

CALL-SLT

Un partenaire de conversation automatique

Pierrette Bouillon
Faculté de Traduction et d'Interprétation
Université de Genève



CALL-SLT

- Plate-forme pour l'apprentissage des langues, fondée sur la reconnaissance vocale
 - **CALL** : « **Computer-Assisted Language Learning** » (= ELAO)
 - **SLT** : « **Spoken Language Translation** » (= TAP)
- Développée à la Faculté de Traduction et d'Interprétation (FTI), dans le cadre de deux projets SNF : <http://callslt.unige.ch/>

Idée générale (1)

Dis ton âge : 10 ans

Dis ton âge : 21 ans

Dis qu'il est : 10:30

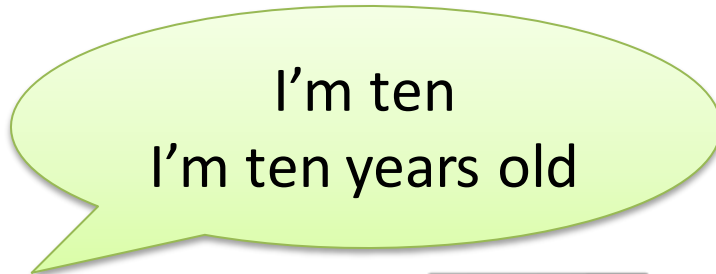
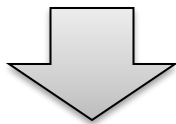
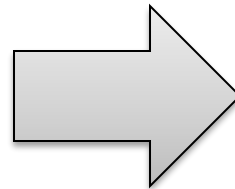
Dis qu'il est : 11:30

Dis que ta sœur est : architecte

Demande : l'addition



Idée générale (2)



L'hypothèse

- En essayant de se faire reconnaître par l'outil de reconnaissance vocale, l'apprenant améliore ses **compétences productives** :
 - Fluidité
 - Prononciation
 - Mémorisation lexicale et syntaxique

Plan de l'exposé

- Contexte : la reconnaissance vocale pour le CALL
- Fonctionnalités de CALL-SLT
- Différentes versions
- Points forts
 - Flexibilité
 - Reconnaissance vocale
 - Accessibilité
 - Accès aux données
- Evaluations
 - Subjective
 - Objective

Vidéo

- CALL-SLT en pratique : [vidéo](#)

Reconnaissance vocale : quid ?

- Convertit l'oral dans l'écrit
- Bcp de progrès ces dernières années
- Logique de l'utiliser pour l'apprentissage des langues
 - Intérêt d'un partenaire de conversation **automatique**
 - «Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)» insiste sur l'importance de la communication orale. Les connaissances peuvent être stockées dans le cerveau, sans pouvoir être réalisées comme compétences productives
- Idéalement, conversation libre, mais.... Impossible aujourd'hui.

Reconnaissance vocale : fonctionnement

- 2 étapes :
 - **Modèle et dictionnaire acoustiques** :
suggèrent toutes les séquences de mots possibles, étant donné les sons :
 - *le chat/chas/shah d'Iran*
 - Entraîné sur des données enregistrées par des natifs

Reconnaissance vocale : fonctionnement

- 2 étapes :
 - **Modèle du langage** : choisit la meilleure séquence : *le shah d'Iran*
 - Statistique : fréquence des mots et des séquences (2 à 4, N-grammes) dans un corpus de référence
 - chat d', chat d'Iran, chas d', chas d'Iran, shah de, shah d'iran, etc.
 - Linguistique : ensemble de phrases complètes (décrites dans des grammaires)

Reconnaissance vocale : difficulté pour le CALL

- Comprendre les accents des apprenants pour donner la réponse appropriée :
 - Plus le reconnaisseur est **général**, plus ce sera difficile de se faire reconnaître
 - Un reconnaisseur **statistique** va faire des fautes grammaticales
 - *Certaines commandes s'avèrent relativement intuitifs.*
 - Si on a un reconnaisseur **linguistique**, il sera plus précis, mais il ne reconnaîtra que les phrases qu'il connaît
 - Il faut trouver **un juste milieu** : il ne faut pas frustrer les apprenants, en rejetant des phrases compréhensibles, mais il ne faut pas non plus accepter des phrases incompréhensibles/incorrectes.

Différents compromis possibles (1)

- Le plus commun : faire lire une phrase spécifique en L2
 - Désavantage
 - N'exerce que la prononciation
 - Avantage
 - Modèle avec lequel on peut comparer les courbes mélodiques

Rosetta Stone

The screenshot displays the Rosetta Stone website interface. At the top, there is a navigation bar with the Rosetta Stone logo and several menu items: "PARTICULIERS", "ENTREPRISES & EDUCATION", "ASSISTANCE", "Inscrivez-vous", and "Essayez la démo GRATUITE". Below the navigation bar, a large video player is centered on the screen. The video player shows two audio waveforms: a blue one at the top and a red one at the bottom. A progress bar at the bottom of the video player indicates the current time is 01:10. The progress bar has several segments labeled: "ACCUEIL", "MÉTHODE", "PRONONCIATION" (which is highlighted with a yellow checkmark), "MOBILE", "TUTEURS", and "COMMUNAUTÉ". A "EN SAVOIR PLUS?" button is located at the bottom right of the video player. To the right of the video player, there is a chat window with a woman's profile picture and the text: "Vous avez des questions ? Laissez-nous vous aider. Cliquez ici pour démarrer le chat". The background of the website is dark with a faint image of a tablet displaying a lesson interface.

Différents compromis possibles

- Wang and Seneff, «translation game», MIT
- On donne une phrase source à traduire :
 - System: (*greet and prompt user*) Welcome! You are playing at level four. Okay, here's your first sentence to translate: **The return date is Tuesday October 25.**
- Avantages :
 - On exerce les capacités productives spontanées
- Mais :
 - On lie une phrase source à une phrase cible

Différents compromis possibles (3)

- Autre possibilité (CALL-SLT) : donner un acte du langage en L1 à exprimer en L2
 - **DEMANDER : CHAMBRE A DEUX LITS**
 - Exerce la prononciation, mais aussi la capacité productive spontanée, puisqu'il y a plusieurs réponses possibles
 - Se combine bien avec la reconnaissance linguistique: 1 reconnaisseur différent par acte du langage

L'interface CALL-SLT : 3 fonctionnalités

The screenshot displays the CALL-SLT interface, a prototype conversation partner. The title "CALL-SLT" is prominently displayed in white on a dark blue background, with the subtitle "A PROTOTYPE CONVERSATION PARTNER" below it. The interface is organized into several vertical panels and a central content area. On the left, four panels are labeled: "INTRODUCTION" (grey), "SELECT DOMAIN" (yellow), "SELECT LESSON" (purple), and "USE THE APPLICATION" (blue). On the right, two panels are labeled: "ADDITIONAL INFO" (green) and "INSTRUCTIONS" (orange). The central area features a white text box containing the Italian phrase "DI' PROFESSIONE DI_TUO_PADRE : MEDICO". Below this, a red-bordered input field contains the user's response: "Mon père est **pharmacien**". To the right of this input field is a microphone icon. Below the input field, two suggested responses are listed: "Mon père est docteur" and "Mon père est médecin", each preceded by a lightbulb icon. To the right of these suggestions is a green question mark icon. Navigation arrows (left and right) are positioned on either side of the top text box.

Entraîne différents processus cognitifs

- Enoncer un acte du langage en L2
- Parler et être compris par la reconnaissance vocale
- S'écouter
- Lire ce qui a été reconnu
- Lire et écouter les réponses correctes

3 versions différentes du logiciel

- Jeux de traduction
- Jeux de dialogue
- Jeux multimédia

Jeux de traduction

- Tirent leur nom du «translation game» (Wang et Seneff, MIT)
- Les actes du langage sont :
 - L1
 - Hors contexte
 - Groupés par but d'apprentissage (drill) :
 - Un ensemble de mots (les noms de métier, ...)
 - Une tournure syntaxique (manière de dire) (demander qqch avec «je voudrais», etc.)
 - Une paire phonétique

Exemple de jeu de traduction :

CILTA (Cervini)

- Module : Se présenter
 - Leçon : Dire le métier des membres de sa famille
 - Exercices :
 - DI' PROFESSIONE DI TUO FRATELLO : FORNAIO
 - DI' PROFESSIONE DI TUO PADRE : MACELLAIO
 - DI' PROFESSIONE DI TUO FRATELLO : GIORNALISTA
 - Leçon : Dire sa nationalité
 - Leçon : Dire son âge



Interface CALL-SLT : Jeu de traduction

CALL-SLT
A PROTOTYPE CONVERSATION PARTNER

Input → DI' PROFESSIONE DI_TUO_PADRE : MEDICO

Feedback → Mon père est **pharmacien**

Aide →

- Mon père est docteur
- Mon père est médecin

INTRODUCTION | SELECT DOMAIN | SELECT LESSON | USE THE APPLICATION | ADDITIONAL INFO | INSTRUCTIONS

Jeux de dialogue

Bonjour

SAY: HELLO

Bonjour

Que puis-je faire
pour vous ?

SAY : I WOULD LIKE A
DOUBLE ROOM

J'aimerais une
chambre double

- Les actes du langage sont mis en contexte, ce qui permet de créer un véritable dialogue.



Interface : jeu de dialogue

Fichier
multimedia :
contexte et
feedback



New badge!



Score



Input



frag : Zimmer für 3 Nächte



Aide



I would like a room for three nights

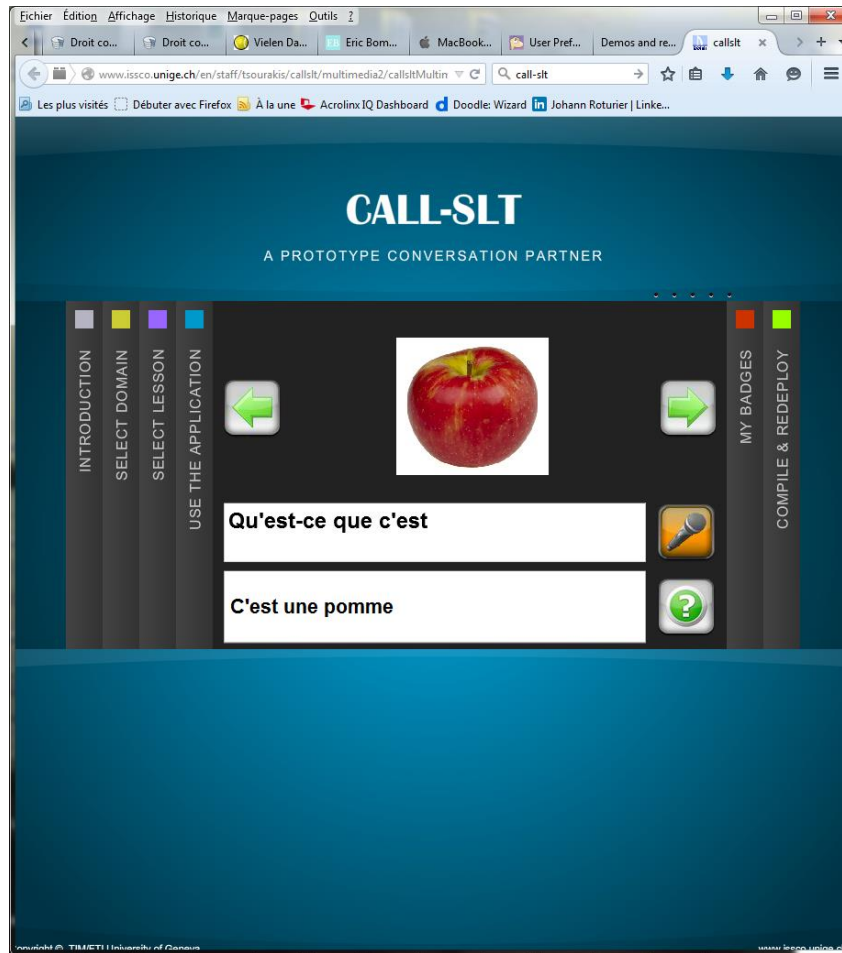


Jeux multimedia

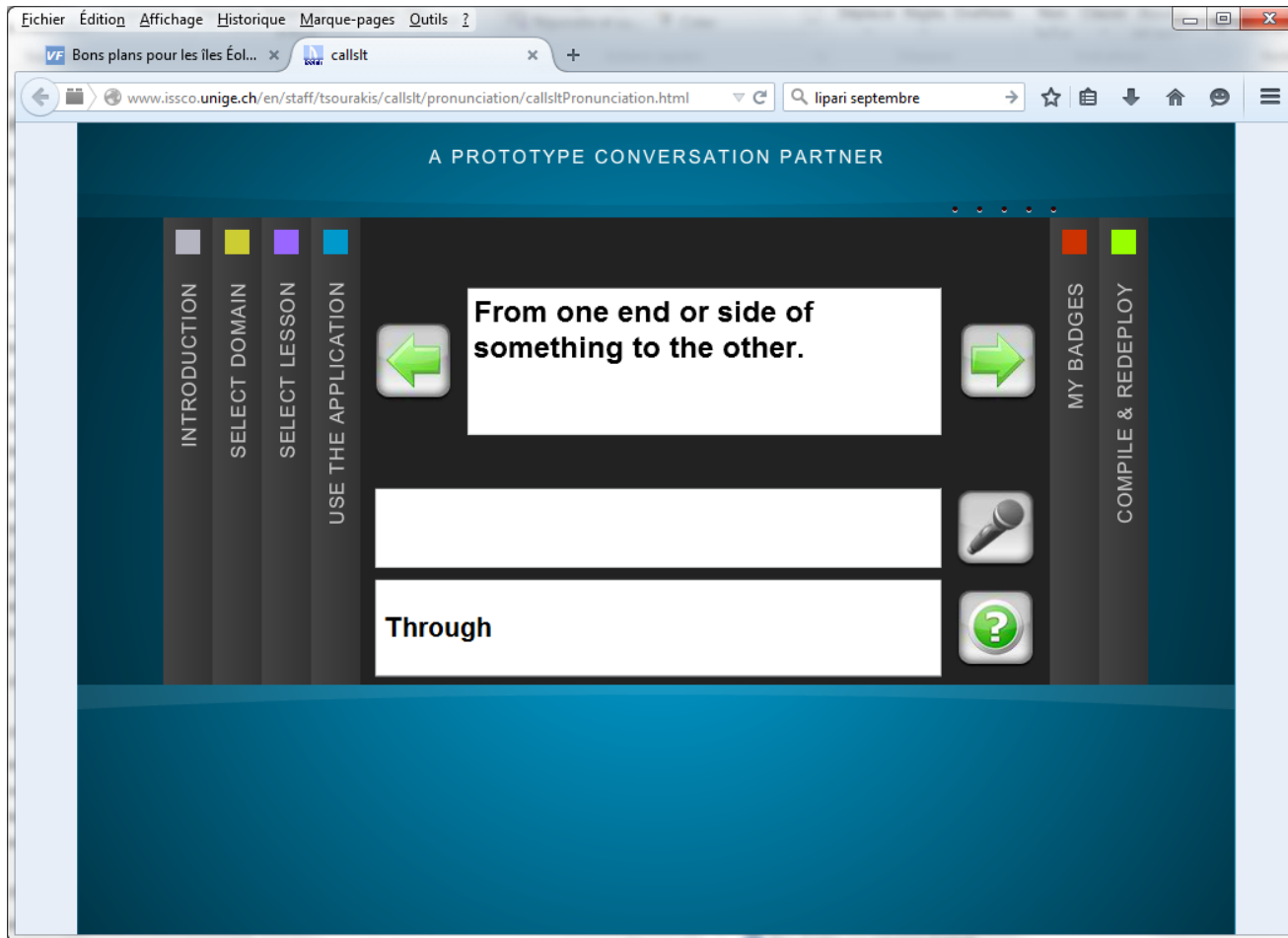
- Combinent :
 - Texte (L1 ou L2)
 - Fichier multimedia



Jeu de vocabulaire (1)



Jeu de vocabulaire (2)



Points forts

- Accessibilité
- Flexibilité
- Reconnaissance vocale
- Sauvegarde des données



Point fort : accessibilité

- Disponible sur le web
 - Rien à installer de la part de l'étudiant / professeur

Point fort : flexibilité

- On peut définir **en ligne** son plan de leçon avec son propre contenu, avec un langage simple (Rayner et al., 2015)

Jeux de traduction

- Prompt
 - Domain bologna_se_presenter
 - Lesson bologna_3
 - Text DI' DOVE ABITI : IN FAMIGLIA
 - Response j'habite avec ma famille
 - Response j'habite en famille
 - Response je vis avec ma famille
 - Response je vis en famille
- EndPrompt
- PromptTemplate j_habite_ville ITALIAN FRENCH
 - Domain bologna_se_presenter
 - Lesson bologna_3
 - Text DI' DOVE ABITI : ITALIAN
 - Response j'habite FRENCH
 - Response (j'habite | je vis) à FRENCH
- EndPromptTemplate
- ApplyTemplate j_habite_ville "ATENE" "athènes"
- ApplyTemplate j_habite_ville "PARIGI" "paris"



Jeux multimedia

- Prompt
 - Lesson vocabulaire
 - **Multimedia** **pomme.jpeg**
 - Text Qu'est-ce que c'est ?
 - Response c'est une pomme
 - Response une pomme
- EndPrompt

Jeux de dialogue

- On peut lier les «prompts» entre eux pour créer un dialogue avec des scripts XML :
 - Script = ensemble d'étapes
 - Id = Questions + acte du langage
 - Limit/Success = Les étapes suivantes en fonction des réponses

```
<step>
```

```
  <id>ask_for_number_nights</id>
```

```
  <group>room_for_number_of_nights</group>
```

```
  <limit>is_one_night_okay</limit>
```

```
  <success>ask_type_of_room</success>
```

```
</step>
```


Point fort : reconnaissance vocale

- Construite **dynamiquement** à partir des leçons avec la plate-forme Regulus :
 - Reconnaissance linguistique vs statistique (avec Nuance)
 - On peut adapter la difficulté de la reconnaissance vocale en fonction du public en dérivant des reconnaisseurs vocaux avec une couverture plus ou moins large
 - Possibilité d'avoir des réponses positives et négatives et d'améliorer ainsi le système au fur et à mesure des collectes de données
 - Prompt
 - Domain bologna_se_presenter
 - Lesson bologna_3
 - Text DI' DOVE ABITI : IN FAMIGLIA
 - Response j'habite avec (*mon | ma) famille
 - Response j'habite en famille
 - Response je vis avec (*mon | ma) famille
 - Response je vis en famille
 - EndPrompt

Point fort : sauvegarde des données

- On peut sauver ce que les étudiants ont fait, pour évaluer leur performance en parallèle
 - Possibilité d'exporter une table Word avec toutes les informations

Exemple de fichiers Word

The image shows a Microsoft Word document with two pages. The left page contains a table with 10 rows of audio file information. The right page contains a table with 10 rows of audio file information. The tables are structured as follows:

File ID	Date	Title	no_help	no_rec	audio file	5	2	3
fq613:	2011-11-30_15-28-43	DI' NAZIONALITA' DI_TUO_PADRE::TURCA	no_help	no_rec	audio file	5	2	3
fq613:	2011-11-30_15-28-50	DI' NAZIONALITA' DI_TUO_PADRE::TURCA	no_help	no_rec	audio file	5	2	4
fq613:	2011-11-30_15-29-01	DI' NAZIONALITA' DI_TUO_PADRE::TURCA	help	rec	audio file	5	4	4
fq613:	2011-11-30_15-29-13	DI' NAZIONALITA' DI_TUA_MADRE::GRECA	no_help	rec	audio file	5	3	4
fq613:	2011-11-30_15-29-24	DI' NAZIONALITA' DI_TUA_MADRE::TURCA	no_help	rec	audio file	5	3	4
fq613:	2011-11-30_15-29-36	DI' NAZIONALITA' DI_TUO_PADRE::GRECA	no_help	no_rec	audio file	5	3	4
fq613:	2011-11-30_15-29-43	DI' NAZIONALITA' DI_TUO_PADRE::GRECA	no_help	no_rec	audio file	5	3	4
fq613:	2011-11-30_15-29-55	DI' NAZIONALITA' DI_TUO_PADRE::GRECA	no_help	no_rec	audio file	3	3	3
	2011-11-	DI' NAZIONALITA'			audio file			

File ID	Date	Title	no_help	no_rec	audio file	4	3	3
fq613:	2011-12-09_09-44-45	DI' DOVE_ABITI::GRECIA	no_help	no_rec	audio file	4	3	3
fq613:	2011-12-09_09-45-02	DI' DOVE_ABITI::GRECIA	no_help	rec	audio file	Cela ne me semble pas sa voix		
fq613:	2011-12-09_09-46-01	DI' DOVE_ABITI::CON_GENITORI	no_help	rec	audio file	5	3	3
fq613:	2011-12-09_09-46-26	DI' DOVE_ABITI::IN_FAMIGLIA	no_help	no_rec	audio file	5	3	3
fq613:	2011-12-09_09-46-36	DI' DOVE_ABITI::IN_FAMIGLIA	no_help	no_rec	audio file	5	4	3
fq613:	2011-12-09_09-46-59	DI' DOVE_ABITI::IN_FAMIGLIA	help	rec	audio file	5	4	4
fq613:	2011-12-09_09-47-15	DI' DOVE_ABITI::CON_MIA_MADRE	no_help	rec	audio file	5	3	4
fq613:	2011-12-09_09-47-30	DI' DOVE_ABITI::SPAGNA	no_help	no_rec	audio file	5	3	4

Evaluations

- Evaluations subjectives : questionnaires
- Evaluations objectives :
 - La reconnaissance vocale motive les étudiants
 - L'aide améliore les performances
 - La reconnaissance accepte les phrases correctes et rejette les incorrectes
- <http://callslt.unige.ch/publications/>

Evaluations subjectives

- Plus de 300 apprenants (différents niveaux et langues)
- Dans toutes les expériences, au moins $\frac{3}{4}$ des étudiants sont **très satisfaits** et pensent **avoir appris** avec le logiciel
- Point **négatif** le plus souvent mentionné : parfois difficile de se faire reconnaître, surtout pour le niveau A1 (apprenants chinois et arabes).

Evaluation objective : motivation

- La reconnaissance vocale **motive** les étudiants, par rapport à une version qui accepte tout ce que les étudiants ont dit
 - Deux versions du logiciel, avec et sans la reconnaissance vocale
 - Ils utilisent plus le logiciel avec la reconnaissance que sans, de manière statistiquement significative (Rayner et al., 2013)

Evaluation objective : aide

- L'aide permet clairement aux étudiants d'améliorer leur performance
 - Ils ont de meilleures évaluations humaines quand ils ont sélectionné l'aide (Bouillon et al., 2012 ; Cervini et al., 2013)

	Total phrases	Lex/Gram	Phonétique	Fluidité
Pas d'aide	761	4,42	2,82	2,84
Aide	304	4,75	3,06	3,14

Evaluations objectives : reconnaissance

- La reconnaissance permet de distinguer ce qui est correct/incorrect dans une certaine mesure
- Elle rejette la plupart des phrases jugées négativement (90%) par un professeur (0-2 sur une échelle de 5) (Cervini et al., 2013)
- Elle accepte plus de phrases avec un score de 5 qu'elle n'en rejette (Cervini et al., 2013)
- Les phrases jugées correctes par le logiciel obtiennent un meilleur score de la part de l'enseignant que les incorrectes (Bouillon et al., 2012)

	Total phrases	Lex/Gram	Phonétique	Fluidité
Phrases considérées comme correctes par le système	505	4,85	3,11	3,11
Phrases considérées comme incorrectes par le système	560	4,31	2,77	2,87

Données authentiques disponibles

- Permet de collecter des données authentiques (non lues)
 - Germanophones apprenant l'anglais : plus de 25.000 interactions, donc 4.000 annotées par des non-spécialistes
 - Italophones apprenant le français : 4.000 interactions, dont 1.000 annotées par un professeur
 - Francophones apprenant l'anglais : 1.000 interactions annotées
 - Chinois apprenant l'anglais : 2.000 interactions

Partenaires

- Différents collèges en Suisse alémanique
- Intercountry, Langues, Business et culture
- CILTA, Université de Bologne
- Innovalangues, projet de l'université Stendhal, Grenoble

Références

- Bouillon P., Gerlach J., Baur C., Cervini C., Gasser R.B., Intégration d'un jeu de traduction orale sur le web pour l'apprentissage d'une langue seconde, in: Proceedings of TICE, Lyon, France, 2012.
- Cervini C., Bouillon P., Gasser G., Apprendre à s'autoformer en langues : approches créatives et outils numériques, in: Les Langues Modernes 4/2013.
- Wang, C., Seneff, S., Automatic assessment of student translations for foreign language tutoring. In Proceedings of NAACL/HLT 2007, Rochester, NY.
- Jolidon, A. Reconnaissance vocale et amélioration de la prononciation : élaboration et évaluation de leçons avec le logiciel CALL-SLT, Master Thesis, University of Geneva, Faculté de Traduction et d'Interprétation, 2013.
- Rayner M., Tsourakis N., Methodological Issues in Evaluating a Spoken CALL Game: Can Crowdsourcing Help Us Perform Controlled Experiments?, in: Proceedings of the SLaTE Workshop, Grenoble, France, 2013.
- Rayner M., Baur C., Bouillon P., Chua C., Tsourakis N., Build Your Own Speech-Enabled Online CALL Course, No Experience Required, in: Proceedings of INTED, Madrid, Spain, 2015.