18%	26%	29%	26%
Intermediate (70% of total)						
1	The game was too difficult	2%	8%	24%	52%	15%
2	The game was too easy	4%	18%	34%	41%	3%
3	I enjoyed learning English with the game	20%	42%	26%	9%	3%
4	The game helped me improve my English	6%	34%	33%	18%	9%
5	I would use this game again if I could	7%	32%	43%	14%	4%
6	I would use a similar game which covered another topic	9%	31%	52%	6%	2%
7	I would recommend the game to my friends	12%	35%	33%	17%	3%
8	I would prefer a game which focussed only on vocabulary	5%	13%	35%	42%	5%
9	I would prefer a game which was more interactive	7%	31%	41%	20%	1%
10	It was easy to understand what I was meant to say	23%	46%	21%	9%	1%
11	The help system helped me find what I was meant to say	44%	30%	22%	4%	0%
12	The help system helped me find what I had mispronounced	14%	30%	26%	25%	5%
13	It was easy to get the machine to understand me	3%	16%	25%	33%	23%
Advanced (15% of total)						
1	The game was too difficult	1%	0%	18%	52%	29%
2	The game was too easy	8%	32%	34%	26%	0%
3	I enjoyed learning English with the game	12%	30%	27%	18%	12%
4	The game helped me improve my English	0%	22%	38%	23%	17%
5	I would use this game again if I could	4%	17%	49%	13%	17%
6	I would use a similar game which covered another topic	8%	27%	44%	7%	14%
7	I would recommend the game to my friends	11%	22%	39%	14%	14%
8	I would prefer a game which focussed only on vocabulary	4%	12%	39%	35%	10%
9	I would prefer a game which was more interactive	9%	31%	43%	16%	1%
10	It was easy to understand what I was meant to say	32%	31%	25%	9%	3%
11	The help system helped me find what I was meant to say	24%	34%	30%	4%	8%
12	The help system helped me find what I had mispronounced	18%	39%	18%	21%	4%
13	It was easy to get the machine to understand me	3%	12%	23%	40%	22%

Table 3: Responses to questionnaire filled out by 600 French-speaking high-school students who used the L1-French/L2-English version of the system. We separate the students into three groups, according to their own assessment of their level of competence in English. The five rightmost columns show the proportion strongly agreeing (++), agreeing (+), neutral (=), disagreeing (–) and strongly disagreeing (– –).

on an example, ask for help only about vocabulary or only about grammar. The student will be able to navigate between lessons using a menu-based structure.

- Deployment on the Web. CALL-SLT will obviously be far more useful if it can be deployed on the Web, so that potential users can access it through their browsers; this is very much in line with popular language-learning sites like `livemocha.com` and `tellmemore.com`, and could in principle be integrated with them. We expect to have a Web version of the system operational before the end of 2010.

- Evaluation. Although we are encouraged by the positive attitude shown by the subjects, it is impossible to draw any hard conclusions from our initial experiment with the Geneva students. A proper evaluation would need to track student progress over a minimum of a couple of weeks of using the system, comparing them against a control group who are not using it. We hope to organise a study of this kind once the Web-based version of the system is available.

- More languages. We have a simple German version of the system operational already, and will soon begin work on a version for Arabic, using the Arabic resource grammar from (Bouillon et al., 2007a). We may also build a version for Mandarin Chinese.

Acknowledgements

This work was carried out under a grant from the Swiss National Science Foundation. We are grateful to Nuance for their kindness in continuing to grant us free access to their recognition software for research purposes.

6. References

P. Bouillon, M. Rayner, N. Chatzichrisafis, B.A. Hockey, M. Santaholma, M. Starlander, Y. Nakao, K. Kanzaki, and H. Isahara. 2005. 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.

P. Bouillon, S. Halimi, M. Rayner, and B.A. Hockey. 2007a. Adapting a medical speech to speech translation system (MedSLT) to Arabic. In *Proceedings of the 2007 Workshop on Computational Approaches to Semitic Languages: Common Issues and Resources*, Prague, Czech Republic, June. Association for Computational Linguistics.

P. Bouillon, M. Rayner, B. Novellas, M. Starlander, M. Santaholma, Y. Nakao, and N. Chatzichrisafis. 2007b. Une grammaire partagée multi-tâche pour le traitement de la parole: application aux langues romanes. *Traitement Automatique des Langues*, 47(3).

P. Bouillon, G. Flores, M. Georgescul, S. Halimi, B.A. Hockey, H. Isahara, K. Kanzaki, Y. Nakao, M. Rayner, M. Santaholma, M. Starlander, and N. Tsourakis. 2008a. 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.

P. Bouillon, S. Halimi, Y. Nakao, K. Kanzaki, H. Isahara, N. Tsourakis, M. Starlander, B.A. Hockey, and M. Rayner. 2008b. Developing non-European translation pairs in a medium-vocabulary medical speech translation system. In *Proceedings of LREC 2008*, Marrakesh, Morocco.

H. Chen. 2001. Evaluating five speech recognition programs for ESL learners. In *Papers from the ITMELT 2001 Conference*.

F. Ehsani and E. Knodt. 1998. Speech technology in computer-aided language learning: Strengths and limitations of a new call paradigm. *Language Learning and Technology*, 2(1):45–60.

D. Goddeau, E. Brill, J.R. Glass, C. Pao, M. Phillips, J. Polifroni, S. Seneff, and V.W. Zue. 1994. Galaxy: A human-language interface to on-line travel information. In *Proceedings 3rd International Conference on Spoken Language Processing (ICSLP)*, Yokohama, Japan.

W.L. Johnson. 2007. Serious use of a serious game for language learning. *Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work*, page 67.

J. Mostow, S. Roth, A.G. Hauptmann, and M. Kane. 1994. A prototype reading coach that listens. In *Proceedings of the 12th National Conference on Artificial Intelligence*, pages 785–792.

J. Nerbonne. 2003. Natural language processing in computer-assisted language learning. In R. Mitkov, editor, *Handbook of Computational Linguistics*, pages 670–698. Oxford University Press.

M. Rayner, N. Chatzichrisafis, P. Bouillon, Y. Nakao, H. Isahara, K. Kanzaki, and B.A. Hockey. 2005. Japanese speech understanding using grammar specialization. In *HLT-NAACL 2005: Demo Session*, Vancouver, British Columbia, Canada. Association for Computational Linguistics.

M. Rayner, P. Bouillon, B.A. Hockey, and N. Chatzichrisafis. 2006a. Regulus: A generic multilingual open source platform for grammar-based speech applications. In *Proceedings of LREC 2006*, Genoa, Italy.

M. Rayner, B.A. Hockey, and P. Bouillon. 2006b. *Putting Linguistics into Speech Recognition: The Regulus Grammar Compiler*. CSLI Press, Chicago.

M. Rayner, P. Bouillon, B.A. Hockey, and Y. Nakao. 2008. Almost flat functional semantics for speech translation. In *Proceedings of COLING-2008*, Manchester, England.

N. Tsourakis, M. Georgescul, P. Bouillon, and M. Rayner. 2008. Building mobile spoken dialogue applications using Regulus. In *Proceedings of LREC 2008*, Marrakesh, Morocco.

C. Wang and S. Seneff. 2007. Automatic assessment of student translations for foreign language tutoring. In *Proceedings of NAACL/HLT 2007*, Rochester, NY.