Speech Recognition For Online Language Learning: Connecting CALL-SLT and DALIA

*Bouillon Pierrette, Cervini Cristiana, Mandich Anna, *Rayner Manny, *Tsourakis Nikos

*ETI/TIM, Université de Genève, Switzlerland
{Pierrette.Bouillon, Emmanuel.Rayner, Nikolaos.Tsouakis}@unige.ch
CILTA/FACOLTA' di LINGUE, Université de Bologne, Italy
{Cristiana.Cervini, Mandich.Anna}@unibo.it

Abstract

In this paper, we show how speech recognition can contribute to the development of a multimedia course for foreign language self-learning that focuses both on receptive and productive skills. Specifically, we describe how we merged two different existing e-leaning projects, DALIA [5] and CALL-SLT ([1], [8]). DALIA is a project for self-learning and blended learning which targets several European languages and proficiency levels. CALL-SLT is a web-based spoken translation game configured to teach productive competences in specific domains. This paper describes the methodology used to merge these two approaches.

1. Introduction

In this paper, we show how speech recognition can contribute to the development of a multimedia course for foreign language self-learning that focuses both on receptive and productive skills. Specifically, we describe how we merged two different existing e-leaning projects, DALIA [5,2] and CALL-SLT [1,8]. DALIA is a project for self-learning and blended learning which targets several European languages and proficiency levels. Using the Moodle platform, it offers a large set of exercises based on lexicon, grammar, and oral/written comprehension, intended to teach receptive skills. CALL-SLT is a web-based spoken translation game configured to teach productive competences in specific domains. The student is presented with a set of meanings in the form of L1 glosses (for example, SAY: AGE OF_YOUR_MOTHER 40) which they are asked to verbalise in the L2. The system uses two main components: a grammar-based speech recognizer that recognizes what the student said and an interlingua-based machine translation system that checks if the meaning of the recognized sentence corresponds to the meaning of the L1 gloss. Meanings can easily be organized into lessons that focus on different linguistic aspects [9].

This paper describes the methodology used to merge these two approaches. It explains how we developed CALL-SLT lessons for Italian students learning French, based on the already validated communicative situations and exercises offered in DALIA, and how we integrated these lessons on the Moodle platform. The emphasis will be on readability and learning value of the linguistic input (glosses). In the rest of the paper, we first present the relevant background, in particular the DALIA project. We then focus on the CALL-SLT translation game. The last section presents our integration of the two projects.

2. From ALTAIR to DALIA: The challenge of foreign language e-learning at the University of Bologna

2.1. Political background

The Reform of the Italian University system in 1999 (*Ministerial Decree no. 509/99*) introduced some important innovations concerning foreign language competences. Since that time, the knowledge of at least one European foreign language and the achievement of specific credits has become compulsory for non-specialist students who wish to obtain an academic degree in Italy. This implied a strong effort to meet the sudden huge increase in foreign language learning demand. In this process of reorganization, ICTs and e-learning have played a central role. Just two years after the Reform, the «Common European Framework of Reference for Languages: Learning, Teaching, Assessment» (CEFR) [4] was published, and quickly became the key document for language policies as well as a standard for educational and professional mobility of European citizens. The concept of the «threshold level» became popular as a way of defining the level of autonomy of people in non-first-language environments. It is in this socio-cultural and economic context that the University of Bologna developed, in 2000, a project called ALTAIR («Apprendimento Linguistico Tecnologicamente Avanzato in Rete»), followed in 2006 and 2009 by two other projects: AIR and, most recently DALIA («Didattica per l'apprendimento delle lingue in autonomia»).

Beyond the primary receptive and productive skills - reading, listening, writing and talking - the CEFR quotes *mediation* and *translation* as the two central integrative competences for the learner who is considered as a "social agent", i.e. "members of society who have tasks (not exclusively language-related) to accomplish in a given set of circumstances, in a specific environment and within a particular field of action" [4]. However, although many instruments proven to be useful to enhance receptive skills are now widely available through e-learning, oral training and oral self-evaluation have been much more challenging to develop in this context. Without oral training, linguistic knowledge (words, grammar rules, structures…) may well be stored in the mind during self-learning study, but cannot then be realized as productive speaking skills.

Every year, the University of Bologna evaluates between 12.000 and 18.000 students in receptive skills (lexis, grammar, reading and listening) using self-corrective computer-based tests. The political will to integrate productive abilities into the current examinations within the near-term future introduces important challenges concerning oral training and testing. The development of the integrative CALL paradigm within DALIA addresses this need, which is both educational and political in nature.

2.2. DALIA: towards the Integrative CALL paradigm

We share the view that the use of educational technologies in foreign language teaching and learning processes spreads a series of benefits, i.e. *diversification in the ways of interacting, helping students with logistic obstacles, enriching face to face traditional lessons, promotion of individualized learning paths etc.*, and the tracks we have chosen to enrich DALIA concern the "integrative" model of e-learning as a more recent shift from the previous behaviouristic and communicative CALL paradigms [11]. Moodle is the Open Source Virtual Learning Environment that we have chosen for online learning. It promotes social-constructivism as a pedagogical approach to language learning, contemplating a focus on interaction thanks to specific embedded tools such as chats, forums and web-meetings, where proactive expert tutoring or peer-to peer tutoring can promote online exchanges with other users though both written and oral modalities.

The integration of DALIA with new exercises like the translation game aims at promoting productive and interactive competences, while limiting the role of humans. It is not only meant to give a boost to the involvement of new cognitive processes [6]; we also think that we can gain some of the benefits that normally characterize interpersonal and intrapersonal communication, namely obtaining enhanced or modified input and directing attention to the linguistic form ([3], p. 55). This will be the focus of the next sections, where we describe the game and its integration into DALIA in more detail.

3. The CALL-SLT system
3.1. The translation game

CALL-SLT is an Open Source speech-based translation game designed for improving productive skills in different situations (visiting a restaurant, introducing yourself, etc.). The idea behind an automatic spoken translation game comes from [10]: the learner receives as input a set of L1 sentences that have to be verbalized in the L2 language. Speech recognition is used to recognize what the student has said, and a machine translation system checks if the translation is correct. This translation exercise can be seen as a preparation phase for later dialogue interaction. With this type of game, the student does not just listen to a given sentence and repeat it, as in many games; he has to produce a translation that can be accepted by a speech recognizer.

In CALL-SLT, we have improved the game in two main ways compared to Wang and Seneff's work. First, the system does not simply show the learner an L1 sentence, but rather an L1 gloss of the linguistic act they have to perform in L2; for example «BOOK 1 TABLE FOR 19:00», «SAY YOUR AGE: 20», etc. The main advantage of the gloss is that it should avoid the undesirable effect of linking the L2 language too closely to the L1 in the student's mind. The focus is thus more on L2 language production rather than translation [1,8]. It is important to add that the glosses are not a frozen language. They are flexible enough to be adapted to teacher's needs as we will see in the next section.

The second innovation is that CALL-SLT includes a powerful mechanism to build lesson plans [7]. This mechanism makes it possible to structure semi-automatically the set of sentences included in the game into fine-grained lessons that pick out subsets of sentences based on predefined lexical, syntactic or semantic properties. Depending on the pedagogical aim, the same sentences can be grouped in different ways to produce different lessons.

3.2. The architecture

The CALLSLT system is based on two main components: a grammar-based speech recognizer and an interlingua-based machine translation (MT) system, both developed using the Regulus platform [9]. In order to check whether the sentence pronounced by the learner is correct or not, the system first

performs speech recognition. The MT system then determines if the recognized sentence corresponds to the meaning of the prompt presented to the learner. To do this, it transforms the sentence into the meaning (interlingua) representation and matches it against the representation of the prompt.

This architecture presents several advantages for a CALL application. On one side, the grammar-based recognition approach is well suited to the web-based CALL task; in particular, it gives good coverage on in-coverage sentences even without speaker adaptation or training data. It is also very rare for recognition to produce undesirable ungrammatical sentences, which could give misleading feedback to students. The interlingua-based MT, on the other side, allows us to produce a language-independent meaning for the sentence which can easily be glossed for different L1s. In addition, the core system is not related to a particular language or domain, as in [10]. The REGULUS platform offers many tools to support addition of new languages and new coverage (vocabulary, grammar) for existing languages: the recognizer used by the students is extracted by specialisation from a general resource grammar in order to get an effective grammar for a specific domain. The specialisation process is driven by a small corpus of sentences, constructed so as to contain at least one example of each required word and grammatical structure. The specialised recogniser is extended to cover new exercises primarily by changing the corpus and lexicon; small changes to the general grammar may sometimes also be required. It is also possible to extract different recognisers, more or less general, that correspond to different levels of difficulty in the translation game - the more general the recognizer, the less "forgiving" recognition becomes. In summary, the approach appears to be appropriate to a limited-domain multilingual system that can be accessed by a wide variety of casual internet users.

3.3. The interface and the interaction with students

The game is accessible via a web-based interface (http://callslt.org). It offers seven main functionalities, explained later in this section and illustrated in Fig, (1): 1) choosing a language pair and specific situations/lessons; 2) getting an exercise in the L1 language, in the form of a gloss; 3) asking for recognition; 4) getting written and spoken help for a given exercise, 5) getting the recognition result, 6) getting lesson help and 7) selecting the level of the game (hard, medium, easy). The interaction with the student is as follows: the student is first provided with a set of lessons, which contain different exercises in the form of a set of glosses. When he receives a gloss, he can either speak immediately, or first ask for spoken or written help that provides different examples of possible responses for the current exercise. These examples come for the initial corpus used to derive the glosses presented to the students, or can be generated automatically by the MT system. In this way, it is possible to provide coherent help for the different glosses, without having to enumerate all the possibilities (for example in French, the system will generate the different ways of asking a question). After speaking, the student can see the result of the recognition. A green bar indicates that the result was correct; orange means it was correct but not in the lesson coverage (for example, the student might use a question in a lesson about the conditional); red shows an incorrect answer.

Figure 1: Call-slt interface

4. CALL-SLT lessons for DALIA

The DALIA Moodle e-learning platform hosts multimedia learning resources for the European languages Italian, English, French, Spanish, German from level A2 to B2 of the CEFR, and also Chinese for beginners. The oral lessons that we have produced so far within the CALL-SLT system are aimed at Italian students of French who are at beginner or low-elementary level and studying French in blended mode learning or total self-learning. The French multimedia course level A2 is structured into 4 modules (1. «Bonjour, vous êtes italiennes», 2. «Qu'est-ce que tu préfères» ? 3. «C'est vraiment délicieux !» and 4. «Mon chez moi»). Each module is composed of 4 units («Grammaire», «Lexique», «Savoir-Faire linguistiques», «Phonétique») and offers a rich variety of self-corrective receptive exercises such as multiple-choice, fill-in, dictation, sentence reordering, listening comprehension, reading comprehension, drag and drop, category, and listen and repeat.

During the elaboration of the oral production exercises for the CALL-SLT system, we have made efforts to merge the principles of the interactionist perspective and task-based language learning with the rules of instructional design; in particular: provide rich input, encourage inductive chunk learning, provide feedback and respect learner syllabi/developmental processes. For example, the DALIA module 1 "Bonjour, vous êtes italiennes" is now linked to 8 CALL-SLT lesson units, illustrated in Table 1. Each unit focuses on a particular speech act («saluer», «dire son âge», etc.) and illustrates different ways of speaking, which correspond to different ways of conveying the same meaning provided in the student help (see Fig. 1). The 8 units do not constitute a real dialog; they decompose it in different subtasks, for example answering the question «what is your age» and giving the answer «10», using a specific way of speaking. In this way, the student can learn different grammatical structures for the same speech act and practice the same structure many times with multiple examples.

Units	Glosses: examples	Student's help/ accepted answers	Number of items per unit
Saluer	TRADUCI : A_PRESTO	À bientôt	14
Saluer et dire son nom	SALUTA UN_AMICO (SALUTO_INIZIALE) DI' IL_TUO_NOME (DI BATTESIMO) : MARC	Bonjour, mon prénom est Marc	7
Dire où on habite	DI' DOVE_ABITI : ATENE	Je vis à Athènes ; J'habile à Athènes ; J'habite Athènes	11
Décrire sa famille	DI' NOME DI_TUO_FRATELLO_MAGGIORE : OCTAVIEN	Le nom de mon frère aîné est Octavien ; Mon frère aîné s'appelle Octavien	12
Dire le métier de ses parents	DI' PROFESSIONE DI_TUA_MADRE : PROFESSORE	Ma mère est professeur	21
Donner sa nationalité	DI' NAZIONALITA DI_TUO_PADRE : GRECA	Mon père est grec ; Mon père est de nationalité grecque ; Mon père a la nationalité grecque	14
Dire son âge	DI' ETA DI_TUO_FRATELLO_MAGGIORE : 27	Mon frère aîné a vingt-sept ans	14
Donner le lieu et la date de naissance	DI' IL_TUO_LUOGO_DI_NASCITA : ATENE	Je suis né à Athènes	10

Table 1: The 8 CALLST units

In the translation game, the glosses play a critical role, which is why the drafting needs to be very accurate. They follow three general rules: (1) the linguistic input they convey has a teaching/learning value (answering a question, performing an act, translating some frozen expression, etc,), (2) They are clear and exhaustive and (3) they avoid ambiguity. They first mirror the specific learning challenge focused in the lesson and implicitly use a contrastive approach stressing the most common difficulties of Italian mother tongue students learning French. For example, in the unit «Saluer et dire son nom», several sentences target the use of the words «prénom» and «nom» because they are not distinguished in Italian (=«nome»). In the unit «Dire les métiers de ses parents», we have similarly stressed the gender of the profession nouns, i.e. regular and irregular feminine and invariable nouns. Glosses also reflect the varieties of forms and expressions while speaking a language (many ways to tell the same things in different communicative contexts). For example, the glosses in lesson 1 «Saluer» contain information on the register (formal vs informal). The gloss is then more than a L1 sentence - it first forces the student to understand the main differences between the L1 and the L2 languages and then also helps him to associate to a meaning a set of alternative oral sentences.

In conclusion, the different combinations of the seven functionalities of CALL-SLT described in section 3.3 open up a powerful and interesting perspective, because different types of new cognitive processes are now involved: i) translation and transformation of a rich linguistic input (gloss) from L1 to L2; ii) speaking aloud and being checked by speech recognition; iii) reading what has been recognised (correctly or incorrectly); iv) listening and reading the correct answers (learner's production vs. the mother tongue version). Communicative language competence relates not only to the range and quality of knowledge (e.g. in terms of phonetic distinctions made or the extent and precision of vocabulary) but also to cognitive organisation and the way this knowledge is stored (e.g. the various associative networks in which the speaker places a lexical item) and to its accessibility (activation, recall and availability). We assume – and we will soon empirically verify the assumption – that all these operations can improve the activation, recall and availability of new stored knowledge.

References

[1] Bouillon P., Rayner M., Tsourakis N. and Zhang Q., A Student-Centered Evaluation of a Web-Based Spoken Translation Game, *Proceedings of the SLaTE Workshop*, Venice, Italy, 2011.

[2] Cervini C., Le dispositif de formation hybride et d'autoapprentissage des langues étrangères pour non-spécialistes : évolution des projets et évolution de la recherche-action au CILTA, *Repères DORIF: autour du français, langue, culture, plurilinguisme*, Roma : DORIF Università, in press.

[3] Chapelle C. A, Interactionist SLA Theory in CALL research, *CALL Research Perspectives,* LEA Publishers, 2005.

[4] Council of Europe, *The Common European Framework of Reference for Languages : Learning, Teaching, Assessment* , Cambridge University Press, 2001.

[5] DALIA Web e-learning platform: http://dalia.cilta.unibo.it

[6] Dornyei Z. and Skehan P., Individual differences in second language learning. In Doughty C. and Long M. (Eds.), *Handbook of Second Language Acquisition.* Oxford: Blackwell, 2000.

[7] Rayner M., Bouillon P., Tsourakis N., Gerlach J., Baur C., Georgescul M. and Nakao Y, A Multilingual Platform for Building Speech-Enabled Language Courses. *Proceedings of the L2WS Workshop*, Tokyo, Japan, 2010a.

[8] Rayner M., Bouillon P., Tsourakis N., Gerlach J., Georgescul M., Nakao Y. and Baur C. A Multilingual CALL Game Based on Speech Translation. *Proceedings of LREC*, Valetta, Malta, 2010b.

[9] Rayner M., Hockey B.A. and Bouillon P. *Putting Linguistics into Speech Recognition*, CSLI, Stanford, 2006.

[10] Wang C. and Seneff S. Automatic assessment of student translations for foreign languages tutoring. *Proceedings of NAACL/HLT*, Rochester, 2007.

[11] Warschauer, M. and Healey, D., Computers and language learning: An overview. *Language Teaching, 31*, 57-71, 1998.