

Alain KAMBER, Maud DUBOIS &
Sara COTELLI KURETH (Eds)

- "La linguistique appliquée à l'ère digitale"
Actes du colloque VALS-ASLA 2020
(Neuchâtel, 12-14 février 2020)
- "Applied Linguistics in the Digital Age"
Proceedings of the VALS-ASLA conference
(Neuchâtel, 12-14 February 2020)



bulletin vals-asla

numéro spécial

Bulletin suisse
de linguistique appliquée
Été 2021 volume 2

Vereinigung für angewandte
Linguistik in der Schweiz

Associazion svizra
da linguistica applitgada

Association suisse
de linguistique appliquée

Associazione svizzera
di linguistica applicata

Vereinigung für Angewandte
Linguistik in der Schweiz

Association Suisse de
Linguistique Appliquée

Associazione Svizzera di
Linguistica Applicata

Associaziun Svizra da
Linguistica Applitgada



Le *Bulletin suisse de linguistique appliquée* est l'organe de publication de l'Association Suisse de Linguistique Appliquée (VALS/ASLA).

Comité de relecture pour ce numéro

Marie-Josée Béguelin (Université de Neuchâtel); Anne-Claude Berthoud (Université de Lausanne); Peter Blumenthal (Universität zu Köln); Alex Boulton (ATILF); Paul Cappeau (Université de Poitiers); Cristelle Cavalla (Université Paris III – Sorbonne Nouvelle); Thomas Cobb (UQAM); Renata Coray (Université de Fribourg); Sara Cotelli Kureth (Université de Neuchâtel); Anna Dal Negro (Universität Zürich); Sabine De Knop (Université Saint-Louis – Bruxelles); Jean-François de Pietro (IRDP); Maud Dubois (Université de Neuchâtel); Daniel Elmiger (Université de Genève); Laurent Fillietaz (Université de Genève); Marianna Fonseca Favre (Université de Genève); Noémie Guérif (Université d'Aix-Marseille); Martin Hilpert (Université de Neuchâtel); Loanne Janin (Université de Neuchâtel); Adam Jones (Université de Neuchâtel); Laure Anne Johnsen (Université de Neuchâtel); Alain Kamber (Université de Neuchâtel); Martin Luginbühl (Universität Basel); Marinette Matthey (Université de Grenoble-Alpes); Gilles Merminod (Université de Lausanne); Johanna Miecznikowski-Fuenfschilling (Università della Svizzera italiana); Lorenza Mondada (Universität Basel); Simona Pekarek Doehler (Université de Neuchâtel); Daniel Perrin (ZHAW); Isabelle Racine (Université de Genève); Regula Schmidlin (Université de Fribourg); Elena Smirnova (Université de Neuchâtel); Isabelle Stabarin (Università degli studi di Trieste); Kathrin Steyer (IDS - Leibniz-Institut für Deutsche Sprache); Ioana Maria Stoenica (Université de Neuchâtel); Anja Stukenbrock (Université de Lausanne); Christian Surcouf (Université de Lausanne); Anita Thomas (Université de Fribourg); Henry Tyne (Université de Perpignan).

Nous remercions vivement les membres du comité de lecture pour leur contribution précieuse à la qualité de ce volume.

Publié avec le soutien financier de l'Académie suisse des sciences humaines et sociales (ASSH / SAGW), le *Bulletin suisse de linguistique appliquée* paraît deux fois par an.

Chaque article paru dans le bulletin a été soumis à un processus de "peer review" anonyme.

Abonnement personnel	Suisse CHF 30.-	Etranger CHF 35.-
Abonnement institutionnel	Suisse CHF 50.-	Etranger CHF 55.-
Prix au numéro	Suisse CHF 25.-	Etranger CHF 25.-
Pas de réduction librairie		

Rédaction

Dr. Sara Cotelli Kureth (Université de Neuchâtel)
E-mail: sara.cotelli@unine.ch
Centre de langues, Av. du 1^{er}-Mars 26, CH-2000 Neuchâtel



Responsables comptes rendus	Dr. Gilles Merminod (Université de Lausanne) E-mail: gilles.merminod@unil.ch
Administration	Florence Waelchli (Université de Neuchâtel)
Abonnements, commandes	Institut des sciences du langage Rue Pierre-à-Mazel 7, CH-2000 Neuchâtel E-mail: bulletin.valsasla@unine.ch / CCP: 20-7427-1

© Cette revue est distribuée selon la licence Creative Commons CC-BY 4.0.



Table des matières

"Applied Linguistics in the Digital Age"

Proceedings of the VALS-ASLA conference (Neuchâtel, 12-14 February 2020)

Sommaire du volume II

Alain KAMBER, Maud DUBOIS & Sara COTELLI KURETH

Introduction 1-10

Thomas COBB

Corpus for courses: Data-driven course design 11-22

Christian SURCOUF & Alain AUSONI

Du phonographe aux corpus numériques oraux: aperçu
historique de la compréhension du français oral quotidien
en FLE 23-39

Fryni KAKOYIANNI-DOA & Dora LOÏZIDOU

Un corpus parallèle français-grec en ligne pour une aide à
la rédaction d'une argumentation 41-58

Hasti NOGHRECHI

Introduire des constructions verbales en classe de français
langue étrangère du niveau A1: une pratique sur corpus..... 59-75

Anita THOMAS

Corpus et exercices en ligne, complémentarité ou
redondance? L'exemple des pronoms relatifs *qui* et *que* 77-97

Carmen ALBERDI & Carole ETIENNE

Apprendre à interagir en classe de FLE: la situation
d'invitation 99-120

Biagio URSI & Virginie ANDRÉ

Corpus: exploration, médiation et autonomisation.
L'utilisation du concordancier de la plateforme FLEURON
en cours de FLE 121-140

Yu KANAZAWA

Replication and extension of an empirical study about
foreign language formulaic familiarity database: Reliability
and emotionality..... 141-156

Mariana FONSECA & Laurent GAJO

Le plurilinguisme dans les MOOCs: de nouvelles
ressources pour la construction des savoirs 157-173

Tanya MCCARTHY & Matthew ARMSTRONG

Overcoming language barriers and boundaries: Video-
mediated eTandem..... 175-194

Sara COTELLI KURETH & Alain KAMBER

Digitalisation des outils de référence: enjeux et
perspectives pour l'enseignement de la traduction vers le
français L2 195-212

Monica MASPERI, Triscia BIAGIOTTI & Sandra CANELAS-TREVISI

Écouter la "voix" du texte: modélisation d'une approche de
l'oral en L2 au sein d'un environnement numérique 213-237

Proceedings of the VALS-ASLA conference "Applied Linguistics in the Digital Age", Neuchâtel, 12-14 February 2020

Introduction

Alain KAMBER & Maud DUBOIS

Université de Neuchâtel
Institut de langue et civilisation françaises
Faubourg de l'Hôpital 61-63, 2000 Neuchâtel, Suisse
alain.kamber@unine.ch; maud.dubois@unine.ch

Sara COTELLI KURETH

Université de Neuchâtel
Centre de langues
Avenue du 1^{er}-Mars 26, 2000 Neuchâtel, Suisse
sara.cotelli@unine.ch

Triggered by the rise of digital technologies in the mid-20th century, the "digital revolution" has gathered momentum in recent years. Its current developments anticipate – and lead to – a profound upheaval of our societies, affecting professional processes as well as our daily interactions, or even teaching/learning practices in the most diverse contexts.

The aim of the biannual VALS-ASLA conference, which took place from 12 to 14 February 2020 in Neuchâtel, was to explore how research in applied linguistics can help us better understand some challenges of the digital revolution, and provide innovative solutions. Scholars addressed central social issues for the digital society. Among these, we can mention the new forms of literacy that are currently emerging under the label of "digital literacy", which have become essential to access information, or even to participate in contemporary debates. The lack of digital literacy can thus lead to the exclusion of some segments of the population. On a more general level, digitalisation is shaping new ways of thinking, acting and learning within our education systems, with fundamental implications for L1 and L2 teaching, among others. Finally, digitalisation has dramatically transformed our interactive and communicative practices, including in professional contexts; the new constraints, opportunities and dynamics characterising these interaction processes will still have to be identified and analysed in detail in the coming years.

Participants were asked to explore how applied linguistics responds to this evolution and the challenges that follow, what role linguistics can play to better understand the social and educational dimensions of this digital transformation, what new objects of study it brings to light and what new research methodologies it offers scholars. The papers presented at the Neuchâtel conference were organised around the following three axes:

1. *New technologies and communication*: How do new digital technologies transform our communicative practices and affect language?

This axis looks at both communication via new technologies (WhatsApp, Facebook, Skype, etc.) and the presence of such technologies in face-to-face communication (e.g. how the handling of a smartphone or a computer influences the interactional dynamics of a professional meeting).

2. *Learning, teaching and digital media*: What new demands does digitalisation make on our education systems, especially – but not only – in terms of digital literacy?

This second axis focuses on the opportunities of digitalisation for teaching practices, including the role of digital tools in teaching and the didactic implications that may arise, particularly in language learning and teaching.

3. *Methodological and analytical opportunities brought by digitalisation*: What are the possibilities offered by using digital media in scientific research?

This axis offers a meta-reflection on the research carried out in applied linguistics specifically and in linguistics in general, by questioning the impact of digital media on research: the study of large data corpora, the increasingly fine-tuned possibilities of processing audio and video data, and the advantages of new visualisation possibilities (for example, dynamic graphical representations, which retrace the evolution of a phenomenon through time in historical linguistics).

The proceedings of the conference, collected in this two-volume special issue of *Bulletin suisse de linguistique appliquée*, do not follow these three axes, but were ordered according to a different logic. The first volume groups papers that explore how oral and written interactions are modified in relation to different digital devices and media (mobile phones, tablets, games and video clips) (part 1). Part 2 focuses on the new practices and forms of communication induced by the Internet or Twitter.

The second volume contains contributions dealing more specifically with the design and use of language corpora for language teaching (part 1), and more broadly with the use of digital resources by both learners and teachers (part 2).

Volume 1

Part 1: Interaction

The first volume opens with the text of the plenary lecture by **Florence Oloff** (University of Oulu). After a state of the art of applied linguistics and communication studies focusing on the use of smartphones, specifically in face-to-face interactions, the article proposes a multimodal interactional analysis of two video excerpts that exemplify some interactional phenomena specific to the use of mobile phones. The author thus proposes to investigate how interactional studies provide a valuable understanding of the new social practices involved in the use of digital devices. Such research is essential to provide today's society with a more objective assessment of the role of technology in social interactions.

In the next paper, **Biagio Ursi** (Aix-Marseille University) & **Heike Baldauf-Quilliatre** (Université Lumière Lyon 2) analyse intergenerational multimodal interactions: parents show their very young children how to play on a giant tablet, helping them complete the required tasks while allowing them to acquire the ability to perform gestures on the screen. This study is a first of its kind and explores the issues raised by this type of interaction. For example, one can highlight how parents use tactile guidance and demarcation of action space to guide children to perform the correct gesture. By focusing on the temporal aspect, the article finally shows that the analysis must consider a double temporality (that of the actions required by the game and that of the child's motor and cognitive acquisition). Such research projects should make it possible to describe in detail how the acquisition of fine gestures and specific tactile configurations is co-constructed in interaction.

Jakub Mlynář (Charles University, Prague) examines a specific situation of group interaction around the collective viewing of a video clip in the classroom: rewatching, where participants re-watch all or part of a video they have already watched. This is one of the first studies on this phenomenon and the author analyses the functions of rewatching (resource or subject) and how these new viewings are brought about (tacitly - only through gestures - and explicitly). The ethnomethodological and conversational approach allows the author to highlight how, during re-watching, the video clips are treated as structured objects with a beginning and an end and the participants navigate through the horizontal timeline. They search for information to resolve questions that were raised after the first viewing. The present study highlights the contextualised objectification of sources such as video clips in classroom interactions.

Mireille Rodi & Thierry Geoffre (Haute école pédagogique de Fribourg) analyse the interactions of child-speech therapist dyads around *L'Orthodyssée des Gram*, an online game that aims to develop metagraphic reflection in grammatical spelling (agreement chains). They seek to highlight interaction

profiles displaying the synergy between scaffolding strategies and the evolution of metagraphic verbalisations. A corpus of 40 filmed sessions between a speech therapist and a French monolingual patient with developmental disorders of written language allows the authors to categorise the support strategies of the professionals and the metagraphic discourse of the children according to a qualitative ascending method and to show their cross evolution: continuous adaptation of the speech therapist's support and progression of the child's explanatory discourse, both being strongly correlated. On this basis, the authors are considering the introduction of the online game in ordinary classrooms and are hoping to find the same synergies at work.

Part 2: Internet, Twitter and new digital platforms

Having adapted her plenary lecture to recent momentous events, **Louise-Amélie Cougnon** (Université catholique de Louvain), collaborating here with **Louis de Viron** (Data Text SPRL), focuses on the crisis communication around COVID-19 in French-speaking Belgian tweets sent during the lockdown initiated in March 2020. To analyse the linguistic aspects of this crisis, the authors built a corpus of 100,000 tweets organised in three sub-corpora: citizen corpus, political corpus and media corpus. These data are processed from three angles: 1) the lexical creativity that accompanied the emergence of the pandemic, specifically name the disease or the new elements that resulted from it; 2) the themes addressed, which reflect the different concerns related to the disease, lock-down or the political management of the crisis; 3) the main actors highlighted in the crisis, including political figures and the media. The results obtained show both similarities with other situations and elements specific to this particular crisis.

The article by **Christian Kriele & Philipp Dreesen** (ZHAW) deals with a project to optimise the search results of a legal information platform. The project explored how linguistic phenomena influence the factors of "precision" and "recall" and how search results can be optimised using applied linguistic methods. In two sub-projects, independent solutions, one oriented towards terminology (controlled vocabulary) and the other using corpus linguistic methods ("word embedding" and collocation profiles), were developed to optimise queries. The paper concludes with a discussion of a possible synthesis of the two approaches and how its implementation in the search engine could improve users' search experience.

Cordula Meissner (Universität Innsbruck) notes that with the expansion of the digital communication and information space, sources for multilingually oriented language teaching have become easier to access. These "digital language resources" include, for example, databases on the characteristics of the world's languages, multilingual thesauri or word lists linked to frequency and distribution. From the point of view of language teaching, their strength lies in the fact that they illustrate selected features of the language(s) and make them

accessible in a concise form. They therefore offer specific opportunities for reflection on the language(s) and the development of language awareness. Focusing on academic language and scientific propaedeutics, the article shows what opportunities digital language resources open up for multilingually oriented foreign language learning/teaching and the challenges associated with them.

In the next article, **Gianni De Nardi, Igor Matic & Felix Steiner** (ZHAW) examine what makes understanding specialised vocabulary in informed consent information more difficult for laypersons and explore some digital solutions to overcome these obstacles. Based on the observation that the basic concepts of scientific research in information texts impede understanding, the authors focus on the use of specialised vocabulary by lay people. An analysis of interactions in Internet forums shows that laymen have their own practices for using specialised vocabulary. In order to make information texts more comprehensible, the authors therefore advocate replacing experts' specialised vocabulary with variants that are more accessible to laymen, which would be systematically listed in a digital dictionary of specialised variants.

Alice Delorme Benites (ZHAW) & **Lorenza Pescia** (Institute for Advanced Study Princeton) focus on the topic of energy, which has been in the spotlight in Switzerland in recent years, particularly in the context of the popular vote on the new Federal Energy Act. Based on a qualitative analysis of the French and Italian texts of the brochure *Popular Vote of 21 mai 2017. Explanations by the Federal Council. Federal Energy Act*, they analyse the emergence of four argumentative topoi using the Swiss-AL multilingual corpus. In order to study the relationship between language and the reception of energy discourse at the local and national levels, the Swiss-AL corpus was subdivided according to the language of writing, i.e. between texts written directly in French or Italian and translated texts. This subdivision reveals linguistic peculiarities that lead to questions of language policy.

Elena Gallo (Universities of Udine and Trieste - Université Paris Sorbonne Nouvelle) reports on a study analysing 15 societal advertising campaigns broadcasted in France on the Internet by ministries, private companies and NGOs between 2018 and 2019. Collected on the websites of these organisations and social networks (Facebook, Twitter and Instagram, specifically), these data (videos and images) allow the author to highlight how the digital channel influences both the construction of the message and its pragmatic scope. To this end, she examines technowords appearing in these advertisements, in particular hashtags and URL links. Both are key elements in this type of communication, and they carry interdiscursivity, as well as the interactive dimension made possible by the digital medium and activated by means of proposals for concrete actions, surveys, or knowledge tests.

In her article, **Fabia Hultin Morger** (Stockholm University) examines the place and role of Internet memes in the political campaigns launched by the Swiss political group *Opération Libero*. To this end, the author assembled a collection of 64 memes published on *Opération Libero*'s Facebook page during several referendum campaigns between February 2016 and February 2020. Their analysis, based on both multimodality theory and a historical approach to discourse, sheds light on how arguments are created between images and text and how they interact more broadly with political discourse. This study thus sheds light on how this specific phenomenon of digital election campaigns works and how it can be used as a tool of political persuasion.

Karina Frick (Universität Zürich) presents an analysis of new mourning practices emerging from digital media such as the Internet or Twitter, she looks at a typical topos of interactions around mourning, speechlessness, and the way it is staged, especially in texts accompanying virtual candles and in a series of tweets related to deaths or tragic events. Thus, some hashtags take up stereotypical expressions of speechlessness (*#Sprachlos*, *#mirfehlendieWorte*), while at the same time participating in the creation of new forms of collective grief online. On the Internet, the topos of speechlessness, when it is exposed at length, is sometimes accompanied by its opposite. Finally, a corpus of press articles examines the reception of these new practices and reveals that they are criticised and considered inauthentic.

Volume 2

Part 1: corpus

The first article in this volume presents the basis of corpus-based learning. Through his focus on a French reading course, **Thomas Cobb** (UQAM) demonstrates the effectiveness and relevance of this approach to language learning and teaching, specifically for the creation of a course. Through numerous concrete examples, he presents the tools developed within the framework of this approach (here the different software available on the website www.lexutor.ca). The author shows how teachers can use these tools to develop a corpus of texts that are coherent in terms of level or specialised vocabulary, for example. These tools allow for the selection of appropriate documents, their possible modification, and the possibility of ensuring that the final assessment test corresponds to the tasks studied during the course. The use of the various tools derived from corpus-based learning makes it possible to address questions and problems that cannot be answered by human eyes and can thus improve authentic material made available to learners.

Through a diachronic and panoramic approach, **Christian Surcouf & Alain Ausoni** (University of Lausanne) examine the technological history of sound recording as well as the use of recorded speech for language learning and

teaching, and more particularly understanding everyday oral communication. The authors cover a time span from the middle of the 19th century to the present day, with the aim of showing that even though there have long been calls for authentic oral communication to be integrated into the learning of French as a foreign language, this has not yet happened satisfactorily, even though it is possible following the creation of oral corpora. The authors explain this lack by the insufficient emphasis placed on oral comprehension, which is attributable to a certain representation of oral language still considered through the normative vision of the written language.

In order to find out to what extent a bilingual parallel corpus can help learners of French to improve their writing skills, **Fryni Kakoyianni-Doa & Dora Loïzidou** (University of Cyprus) present the results of an experiment conducted in November 2019 with 22 Greek-speaking Cypriot students of B1-B2 level at the University of Cyprus. The students were asked to use the parallel French-Greek SOURCE Corpus to write argumentative texts. By comparing the writing of texts with simple online dictionaries, in the first instance, and with the parallel corpus in the second instance, the authors are particularly interested in the presence of modalising adverbials in the two texts produced by each student. Students were handed a list of such adverbials and asked to use them. The authors found a positive effect of using the parallel corpus, which has the advantage of showing the words in context. They confirm this favourable result by also relying on questionnaires completed by the learners about the tools that were used.

Using a corpus-based approach, **Hasti Noghrechi** (University of Neuchâtel) shows how to introduce verbal constructions to A1 learners of French, by answering the question of what to teach and how. A quick review of the reference books, methods and grammar books for A1 level allows the author to note that the emphasis is rather on verbal morphology and that if constructions are present, the subject is never thematized as such. Moreover, the verbs that tend to be presented first have complicated constructions, notably *être* and *avoir*. Based on criteria of affinity and frequency that she explains, the author chooses four verbs whose constructions are analysed in Emolex to find prototypical statements. She then proposes three-step didactic activities to introduce verbal constructions in the language classroom, including comparison with the mother tongue, parallel corpora and work on online platforms.

In her article, **Anita Thomas** (University of Fribourg) discusses the use of online corpora as a resource for foreign language teaching. Her study is based on an analysis of the French relative pronouns *qui* and *que* in their most basic uses. Analysis of data from L2 French learners indicates a preference for the use of *qui* with animate antecedents and *que* for inanimate ones, but there are differences depending on the task and the learners' L1. A comparison between the teaching material available on the website *Le Point du FLE* and a sample of

the ScienQuest written corpus reveals a similar distribution of the two relative pronouns. The discussion focuses on the contribution of the comparison of different data sources to the development of corpus-based learning materials.

Carmen Alberdi (University of Granada) & **Carole Etienne** (ENS Lyon) note in their text that the data collected and documented by oral language specialists is rarely used in the classroom, despite the availability of pedagogical resources online. They present the CLAPI-FLE application, designed by the ICAR/LIS team in collaboration with a network of language specialists and teachers, and dedicated to the teaching of oral French, which offers some forty extracts described, transcribed, and designed in a didactic way from everyday situations, with collections and explanatory sheets. The authors aim to show the extent to which CLAPI-FLE enables learners to feel immersed in French interactions in their classrooms, whatever their knowledge or objectives. They explain the pragmatic functions and processes used by speakers during a telephone invitation and share their experience of teaching interactions in French.

Virginie André (ATILF) & **Biagio Ursi** (Aix-Marseille University) explore another digital device: FLEURON (Français Langue Etrangère Universitaire: Ressources et Outils Numériques), an online multimodal corpus which shows interactions experienced by students on and off campus. The authors propose a sequential and multimodal analysis of three excerpts of interaction where pairs of learners use the FLEURON concordancer, with the mediation of a tutor. Thanks to this multimodal corpus and its functionalities (e.g. how easy it is to get access to the wider context by a simple click), these learners manage to develop their interactional skills, skills which are often absent from language textbooks. In this way, learners discover typical phenomena of spoken language (e.g. self-repair). This exploratory study also explores how learners and teachers co-construct the learning sequence, using multimodal resources that include both linguistic elements and gestures.

Part 2: learning and teaching

Yu Kanazawa (Kwansei Gakuin University) investigates the acquisition of formulas by distant language learners (Japanese studying English). His hypothesis is that the learner's familiarity and emotionality with these formulas enables them to learn better. Thus, the author replicates a study that measured the familiarity of a list of 500 English formulas (PHRASE list) and adds a task to study their emotionality: each formula is judged by the respondents in terms of valence (very positive or very negative on a 7-point Likert scale). The results of the study are consistent with those of the baseline study and the validity of the familiarity values obtained is therefore reinforced. Finally, the author calls for further studies on emotionality to verify the first results obtained and to allow for the integration of these two factors in the learning and teaching of vocabulary.

According to **Mariana Fonseca & Laurent Gajo** (University of Geneva), MOOCs (Massive Open Online Courses) are based on digital technologies that make them not only a new medium, but also a new didactic genre. By allowing a "massive" diffusion of knowledge, these courses reexamine the languages used in the construction and diffusion of knowledge as well as their cultural dimension (embedded knowledge). The authors want to show the tension between globalization and contextualization of knowledge that animates MOOCs, by focusing on several discursive functions (constitutive, vehicular and indexical functions). Through a multilingual perspective, they illustrate how these functions are found in MOOCs belonging to different disciplinary fields and having different working languages and subtitles. In conclusion, they discuss how technology can function not only as a tool, but also as a revealer of the complex links between discourse and knowledge.

Tanya McCarthy (Kyoto University) & **Matthew Armstrong** (Kyushu University) present and discuss a type of e-Tandem programme that was set up to promote scientific exchange and improve language and interaction skills in the academic field (poster, conference, terminology, etc.). The article introduces the chosen concept and describes the models that have been adapted by the authors. Finally, a detailed analysis of some extracts of interactions shows the problems encountered by the participants and which the programme seeks to address, but also the evolution of the participants' language skills and fluency.

Based on the observation that digital resources have become easier to access and use in L2 translation courses, **Sara Cotelli Kureth & Alain Kamber** (University of Neuchâtel) report on an action research conducted over two semesters in 2019 with female students of the University of Neuchâtel engaged in English-French and German-French translation courses. Data collected are semi-directed interviews on the representations of 26 learners in relation to the various tools and on their self-declared use, as well as recordings of practices during the end-of-term exam. The comparison of both sets of data enable to identify the criteria used by users to evaluate the tools, how well they know them, the discrepancies between representation and practice, and user profiles. Based on these results, teachers should not only present the different types of digital tools, but also thematise their use.

Monica Masperi (University of Grenoble), **Triscia Biagiotti** (Institut d'Études Politiques de Grenoble) & **Sandra Canelas-Trevisi** (University of Geneva) present a didactic programme developed within the framework of the Innovalangues project for the teaching of listening comprehension in L2, in this case in Italian. This system uses video clips offering related activities to the learners. After situating their approach within the theoretical framework of social-discursive interactionism, the authors explain the criteria for video selection. They then present the characteristics of this device, which works by textual

genres and types of discourse, the tools used explain speech with speech and the concrete implementation of the programme, in 2019-2020 at the University of Grenoble Alpes, with learners of Italian from level B1 to C1 with different profiles. After having used the device independently, these students filled in an opinion questionnaire, the results of which are analysed by the authors.

Acknowledgements

The editorial committee would like to thank the Swiss Academy of Humanities and Social Sciences (SAGW), the Swiss National Science Foundation (SNSF) and the Faculty of Arts and Humanities of the University of Neuchâtel for their support.

Corpus for courses: Data-driven course design

Thomas COBB

Université du Québec à Montréal (UQAM)

Faculté des sciences de l'éducation

1205, rue Saint-Denis, Montréal (Québec) H2X 3R9, Canada

cobb.tom@uqam.ca; www.lexutor.ca

L'apprentissage sur corpus, en tant qu'ensemble de principes et de technologies, joue un rôle bien établi dans l'apprentissage des langues, et cet article montre comment ceux-ci peuvent s'appliquer également à la conception de cours ou de curriculum de langue. Si les supports de cours sont reformulés sous forme de corpus, on a alors à disposition un certain nombre de moyens pour amener la recherche sur l'apprentissage directement dans la salle de classe. Cet article commence par une définition des termes, un examen de la place de l'apprentissage sur corpus dans l'acquisition de la langue, et montre des moyens d'appliquer cette approche à la conception, au testing, et à l'évaluation de l'enseignement des langues. Le contexte est la refonte d'un cours de lecture en cours pour adultes en langue seconde basé sur une collection de documents authentiques trouvés sur Internet. Les principes et technologies de la conception de cours sur corpus ont été utilisés, premièrement, pour exposer les faiblesses de cette collection de matériels en tant que cours, et, deuxièmement, pour montrer des moyens concrets de les améliorer. Tous les logiciels impliqués dans ce travail sont accessibles au public.

Mots-clés:

Apprentissage basé sur les consultations d'un corpus, conception pédagogique, analyse de besoins, vocabulaire.

Keywords:

Corpus-based learning, data-driven learning, instructional design, needs analysis, text analysis, vocabulary.

1. Background & proposition

Data-driven learning (DDL) is an input and comprehension based approach to language learning, but with the proviso that second language (L2) input can be made more comprehensible to learners with its patterns exposed or highlighted by computer programs. An example of a DDL research finding is that new word meanings are inferred more successfully by learners consulting several examples of words assembled by software than by learners consulting single instances over a natural timespan (Cobb 1999). Other types of computer-assisted pattern exposure that are relevant to learning include text-to-speech algorithms that transform written language to spoken; search software that reveals distant and low-frequency collocations in a text (Greaves 2009); discourse tracking software that highlights the coherence threads through a text (Crossley et al 2016); error-tracking software for tagging corpora of learner writing (Granger 2003); corpus frequency lists for vocabulary sequencing and testing (Webb & Nation 2017); and many others. The power of data-driven learning in an array of applications involving corpus consultation by learners is shown in a meta-analysis of empirical studies performed by Boulton & Cobb

(2017). The overall finding of a strong learning effect compared to other approaches is currently being unpacked in a series of finer cut studies (e.g. Lee et al. 2020).

The topic of this chapter, however, is not the learning power of DDL, but rather its potential use in course or curriculum design (hence data-driven course design, DDCD). The idea is that just as learners can profit from the computer's ability to expose patterns in language, so at another level can course designers and teachers.

The first step in DDCD is to assemble learners' inputs or materials into a corpus. A corpus is a large structured text with discernible sub-sections, which is representative of language beyond itself, whether on the scale of an entire language (the British National Corpus, BNC 2001; the Corpus of Contemporary American for U.S. English, COCA, Davis, 2008) or the language used by a defined subset of users. From a modest corpus of classroom materials, combined with insights from research on L2 acquisition, it will be possible to predict whether these materials are useful for these particular learners; what can be done to make them more so; what learning challenges and outcomes can be expected; what supplementary materials will be required; and what if anything will be examinable from what the learners have been exposed to.

The present case is an elaborated example focusing on an existing setting, namely an upper-intermediate remedial ESL reading course for adult Francophone learners in Quebec, Canada. The students were returning to the classroom to obtain high school graduation, which includes basic literacy in English. Their reading course consists almost exclusively of materials found on the internet, which is an increasingly common format for such courses. The typical design of such a course is to read, discuss, and answer questions about a series of such texts, then take a test based on a similar text. The present study stems from a request from the course designers for research-informed assistance to make the course more interesting and its outcomes more reliable. With data-driven tools, it should be possible, for example, to develop more varied content from the texts in the learning phase, and to check the similarity between course and test in the testing phase, and perhaps more.

To limit the topic, the focus of the intervention was primarily on the vocabulary component of the course, and the computer tools are an assortment from the author's website, Lexical Tutor (Lextutor, www.lextutor.ca), a suite of free online text analysis tools mainly relevant to learning English but with adaptation to French and other languages. Lextutor is a reverse engineering of some standard text analysis tools such that they can be (1) easily accessed by practitioners, (2) worked into coherent sequences of application, and (3) used to identify and resolve specific practical problems in language learning.

The plan of the paper is to show readers how to do the following: make and format a corpus; use vocabulary profiling tools in conjunction with vocabulary testing to judge the match between corpus and learners; and deploy a variety of related tools to improve the match between course, learners, and assessment. A sub-goal of the paper is to model one possible way of sequencing text analysis tools coherently. Until now, these have been validated separately and out of context, and, if ever used by practitioners, used separately. Though this is not an empirical study, the paper's hypothesis is that there is a *prima facie* case for designing reading courses as corpora, and the conclusion will include proposals for empirical validation and a discussion of the issues involved.

2. Making & analyzing a corpus

Putting together an Internet selection of reading materials has become easy to do and is an increasingly common practice. The advantages of doing this are obvious: A wide range of text types is readily available, in a size that can be read or re-read in one sitting, of a variety that is unlikely to provoke boredom, and that can be fairly easily matched to learners' environments and interests. Only a small amount of editing is likely to be required, e.g., to remove advertisements from news stories. But putting these texts together as a corpus can reveal some issues that are not obvious.

To become a smoothly functioning corpus, these texts have to be changed into text files (with *.txt* extension) so they do not carry formatting information. Doing this involves either saving the file as text or copy-pasting it into a text document. For the course in question, Fig. 1 shows the text files involved, a de-formatted collection of news stories from the learners' environment. These are then combined in a compressed *zip* file, which maintains individual file identity, using either Mac or PC system software. They are further collected together as a large single text file by a Lextutor routine called *Corpus Builder* (www.lexutor.ca/cgi-bin/tools/corp_build). The corpus thus exists in three formats, as a collection of separate files and in two single file versions, and each is used for different purposes in the procedures that follow. The combined versions appear at the bottom of Fig. 1. The combined corpus consists of 7,198 individual words in 1,460 word families distributed across fifteen 1000-family levels (in the word-frequency scheme of Nation 2016, to be discussed below). With data thus assembled, we are in a position to ask some interesting questions of it.

 Muhammad Yunus.txt	2 KB
 Social Media Booklet.txt	4 KB
 Understanding Poverty.txt	2 KB
 Camouflage pants.txt	4 KB
 Your Clothes.txt	5 KB
 Certification Prep.txt	8 KB
 Protest or threat.txt	3 KB
 Importance of English.txt	8 KB
 Music Couple.txt	4 KB
 Polling station.txt	2 KB
 Child Labour.txt	2 KB
 Women and Vote.txt	3 KB
 Unprepared voters.txt	2 KB
 Charity.txt	2 KB
 corpus.zip	23 KB
 corpus.txt	43 KB

Fig. 1: Reading course as corpus

2.1 Can these learners read these materials?

The first thing to know about this corpus is whether it is readable by our learners, and following that whether it has anything for them to learn.

The primary assumption of the present analysis is that vocabulary knowledge plays a key role in reading comprehension and development (though grammar will also get some attention). A vocabulary focus could be seen as merely reflecting what text analysis happens to be good at, i.e., counting up the bits of text between spaces, which is true, but it is more than that. Of all the components of L2 reading comprehension (like topic familiarity, grammar knowledge, first language distance, first language reading ability, working memory capacity, and others), and despite the overlap between vocabulary knowledge and many of these, vocabulary knowledge reliably predicts the major share of the explained variance in reading comprehension (Bernhardt 2005). Thus readability is largely a matter of knowing vocabulary, with more common words normally being more readable than less common words, owing to the likelihood they have been met and processed more often and in a wider variety of contexts.

A tool for evaluating the vocabulary level of texts and small corpora is *Vocabprofile* (VP; at www.lex tutor.ca/vp/comp). The single-file version of the corpus is used for this step in the analysis. VP traces every word in the corpus

to its frequency rating in a much larger corpus of English (in this case, the BNC-COCA lists developed by Nation 2016). It then puts these ratings together into bands of 1,000-word families, as shown in Fig. 1, where for example in our corpus, first-thousand families account for 82.1% of the total number of individual words, or word tokens. A word family is a headword (*read*) plus all its inflections (*reads*, *reading*) and its most obvious derivations (*reader*), as elaborated in Laufer & Cobb (2019). For brevity, the notation K-1 refers to words in the first 1,000 families (function words, proper nouns, and common lexical words); K-2 to the second (slightly less common words); and so on. The K-levels are arbitrary cut-offs (though they correspond roughly to what an avid learner might acquire in a year, and K-1 is a functional definition of the lexis of everyday speech).

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token (%)
K-1 :	823 (56.2)	1167 (57.52)	5951 (82.1)	82.1
K-2 :	296 (20.2)	412 (20.31)	687 (9.5)	91.6
K-3 :	184 (12.6)	207 (10.20)	311 (4.3)	95.9
Coverage 95				
K-4 :	59 (4.0)	65 (3.20)	76 (1.0)	96.9
K-5 :	36 (2.5)	36 (1.77)	40 (0.6)	97.5
K-6 :	21 (1.4)	23 (1.13)	27 (0.4)	97.9
Coverage 98				
K-7 :	14 (1.0)	14 (0.69)	16 (0.2)	98.1
K-8 :	6 (0.4)	6 (0.30)	6 (0.1)	98.2
K-9 :	8 (0.5)	8 (0.39)	9 (0.1)	98.3
K-10 :	6 (0.4)	6 (0.30)	6 (0.1)	98.4
K-11 :	1 (0.1)	1 (0.05)	1 (0.0)	
K-12 :	2 (0.1)	2 (0.10)	2 (0.0)	
K-13 :	1 (0.1)	1 (0.05)	1 (0.0)	
K-14 :	2 (0.1)	3 (0.15)	4 (0.1)	98.5
K-15 :	1 (0.1)	1 (0.05)	1 (0.0)	
K-16 :				
K-17 :				

Fig. 2: Lexical profile of the course corpus

VP further matches K-levels to the percentages of words known that empirical research has shown to correspond to different comprehension levels. In Fig. 2, 'Coverage 95' is drawn at the point where 95% coverage is reached, K-3 in this case. Empirical research has shown that 95% known-word coverage typically corresponds to ability to read and comprehend with the help of resources (dictionary, Google, etc.). Reading with 95% coverage also enables learners to

acquire some of the remaining unknown vocabulary through contextual inference. 'Coverage 98' is reached at K-6 in the corpus, with 98% known words typically corresponding to independent reading and native speaker inferencing ability without resources (Laufer & Ravenhorst 2010; Schmitt et al. 2011). Extending the framework downward, reading with much less than 90% coverage is likely to be arduous, error prone, and lead to little further vocabulary growth through inference (Hu & Nation 2000). This leaves reading with 85%-95% known-word coverage as the zone where reading development is feasible and instruction justified. Without this minimal knowledge of what the words in a text mean, the usual 'high level' activities of the reading classroom (find the main idea, track the references and transitions, distinguish fact from opinion, articulate the implicit) are not doable (Alderson 1984).

VP analysis is thus interesting, but it would remain somewhat theoretical if our learners' knowledge could not be brought into relation with the corpus profile. In practice, this can be done because the K-levels scheme has been employed in a number of vocabulary tests that fully correspond to the text profile framework. Several such tests, in a variety of formats produced by a variety of researchers, are available in multi-platform versions with score recording at www.lexutor.ca/tests. A pilot sample of 11 of the learners for whom these reading materials had been developed were tested with the most comprehensive of these tests, the Vocabulary Size Test (VST) developed by Nation & Beglar (2007). The VST tests meaning recognition by sampling 10 random words at each K-level, from K-1 (*jump, shoe, stone*) to K-14 (*plankton, skylark, beagle*). The score is a percentage of correct answers at each level. The scores in this case showed that six of the learners were strong all the way from K-1 to K-5 (*devious, threshold, veer*), meaning they could almost certainly read these texts easily (with almost 98% of words known), but with little to gain other than fluency practice, deepening of knowledge for words already known, and possibly a few new acquisitions from the 73 available beyond K-5 (all worthwhile achievements, but beyond the stated brief of this basic course – these learners should have had a different course).

Another two of the learners had strong knowledge at K-1, but less than 50% knowledge across the next three levels, meaning they would have knowledge of about 85% of the words in most of the corpus (K-1 is 82%, Fig. 2). These learners would probably have difficulty reading these texts by themselves, but could benefit from going through them with a teacher; they were in the instructable zone. For learners in the 85-95% known-word zone, these texts offer two important things: adequate or near-adequate contextual support for inferring the meanings of unknown words and plenty of unknown words to learn. The remaining four learners had less than 50% knowledge across the board, from K-1 to K-4, meaning they were in no position to profit from engaging with

these materials. They should have been reading the most basic simplified readers.

To sum up, this course was effectively useless for all but two of the 11 class members that were tested. Would any of this discrepancy be noticed in a classroom without text profiling and vocabulary testing?

2.2 *Can these learners learn anything from these materials?*

From the profiling outcome, it is clear that little word learning or other skill development will result for most of these learners from engaging with these materials. For display purposes, however, let us suppose that the course corpus had been more or less within the learners' range – its vocabulary a bit more basic, or the learners a bit more advanced, so that lexical growth was a possibility for more of them. Even so, words are not learned unless encountered more than once or twice. Is that the case in this corpus?

To answer this question, we turn to another Lextutor routine called *Range for Texts* (www.lex tutor.ca/cgi-bin/range/texts) which looks at word distributions over sets of texts. The zip file version of the corpus is entered into the program and then dissected family by family to see how many of the 14 texts each family appears in. (*Range* will count different family members as recurrences of the same word.) How often will the words in these materials be re-encountered? Empirical research suggests that 10 occurrences are typically needed for even rudimentary learning (Cobb 2007) and many more for productive use.

Fig. 3 shows that for the 1,299 word families present in the course corpus, more than 75% have a range of just one or two out of 14, that is, they appear in only one or two of the texts. Abundant research shows that with such infrequent occurrences they are unlikely to be noticed or learned. Of course, as a reviewer usefully points out, a new word may well appear multiple times in one or two texts, and even if not reinforced in a subsequent text within the course nevertheless establish an initial representation in the lexicon that could be developed at some future point. However, *Range* also tallies frequency data, though not text by text, and the average frequency of the words appearing in these 1 to 2 texts is just 1.8 occurrences (SD=2.15). In other words, the vast majority of word families appear just once per text. At the other end of the distribution, the roughly 18 words with a range of 12 to 14 are function words (*the, of, and*) without lexical meaning. This leaves a potentially learnable vocabulary load (content words encountered in from three to 11 of the texts) of just 305 word families. This contrasts to the 3,000 families needed to read these texts (at 95% coverage, with resources) and the learners' with knowledge of fewer than 1,000.

Such a distribution of learning opportunities is not sufficient for either the weak or the strong learners in the present sample. Strong learners need a rich reading diet to meet the words they do not already know; the VP analysis showed these

learners meeting only 2.5% unknown items from 1,299 families, or 33 families. Weak or beginning learners need to meet high frequency words over and over to start building a lexicon, which *Range* analysis shows can not happen with this corpus. It is unlikely either level of learner would show any difference in vocabulary size if VST-tested before and after this varied diet of texts. (Sequential texts, such as chapters of a book or theme-based selections, typically have far more recurrence across texts than a set of unrelated texts.)

RANGE PROFILE FOR 14 TEXTS

1299 fams, 6,421 tokens

Range	Number fams at this range	% fams at this range	Cumulative percent
range=1	754 fams	58.04	58.04
range=2	222 fams	17.09	75.13
range=3	134 fams	10.32	85.45
range=4	66 fams	5.08	90.53
range=5	42 fams	3.23	93.76
range=6	24 fams	1.85	95.61
range=7	16 fams	1.23	96.84
range=8	9 fams	0.69	97.53
range=9	6 fams	0.46	97.99
range=10	5 fams	0.38	98.37
range=11	3 fams	0.23	98.60
range=12	4 fams	0.31	98.91
range=13	2 fams	0.15	99.06
range=14	12 fams	0.92	99.98

0 items unclassified

Fig. 3: Range profile of the reading corpus

So we see that these materials have a number of limitations. The analysis at this point could proceed to several further aspects of their usability and learnability, but instead turns to data-driven strategies for relieving the problems found in them thus far.

3. Data to the rescue

It would be possible to rewrite these materials substantially to make them match their intended learners' abilities better and offer them better learning opportunities. *Vocabprofile* has an Edit-to-a-Profile facility to help course writers

write words up or down, that is, replace particular words with simpler or more complex synonyms, but this is time-consuming work. However, there are a number of simpler things that can be done to add value to the existing materials.

3.1 Sequencing from easier to harder

The lexical profile presented in Section 2.1 was for the corpus as a whole, but in fact the profiles differ from text to text. *Vocabprofile* offers a profile summary that can be extracted from the analysis one text at a time and then sorted in Excel or other spreadsheet software. Since the texts are in no particular sequence, they might as well be organized from easier to more difficult – easier in the sense of having a greater component of higher frequency, particularly K-1, words. Fig. 4 shows the collected individual profiles of the texts in the corpus with the percentage of words at each K-level from K-1 to K-6 and highlighting the point where 95% (and plausible learnability) is reached. In Text 8 ('Polling Station'), the 95% criterion is reached at K-2; in Text 2 ('Certification'), it is reached at K-6. Like the topics, the lexical sophistication is random.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1																			
2	(1) Camouflage Pants	K-1	87.5		K-2	91		K-3	94.7		K-4	95.3		K-5	95.7		K-6	96	
3	(2) Certification	K-1	82.5		K-2	89.3		K-3	91.1		K-4	93.2		K-5	94.4		K-6	95	
4	(3) Charity	K-1	84.5		K-2	93.1		K-3	94.1		K-4	95.1		K-5	96.4		K-6	97	
5	(4) Child Labour	K-1	81.6		K-2	93.8		K-3	95.4		K-4	96.6		K-5	99.3		K-9	100	
6	(5) Importance English	K-1	88.2		K-2	94.1		K-3	95.3		K-4	96.1		K-5	96.5		K-6	97	
7	(6) Muhammad Yunus	K-1	79.4		K-2	89.7		K-3	90.7		K-4	93.2		K-6	94.7		K-7	96	
8	(7) Music Couple	K-1	88		K-2	94.5		K-3	96.3		K-4	96.7		K-5	97.6		K-6	98	
9	(8) Polling Station	K-1	88.2		K-2	95.4		K-3	96.3		K-6	97.2		K-7	97.7		K-8	98	
10	(9) Protest Threat	K-1	81.1		K-2	86.4		K-3	89.1		K-4	93.8		K-5	95		K-6	95	
11	(10) Social Media	K-1	81.1		K-2	90.4		K-3	92.6		K-4	96		K-5	96.5		K-8	97	
12	(11) Understand Poverty	K-1	88		K-2	94.4		K-3	95.7		K-4	97.4		K-5	99.1		K-6	99	
13	(12) Unprepared Voters	K-1	90.2		K-2	96.3		K-3	96.8		K-4	99.1		K-7	99.6		K-8	100	
14	(13) Women & Vote	K-1	90.8		K-2	95.3		K-3	96		K-4	98.7		K-5	99.2		K-7	99	
15	(14) Your Clothes	K-1	87.5		K-2	94.3		K-3	95.4		K-4	97.3		K-5	97.6		K-6	98	
16																			

Fig. 4: Random arrivals at the 95% point

With this knowledge, it is a simple matter to re-sort the worksheet by Column "C" (K-1 coverage percentage) to obtain a more coherent learnability sequence, as shown in Fig. 5. If desired, a secondary sort could be requested for Column "E" (K-2), and so on, but this was not done here. Once re-sorted, the texts met first will be those with a higher proportion of common vocabulary. Then the numbering can be changed and the materials reassembled. In this case, three of the first four texts encountered meet their 95% criterion with 1,000 word families. This procedure clearly accommodates the weaker learners; it could easily be flipped to accommodate the stronger – starting with the more sophisticated texts, or replacing easier with more difficult texts.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	>>>> better sequence SORTED BY COL "C" (1k COVERAGE %)																		
(13) Women & Vote	K-1	90.8		K-2	95.3		K-3	96		K-4	98.7		K-5	99.2		K-7	99		
(12) Unprepared Voters	K-1	90.2		K-2	96.3		K-3	96.8		K-4	99.1		K-7	99.6		K-8	100		
(5) Importance English	K-1	88.2		K-2	94.1		K-3	95.3		K-4	96.1		K-5	96.5		K-6	97		
(8) Polling Station	K-1	88.2		K-2	95.4		K-3	96.3		K-6	97.2		K-7	97.7		K-8	98		
(7) Music Couple	K-1	88		K-2	94.5		K-3	96.3		K-4	96.7		K-5	97.6		K-6	98		
(11) Understand Poverty	K-1	88		K-2	94.4		K-3	95.7		K-4	97.4		K-5	99.1		K-6	99		
(1) Camouflage Pants	K-1	87.5		K-2	91		K-3	94.7		K-4	95.3		K-5	95.7		K-6	96		
(14) Your Clothes	K-1	87.5		K-2	94.3		K-3	95.4		K-4	97.3		K-5	97.6		K-6	98		
(3) Charity	K-1	84.5		K-2	93.1		K-3	94.1		K-4	95.1		K-5	96.4		K-6	97		
(2) Certification	K-1	82.5		K-2	89.3		K-3	91.1		K-4	93.2		K-5	94.4		K-6	95		
(4) Child Labour	K-1	81.6		K-2	93.8		K-3	95.4		K-4	96.6		K-5	99.3		K-9	100		
(9) Protest Threat	K-1	81.1		K-2	86.4		K-3	89.1		K-4	93.8		K-5	95		K-6	95		
(10) Social Media	K-1	81.1		K-2	90.4		K-3	92.6		K-4	96		K-5	96.5		K-8	97		
(6) Muhammad Yunus	K-1	79.4		K-2	89.7		K-3	90.7		K-4	93.2		K-6	94.7		K-7	96		

Fig. 5: Staged arrivals at the 95% point

3.2 A corpus-based vocabulary supplement

For the very weak learners, however, just meeting words in a more rational sequence will still not be sufficient either for independent reading or further vocabulary growth. These learners need more encounters, such as would be provided by a vocabulary supplement, where new words met in the text are met again in a different or more memorable format or more comprehensible context. From a corpus *Range* can easily assemble a vocabulary list based on both frequency and range, so that words with a high range across the corpus and frequency within the corpus can be extracted for further more targeted work. Fig. 6 is a list of 46 words pulled out of the corpus with the following specifications selected in the checkboxes:

- Not a function word
- Has at least 7 corpus occurrences
- Appears in five texts or more (from 14)

able	age	allow	also	become	better
child	come	country	day	demand	economy
even	family	find	force	give	govern
help	important	issue	last	like	live
need	now	order	part	people	poor
problem	public	reason	say	school	service
show	situation	think	time	use	want
way	world	work			

Fig. 6: A basic data-driven word bank

These are the words that can be considered candidates for input-based or incidental learning. Many of the words will already be known to some class members, but with four out of 11 in the sample group (36%) having little vocabulary at all, they can profit from having their attention drawn to common words (*want, show, way, need, think*). The list will make teachers aware of which common words their learners will be meeting in the readings. For more advanced learners, the corpus can be re-run through *Range* excluding not just function words but all K-1 words. This produces the list shown in Fig. 7, which, though similar to Fig. 6 from a native speaker's perspective, contains a number of interesting challenges for the upper-intermediate Francophone learner. These include words that have no related form in French (*threat, increase, remove*), or have a misleading similarity of form (*union, due, demand*), or present a pronunciation challenge (*develop, economy*). Such a list allows a teacher not just to respond to learners' needs as they arise but to anticipate them. Basic and advanced lists could be used by the teacher or given to the relevant learners themselves, whether as flashcards, example sentences from the course corpus or another corpus, or even as raw lists.

active	advance	create	demand	develop	due
economy	effort	etc	example	extreme	include
increase	individual	industry	interact	labour	organize
politics	population	poverty	provide	province	remove
society	text	threat	union	vote	

Fig. 7: A less basic data-driven word bank from the same corpus

Yet it could be argued that single family headwords (as shown in Fig. 7 and 8) without associated family members in their various morphologies are of limited value. Additional forms of the same word can often be learned for little additional effort (Laufer & Cobb 2019) and thus should be available when or shortly after a new word is encountered. Any headword can be fleshed out to its full family at Fami/Lemmatizer (www.lex Tutor.ca/familizer), which can then be used in, for example, a word bank to appear at the end of the course book. Such a glossary is shown in Fig. 8 for the first few items of Fig. 7.

active actively activism activist activists activities activity inactive inactivity
advance advanced advancement advances advancing
create created creates creating creation creations creative creatively creativity creator creators recreate recreated recreates recreating
demand demanded demanding demands undemanding

Fig. 8: Headwords fleshed out

However, a complete listing of all possible family members may be more than is required or useful, especially in the case of derived forms that are either infrequent (*active*, *activism*) or involve substantial change to the base form or pronunciation (*create*, *creativity*). A better approach to creating a course word bank is to start with a complete vocabulary list and pare it down to a nucleus of just those members that actually appear a certain number of times in the course corpus. *Nuclear List Builder* (www.lex tutor.ca/freq/nuclear, Cobb & Laufer, in press) performs exactly this function, reducing Nation's (2016) BNC-COCA lists by K-level to just those forms found in a corpus (of, e.g., the learners' course materials). A way to tailor this concept and software to these learners' needs would be to draw K-1, K-2, and K-3 nuclear lists out of the full lists and match them to the three populations of learners identified by vocabulary testing. Part of the K-1 version of this list is shown in Fig. 9a, K-3 in Fig. 9b.

These tailored nuclear lists are clearly reduced in both number and size of families, compared to the original lists. Both lists originally comprised 1,000 families, as the 'K' in their names indicates, or 6,862 words including all family members at K-1 and 5,884 members at K-3. The nuclearized K-1 list is 369 families with 479 members; K-3 is 60 families with 68 members.

A further advantage to nuclear lists is that in subsequent reading courses for the same learners, new and probably somewhat overlapping lists could be drawn from their new learning materials, such that the vocabulary component was spiraling through new and familiar items and eventually covering a substantial number of them in a substantial number of occurrences and contexts.

a	age
an	ages
able	underage
ability	ago
about	agree
across	ahead
act	all
action	allowance
actually	allowed
advertisements	almost
afford	along
after	also
again	although
against	always
	amount

Fig. 9a: Part of BNC-COCA K-1 Nuclear

accomplished	communication
addiction	concert
authority	consumer
awards	content
charity	controversy
client	controversial
clients	controversies
communicate	

Fig. 9b: Part of BNC-COCA K-3 Nuclear

Once created, such lists could be used in a number of ways. They might be provided just to teachers, who could then anticipate which words would be most usefully emphasized in discussing the texts or constructing worksheets, or they could be given to students as the basis of a learning activity to supplement input-based learning. For the latter, one approach would be to choose one of the data-driven learning tools featured in Boulton & Cobb's (2017) meta-analysis of DDL tools, such as Lextutor's *Group Lex* (www.lex tutor.ca/group_lex). With this routine, learners can enter, say, 10 words per week from their list into an app on their computers, pads, or mobile phones, accompanied by an example and

a brief meaning, and then quiz themselves on their own and their classmates' entries as they effectively co-construct their own K-1 and K-3 level-appropriate lexicons or dictionaries. Or the teacher can use the software to create whole-class paper quizzes. The phone input and a full-size sample quiz are shown in Fig. 10.

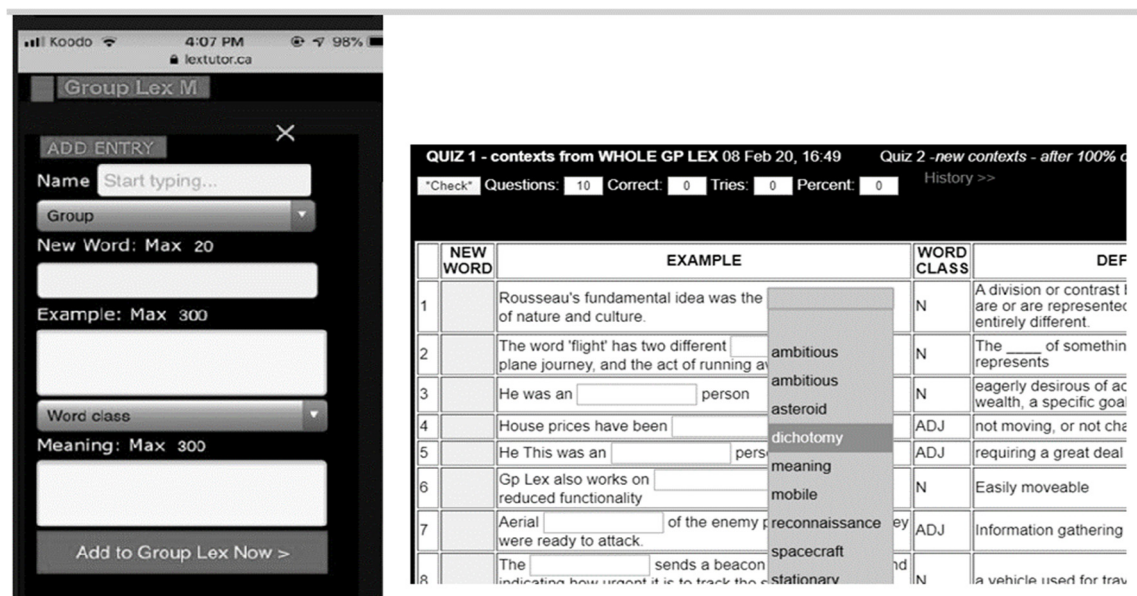


Fig. 10: Lists to lexicons

Another focus on key words as they actually appear in context in the corpus would be to put lexically rich portions of course texts into Lextutor's *Clozebuilder* (www.lex tutor.ca/cgi-bin/cloze/n/) producing an online or paper activity like the one shown in Fig. 11 from the course text 'Your Clothes.' The teacher can select which words to focus on, by principle instead of by guessing, and can add extra resources like dictionary or text-to-speech.

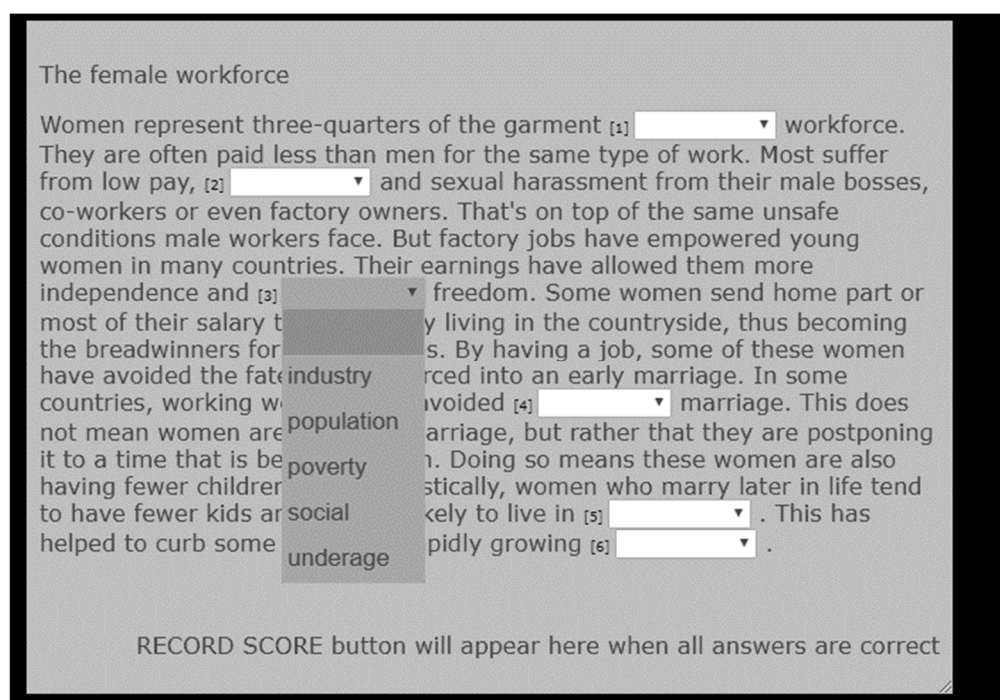


Fig. 11: List words in context

4. Adding a grammar component

While vocabulary knowledge is the main component of reading comprehension, comprehension can also be limited by grammar knowledge. It is worth knowing what grammatical features are present or predominant in the course corpus and planning some course time around them. As is well known, different text types have their own grammatical profiles: news stories often rely on past tense, opinion pieces on present tense, scientific pieces on passive voice, etc. To determine the grammar profile of the corpus, it can be entered as a single file into a Lextutor's *Text Concordance* (www.lex tutor.ca/conc/text/).

This program provides a teacher or course designer with a grammatical snapshot of the corpus as a whole. Frequent words from the listing on the left in Fig. 12a can be clicked to reveal the type of environment each comes from, grammatical as well as lexical. Clicking *been* for example leads to 15 instances of present perfect verbs, of which nine are passives and one continuous, suggesting that both present perfect and passivization could at some point be worth reviewing or anticipating as a source of miscomprehension.

Home > Concordancers > Text Concordance input > Concordance output (Last Tune-Up FEB 2020 ... Concordance index for text: **COURSE CORPUS** With secondary sort *left* of keywo

BECOMING 2	1012.	am off For Montréal natives the pants protest has	BECOME such a regular fixture 1
BEEN 15	1013.	s supposed to be free and equal these issues have	BECOME cause for alarm and Inte
BEFORE 1	1014.	it Moreover many files uploaded to such Web sites	BECOME the Web sites' property
BEGAN 1	1015.	tastes With enough traces of online activities it	BECOMES easy to find who the ci
BEGIN 2	1016.	nformation at our fingertips and our horizons are	BECOMING ever more expanded per
BEGINNING 1	1017.	ry to their family living in the countryside thus	BECOMING the breadwinners for 1
BEHALF 1	1018.	y part of their public image and work It s always	BEEN this intrinsic part of art
BEHAVIOR 1	1019.	ning Issues TEXTS ONLY Situation Social media has	BEEN present since the early 24
BEHIND 1	1020.	such as Facebook Instagram Snapchat etc There has	BEEN an ongoing debate on wheti
BEING 8	1021.	access to certain books or music Some sites have	BEEN shut down due to unaccepte
BELIEVE 2	1022.	rs in such countries as France and Australia have	BEEN revealed Totalitarian gove
BELIEVED 1	1023.	tacks criminal acts and government takeovers have	BEEN documented by everyday cit
BELOEIL 1	1024.	Awards is a true honour for us As a band we have	BEEN very fortunate to receive
BELONG 1	1025.	cturing mining and domestic service Children have	BEEN found harvesting bananas i
BENEFIT 1	1026.	is easier That is the standards my children have	BEEN subjected to in the Quebec
BENEFITS 2	1027.	at wasn't working well The organizers should have	BEEN encouraged by the fact th
BEST 1	1028.	involved in their community and they should have	BEEN treated better The organi
BETTER 10	1029.	ice is outrageous pants Visitors to Montréal have	BEEN scratching their heads at
BETWEEN 2	1030.	face any grievances Unconventional trousers have	BEEN a protest tactic in Montré
BEYOND 1	1031.	sense Up to now public health and safety have not	BEEN threatened said Michel Gr
BIEBER 3	1032.	ildren in conflict zones for which they have now	BEEN recognized with a joint a
BIG 4	1033.	room where we waited in line for almost two hours	BEFORE we could mark our ballot
BIGGER 1	1034.	viable business model In 1 983 the pilot project	BEGAN operation as a full fledg
BILINGUAL 2			
BILINGUALISM 1			
BILL 4			
BILLY 1			

Fig. 12a: Grammar in the course corpus

Home > Concordancers > Text Concordance input > Concordance output (Last Tune-Up FEB 2020 ... Concordance index for text: **COURSE CORPUS** With secondary sort *right* of k

RHETORIC 1	4921.	lly stunting the growth of these kids if they got	RID of the program Perhaps
RID 1	4922.	is instrumental doing so by devaluing English is	RIDICULOUS Quebec industrie
RIDICULOUS 1	4923.	at the old Maple Leaf Gardens in Toronto Sitting	RIGHT behind her was Raine
RIGHT 8	4924.	ing access to resources that are supposedly their	RIGHT to dispose of On scen
RIGHTS 1	4925.	ants If you re a unionized police officer with no	RIGHT to strike you have to
RIGOROUS 1	4926.	appointment Mary Karlie Sand Dorval Women and the	RIGHT to Vote in Canada: An
RISK 1	4927.	en voting in Canada: In 1916 women had earned the	RIGHT to vote in Manitoba E
ROADS 1	4928.	er we said was 1921 It is true that women had the	RIGHT to vote in the 1921 f
ROBERT 1	4929.	ntil 1960 that ALL Canadian women finally had the	RIGHT to vote Just over 50
ROBIN 1	4930.	In the spring of 1918 the government extended the	RIGHT to vote to Canadian w
ROCK 1	4931.	d to monetary gain while others touch on personal	RIGHTS and freedoms The Int
RODEO 1	4932.	intentioned program be assessed in a transparent	RIGOROUS and independent fa
ROLE 1	4933.	t reasonable interest to the poor due to the high	RISK of default Yunus belie
ROLLING 1	4934.	isolated and protected man saw the need to build	ROADS to broadened both phy
ROMANCE 1	4935.	e still not allowed to vote in national elections	ROBERT Borden was Prime Min
ROOM 1	4936.	going to help them in establishing this authority	ROBIN Côté President of She
ROOTS 1	4937.	ght behind her was Raine Maida lead singer of alt	ROCK heavyweights Our Lady
RUE 1	4938.	sts such as the Rolling Stones Justin Bieber Blue	RODEO Nikki Yanofsky Michae
RULED 1	4939.	he Pathways of the Internet are playing a similar	ROLE in today's society Yet
RULING 1	4940.	unts told CharitvVillage that artists such as the	ROLLING Stones Justin Biebe
RULINGS 1			
S 42			
SAFETY 9			

Fig. 12b: Collocation in the course corpus

Any recurring collocation pattern can also be easily identified. Fig. 12b shows that the polysemous word *right* appears in the corpus only in the context of *the right to* do something, but that it never appears in the contexts of *right and left* or *right and wrong*. However, a teacher who was bilingual would probably notice that constructions employing this word are subtly different from ostensible equivalents in French. *Have a right to* in English tends to be followed by a verb, while the similar-looking *avoir droit à* is invariably followed by a noun; for a verb the expression is *avoir le droit de (partir)*, while in English *the right of* prepares for a noun (like *passage* or *way*). *Text Concordance* alerts a teacher of Francophones to think about a worksheet on this.

4. Examination control

The final use for the course corpus that we will discuss here is to assure that the examination text, usually a new text, and the questions about it line up with what the learners have actually read in the course or could reasonably be expected to have inferred from what they have read. Such an assurance is what Biggs (2014) calls 'constructive alignment.' For a language course at this level, alignment means the test should contain only the words and structures the learners had actually encountered in the course, plus some small space for inference if that was part of the training program. Providing such assurance is not only a pedagogical but also an ethical issue. It applies not only to language examinations but to any that involve reading comprehension (like story problems in mathematics or science).

Alignment of language courses and tests, even on the level of words and phrases, is often not achieved and indeed is not simple to achieve without computation. A routine called *Text Lex Compare* (www.lex tutor.ca/cgi-bin/tl_compare/) compares all the words or phrases of an examination text to the collected course texts the learners have worked with. Fig. 13 shows part of a comparison between the course corpus under discussion and a passage that was actually used as the basis for an examination in an earlier run of the course. The output consists of the words that are unique to the course (left column), the words that are shared between course and examination text (middle column), and the words that are unique to the examination text (right column). As shown in the heading information, just under 15% of the word tokens (individual words) or 24% of word families appearing on the test had not previously been met in the course at all. This is a far higher proportion than learners can be expected to work out making inferences from context (native speakers can manage 2%; Schmitt et al. 2011).

New words in second text Index-Edit-Area at bottom
First text: COURSE CORPUS (1504 families)
Second text(s): EXAMINATION (234 families)

TOKEN Recycling Index: (713 repeated tokens : 840 tokens in second/last text) = 84.88%
FAMILIES Recycling Index: (177 repeated families : 234 families in second/last text) = 75.64%

<u>Unique to first(s)</u>	<u>Shared</u>	<u>Unique to second/last</u>
3607 tokens	713 tokens	127 tokens
1327 families	177 families	57 families
001. they 124	001. the 47	001. bus 17
002. as 47	002. i 34	002. emit 9
003. social 36	003. be 33	003. transport 9
004. by 32	004. you 32	004. bicycle 7
005. english 32	005. to 28	005. drive 6
006. vote 32	006. and 24	006. electric 5
007. woman 32	007. car 20	007. compare 4
008. language 28	008. a 19	008. arrive 3
009. site 28	009. for 19	009. bike 3
010. say 23	010. of 13	010. cycle 3
011. police 22	011. gas 11	011. distance 3
012. who 22	012. in 11	012. watch 3

Fig.13: Data-driven test control

From here three remedial interventions are possible. One is to re-write the examination text, focusing on the words in Column 3, either deciding these are acceptable as known cognates with French (*bus, transport, bicycle*) or else glossing/elaborating their meanings inside the text or in a supplementary glossary for non-cognate items (*drive, watch, bike*). Another possibility is to leave 2% of non-cognate items for contextual inference, assuming this has been a focus of the course and is thus testable.

5. Conclusion

The main point of interest in this sequence of DDCCD measures is how few of the patterns discussed would have been observed without the data aggregation of the corpus and the computational tools to look at it. There is too much information in a set of texts to be gleaned by the naked eye; a technology is required to extend the pedagogical sensorium.

The hypothesis of the paper was that a reading course of online texts can be usefully treated as a corpus and investigated then improved with text analysis tools. Readers can judge whether the argument has been made, but for the writer it seems clear that following the procedures outlined above will produce a more successful reading course. This proposition would have to be tested empirically, in the context of its integration in a real course, though many of the procedures have been tested in smaller and more controlled contexts already. Well established research findings include these: Reading instruction is unsuccessful with less than 90% word knowledge; vocabulary growth is weak

without numerous encounters with words; collocational *faux-amis* can impede comprehension; assessment is regularly misaligned with what was taught. These are separate research insights, however, and are only potentially useful until they are integrated into what learners actually do over extended periods of instruction - in other words, into courses. Such an integration is extremely difficult to achieve by individual teachers with a handful of random texts and a roomful of roughly placed learners; it can only be done on a system-wide scale with the aid of instrumentation. Validated instruments are needed to test and place the learners, analyse and sequence the materials, scan materials for obstacles, and correct or supplement them in line with what is known about learning. The particular sequence of tool deployment modeled here – build corpus, profile vocabulary, test learners, re-sequence materials, scan for grammar and collocations, supplement materials, test in line with course characteristics - is probably the most plausible sequence, though others are possible and the software is set up to encourage experimentation.

Would such a redesigned course lead to significant improvements in achievement over the course as originally conceived? It is plausible that it would, because the course would be built from pieces that have already been validated individually. But the pieces would have to be empirically validated again in this larger context, and doing this involves a number of challenges. Course comparison research or any large scale comparison is rarely attempted in applied linguistics, or education research generally, because of the number of variables, the extended time frame, the potential interactions, and the likelihood of confounding. Nevertheless, it is arguably time to expand course evaluation methodologies, which measure courses against objectives, into course comparison methodologies. Most applied linguists would probably agree that research and implementation are out of balance. Without course/curriculum comparisons, large scale implementations of research findings will be slower than they need to be or may never happen at all. The challenge is to maintain experimental rigour while scaling up to real-world settings.

To summarize, evidence-based pedagogy has been difficult to implement at the course or curriculum level, despite the wealth of single-study research results, whether in vocabulary or other areas of language learning. Vocabulary just makes a particularly clear case. Between the research and the learner lies a void that has seemed difficult to fill. Data-driven course design is a strong contender to fill it.

REFERENCES

- Alderson, J.C. (1984). Reading in a foreign language: A reading problem or a language problem? In J.C. Alderson & A.H. Urquhart (eds), *Reading in a foreign language* (pp. 1-27). New York: Longman.
- Bernhardt, E. (2005). Progress and procrastination in second language reading. *Annual Review of Applied Linguistics*, 25(1), 133-150.
- Biggs, J. (2014). Constructive alignment in university teaching. *HERDSA Review of Higher Education* 1, 5-22.
- Boulton, A. & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 65(2), 1-46.
- British National Corpus, version 2 (BNC World). (2001). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Available at: <http://www.natcorp.ox.ac.uk/>.
- Cobb, T. (1999). Applying constructivism: A test for the learner as scientist. *Educational Technology Research and Development*, 47(3), 15-31.
- Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, 11(3), 38-63.
- Cobb, T. & Laufer, B. (In press.) A Nuclear word family list: A list of most frequent family members - base words and derived words. *Language Learning*, 71(3).
- Crossley, S., Kyle, K. & McNamara, D. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*, 48(4), 1227-1237.
- Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 600 million words, 1990-present*. Available at <https://www.english-corpora.org/coca/>.
- Granger, S. (2003). Error-tagged learner corpora and CALL: A promising synergy. *CALICO Journal*, 20(3), 465-480.
- Greaves, C. (2009). *Concgram 1.0: A phraseological search engine*. Amsterdam: John Benjamins.
- Hu, M. & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403-430.
- Laufer, B. & Cobb, T. (2019). How much knowledge of derived words is needed for reading? *Applied Linguistics*, November.*
- Laufer, B. & Ravenhorst-Kalovski, G. (2010). Lexical threshold revisited: Lexical text coverage, learner's vocabulary size and reading comprehension, *Reading in a Foreign Language*, 22, 15-30.
- Lee, H., Warschauer, M. & Lee, J. (2020). Toward the establishment of a data-driven learning model: Role of learner factors in corpus-based second language vocabulary learning. *Modern Language Journal*, 104, 2.*
- Nation, P. (2016). *Making and using word lists for language learning and testing*. Amsterdam, Netherlands: Benjamins.*
- Nation, P. & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31, 9-13.
- Schmitt, N., Jiang, X. & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, 95, 26-43.
- Webb, S. & Nation, P. (2017). *How vocabulary is learned*. New York: Oxford University Press.

* These references can be found online.

Du phonographe aux corpus numériques oraux: aperçu historique de la compréhension du français oral quotidien en FLE

Christian SURCOUF & Alain AUSONI

Université de Lausanne
École de français langue étrangère
Anthropole, 1015 Lausanne, Suisse
christian.surcouf@unil.ch; alain.ausoni@unil.ch

In the 19th century, when the traditional (grammar-translation) methodology prevailed, language learning specialists were already proposing that priority should be given to familiarising learners with the spoken language of daily interactions. In the first part of this article, while briefly reviewing the emergence and development of speech recording technologies, we will show how they received wide approval among advocates of such an approach. But, did this “perfect match” between technology and pedagogy ever take place in language learning classrooms? Not really, as the historical overview of the uses of audio materials (with a special focus on French as a foreign language) presented in the second part of the article indicates.

Keywords:

technology, listening comprehension, spoken French, history, French as a foreign language.

Mots-clés:

technologie, compréhension orale, français parlé, histoire, FLE.

1. Introduction

Parmi les compétences langagières, Porcher (1995: 45-46) considère que la "réception orale est de loin la plus difficile à acquérir" bien qu'elle soit "la plus indispensable". Field (2008: 5) relève également: "it is listening which is arguably the more important since it is listening which enriches the learner's spoken competence with new syntactic, lexical, phonological and pragmatic information". De tels constats devraient faire du développement de la compréhension de l'oral une priorité dans l'enseignement, d'autant plus que les recherches établissent la hiérarchisation suivante en termes quantitatifs pour les locuteurs natifs: écouter > parler > lire > écrire (pour une synthèse, voir Worthington & Fitch-Hauser 2018: 6). Pourtant, selon Vandergrift & Goh (2012: xiii), "listening receives limited attention in many classes", et Carette (2008: 144) déplore également qu'en général, l'enseignement "ne donne aucune ou que peu de place à l'apprentissage de la compréhension orale en tant qu'aptitude à part entière". On peut évoquer plusieurs raisons à cette situation. Du point de vue de la recherche, aucun des modèles proposés (Cornaire 1998: 35-51; Worthington 2018) ne fait l'unanimité ou n'explique clairement les processus



cognitifs impliqués dans la compréhension de l'oral¹, à fortiori lorsqu'il s'agit d'une langue étrangère. Du point de vue pédagogique, cette compétence échappe en grande partie à l'observation directe², ce qui la rend plus difficile à enseigner et à évaluer (Field 2008: 1-2; Weir & Vidaković 2013: 347), conduisant assez naturellement à l'idée, résignée, que les apprenants "apprendrai[en]t par osmose", ne "p[ouvant] faire autrement qu'améliorer leur compréhension, puisqu'ils doivent écouter l'enseignant durant de longues périodes" (Cornaire 1998: 199). Par ailleurs, si la production orale a attiré l'attention des didacticiens depuis plus d'un siècle (voir §3), la compréhension orale n'a pas connu un sort aussi favorable:

Unlike speaking, listening never had its own powerful advocates [...]. In fact, until the late 20th century listening was the skill least practised in the language classroom, the least researched in the literature, and the least understood in the language testing field, and [...] it remains uniquely undervalued to this day. (Weir & Vidaković 2013: 347)

La compréhension orale jouit d'une tradition moins longue et, en classe, avant l'invention puis la commercialisation du phonographe (voir Tab. 1), la production orale de l'enseignant constituait la seule source de compréhension pour l'apprenant (Field 2019a: 1; Marchand 1929: 18, note 6), source qui, en contexte pédagogique, s'avère peu représentative des pratiques langagières quotidiennes des locuteurs natifs, en termes de débit, de complexité syntaxique, de richesse lexicale, etc. (Brown 2011: 63; Chaudron 1988: 54-89). L'apport de la technologie est en ce sens déterminant. En 1889 apparaissaient les premiers cylindres préenregistrés destinés à l'apprentissage des langues vivantes (Benson 2005: 595; Peterson 1974). Deux ans plus tard, le premier numéro de *The Phonogram* évoquait certains de ses avantages pédagogiques:

The phonograph is used in the College of Milwaukee as an aid to the Professor of Languages. It never tires, and can be made to repeat the same phrase or the same word hundreds of times. The teacher, while addressing the class, speaks into the phonograph, which in turn, repeats the lesson as often as desired. (Anonyme 1891: 11)

Et Bonnaffé de s'enthousiasmer:

l'invention d'Edison dont les applications pratiques étaient jusqu'à présent assez limitées, permettra aux générations futures d'apprendre n'importe quelle langue aussi rapidement qu'agréablement, sans avoir à s'expatrier, ni même à quitter le coin de son feu. C'est l'enseignement par l'oreille, comme à l'étranger, et cependant chez soi, dans un fauteuil! (Bonnaffé 1905: 80)

Cette innovation technologique présentait l'avantage de compenser certaines des insuffisances de l'enseignant:

¹ "currently, there is no comprehensive theory that fully explains either the production or comprehension sides" (Vandergrift & Goh 2012: 38).

² Son explicitation ne peut s'effectuer qu'en sollicitant une autre compétence, qu'il s'agisse de production écrite ou orale.

It is evident that we have at our disposal now many of the advantages to be obtained from having a native teacher, including that authoritative quality, both in the form of what is said and the way of pronouncing it, which the best trained American inevitably lacks. (Clarke 1918: 117)

Tout comme la démocratisation de l'accès au livre imprimé avait permis de restituer ailleurs et à une autre époque les pensées d'un individu, le phonographe introduisait dans la classe des productions orales venues d'ailleurs, ouvrant des perspectives pédagogiques auparavant inconcevables sans l'apport de cette première technologie de l'enregistrement sonore, dont nous allons maintenant retracer brièvement l'évolution.

2. Aperçu chronologique des inventions déterminantes du traitement du son

Nature	Date	Technologie	Concepteur	Apport essentiel par rapport à la technologie précédente
MECANIQUE				
	1857	phonautographe	Scott de Martinville	inscription graphique de la parole sur un papier ou un tissu
	1877	phonographe (cylindre)	Edison	restitution du signal sonore
	1887	gramophone (disque)	Berliner	reproduction aisée des disques grâce à la galvanoplastie
ELECTRIQUE				
	1907	amplificateur électronique à tubes	De Forest	amplification du signal sonore
	1919	microphone	Laboratoires Bell	enregistrement électrique, réglage du volume, amélioration de la qualité sonore
	1921	première radio éducative	Université du Wisconsin	
	1925	enregistrements électriques commercialisés	Victor et Columbia	
	1934	magnétophone à bandes	AEG	amélioration de la prise de son, mixage et montage possibles, indexation temporelle (compteur), enregistrement de nouveau possible
	1947	transistor (composant)	Laboratoires Bell	réduction de la taille des appareils, moindre consommation, plus grande autonomie
	1962	magnétophone à cassette	Philips	portabilité, facilité d'usage, cout réduit
NUMERIQUE				
	1977	ordinateur personnel grand public		démocratisation de l'usage de données textuelles et audiovisuelles, cherchabilité, possibilité de statistiques, facilité de reproduction, de transfert, de sauvegarde
	1982	disque compact (CD)	Philips et Sony	réduction du bruit, pas d'usure par frottement
	1984	CD-ROM	Denon et Sony	multimédia (son, image, texte), hypertextualité (liens cliquables),

			interactivité, intégration et exploitation des métadonnées
1989	format mp3	Fraunhofer-Institut	compression, facilité de stockage, de reproduction, de découpage
1990	Internet		"dématérialisation" des données, partage, disponibilité internationale, gratuité
1998	Transcriber	(Barras <i>et al.</i> 1998)	possibilité de transcrire, aligner, annoter un fichier sonore

Tab. 1: Chronologie des quelques événements marquants dans la reproduction technologique et le traitement du son³.

Durant ces 144 années d'existence (jusqu'à aujourd'hui), toutes ces technologies d'enregistrement et de restitution du son et de la voix s'offraient comme outils pédagogiques *potentiels* pour travailler la compréhension de l'oral dans les cours de langue vivante. Mais avant d'envisager leur usage effectif, voyons rapidement en quoi ont consisté les progrès successifs de ces développements technologiques.

Signalons en premier lieu que les premiers phonographes imposaient de fortes contraintes sur l'enregistrement, interdisant de saisir la voix naturelle, comme en témoignent les recommandations suivantes, fournies en 1903 aux apprenants de langue enregistrant leur propre production orale:

Speak loud — louder than in ordinary conversation — but do not yell. Be particularly careful not to move your mouth away from the mouthpiece of the tube while you are recording. [...] Pronounce every sound with perfect distinctness. Final consonants, such as *d*, *t*, *r*, should be pronounced very clearly. You should lay special stress on the sound of the letter *s*, which is very difficult to record well. (International Correspondence Schools 1903: xxviii)

Ces contraintes résultaient de la nature mécanique de l'enregistrement: "les ondes sonores ne se propageaient que par leur énergie propre, et seules les plus fortes parvenaient au diaphragme" (Coeuroy & Clarence 1932: 935). En dehors des problèmes de qualité des premiers enregistrements (voir Marchand 1937: 152), de telles précautions dans la prise de son et la difficulté de manipulation des instruments rendaient difficile l'enregistrement de la langue spontanée. Même si, alternativement, il était possible d'enregistrer des textes préparés – notamment à des fins pédagogiques –, toute approximation ou erreur d'articulation exigeait le réenregistrement intégral du texte prévu⁴. Cependant la prise de son allait être considérablement améliorée et facilitée avec l'amplification électrique et l'utilisation du microphone à partir de 1919. Quinze ans plus tard, en 1934, le magnétophone à bandes autorisera le

³ Plusieurs sources ont été croisées pour établir cette chronologie indicative (Almquist 1987; Benoit et al. 2009; Cuban 1986; Filteau 2006; Kimizuka 2012; Lazard & Mounier-Kuhn 2019; Maisonneuve 2006; Skornia 1962; Tournès 2008), d'autant plus difficile à élaborer que les concrétisations technologiques sont toujours issues de développements plus ou moins longs. Signalons que dans le cadre de cet article, ce tableau n'a aucune prétention exhaustive.

⁴ La durée des premiers cylindres n'excédait pas deux minutes (Maisonneuve 2006: 23).

découpage et le montage. En 1947, l'invention du transistor allait réduire la consommation électrique (Lazard & Mounier-Kuhn 2019: 97) et surtout marquer le tout début de la miniaturisation, conduisant au cours des décennies suivantes à une portabilité accrue. C'est ainsi que sera notamment conçu, en 1962, le magnétophone à cassette, bon marché et facile à utiliser, autorisant une prise de son sur le vif, même à l'insu des locuteurs, par exemple à l'aide d'"un micro-cravate que nous camouflions sous la manche et un magnétophone caché dans un sac en bandoulière, commandé par un interrupteur placé dans une poche" (Luzzati 1985: 39). L'ère numérique constituera la prochaine grande étape fondamentale, plus particulièrement par la démocratisation de l'ordinateur personnel à la fin des années 1970⁵ et, parallèlement, des supports numériques, qui malgré leur taille réduite, pouvaient contenir une grande quantité de données faciles à reproduire, transférer et sauvegarder. Inventé en 1984, deux ans après le CD, le CD-ROM offrira la possibilité de rassembler son, texte, image et vidéo, en alignant le signal sonore sur la transcription (Chafe et al. 1991: 66), et en autorisant la navigation à l'aide de l'hypertextualité et la recherche par mots clés ou opérateurs booléens, fonctionnements que l'on retrouve également dans les corpus oraux, dont l'émergence d'Internet dans les années 1990 (Lazard & Mounier-Kuhn 2019: 256-259) contribuera à faciliter le partage au niveau international. Durant ces décennies de développement, l'informatique aura notamment permis d'effectuer rapidement des calculs de fréquence, de rechercher des chaînes de caractères dans les transcriptions, de procéder à l'étiquetage grammatical automatique massif, et de présenter les résultats à l'aide de concordanciers, ouvrant de nouvelles possibilités pour la linguistique (pour un bref historique, voir McEnery & Hardie 2012: 37-48) et la didactique des langues (pour l'oral, voir Boulton & Tyne 2014: 46-50). Dans la problématique qui est la nôtre, mentionnons également la conception de logiciels d'annotation et de transcription de données audio(visuelles) autorisant l'alignement texte-son (par ex. Transcriber, Barras et al. 1998; Praat, Boersma & van Heuven 2001; Elan, Brugman & Russel 2004; Winpitch, Martin 1999) et la création de corpus oraux comme C-ORAL-ROM (Cresti & Moneglia 2005), CFPP (Branca-Rosoff et al. 2011), Clapi ou PFC (voir 3.3).

Comme l'illustre cet aperçu chronologique, depuis l'avènement des premiers enregistrements sonores à visée pédagogique il y a plus de 130 ans, les progrès technologiques ont permis, au moins depuis un siècle, d'enregistrer et de restituer la langue parlée de manière tout à fait satisfaisante et, pour cette raison, auraient pu servir de soutien pédagogique pour travailler la compréhension orale dans l'enseignement-apprentissage des langues. Mais qu'en est-il exactement? Les potentialités de la technologie ont-elles été exploitées pour servir cette finalité?

⁵ En 1977, l'Apple II, le PET Commodore et le TRS-80 "totaliseront des millions de vente et lanceront l'industrie de l'ordinateur personnel" (Lazard & Mounier-Kuhn 2019: 218).

3. Développements technologiques et usages des documents sonores dans l'enseignement-apprentissage des langues vivantes étrangères

En 1855, dans ses *Premiers principes d'éducation avec leur application spéciale à l'étude des langues*, Marcel (1855: 270), inspiré par le déroulement de l'acquisition de la L1, préconise de commencer par mettre l'accent sur ce qu'on appellerait aujourd'hui la *compréhension orale*, proposant que l'apprenant passe successivement "de l'intelligence de la langue parlée à l'art de parler, de l'art de lire à l'art d'écrire". À la fin du XIX^e siècle, la revue *le Maître phonétique* arborait sur sa couverture six principes pédagogiques fondamentaux, dont le premier confère à la langue parlée un rôle crucial: "ce qu'il faut étudier d'abord dans une langue étrangère, ce n'est pas le langage plus ou moins archaïque de la littérature, mais le langage parlé de tous les jours" (Passy & Rambeau 1897: 3). En somme, même durant la période de "plein épanouissement" de la méthodologie traditionnelle (grammaire-traduction) (Besse 1985: 27; voir aussi Howatt 2006: 642), des didacticiens proposaient déjà de familiariser prioritairement les apprenants à la langue parlée quotidienne, et les développements rapides des technologies de l'enregistrement de la parole ne pouvaient qu'enthousiasmer les promoteurs d'une telle approche. Désormais, des productions orales extérieures à la classe et présentant une certaine diversité de situations quotidiennes de communication allaient pouvoir compléter les interventions orales ou les exercices de lecture à voix haute de l'enseignant et des apprenants. Mais ce mariage entre technologie et pédagogie a-t-il vraiment eu lieu? Pas véritablement, comme le montrera notre rapide survol historique de la question. En effet, longtemps érigés comme modèles phonétiques et grammaticaux à imiter et reproduire, les documents sonores n'ont guère servi à travailler "l'intelligence de la langue parlée" prônée par Marcel (1855: 270). Et, bien que la compréhension orale ait été reconnue comme activité à part entière depuis l'avènement de la méthodologie audiovisuelle vers la fin des années 1950, les documents sonores utilisés restent pour la plupart fabriqués et interprétés par des comédiens sur une base écrite (Ravazzolo *et al.* 2015: 111), et pour cette raison s'avèrent peu représentatifs du "langage parlé de tous les jours", auquel voulaient sensibiliser Passy & Rambeau (1897: 3).

3.1 Les documents sonores, une nouveauté complémentaire

Pour convoquer des acteurs de l'utilisation précoce du phonographe à des fins pédagogiques et donner à voir les défis et les résistances suscités par l'utilisation de cette nouvelle technologie dans l'apprentissage des langues, arrêtons-nous sur deux témoignages rétrospectifs de pédagogues, parus en 1936 et 1937 dans *Les Langues modernes*. Constatant que deux-mille phonographes sont en service pour l'enseignement des langues vivantes dans les collèges et lycées français, Varenne (1936: 290) tient à honorer "le courage qu'il fallut aux précurseurs [...], il y a plus de trente ans, pour tenter à leurs

risques et périls et avec leurs seules ressources, l'expérience maintenant banale de l'emploi du disque dans l'enseignement". En plus des réticences de l'administration scolaire, qui "se méfiait et jugeait assez ridicule une fantaisie qui introduisait dans la classe ce qu'elle croyait encore un vain amusement dispensant de tout travail sérieux" (Varenne 1936: 291), ces pionniers devaient faire face à des contraintes pratiques, évoquées à travers le souvenir de la vision d'un collègue "traversant la cour, précédé ou suivi d'un élève portant la boîte du phono, d'un autre brandissant l'énorme pavillon nécessaire aux auditions, d'un troisième chargé des disques" (Varenne 1936: 291). En réponse à l'article de Varenne, Marchand (1937: 152) revient sur son propre usage pédagogique du phonographe en classe d'allemand au début de l'année 1900, dès l'apparition des premiers phonographes à cylindres, quand "le phonographe était dans l'enfance, les cylindres fragiles, les enregistrements nasillards et souvent défectueux". Il avait bravé sa hiérarchie pour apporter en classe, avec son phonographe, "un cylindre que j'avais enregistré moi-même et un diaphragme de mon invention que j'avais perfectionné avec un saphir monté sur une moitié de noisette" (Marchand 1937: 152). Que tenait-il tant à faire écouter à ses élèves? Sa lecture d'un poème étudié la veille. Ce furent, plus tard, les enregistrements des dialogues et des chansons de ses manuels qui, grâce à de fréquentes séances de "répétitions phonographiques" (Marchand 1937: 154), devaient permettre le travail de la prononciation et l'acquisition de nouvelles structures langagières.

Comme en témoignent ces exemples, la compréhension de l'oral n'était pas au centre des préoccupations didactiques de l'époque. Au sein de la méthodologie dite *directe*, telle qu'elle se déploie notamment dans différentes instructions officielles françaises parues entre 1890 et 1908, l'importance donnée à l'approche orale se justifiait "par la nécessité d'assurer en premier lieu la maîtrise de la prononciation", le phonographe faisant alors office de "répétiteur phonétique" (Puren 1988: 235). Conservant largement cette fonction entre 1920 et 1960, période de prédominance d'une méthodologie parfois dite *active* – cherchant à opérer une forme de synthèse des méthodologies traditionnelle et directe (Puren 1988: 235-237 & 213) –, le phonographe, le gramophone, puis les magnétophones à bandes et à cassettes furent avant tout utilisés pour présenter de nouvelles structures que les apprenants devaient ensuite mobiliser en production (Field 2008: 1). La priorité de la langue orale en classe n'était en réalité le plus souvent qu'une "priorité chronologique", et le dialogue donnant forme au cours de langue "un mode pédagogique d'assimilation des formes écrites" (Puren 1988: 238). Dans ce contexte, la technologie de restitution sonore avait essentiellement un rôle de complément, ce qui explique qu'au milieu des années trente, par exemple, on ait pu concevoir le phonographe "comme un moyen auxiliaire qui ne devrait pas modifier les principes et la pratique de notre pédagogie" (Fouret 1936: 66).

3.2 Enseignement audio-oral ou compréhension de la langue parlée?

Selon Field (2019b: 283), l'enseignement de la compréhension orale en tant que telle n'a commencé qu'à la fin des années 1960. En Europe, cette période est caractérisée par l'influence de la méthodologie audiovisuelle française, dite SGAV (structuro-globale audiovisuelle), dont l'avènement est intimement lié à la disponibilité du magnétophone à bande et du film fixe (Puren 2004: 247). De nombreux ensembles pédagogiques SGAV se sont basés sur les résultats du *Français fondamental* (Gougenheim et al. 1964)⁶, première analyse systématique du français parlé à visée pédagogique, établie au fil des années 1950 à partir d'un corpus de 75 heures de conversations enregistrées en grande partie sur des disques de papier magnétique de six minutes à l'aide d'appareils Recordon:

Pour la première fois, en France, de manière extensive, on recueille sur bande magnétique des conversations. De l'oral à l'état pur et non pas un texte dit, oralisé. Et ce corpus va permettre d'établir une table de fréquence du vocabulaire français [...]; d'esquisser le classement des structures syntaxiques les plus employées; de proposer une pédagogie à base d'audition/articulation où le dialogue, en oral, est point de départ, où l'image prend fonction de code adjuvant. Le but visé étant que l'étudiant s'exprime d'abord et qu'il parle. L'oral est au cœur de l'enseignement de la langue. (Peytard 1977: 194)

Une telle approche allait considérablement influencer l'étude de la langue parlée et constituait "à plus d'un titre un événement marquant de la didactique des langues et de la didactique moderne du français langue étrangère" (Cortier 2006: §1). Cependant, à l'apport technologique s'opposait l'inertie des représentations sur la langue parlée, même parmi les initiateurs du *Français fondamental*:

En prenant connaissance des premiers enregistrements recueillis, Sauvageot déclara péremptoirement et jusque dans sa famille que "les gens cultivés" s'exprimaient tout autrement et beaucoup mieux, même dans leurs conversations les plus familières. C'est alors que son fils Serge [...] s'empara du magnétophone paternel et le cacha sous la table au cours d'un repas animé qui réunissait autour de son père plusieurs linguistes renommés, et enregistra les conversations. [...] Sauvageot, beau joueur mais très perplexe, nous apporta cet enregistrement à Saint-Cloud, et chacun dut faire son mea-culpa et réviser ses hypothèses de linguiste. (Rivenc 2006: §14)

En dépit de ce mea-culpa, les réticences initiales ne sont pas restées sans conséquences sur le projet lui-même. Même si elle prétendait fonctionner "plus rationnellement" que ses prédécesseurs en tenant compte notamment des "dépouillements grammaticaux", de la "liste générale des fréquences du français parlé" et de l'"étude statistique des formes verbales" (Gougenheim et al. 1964: 215), la commission en charge de la création de *La Grammaire du français fondamental (1^{er} degré)* décida d'écarter certains traits de la langue parlée:

⁶ Pour une liste d'ouvrages pédagogiques faisant appel au *Français fondamental*, voir Gougenheim et al. (1964: 298-300) et pour une analyse des phases de la leçon audiovisuelle dans les différentes éditions du premier manuel SGAV *Voix et images de France* (1958, 1961, 1971), voir Besse (1985: 59-97).

De nombreux indices existaient bien dans la liste des fréquences, mais ils ne furent pas retenus, ni exploités. Par exemple:

[...]

- l'effacement du *ne* dans l'expression familière de la négation [...], rejeté par la commission;

- l'emploi systématique de *on* à la place de *nous*, ("Michèle et moi, on est partis à 9 h.") rejeté lui aussi comme "incorrect";

[...]

- le rejet systématique par la Commission de toutes les marques verbales de l'affectivité dans les dialogues: *ben! bon! eh ben! ma foi! quoi! bref! d'ailleurs!* (Rivenc 2006: §33)

Pour négligeable qu'elle puisse paraître, cette forme d'expurgation n'en indique pas moins que, même dans le cadre d'une démarche novatrice appuyée sur des développements technologiques favorisant l'analyse de la langue parlée, "les modèles de référence restaient bien, secrètement mais fermement, ceux de la langue écrite" (Rivenc 2006: §36).

S'ils n'ont pas reproduit les enregistrements de l'enquête, dont on n'a semble-t-il jamais imaginé qu'ils puissent servir une fonction pédagogique directe, et qui ont d'ailleurs été détruits après leur exploitation statistique (Gougenheim et al. 1964: 215), de nombreux manuels SGAV (voir note 6) ont tiré profit des résultats du *Français fondamental* pour la sélection et la présentation des mots et des structures en fonction de leur fréquence d'usage et de leur disponibilité. Le manuel *Voix et images de France* s'inscrit dans ce courant. Même si les auteurs expliquent dans leur préface qu'ils ont "cherché à enseigner, dès le début, la langue comme un moyen d'expression et de communication faisant appel à toutes les ressources de notre être: attitudes, gestes, mimiques, intonations et rythmes du dialogue parlé" (Credif 1971: 23), Moirand (1974: 8) juge "les dialogues peu naturels, malgré un certain effort pour introduire les mots et les expressions relevant de la fonction émotive du langage", arguant qu'il "s'agit souvent d'un récit (ou même d'une description) artificiellement transformé en dialogue et non d'un échange véritable entre des locuteurs désirant communiquer". Bien qu'ils reconnaissent l'ambition originelle des concepteurs de tels manuels d'"enseigner la langue commune, comprise et pratiquée par la majorité des Français dans la communication orale usuelle", Porquier & Vivès (1974: 116) concluent également qu'"il ne suffit pas d'une situation vraisemblable pour donner vie à un dialogue artificiel". La technologie permettait déjà d'enregistrer sans difficulté des documents sonores reflétant cette "communication orale usuelle". Mais la rencontre entre potentialités technologiques et enseignement de la langue parlée n'aura pas lieu. À titre d'illustration, observons l'extrait de dialogue suivant provenant de *Voix et images de France*:

- [1] M. Thibault: Nous sommes dans la salle à manger / Venez Jean.
 Jean: Avec plaisir. Vous êtes bien ici! / Vous regardez des photos?
 M. Thibault: Nous regardons des photos de vacances.
 F. Thibault: Ici, voilà nos enfants; / ils courent, / ils jouent au ballon.
 Jean: Je ne vois pas leur ballon.
 F. Thibault: Non, il n'est pas sur la photo. (Credif 1970: 18)



Fig. 1: Images des films fixes de la partie "Mécanismes" de la leçon 8 de *Voix et images de France* (Credif 1960: 41-42).

Outre le décalage manifeste entre la situation de communication et l'allure de l'échange langagier, notamment au niveau de l'enchaînement discursif, remarquons simplement, parmi les caractéristiques du français parlé que Rivenc (2006: §33) déplorait d'avoir écartées du *Français fondamental* (voir plus haut), l'utilisation de *nous* à la place de *on* et le maintien du *ne* dans les phrases négatives, rendant le dialogue statistiquement – contre l'entreprise du *Français fondamental* – peu représentatif du français parlé de l'époque.

Alors qu'autour de 1970, les débats méthodologiques en FLE se focalisent sur la question de l'intégration des documents authentiques (Cuq 2003: 29), un second grand corpus oral conçu pour "être valable à la fois pour l'application pédagogique et pour la recherche sur la langue parlée" (Blanc & Biggs 1971: 16) finit d'être collecté par des chercheurs britanniques: l'Étude sociolinguistique sur Orléans (ESLO). Pour ses concepteurs, il fallait familiariser les apprenants au français parlé tout en les émancipant de "ce langage spécialisé qu'ils ont appris de leurs textes écrits" et sans nécessairement les "envoyer [...] en France, pour un séjour prolongé" (Blanc & Biggs 1971: 22). L'apport de la technologie s'avérait déterminant:

la technique de l'enregistrement sur bande magnétique a transformé profondément les possibilités pour l'apprentissage et l'étude de la langue parlée [...] nous avons maintenant la possibilité [...] non seulement d'écouter mais, pour la première fois, d'étudier aussi la langue étrangère. (Blanc & Biggs 1971: 22)

En résultera le manuel *Les Orléanais ont la parole* (Biggs & Dalwood 1978), présentant un français parlé différant largement de celui des dialogues fabriqués de l'approche audiovisuelle. En témoigne le début de l'échange de la première leçon⁷:

- [2] – quelqu'un est d[] passage à Orléans – il[] vous d[]mande c[] qu'il[] pourrait faire pendant son séjour: et qu'est-ce que vous lui dites
 – ah ben alors là on aborde un peu l[] tourisme quoi –
 – oui
 – ben euh: il[] y a pas mal de choses à voir d'abord euh – euh: – la euh il[] y a une région la Sologne qui est très jolie – c'est c'est la la plus belle région que j[] trouve – dans. dans la région si on veut (Biggs & Dalwood 1978: 19)

On remarquera d'emblée la présence d'hésitations (*euh, c'est c'est la la plus belle région*), de marqueurs discursifs (*ben, alors, quoi*), de *il* prononcés [i], du présentatif existentiel *il y a X qui*, etc. Les auteurs profitent de ces documents sonores authentiques pour sensibiliser les apprenants aux caractéristiques du français parlé, au travers de textes lacunaires, d'exercices consistant à retrouver "de mémoire" en les oralisant "certaines expressions familières" remplacées dans le texte de l'exercice "par des tournures de la langue écrite" (par ex. "Il y a *bien* des choses à voir" à reformuler en "il y a *pas mal* de choses à voir") (Biggs & Dalwood 1978: 19). Mais, là encore, leur entreprise, novatrice, suscite des réticences parmi les enseignants:

Certains professeurs anglais qui ont écouté pour la première fois quelques extraits en ont été scandalisés. C'est que nous enseignons, même par les cours audiovisuels, un français ou tout est en place, comme un scénario perfectionné – alors que la langue parlée [...] possède une syntaxe [...] très différente de celle qui est extraite et raffinée pour les manuels traditionnels. (Blanc & Biggs 1971: 24)

La réaction scandalisée de ces enseignants met une fois de plus en évidence les représentations négatives face au français parlé. Rien d'étonnant en somme

⁷ Pour plus de lisibilité, les lettres originellement biffées dans le manuel sont ici encadrées. Elles indiquent les élisions et les réductions.

à ce que les manuels basés sur l'ESLO n'aient eu "finalement que peu d'impact au niveau éditorial" (Delahaie 2013: 99-100).

Si l'utilisation du magnétophone s'est révélée déterminante pour la collecte de données authentiques du français parlé⁸, sans qu'elles n'aient pour autant un impact notable en didactique, l'entrée dans l'ère de la démocratisation du numérique allait-elle offrir une nouvelle chance au français parlé dans l'enseignement-apprentissage du FLE?

3.3 *Les apports du numérique*

Comme le montre, parmi d'autres, le cas de l'ESLO, didactique des langues et réflexion linguistique entretiennent des relations étroites, que les avancées technologiques ne feront que renforcer durant l'ère numérique. En témoignent les débouchés pédagogiques qu'ont connus des corpus de recherche sur le français oral comme Clapi ou PFC, pouvant désormais servir à l'enseignement-apprentissage du FLE: Clapi-FLE (Ravazzolo & Etienne 2019) et PFC-EF (Detey et al. 2009). De telles évolutions puisent dans le développement et la démocratisation de l'ordinateur personnel depuis les années 1980 (voir Tab. 1), la possibilité de numériser, éditer, annoter, reproduire, stocker et partager facilement des données audiovisuelles (notamment via Internet), ouvrant des horizons insoupçonnés jusqu'alors pour la linguistique et la didactique des langues. Ainsi en est-il de la constitution de bases de données massives, interrogeables à volonté à l'aide d'interfaces de plus en plus intuitives, et pouvant présenter la langue orale en alignant transcription, annotations et signal sonore. Si ces innovations logicielles ont avant tout servi la linguistique de corpus, elles peuvent néanmoins répondre à des attentes pédagogiques concernant l'enseignement-apprentissage de la langue parlée et plus particulièrement, pour ce qui nous concerne ici, le travail sur la compréhension de l'oral.

De nouvelles possibilités technologiques développées à des fins de recherche rejoignent ainsi des préoccupations didactiques plus ou moins anciennes comme l'intégration et l'usage des documents authentiques oraux en FLE (Chambers 2009), l'autonomisation de l'apprenant (Boulton & Tyne 2014: 16-21), la nature inductive de l'apprentissage – déjà présente dans la méthode directe –, le rôle qu'y jouerait plus particulièrement la compréhension orale, etc. S'est ainsi progressivement constituée et théorisée à partir des années 1990 une nouvelle approche didactique fondée sur les outils de la linguistique de corpus. L'apprenant, investi d'un rôle de "chercheur" (Boulton & Tyne 2014: 76; Johns 1986: 160) devait formuler des hypothèses, et dans une démarche inductive "mettre à profit ses différentes observations de la langue à partir de données qui se présentent sous forme de corpus" (Boulton & Tyne 2014: 6) et

⁸ Sur l'intérêt de l'usage des documents authentiques et de leur utilisation en FLE, voir Holec (1990).

dont les résultats s'affichaient dans un concordancier (Johns 1986). Naissait l'*apprentissage sur corpus* ou *data-driven learning*.

Si tout corpus oral peut à priori nourrir des démarches d'apprentissage, ces dernières années ont vu naître la création d'interfaces pédagogiques ciblant les apprenants du FLE. C'est le cas notamment de PFC-EF, CLAPI-FLE, FLEURON et FLORALE⁹. Alors que PFC se proposait de "décrire la prononciation du français dans sa diversité géographique, sociale et stylistique" sur la base d'"un vaste corpus de français parlé à travers le monde" (Detey et al. 2009: 226), sa déclinaison pédagogique, PFC-EF, avait pour "but initial [...] de donner accès à des ressources linguistiques, pré-pédagogiques [...] accessibles, en particulier via un moteur de recherche qui permet d'extraire, selon plusieurs paramètres, des unités textuelles alignées au son, tout en offrant la possibilité [...] de la resituer dans son contexte, avec possibilité d'étendre la récupération des contextes gauche et droit" (Detey et al. 2009: 232). Originellement conçu pour "l'étude des interactions", le projet CLAPI "recueille, documente et transcrit depuis plus d'une quarantaine d'années des situations sociales ordinaires dans une grande variété de contextes" (Ravazzolo & Etienne 2019: §1). Dans la lignée de ces préoccupations a été créée la "plateforme CLAPI-FLE dédiée à l'apprentissage des interactions" (2019: §1), permettant d'accéder à des "extraits vidéo et audio, téléchargeables, contextualisés et décrits par une fiche pré-pédagogique, didactisés avec quelques pistes d'exploitation, transcrits avec différents niveaux possibles d'annotations, dotés d'un lexique et d'un dictionnaire en ligne" (2019: §18). Si FLEURON recourt également à des vidéos, contrairement à PFC-EF et CLAPI-FLE, il a dès l'origine été conçu comme "dispositif d'apprentissage du français [...] destiné à des étudiants ou futurs étudiants étrangers qui souhaitent faire un séjour universitaire en France", et se compose "de ressources multimédia de situations de communication authentiques et d'un concordancier, permettant des allers-retours entre les occurrences recherchées et les ressources" (André 2016: 70). Plus récent et également conçu à des fins pédagogiques dès son origine (Surcouf 2020; Surcouf à paraître; Surcouf & Ausoni 2018), FLORALE permet d'accéder à plusieurs milliers de courts extraits sonores (moins de 4", et extensibles à souhait) d'entretiens et de documentaires radiophoniques pris sur le vif, dont la transcription met en évidence – via un code de couleurs – plus de 200 caractéristiques phonétiques, syntaxiques, discursives et lexicales du français parlé quotidien, annotées manuellement dans le corpus.

L'apprentissage informatisé sur corpus occupe toutefois une place marginale en didactique des langues (Boulton 2017: 483). Et l'oral y est de surcroît très

⁹ <https://www.projet-pfc.net/le-projet-pfc-ef/ressources-didactiques/fichespedago/>; <http://clapi.ish-lyon.cnrs.fr/FLE/accueil.php>; <https://fleuron.atilf.fr/>; <https://florale.unil.ch/>.

minoritaire (Boulton & Tyne 2014: 46), à fortiori lorsqu'il s'agit de travailler la compréhension orale...

4. Présentation ou représentation?

Véritable révolution anthropologique, l'invention du phonographe en 1877 a constitué la première étape d'une série d'innovations technologiques sans précédent permettant d'enregistrer, d'écouter et d'analyser scientifiquement la langue parlée comme jamais auparavant. Cependant, linguistique et didactique n'ont pas tiré le même profit du potentiel offert par ces avancées. Bien que la didactique des langues se soit très tôt approprié l'usage de l'enregistrement sonore (1889), les représentations sur la langue parlée restent empreintes des normes de l'écrit (voir à cet égard l'analyse des dialogues de manuels de FLE dans Giroud & Surcouf 2016; Surcouf & Giroud 2016; Vialleton & Lewis 2014). Pourtant, l'enregistrement sonore est la seule technologie permettant de capturer la langue parlée dans sa matérialité et dans la diversité de ses fonctionnements, au risque de "scandalis[er]" et de "boulevers[er] les idées reçues" (Blanc & Biggs 1971: 24) des locuteurs imaginant "s'exprim[er] tout autrement et beaucoup mieux, même dans leurs conversations les plus familières" (Rivenc 2006: §14). Les corpus informatisés de français parlé se prêtent particulièrement au travail de la compréhension orale. Reste toutefois à savoir si, grâce à leur usage, la présentation *de* la langue parlée parviendra à triompher des représentations *sur* la langue parlée...

BIBLIOGRAPHIE

- Almquist, S. G. (1987). Sound recordings and the library. *Occasional papers*, 179, 1-37.
- André, V. (2016). FLEURON: Français Langue Étrangère Universitaire – Ressources et Outils Numériques. Origine, démarches et perspectives. *Mélanges Crapel*, 37, 69-92.
- Anonyme (1891). The phonograph in teaching. *The Phonogram*, 1(1), 11.
- Barras, C., Geoffrois, E., Wu, Z. & Liberman, M. (1998). Transcriber: a free tool for segmenting, labeling and transcribing speech. *First international conference on language resources and evaluation (LREC)*, 1373-1376.
- Benoit, S., Blouin, D., Dupont, J.-Y. & Emptoz, G. (2009). Chronique d'une invention: le phonautographe d'Édouard-Léon Scott de Martinville (1817-1879) et les cercles parisiens de la science et de la technique. *Documents pour l'histoire des techniques. Nouvelle série*, 17, 69-89.
- Benson, C. (2005). Language instruction records. In F. W. Hoffmann & H. Ferstler (éds.), *Encyclopedia of recorded sound* (p. 595). New York: Routledge.
- Besse, H. (1985). *Méthodes et pratiques des manuels de langue*. Paris: Didier.
- Biggs, P. & Dalwood, M. (1978). *Les Orléanais ont la parole (Arbeitsbuch)*. München: Langenscheidt-Hachette.
- Blanc, M. & Biggs, P. (1971). L'enquête socio-linguistique sur le français parlé à Orléans. *Le français dans le monde*, 85, 16-25.
- Boersma, P. & van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glott International*, 5(9/10), 341-347.

- Bonnaffé, É. (1905). L'enseignement des langues par le phonographe. *Le Magasin pittoresque*, 80-81.
- Boulton, A. (2017). Corpora in language teaching and learning. *Language Teaching*, 50(4), 483-506.
- Boulton, A. & Tyne, H. (2014). *Des documents authentiques aux corpus. Démarches pour l'apprentissage des langues*. Paris: Didier.
- Branca-Rosoff, S., Fleury, S., Lefeuve, F. & Pires, M. (2011). Constitution et exploitation d'un corpus de français parlé parisien. *Corpus*, 10, 81-98.
- Brown, S. (2011). *Listening myths. Applying second language research to classroom teaching*. Ann Arbor: University of Michigan Press.
- Brugman, H. & Russel, A. (2004). Annotating Multi-media/Multi-modal Resources with ELAN. *LREC 2004. Fourth International Conference on Language Resources and Evaluation*. Lisboa: European Language Resources Association.
- Carette, E. (2008). Mieux comprendre l'oral: formation des formateurs. *Le français dans le monde. Recherches et applications*, 43, 144-157.
- Chafe, W. L., Du Bois, J. W. & Thompson, S. A. (1991). Towards a new corpus of spoken American English. In K. Aijmer & B. Altenberg (éds.), *English corpus linguistics. Studies in honour of Jan Svartvik* (pp. 64-82). London: Longman.
- Chambers, A. (2009). Les corpus oraux en français langue étrangère: authenticité et pédagogie. *Mélanges Crapel*, 31, 15-33.
- Chaudron, C. (1988). *Second language classrooms: research on teaching and learning*. Cambridge: Cambridge University Press.
- Clarke, C. C. (1918). The phonograph in modern language teaching. *The Modern Language Journal*, 3(3), 116-122.
- Coeuroy, A. & Clarence, G. (1932). Le phonographe en 1932 (I). *La Revue de Paris*, 5, 930-943.
- Cornaire, C. (1998). *La compréhension orale*. Paris: CLE International.
- Cortier, C. (2006). De quelques enjeux et usages historiques du Français fondamental - Présentation. *Documents pour l'histoire du français langue étrangère ou seconde*, 36, en ligne.
- Credif (Centre de recherche et d'étude pour la diffusion du français) (1960). *Voix et images de France. Livre de l'élève*. Paris: Didier.
- Credif (1970). *Voix et images de France. Première partie, Texte des dialogues*. Paris: Didier.
- Credif (1971). *Voix et images de France. Livre du maître*. Paris: Didier.
- Cresti, E. & Moneglia, M. (éds.) (2005). *C-ORAL-ROM: integrated reference corpora for spoken Romance languages*. Amsterdam/Philadelphia: John Benjamins.
- Cuban, L. (1986). *Teachers and machines: the classroom use of technology since 1920*. New York: Teachers College Press.
- Cuq, J.-P. (éd.) (2003). *Dictionnaire de didactique du français langue étrangère et seconde*. Paris: CLE international.
- Delahaie, J. (2013). Constitution et exploitation de corpus d'interactions verbales pour le FLE: problèmes et programme. *Linx*, 68-69, 95-114.
- Detey, S., Lyche, C., Tchobanov, A., Durand, J. & Laks, B. (2009). Ressources phonologiques au service de la didactique de l'oral: le projet PFC-EF. *Mélanges Crapel*, 31, 223-236.
- Field, J. (2008). *Listening in the language classroom*. Cambridge: Cambridge University Press.
- Field, J. (2019a). *Rethinking the second language listening test: from theory to practice*. Sheffield: Equinox.
- Field, J. (2019b). Second language listening: current ideas, current issues. In J. W. Schwieter & A. G. Benati (éds.), *The Cambridge handbook of language learning* (pp. 283-319). Cambridge: Cambridge University Press.

- Filteau, P. (2006). Un historique des formats de reproduction. *Circuit: musiques contemporaines*, 16(3), 17-32.
- Fouret, L.-A. (1936). Le phonographe, auxiliaire de l'enseignement. *Les Langues modernes*, 34(1/2), 66-81.
- Giroud, A. & Surcouf, C. (2016). De "Pierre, combien de membres avez-vous ?" à "Nous nous appelons Marc et Christian": réflexions autour de l'authenticité dans les documents oraux des manuels de FLE pour débutants. *Congrès Mondial de Linguistique Française*, 2016, 1-18.
- Gougenheim, G., Michéa, R., Rivenc, P. & Sauvageot, A. (1964). *L'élaboration du français fondamental: étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Paris: Didier.
- Holec, H. (1990). Des documents authentiques, pour quoi faire? *Mélanges Crapel*, 20, 65-74.
- Howatt, A. P. R. (2006). Language teaching: History. In K. Brown (éd.), *The encyclopedia of language & linguistics* (pp. 634-646). Amsterdam/Boston: Elsevier.
- International Correspondence Schools, S. (1903). *A textbook on French. Conversational lessons*. Scranton: International Textbook Company.
- Johns, T. (1986). Micro-concord: A language learner's research tool. *System*, 14(2), 151-162.
- Kimizuka, M. (2012). Historical development of magnetic recording and tape recorder. *Survey reports on the systemization of technologies*, 17, 185-274.
- Lazard, E. & Mounier-Kuhn, P. (2019). *Histoire illustrée de l'informatique*. Les Ulis: EDP Sciences.
- Luzzati, D. (1985). Recherches sur la structure du discours oral spontané. *L'information grammaticale*, 27, 39-41.
- Maisonneuve, S. (2006). De la machine parlante au disque. *Vingtième Siècle. Revue d'histoire*, 92, 17-31.
- Marcel, C. (1855). *Premiers principes d'éducation avec leur application spéciale à l'étude des langues*. Paris: Borrani et Droz.
- Marchand, L. (1929). Teaching modern languages: the scientific method. *The Modern Language Journal*, 14(1), 1-19.
- Marchand, L. (1937). Les débuts du phonographe d'enseignement en France. *Les langues modernes*, 35(3), 151-157.
- Martin, P. (1999). Prosodie des langues romanes: analyse phonétique et phonologie. *Recherches sur le français parlé*, 15, 233-253.
- McEnery, T. & Hardie, A. (2012). *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.
- Moirand, S. (1974). "Audio-visuel intégré" et communication(s). *Langue française*, 24, 5-26.
- Passy, J. & Rambeau, A. (1897). Coup d'œil sur nos principes. *Le Maître Phonétique*, 12(1), 3-10.
- Peterson, P. (1974). Origins of the language laboratory. *IALLT Journal of Language Learning Technologies*, 8(4), 5-17.
- Peytard, J. (1977). Le français parlé, langue et usage. Situer l'oral. *Le français moderne*, 45(3), 194-203.
- Porcher, L. (1995). *Le français langue étrangère: émergence et enseignement d'une discipline*. Paris: Hachette.
- Porquier, R. & Vivès, R. (1974). Sur quatre méthodes audio-visuelles. Essai d'analyse critique. *Langue française*, 24, 105-122.
- Puren, C. (1988). *Histoire des méthodologies de l'enseignement des langues*. Paris: CLE international.
- Puren, C. (2004). Quels modèles didactiques pour la conception de dispositifs d'enseignement/apprentissage en environnement numérique? *Ela. Études de linguistique appliquée*, 134, 235-249.

- Ravazzolo, E. & Etienne, C. (2019). Nouvelles ressources pour le FLE à partir des études en interaction. *Linx*, 79, en ligne.
- Ravazzolo, E.; Traverso, V.; Jouin, É. & Vigner, G. (2015). *Interactions, dialogues, conversations: l'oral en français langue étrangère*. Paris: Hachette.
- Rivenc, P. (2006). Les auteurs du *Français fondamental* face à un objet nouveau et insolite: l'interaction orale. *Documents pour l'histoire du français langue étrangère ou seconde*, 36, en ligne.
- Skornia, H. J. (1962). Educational radio: Its past and its future. In U.S. Department of health, education, and welfare. Office of education (éd.), *Educational television, the next ten years* (pp. 354-375). Stanford: Stanford University. Institute for Communication Research.
- Surcouf, C. (2020). Les enjeux de la compréhension du français oral quotidien en FLE: atouts possibles d'un corpus de français parlé annoté à des fins pédagogiques. *Études de Linguistique Appliquée*, 198, 241-256.
- Surcouf, C. (à paraître). Le français oral quotidien, un objectif spécifique en FLE? Retour sur les défis de la création d'un corpus de français parlé annoté à visée pédagogique. In C. Frérot & M. Pecman (éds.), *Des corpus numériques à l'analyse linguistique en langues de spécialité*. Grenoble: UGA Éditions.
- Surcouf, C. & Giroud, A. (2016). À quelle langue accède l'apprenant? Examen critique du traitement de l'oral dans les premières leçons de manuels de français langue étrangère pour débutants. *Linguistik Online*, 78(4), 11-27.
- Surcouf, C. & Ausoni, A. (2018). Création d'un corpus de français parlé à des fins pédagogiques en FLE: la genèse du projet FLORALE. *EDL (Études en didactique des langues)*, 31, 71-91.
- Tournès, L. (2008). *Du phonographe au MP3. Une histoire de la musique enregistrée XIX^e-XXI^e siècle*. Paris: Autrement.
- Vandergrift, L. & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. New York: Routledge.
- Varenne, G. (1936). Un précurseur de l'emploi du phonographe dans les classes de langues vivantes. *Les langues modernes*, 34(5), 290-292.
- Vialleton, É. & Lewis, T. (2014). Reconsidering the authenticity of speech in French language teaching: theory, data, methodology, and practice. In H. Tyne, V. André, C. Benoitoun, A. Boulton & Y. Greub (éds.), *French through Corpora: Ecological and Data-Driven Perspectives in French Language Studies* (pp. 293-316). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Weir, C. J. & Vidaković, I. (2013). The Measurement of listening ability 1913-2012. In C. J. Weir, I. Vidakovic & E. D. Galaczi (éds.), *Measured constructs: a history of Cambridge English language examinations 1913-2012* (pp. 347-419). Cambridge: Cambridge University Press.
- Worthington, D. L. (2018). Modeling and measuring cognitive components of listening. In D. L. Worthington & G. Bodie (éds.), *The sourcebook of listening research: methodology and measures* (pp. 70-96). Malden: John Wiley & Sons.
- Worthington, D. L. & Fitch-Hauser, M. E. (2018). *Listening: processes, functions, and competency*. Oxon: Routledge.



Un corpus parallèle français-grec en ligne pour une aide à la rédaction d'une argumentation

Fryni KAKOYIANNI-DOA & Dora LOÏZIDOU

Université de Chypre

Département d'Études françaises et européennes

B.P. 20537, CY-1678, Nicosie, Chypre

frynidoa@ucy.ac.cy; loizidou.dora@ucy.ac.cy

In this article, according to the approach of "data driven learning" (Johns 1991), we show the impact of a parallel French-Greek online corpus, which is used to teach non-computer specialists to learn argumentative written expression. We present the contents of the online corpus, then we show the results of an experiment carried out with students (an undergraduate class of French as Foreign Language at the University of Cyprus). The aim of this teaching unit was to strengthen students' written production. Specifically, in this experiment, the students had to write a text which included specific adverbs expressing the modalisation of the assertion such as *assurément*, *certainement*, *effectivement* using the *SOURCE corpus*. At the end of this experiment, a survey was conducted to gather information on the evaluation of the bilingual online corpus and students' opinion during this procedure.

Keywords:

Parallel corpus, French-Greek, writing, adverbial modifiers, experiment.

Mots-clés:

Corpus parallèle, français-grec, production écrite, adverbiaux modalisateurs, expérimentation.

1. Introduction

Notre article s'intéresse à un domaine particulièrement délicat de l'enseignement/apprentissage des langues, le processus d'écriture. Dans le but d'aider les apprenants, nous avons exploré comment l'utilisation d'un corpus parallèle bilingue permettait de faciliter la rédaction d'un texte argumentatif en classe de FLE. Plus particulièrement, nous nous posons les questions suivantes: un corpus bilingue pourrait-il aider un apprenant de niveau B2 à observer, explorer, améliorer son niveau et, selon les référentiels du *Cadre Européen Commun de Référence pour les Langues* (CECR), lui permettre d'écrire

des textes clairs et détaillés sur une gamme étendue de sujets relatifs à son domaine d'intérêt en faisant la synthèse et l'évaluation d'informations et d'arguments empruntés à des sources diverses (Conseil de l'Europe 2001: 51)?

Dans l'affirmative, quel serait le taux d'amélioration de la compétence discursive au niveau lexical général ? De même, quel serait ce taux au niveau des marques de modalités reflétant l'opinion de l'énonciateur ? Quel serait enfin le bénéfice didactique du côté des apprenants ayant comme aide à la rédaction un corpus parallèle bilingue ?



Les réponses qui sont données à ces questions dépendent aussi d'études issues de la recherche sur le terrain. C'est le cas de la présente étude qui part du constat que la plupart des apprenants de FLE (mais aussi d'autres langues étrangères) ont d'une part, tendance à traduire mot à mot suivant les règles et la syntaxe de leur langue maternelle et recourent, d'autre part, de plus en plus à la traduction automatique en ligne.

Ainsi, dans le but de contribuer à l'apprentissage de la rédaction de meilleurs textes, le présent article porte sur une expérimentation menée auprès d'étudiants chypriotes hellénophones de niveau B1-B2. Son objectif est d'examiner l'impact de l'utilisation d'un corpus parallèle bilingue sur les apprenants, au moment de la rédaction d'un texte argumentatif soumise à des contraintes spécifiques: comme le texte est une argumentation, il doit entre autres comporter des marqueurs discursifs servant à le structurer, tels que les adverbiaux de phrase modalisateurs de l'assertion *apparemment*, *assurément*, *certainement* mais aussi des adverbiaux polylexicaux tels que *à coup sûr*, *bien entendu*, *bien sûr*, *peut-être*, *sans aucun doute*, etc¹.

Après avoir présenté le contexte de l'étude et la complexité de la tâche d'écriture, nous rappellerons l'impact des corpus multilingues et bilingues pour la recherche contrastive, l'enseignement et l'acquisition d'une langue étrangère. Ensuite, nous présentons le corpus parallèle français-grec-SOURCE qui a servi de base à notre étude. Nous montrons ensuite les différentes phases du déroulement de l'expérimentation, puis les résultats obtenus sur la base de calculs d'utilisation aussi bien des diverses parties du discours que des adverbiaux modalisateurs de l'assertion. Enfin, nous présentons l'avis des apprenants sur les divers outils utilisés au cours de cette tâche expérimentale y compris le corpus bilingue SOURCE.

2. Cadre de l'étude

2.1 Le contexte chypriote

À Chypre, la langue maternelle dominante est le grec moderne, l'anglais est la première langue étrangère apprise et la langue française, la seconde. Cette dernière fait l'objet d'un enseignement obligatoire dans les écoles du secteur public pendant le cycle secondaire (4 ans). Elle est également présente à l'Université de Chypre, l'université d'État du pays, comme langue étrangère au Département d'Études françaises et européennes et au Centre de Langues de cette institution. Les méthodes d'enseignement du français adoptées s'inspirent essentiellement de l'approche communicative et de la perspective actionnelle issues des recommandations du CECR. Malgré cela, le plus souvent, les

¹ En français, ces adverbiaux ont fait l'objet de nombreuses recherches qui proposent des traitements linguistiques assez fins (Borillo 1976; Gardies 1979; Molinier 2000; Le Querler 2004). En grec, ces adverbiaux ont été étudiés et analysés dans une étude contrastive des adverbiaux de phrase français et grecs (Kakoyianni-Doa 2008).

évaluations réalisées durant le cursus d'étude ne portent que sur des épreuves écrites. C'est pour cette raison que nous nous sommes penchées dans cette étude sur la rédaction et sur l'aide que peut apporter l'utilisation d'un corpus parallèle français-grec.

2.2 *La complexité de la tâche d'expression écrite*

Comme plusieurs études l'ont déjà montré, la rédaction de productions écrites en langue non maternelle n'est pas un simple geste graphique mais une tâche ardue pour les apprenants d'une langue étrangère (Beacco 2007; Paquot 2010; Hidden 2014). On sait également que les difficultés sont de tout ordre (orthographique, morphologique, sémantique ou syntaxique) et que "cela implique une longue période d'appropriation, avant d'obtenir des textes qui soient corrects (ou acceptables) et appropriés" (Beacco 2007: 213). Les auteurs du CECR déclarent cependant que les apprenants pourraient surmonter ces difficultés en mettant en œuvre des stratégies de production qui mobilisent "des ressources et [...] des compétences différentes – en exploitant les points forts et en minimisant les points faibles – afin d'assortir le potentiel disponible à la nature de la tâche" (Conseil de l'Europe 2001: 53). Du côté des enseignants, l'enseignement de l'écrit pose quelque fois un problème du fait qu'ils se voient souvent obligés d'adopter "une méthodologie éclectique", méthodologie complexe formée d'éléments empruntés à différentes méthodologies (Puren 2013). Allouche & Maurer affirment à ce sujet que le manque de compétences de communication écrite relève de l'histoire des méthodologies, arguant que "les méthodes d'enseignement-traduction se souciaient peu de la production d'écrits, attendu que l'apprenant n'avait qu'à transcoder – de la façon la plus exacte et élégante qui soit – des mots, des phrases puis des textes déjà là" (2010: 4). Si les méthodes qui se sont succédé – directes, audio-orales et audiovisuelles – ont tenté de faire valoir un écrit moins formel, le travail de transcodage centré sur l'écrit dès les débuts de l'apprentissage est toutefois resté présent (Puren 2013). On n'a ainsi jamais rompu avec les traditions didactiques marquées par "des méthodes d'intervention portant sur [...] la rhétorique ancienne (manipulation des connaissances, mise en langage, puis révision du texte)" (Favart & Olive 2005: 274).

C'est avec l'avènement de l'approche dite communicative que la tâche de rédaction s'appuie sur la pragmatique, l'observation et la réflexion de l'apprenant. "C'est un acte de communication fonctionnel, un savoir et un savoir-faire spécifiques permettant à l'apprenant de s'exprimer et de communiquer au moyen d'un système de signes spécifiques, de signes graphiques" (Anh 2011: 256). C'est à ce moment que l'on cherche aussi à comprendre les différents processus cognitifs mis en œuvre au cours d'une tâche de rédaction. Comme le déclare l'auteur précité, on "est passé d'une centration sur le produit de l'activité à une centration sur les processus cognitifs permettant cette activité" (*Ibid.*). Plusieurs chercheurs ont d'ailleurs tenté d'analyser les diverses

composantes de l'activité de rédaction et ont proposé divers modèles (Moirand 1979; Hayers & Flower 1980; Groupe DIEPE 1995). Enfin le CECR mentionne les quatre étapes du processus d'écriture: la planification, l'écriture (exécution), la réécriture (évaluation) et la révision (remédiation et autocorrection). Malgré ces avancées, ce processus d'écriture reste complexe pour les apprenants. Sa complexité a d'ailleurs souvent été évoquée comme un frein à son investigation (Fayol 1997).

3. Impact et usages des corpus

3.1 Corpus multilingues

Rappelons brièvement qu'il existe en sciences du langage différents types de corpus. Mayaffre (2005) présente une classification en trois catégories: 1) les corpus lexicographiques (des corpus clos), phrastiques de grammairiens ou de syntacticiens, fabriqués pour vérifier ou non des hypothèses linguistiques émises par le chercheur, 2) les corpus d'apprenants et 3) les corpus textuels (monolingues, spécialisés, comparables, multilingues, parallèles) qui reprennent des textes dans leur ensemble. Ce dernier type de corpus nous renvoie à ceux réalisés par Johns (1991) et Sinclair (2004), qui ont été parmi les premiers, après Firth (1957), à penser à la linguistique de corpus au sein d'un "paradigme contextualiste" (Giuliani & Hannachi 2010).

Les corpus multilingues ont beaucoup contribué à des recherches contrastives (Johansson & Oksefjell 1998; Altenberg & Granger 2002). Comme le font remarquer Altenberg & Granger les corpus multilingues dévoilent aussi bien des notions que des différences qui n'auraient pas pu apparaître dans une étude basée sur des corpus monolingues (*Ibid.*). Les corpus multilingues sont en effet utiles dans l'apprentissage des langues car ils fournissent une ressource pour la construction du vocabulaire dans laquelle le mot est traité comme faisant partie d'une phrase. Toutefois comme l'affirment Tran & Falaise (2016), ces descriptions théoriques fort utiles sont peu exploitées par les enseignants de FLE et encore moins par les apprenants qui les connaissent mal. Les enseignants de FLE s'intéressent peu, et restent peu formés quant à leur utilité et à leur utilisation (Cavalla & Loiseau 2013). Enfin, même si un nombre conséquent de corpus a été créé ces dernières années (Leech 1997), les applications de la recherche en linguistique de corpus parviennent encore rarement à la salle de classe, ainsi que Boulton le précise: "les corpus ont peu d'impact [...] dans le domaine de l'apprentissage/enseignement des langues" (2008: 37), ajoutant qu'ils doivent encore faire leurs preuves dans les trois usages de base décrits par Fligelstone (1993), à savoir "*teaching about*" (enseigner la linguistique de corpus), "*exploiting to teach*" (exploiter les corpus pour enseigner les langues) et "*teaching to exploit*" (former les apprenants à exploiter les corpus pour apprendre une langue).

Il est important de noter, à ce point, que l'influence de la langue maternelle des apprenants lors de l'apprentissage d'une langue étrangère a aussi suscité de nombreux débats. Plusieurs travaux conduits dans ce domaine attestent de la place incontournable de la langue maternelle (Corder 1978; Harley *et al.* 1990; Hawkins 1998). Selon Audin, "la langue maternelle est vraiment le seul élément de repérage dont le débutant dispose et elle joue un rôle essentiel" (2004: 68). L'apprenant a alors, même dans les niveaux plus avancés, recours au dictionnaire bilingue qui, suivant Lépinette:

suppose d'une manière toute aussi implicite, l'existence d'une organisation lexico-sémantique en langue source fondamentalement parallèle à celle de langue cible. Cette mise en regard des deux unités lexicales appartenant à deux langues différentes concerne uniquement l'axe paradigmatique et exclut la prise en considération des possibilités de combinaisons de ces items de langue source et langue cible dans des phrases ou dans des syntagmes, sauf dans le cas particulier des constructions figées. (1996: 54)

3.2 Corpus bilingues parallèles: le cas de SOURCe Corpus

Les corpus bilingues et parallèles prennent justement en considération des possibilités de combinaisons des items de langue maternelle et de langue cible dans des phrases ou dans des syntagmes grâce à la mise en regard des deux langues dans un contexte particulier. Suivant Kraif,

la fascination liée à ce type de texte provient [...] de ce qu'elle met en présence deux langues, source et cible, et de ce fait révèle leur nature même de code – au sens saussurien et cryptographique à la fois – la version traduite permettant en quelque sorte de décoder un message initialement chiffré. (2014: 14)

De plus comme dit Boulton

un atout des corpus est d'observer les mots en contexte. Par exemple, quels adverbess en anglais servent ordinairement à intensifier le sens du verbe *advise* (conseiller) ? Une simple requête ADVERBE+advise dans le COCA montre que la seule option fréquente est *strongly advise*. (2018:12)

SOURCe corpus entre dans cette catégorie dans la mesure où il contient des textes complets parallèles issus de la littérature, de discours juridique et politique, scientifique et cinématographique dont l'une des deux langues en présence est la langue maternelle des apprenants (Kakoyianni-Doa & Tziafa 2013). Ces corpus sont dits aussi *alignés* car l'on fait correspondre chaque unité du texte en langue source avec une unité de texte en langue cible, au niveau des paragraphes, des phrases ou des mots (Duchet et al. 2008). Comme le constatent Altenberg & Granger, d'une part, les corpus parallèles ont donné un nouvel essor aux recherches contrastives en leur offrant de larges et riches données authentiques (2002). De l'autre, Duchet et al. affirment pour leur part que:

les avancées permises par l'alignement s'avèrent très utiles dans le domaine de la traduction professionnelle, améliorant la qualité et la productivité de l'activité traduisante. (2008: 131)

Ce type de corpus sert effectivement à sensibiliser les apprenants de langues étrangères aux différences linguistiques entre deux langues, à fournir un

matériel linguistique authentique, des mots et des phrases précises et idiomatiques dans leur contexte.

Le corpus parallèle français-grec SOURCE en question, bilingue grec ↔ français accessible en ligne, a été créé par nos soins (Kakoyianni-Doa et al. 2013) et se destine aussi bien à la recherche qu'à l'enseignement du FLE en milieu hellénophone². Le corpus est constitué de textes traduits dans les deux langues de genres divers, politiques, littéraires, scientifiques, éducatifs (discours TED) et de sous-titrages de films. Il comprend à ce jour des textes dont 760.282 phrases alignées. Il comprend donc aussi bien des textes parallèles existants tels que EUROPARL qui sont un ensemble de textes du Parlement Européen (corpus EUROPARL), ou encore des sous-titrages de films (tirés du corpus OpenSubtitles) que des œuvres du canon classique français ou grec libres de droits, numérisées et parallélisées par les soins de l'équipe créatrice du corpus tels que *Le bourgeois gentilhomme*, *Eugénie Grandet*, mais aussi des textes littéraires grecs traduits en français tels que *Les mystères du Bosphore* de Dimitracopoulos, *Skiathos, île grecque* de Papadiamandis ou *Le Condamné* de Theotokis. On peut explorer chacun de ces sous-corpus séparément ou tous ensemble pour effectuer des recherches non seulement au niveau de l'unité lexicale mais aussi au niveau d'expressions diverses. De nombreuses phrases-exemples peuvent apparaître, avec les références des sources. Il est aussi possible d'exporter des données et de combiner ou d'affiner les critères de recherche. Ce corpus diffère d'autres corpus existant en ligne (tels que Linguee et Glosby) en ce que les genres des textes sont séparés (on peut effectuer des recherches sur un texte spécifique, par exemple *politique*), les textes sont choisis pour l'usage pédagogique et éducatif, les phrases apparaissent dans leur contexte, soit dans les deux langues soit dans l'une ou l'autre séparément, caractéristiques qui permettent un travail pointu, aussi bien monolingue que bilingue (Kakoyianni-Doa et al. 2013). Il est à noter que, si pour la langue française et d'autres langues, il existe un nombre satisfaisant de textes (tels que les corpus Frantext, NBC), en revanche, en ce qui concerne la langue grecque, le nombre et la disponibilité des textes sont plutôt réduits. De plus, comme le déclare Kamber, un corpus tel que Frantext:

est certes un outil précieux et de dimensions appréciables avec ses 210 millions de mots, mais il présente un certain nombre de limites: cet instrument est constitué en grande partie de références littéraires, et son accès est limité par un système d'abonnement. En outre, il présente une dimension diachronique qui s'accorde mal avec les besoins d'apprenants du FLE. (2011: 200)

Enfin le corpus SOURCE est à la base d'une plateforme numérique bilingue français-grec qui héberge divers outils servant à explorer, de manières différentes, soit le corpus dans son intégralité soit par genre textuel. L'utilisateur

² Source Corpus (cf. <http://sourcecorpus.ucy.ac.cy>) est à la fois libre d'accès (Decision No EX-19-4), et dynamique pour être toujours en voie de complétude.

peut ainsi consulter des phrases alignées du corpus dans les deux langues grâce à un ou plusieurs mots-clés, visualiser des concordances, chercher des synonymes, créer des exercices mais aussi aligner ses propres textes ou encore faire des recherches linguistiques dans l'une ou l'autre langue séparément (cf. Kakoyianni-Doa, Eleni Tziafa, 2021).

4. Déroulement de l'expérimentation

C'est à partir de ce corpus que nous avons mené notre expérimentation de rédaction de texte argumentatif avec des étudiants hellénophones³ (N=22) en 2^e année de licence au Département d'Études françaises et européennes de l'Université de Chypre, dans un cours de français sur objectifs universitaires au sein duquel ils travaillent la production écrite en FLE des textes argumentatifs académiques. Quinze étudiants ont comme langue maternelle le grec (N=15). Les autres étudiants ont comme langues maternelles les suivantes: trois étudiants l'anglais, un étudiant le roumain, un autre l'arabe; et deux étudiants étaient bilingues, grec-arabe et grec-russe (N=7). Ces derniers maîtrisent parfaitement le grec car ils ont vécu et ont été scolarisés dans des écoles hellénophones à Chypre depuis leur plus jeune âge et peuvent donc être considéré comme hellénophones. Tous les étudiants ont en français un niveau B1-B2 sur l'échelle du *CECR*.

L'expérimentation s'est faite en trois phases:

- 1) rédaction d'un texte argumentatif à l'aide de dictionnaires divers en ligne, de leur choix (monolingues ou bilingues).
- 2) rédaction d'un texte similaire exclusivement à l'aide du corpus bilingue français-grec SOURCE.
- 3) autoévaluation pédagogique des étudiants sous la forme de réponses à un questionnaire sur l'utilisation des sources électroniques utilisées.

Les trois phases se sont déroulées entre le 12 et le 21 novembre 2019, dans une salle informatique de l'université de Chypre. Chaque phase a duré 75 minutes.

4.1 Première phase

Dans le cadre de cette expérimentation, les étudiants ont dû rédiger un texte qui devait discuter des atouts et des inconvénients de la technologie. Ils étaient autorisés, lors de la première phase, à rédiger librement. Le sujet de la première production était le suivant:

De nos jours la technologie a un grand impact dans nos vies. La révolution digitale a changé drastiquement nos sociétés et les moyens de communication avec les autres. Cette révolution a ses avantages et ses désavantages. Qu'en pensez-vous?

³ Tous les étudiants sont élevés et scolarisés à Chypre.

Les étudiants devaient par ailleurs tenter d'utiliser un nombre maximum d'adverbiaux monolexicaux et polylexicaux exprimant la modalisation de l'assertion tels que: *assurément, certainement, effectivement, sans doute*, etc. Ce sont des adverbiaux qui marquent un rapport de sens entre des propositions ou entre les phrases d'un texte, des relations établies entre les idées par celui qui parle. Ils permettent ainsi à l'énonciateur d'exprimer son sentiment ou son jugement sur ce dont il parle. Sont considérés pour le français et le grec moderne comme adverbiaux modaux une trentaine d'adverbiaux monolexicaux (cf. exemples en section 2). Ils ont généralement les mêmes propriétés en grec et en français. Ils peuvent figurer dans les deux langues en position initiale dans une phrase positive ou négative, suivis ou pas d'une pause, mais jouissent parfois également d'une liberté positionnelle. Sémantiquement, ils ont pour fonction d'évaluer la vérité ou le degré de certitude, sur une échelle nécessairement positive, de la proposition à laquelle ils s'intègrent (Borillo 1976; Molinier & Lévrier 2000). Ainsi, les adverbiaux modalisateurs donnent par exemple une valeur aléthique du nécessaire avec *fatalement, forcément, inévitablement, nécessairement, obligatoirement*, la valeur épistémique du certain avec *certainement, incontestablement, indiscutablement, sûrement*, la valeur aléthique du possible avec *éventuellement, possiblement, probablement*, etc. Selon Borillo (1976), le trait commun de ces adverbiaux serait précisément "d'exprimer une affirmation" d'où le choix du terme adverbiaux assertifs dans le sens d'adverbiaux modalisateurs de l'assertion. La liste bilingue français-grec comprenant 38 adverbiaux leur a été fournie en classe lors de l'expérimentation (cf. tableau en annexe). Le but de cette première tâche était donc de récupérer les productions écrites / argumentations des étudiants ayant travaillé soit librement soit à l'aide de ressources électroniques, afin de les comparer avec le produit de la phase suivante.

4.2 Deuxième phase

Le but de la deuxième tâche était le même, à savoir la rédaction d'une argumentation où il fallait utiliser le maximum d'adverbiaux modalisateurs affichés sur la liste fournie en phase 1 mais cette fois-ci à l'aide exclusive du corpus SOURCe bilingue français-grec. Le sujet proposé aux apprenants lors de la deuxième phase était le même, mais formulé différemment: "La technologie a-t-elle un impact positif ou négatif sur nos vies? " La formulation différente avait pour but d'éviter la tentation de reproduire les textes rédigés en première phase. Pour cette étape donc, trois contraintes ont été imposées aux étudiants: le sujet à traiter, les adverbiaux modalisateurs à utiliser et l'utilisation du corpus⁴. Les étudiants avaient une totale liberté concernant la longueur de leurs productions au cours des deux phases.

⁴ Notons que dans le cadre du cours les étudiants n'avaient aucune formation à l'usage des corpus de ce type.

4.3 Troisième phase

À la fin de la dernière phase, les étudiants ont répondu à un questionnaire en ligne. Le questionnaire, constitué de 23 questions en français, alternait questions fermées et ouvertes tant sur l'utilisation des sources électroniques utilisées que du corpus bilingue SOURCE, ainsi que les modalités de consultation et de travail pour chaque phase⁵. Plus particulièrement, le questionnaire comportait des questions de jugement concernant l'interface, la facilité d'emploi, la capacité de recherche, l'utilité des sources électroniques et la qualité des résultats obtenus.

5. Calculs, résultats et constats de l'expérimentation

5.1 Méthode de calculs et données chiffrées

Le corpus de cette expérimentation nous a fourni au total 44 textes argumentatifs (un texte par étudiant lors des phases 1 et 2). Comme notre recherche expérimentale a un caractère quantitatif (Narcy-Combes 2005), nous avons recherché la validité statistique des résultats obtenus en nous appuyant sur une analyse quantitative de nos données par une méthodologie directe (Adler & Adler 1987). Comme le déclare Van der Maren (1996), les données recueillies sont provoquées, c'est-à-dire nous n'avons pas recueilli des données sur un terrain préexistant, mais nous avons observé des données provoquées par nous-même. De ce fait, pour recenser les différences entre les textes de la première et de la deuxième phase, nous avons recouru à un outil de textométrie appelé AnaText, conçu au Lidilem par Olivier Kraif de l'Université Grenoble Alpes⁶. AnaText nous a permis de calculer la fréquence des lemmes par catégorie. Nous avons ainsi pu mesurer aussi bien la fréquence de chacune des parties du discours que celle des adverbiaux modalisateurs pour voir si ce corpus bilingue a finalement pu aider nos apprenants à écrire des textes argumentatifs plus clairs et plus étendus. Cette analyse nous a en effet montré que le corpus recensé de la première phase contenait 6.498 mots alors que celui de la deuxième phase en contenait 7.262. Le tableau ci-dessous (tab. 1) présente, par phase, le nombre de mots de chaque partie du discours identifiée dans les textes des étudiants. Pour l'analyse des données recueillies à l'aide du questionnaire, nous avons recouru à l'analyse de contenu.

⁵ Pour ce qui est des questions ouvertes du questionnaire, les étudiants étaient libres de répondre tant en français, qu'en grec ou en anglais.

⁶ Nous avons utilisé la version 2.3 (<http://olivier.kraif.u-grenoble3.fr/anaText>).

<i>Partie du discours</i>	<i>Étiquette</i>	<i>Nombre d'occurrences</i>	
		PHASE 1	PHASE 2
Abréviation	ABR	8	9
Adjectif	ADJ	431	495
Adverbe	ADV	531	609
Déterminant	DET	805	936
Conjonction	KON	340	395
Nom	NOM	1256	1426
Nom propre récurrent	NPR	19	8
Numéral	NUM	9	14
Préposition	PRE	837	956
Pronom	PRO	657	683
Ponctuation	PUN	393	370
Phrase	SENT	299	343
Verbe	VER	913	1018
Total		6498	7262

Tab. 1: Nombre de mots de chaque partie du discours, aux phases 1 et 2

5.2 Constats concernant l'emploi des différentes parties du discours

À l'issue de ce calcul textométrique (cf. Fig. 1), nous avons constaté une augmentation générale du nombre d'items à la deuxième phase, avec consultation de SOURCE corpus⁷. Certaines productions arrivent à atteindre une hausse de 40% (313 mots en première phase vs 431 à la deuxième) et 60% (330 mots en première phase vs 515 à la deuxième). Nous avons également constaté une augmentation du nombre d'items des différentes parties du discours (verbes, noms, adverbiaux, adjectifs, déterminants, etc.). Un plus grand nombre d'occurrences des parties du discours apparaît plus particulièrement dans quatre catégories, à savoir par ordre ascendant les noms (1256 en première et 1426 en deuxième), les verbes (913 en première et 1018 en deuxième), les prépositions (837 en première et 956 en deuxième) et les déterminants (805 en première et 936 en deuxième).

⁷ Il faudrait ici mentionner que cette augmentation pourrait ne pas être due à l'utilisation du corpus. En effet, le fait de devoir écrire un deuxième texte sur une thématique très similaire pourrait tout aussi bien l'expliquer. Toutefois l'utilisation d'un corpus devient source d'inspiration grâce aux données consultables, en contexte. Écrire c'est "copier, encoder et produire" déclare Bernadette Kervyn, maître de conférences à l'université de Bordeaux lors de son intervention en formation de formateurs à l'IFÉ en novembre 2017 (<https://bit.ly/2TNvSh4>).

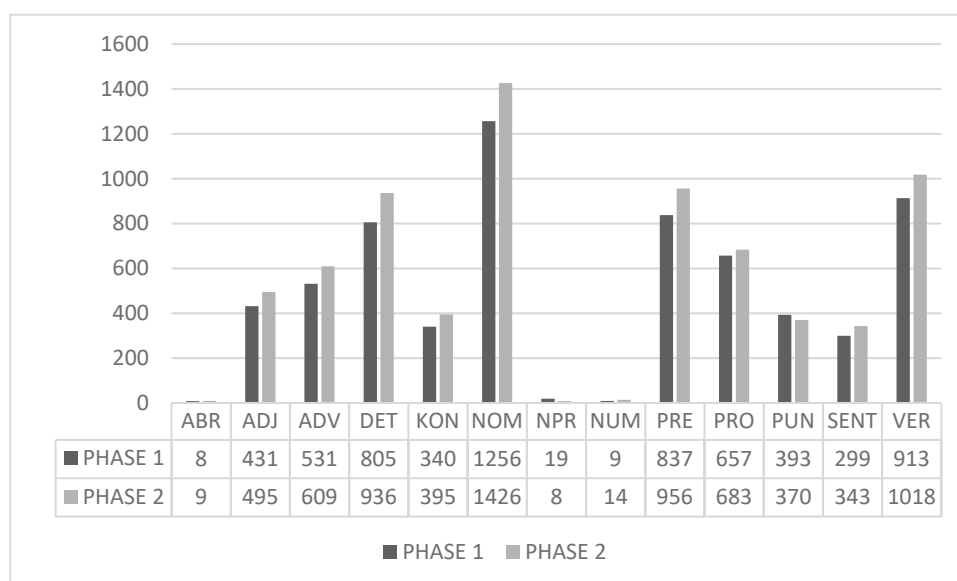


Fig. 1: Nombre de mots par partie du discours, aux phases 1 et 2

De plus, on constate, en phase 2, une augmentation d'emploi des adverbiaux en général (cf. Fig. 1). Certains expriment:

- l'addition ou le classement, comme *d'abord*, *premièrement*, *deuxièmement*, *dernièrement*, *ensuite*, *enfin*, *finalement*, *seulement*, *complètement*, *régulièrement*, *largement*, *quotidiennement*, *instantanément*, *notamment*;
- la restriction, comme *cependant*, *entre-temps*, *d'ailleurs*;
- la comparaison, comme *également*, *excessivement*, *extrêmement*;
- mais aussi des adverbiaux de phrase tels que *franchement*, *généralement*, *paradoxalement*.

Ceci montre que les étudiants ont pu profiter des phrases modèles du corpus pour enrichir leur répertoire adverbial. Comme le dit Di Vito à propos de la linguistique du corpus, "avec cette approche l'étudiant interroge ses propres intuitions, pose ses propres hypothèses et les vérifie par l'analyse de données réelles" (2013: 160). On peut ainsi considérer que l'apprenant découvre la langue étrangère tout en la comparant de façon inductive avec sa langue maternelle.

5.3 Constats concernant l'emploi des adverbiaux modalisateurs

Concernant l'emploi des adverbiaux modalisateurs, nous avons constaté qu'en phase 1 de l'expérimentation, les étudiants

- 1) ont utilisé 23 sur les 38 adverbiaux monolexicaux en *-ment*, soit le 60% des occurrences affichées sur la liste distribuée; nous avons repéré au total 109 emplois de ces occurrences dans l'ensemble du corpus de productions écrite des étudiants.

- 2) n'ont pas utilisé les formes polylexicales ci-dessous, évitées ou peut-être non considérés par les étudiants, à priori, comme des adverbiaux;

Adverbiaux modalisateurs absents		
1.	à coup sûr	0
2.	bien entendu	0
3.	bien sûr	0
4.	peut-être	0
5.	sans aucun doute	0
6.	sans conteste	0
7.	sans doute	0
8.	sans l'ombre d'un doute	0
9.	sans nul doute	0

Tab. 2: Adverbiaux modalisateurs absents en phase 2

En phase 2, les étudiants

- 1) ont utilisé 27 sur 38 adverbiaux monolexicaux en *-ment*, soit le 71% des occurrences imposées; nous avons repéré au total 93 emplois des adverbiaux utilisés dans l'ensemble du corpus de productions écrite des étudiants.
- 2) n'ont pas utilisé les formes polylexicales de la liste distribuée (cf. Tab. 2);
- 3) ont abandonné les adverbiaux *éventuellement*, *indiscutablement*, *probablement* alors qu'ils ont privilégié les adverbiaux *manifestement* et *possiblement* ;
- 4) ont utilisé pour la première fois les adverbiaux *irréfutablement*, *manifestement*, *naturellement* et *nécessairement* ;

En somme durant la première phase, l'ensemble des textes argumentatifs présente moins d'occurrences mais en plus grand nombre. Au contraire la deuxième phase s'est caractérisée par une variation plus importante dans le choix des adverbiaux modalisateurs/des occurrences mais en moins grand nombre (cf. Fig. 2).

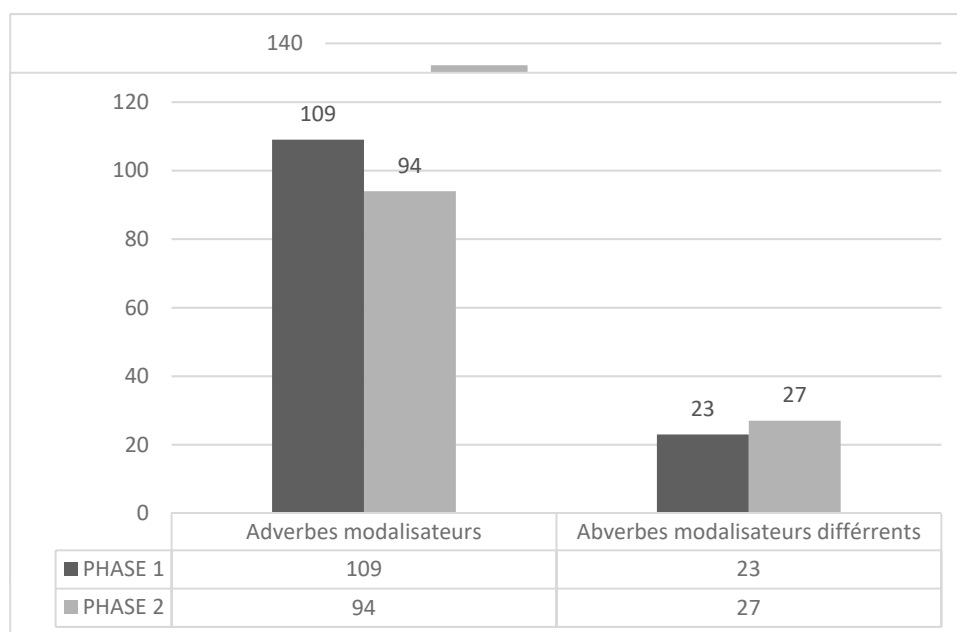


Fig. 2: Nombre total d'adverbiaux utilisés et nombre de leurs occurrences aux phases 1 et 2

Ce constat nous permet de confirmer une plus-value encourageante pour une utilisation plus systématique de SOURCe corpus. En outre, le fait de ne pas avoir répété plusieurs fois les mêmes adverbiaux en phase 2, d'avoir abandonné certains adverbiaux ou d'avoir utilisé d'autres pour la première fois pourraient être des indices que les étudiants ont pu avoir été détournés ou absorbés par les données du corpus ou par les traductions différentes de celles indiquées sur la liste distribuée. Nádvorníková par exemple a longuement discuté des pièges méthodologiques des corpus parallèles, dont celui de la question des équivalences des versions parallèles. L'auteure affirme que "les équivalents considérés comme systémiques mènent certains linguistes travaillant sur corpus parallèles à sous-estimer l'importance du rôle joué par les traducteurs et par tout le contexte de la création de la traduction" (2017: 10). Ceci serait la preuve que les étudiants ont exploité le corpus avec un esprit à la fois critique, pragmatique et créatif.

5.4 Bénéfices didactiques concernant l'usage SOURCe corpus

Les témoignages tirés du questionnaire ont montré que les étudiants ont l'habitude de consulter des ressources en ligne (18 sur 22) telles que *Google translate*, *Bon patron*, *Linguee*, *Thesaurus*, *Wikipedia*, *Wordreference* et le *Larousse* en ligne. Une étude précédente effectuée dans le même contexte a également montré que les étudiants du Département utilisent davantage les sources linguistiques numériques que celles sur papier (Kakoyianni-Doa et al. 2013). SOURCe corpus a également été apprécié en tant que ressource complémentaire en termes d'interface (sur 22 étudiants, 8 l'ont trouvé très bonne, 11 bonne et 3 moyenne), de facilité d'emploi (9 l'ont trouvé très bonne,

11 bonne et 2 moyenne), d'utilité (11 l'ont trouvé très bonne, 8 bonne et 3 moyenne) et de résultats attendus (8 l'ont trouvé très bonne, 10 bonne, 3 moyenne et 1 mauvaise). Les étudiants sont d'avis que SOURCe corpus est une ressource utile à une meilleure utilisation du vocabulaire, car le mot est traité comme faisant partie d'une phrase. Des étudiants déclarent: "Une variété très riche de mots et de significations", "J'aime utiliser le corpus SOURCe parce qu'on trouve des phrases qui montrent la signification d'un mot, ce qui, je pense, est utile pour nous" et "C'est génial pour avoir des idées et savoir comment la structure d'une phrase fonctionne en fonction du mot recherché". D'autres déclarent que le corpus leur a permis de gagner du temps. Ceci rejoint les conclusions d'autres études sur l'usage des outils linguistiques: les étudiants s'approprient les outils linguistiques utiles et faciles à utiliser (Nissen 2007; Kakoyianni-Doa et al. 2013). Par ailleurs, les étudiants soulignent le fait que le corpus donne beaucoup plus d'exemples en grec que d'autres sources disponibles en ligne. Ils affirment enfin que l'outil *Synonyme* de la plateforme les a également aidés à rechercher des occurrences synonymes d'un mot dans les deux langues. Cependant, certains ont déclaré préférer utiliser le dictionnaire en ligne et n'ont pas senti avoir eu de meilleurs résultats. En bref, nous avons observé une tendance générale qui aurait montré que les étudiants étaient globalement intéressés par l'utilisation du corpus SOURCe.

6. Conclusion

Dans le présent article nous avons pu montrer que le recours à un corpus parallèle français-grec en ligne pouvait s'avérer bénéfique pour l'enseignement-apprentissage de la rédaction de l'écrit du FLE. Un tel corpus pourrait donc, à notre sens, aider les apprenants à dépasser le niveau ponctuel des mots et à mieux comprendre comment ils sont utilisés dans différents contextes. Avec le recours à des corpus, les apprenants sont invités à découvrir la langue étrangère, tout en comparant de façon inductive avec leur langue maternelle "des tendances plus ou moins typiques au sein de corpus de plusieurs millions de mots, de la même façon que Firth recommandait pour les linguistes de découvrir à partir de données authentiques ce qui était typique dans une langue" (Giuliani & Hannachi 2010: 151). Enfin, comme Debras (2018) le prône, lorsque les apprenants s'adonnent à des recherches textuelles, ils devraient avoir à leur disposition un grand nombre de textes, nécessaires pour comprendre en contexte, et par la suite (re)produire. Les grands corpus, dit-elle, sont "plus robustes et plus fiables car ils permettent de fonder l'analyse sur un grand nombre d'occurrences attestées [...] qui ne dépendent pas de la subjectivité" (*Ibid.* 2018: 2). Ces recherches et conclusions conduisent à (ré)adopter une nouvelle attitude à l'égard de la consultation des textes, celle d'un apprenant-chercheur en quête d'une information d'ordre aussi bien sémantique que syntaxique. Le corpus parallèle facilite donc effectivement la contrastivité et la compréhension des mécanismes linguistiques propres à chaque langue. Enfin,

l'apprenant se met en position de chercheur autonome dans son apprentissage. Il apprend ainsi à découvrir, réfléchir, comparer, se poser des questions et prendre des décisions et participe plus activement, donc plus efficacement, au processus d'apprentissage. L'expérimentation que nous avons conduite, avec l'analyse des attitudes des participants envers l'exploitation d'un corpus parallèle, a donc donné des réponses à nos questions de départ mais a aussi suscité de nouvelles interrogations. Les apprenants ont pu amplifier leurs textes, enrichir leur vocabulaire et leur répertoire adverbial, mais ils ont moins utilisé en quantité les adverbiaux modalisateurs. Est-ce par une diffraction de l'attention, causée par la pléthore de phrases entières et parallèles à observer et à gérer? Par manque d'usage et d'entraînement de travail sur corpus dans les classes de langue (Boulton 2007)? D'autres raisons pourraient être recherchées pour expliquer ces restrictions mais la route est sinueuse et, comme le disait Murison-Bowie (1993), quand on explore des corpus, on part d'une question qui est ensuite relayée par une autre et on finit par arriver à un endroit que l'on ne pouvait pas prévoir. Cependant, au lieu d'y voir un inconvénient, nous pouvons le considérer comme un atout, une raison de plus pour continuer à explorer et à aller plus loin dans cette réflexion.

BIBLIOGRAPHIE

- Adler, P. & Adler P. (1987). *Membership roles in field research*, Newbury Park, CA: Sage.
- Allouche, V. & Maurer, B. (2010). *L'écrit en FLE: Travail du style et Maîtrise de la langue*. Montpellier: Presses universitaires de la Méditerranée.
- Altenberg, B. & Granger, S. (2002). *Lexis in Contrast. Corpus-based approaches*. Amsterdam: Benjamins.
- Anh, N. V. (2011). La compétence de production écrite dans l'approche par compétences. *VNU Journal of Foreign Studies, [S.l.]*, 27, 4. <https://js.vnu.edu.vn/FS/article/view/1491>
- Audin, L. (2004). Apprentissage d'une langue étrangère et français: pour une dialectique métalinguistique pertinente dès le cycle 3. In *Repères, recherches en didactique du français langue maternelle, Français et langues étrangères et régionales à l'école. Quelles interactions?* 29. 63-80. doi: <https://doi.org/10.3406/reper.2004.2612>
- Beacco, J.-C. (2007). *L'approche par compétences dans l'enseignement des langues*. Paris: Didier, coll. Langues et didactique.
- Borillo, A. (1976). Les adverbes et la modalisation de l'assertion. *Langue française, Lexique et Grammaire*, 30, 74-89. doi: <https://doi.org/10.3406/lfr.1976.6113>
- Boulton, A. (2007). Esprit de corpus: Promouvoir l'exploitation de corpus en apprentissage des langues. *Texte et corpus*, 3, 37-46. URL: <https://hal.archives-ouvertes.fr/hal-00326988/document>
- Boulton, A. (2008). Esprit de corpus: Promouvoir l'exploitation de corpus en apprentissage des langues. *Texte et corpus*, 3, *Actes des Journées de la linguistique de Corpus 2007*, 37-46.
- Boulton A (2018). Les corpus et les TIC comme aide à la découverte des langues. *Les Langues Modernes, Association des professeurs de langues vivantes (APLV)*, 71-84. fhal-01850696f
- Cavalla, C. & Loiseau, M. (2013). Scientext comme corpus pour l'enseignement. In A. Tutin & F. Grossmann (éds.), *L'écrit scientifique: du lexique au discours. Autour de Scientext*, 163-180, Rennes: Presses universitaires de Rennes.



- Conseil de l'Europe. (2001). *Un cadre européen commun de référence pour les langues: apprendre, enseigner, évaluer*. Paris: Didier
- Corder, P. (1978). *Understanding second and foreign language acquisition*. London: Longman.
- Debras, C. (2018) Petits et grands corpus en analyse linguistique des gestes. *Corpus*, 18. URL: <http://journals.openedition.org/corpus/3287>
- Di Vito, S. (2013). L'utilisation des corpus dans l'analyse linguistique et dans l'apprentissage du FLE. *Linx*, 68-69. URL: <http://journals.openedition.org/linx/1519>; doi: <https://doi.org/10.4000/linx.1519>
- Duchet, J.-L., Kraif, O. & Castillo, M.-T. (2008). Corpus massifs et corpus bilingues alignés: leur impact sur la recherche linguistique, *Bulletin de la société linguistique de Paris*, t. CIII, fasc. 1, 129-150. doi: 10.2143/BSL.103.1.2033486
- Ducrot, O. (1993). À quoi sert le concept de modalité. In N. Dittmar & A. Reich (éds.), *Modalité et Acquisition des langues* (pp. 111-129). Berlin: Walter de Gruyter
- Dulay, H. C. & Burt, M. K. (1976). Creative construction in second language learning and teaching. *Language Learning, Special Issue*, 4, 65-79.
- Favart, M. & Olive, T. (2005). Modèles et méthodes d'étude de la production écrite. *Psychologie Française*, 50(3), 273-285. doi: 10.1016/j.psfr.2005.05.012
- Fayol, M. (1997). *Des idées au texte. Psychologie cognitive de la production verbale, orale et écrite*. Paris: P.U.F.
- Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. In J. R. Firth (éd.), *Studies in linguistic analysis. Special volume of the Philological Society* (pp. 1-32). Oxford, England: Blackwell
- Gardies, J.-L. (1979). *Essai sur la logique des modalités*. Paris: P.U.F.
- Gerard, H. (1998). *Quantitatif, qualitatif: même combat!* Communication présentée au Séminaire méthodologique en sciences humaines et sociales, 6-10 avril. Ouagadougou, Côte-d'Ivoire.
- Giuliani, D. & Hannachi, R. (2010). Linguistique de corpus et didactique du F.L.E. Une exploitation du corpus IntUne. *Cahiers de praxématique* 54-55, 145-160. doi: <https://doi.org/10.4000/praxématique.1136>
- Groupe DIEPE, (1995). *Savoir écrire au secondaire, Étude comparative auprès de quatre populations francophones d'Europe et d'Amérique*. Bruxelles: de Boeck Université
- Harley, B., Allen, P., Cummins, J. & Swain, M. (éd.). (1990). *The development of second language proficiency*. Cambridge: CUP
- Hawkins, E. (1998). *Foreign language study and Language awareness, communication de l'auteur*. Québec: ALA
- Hayes, J. R. & Flower, L. (1980). Identifying the Organization of Writing Processes. In L. W. Gregg & E. R. Steinberg (éds.), *Cognitive Processes in Writing: An Interdisciplinary Approach* (pp. 3-30). Hillsdale, NJ: Lawrence Erlbaum
- Hidden, M.-O. (2014). Regards des enseignants sur l'enseignement de la production écrite en FLE. In J. L. Ignacio Aguilar Rio, C. Brudermann & M. Leclère (éds.), *Complexité, diversité et spécificité: Pratiques didactiques en contexte* (pp. 133-150). URL: <https://halshs.archives-ouvertes.fr/halshs-01099170>
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press
- Johansson, S. & Oksefjell, S. (éds.) (1998). *Corpora and Cross-linguistic Research: Theory, Method, and Case Studies*. Amsterdam: Rodopi.
- Johns, T. (1991). Should you be persuaded: two examples of data-driven learning? *English Language Research, University of Birmingham*, 4, 1-16
- Kakoyianni-Doa, F (2008). *Adverbes de phrases français et grecs: étude contrastive et perspectives didactiques*. Thèse de doctorat. Université Toulouse II Le Mirail.



- Kakoyianni-Doa, F. & Tziafa, E. (2013). SOURCe: Building a searchable online french greek parallel corpus for the University of Cyprus. In *Revista Nebrija de Lingüística Aplicada*, 13 (número especial).
- Kakoyianni-Doa, F., Antaris, S. & Tziafa, E. (2013). A Free Online Parallel Corpus Construction Tool for Language Teachers and Learners. *Procedia - Social and Behavioral Sciences* 95, 5th International Conference on Corpus Linguistics (CILC2013), 535-541.
- Kakoyianni-Doa, F., Antaris, S. & Tziafa, E. (2021). Overview of the SOURCe Project as an Open Educational Resource. *Journal of Applied Linguistics* (JAL), 33, 124-142. En ligne depuis le 3 février 2021 URL: <https://ejournals.lib.auth.gr/jal/issue/view/1065>
- Kamber, A. (2011). Contexte et sens: utilisation d'un corpus écrit dans l'enseignement / apprentissage du FLE. *Travaux neuchâtelois de linguistique*, 55, 199-218
- Kern, R. G. (1994). The Role of Mental Translation in Second Language Reading, *Studies in Second Language Acquisition*, 16, 441-461.
- Kraif, O. (2014). Corpus parallèles, corpus comparables: quels contrastes?. *Informatique et langage [cs.CL]*. Université de Poitiers
- Krashen, S.D. (1982). *Principles and practice of second language acquisition*. Oxford: Pergamon
- Le Querler, N. (2004). Les modalités en français. *Revue belge de philologie et d'histoire*, 82(3), 643-656. Disponible sur: www.persee.fr/doc/rbph_0035-0818_2004_num_82_3_4850
- Leech (1997). Teaching and language corpora: a convergence. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (éds.), *Teaching and language corpora* (pp. 1-23). Harlow: Addison Wesley Longman.
- Lepinette, B. (1996). Le rôle de la syntaxe dans la lexicographie bilingue. In H. Béjoint, & P. Thoiron (éds.), *Les dictionnaires bilingues* (pp. 53-69). Louvain-la-Neuve: Duculot
- Mayaffre, D. (2005). Rôle et place des corpus en linguistique. Réflexions introductives. *Actes du colloque JETOU'2005*, 5-17. Université de Toulouse-Le Mirail.
- Moirand, S. (1979). *Situations d'écrit: Compréhension, production en langue étrangère*, Paris: Clé international
- Molinier, C. & Lévrier, F. (2000). *Grammaire des adverbes. Description des formes en -ment*. Genève: Librairie Droz
- Murison-Bowie, S. (1993). *MicroConcord Manual*. Oxford: Oxford University Press
- Nádvorníková, O. (2017). Pièges méthodologiques des corpus parallèles et comment les éviter. *Corela*, HS-21. doi: 10.4000/corela.4810
- Narcy-Combes, J.-P. (2005). *Didactique des langues et TIC. Vers une recherche-action responsable*. Paris: Ophrys.
- Nissen, E. (2007). Quelles aides les formations hybrides en langues proposent-elles à l'apprenant pour favoriser son autonomie? *Apprentissage des Langues et Systèmes d'Information et de Communication (Alsic)*, 10(1), 129-144. URL: <http://alsic.revues.org/617>
- Paquot, M. (2010). *Academic vocabulary in learner writing: from extraction to analysis*. New York: Continuum
- Puren, Ch. (2013). *La didactique des langues étrangères à la croisée des méthodes. Essai sur l'éclectisme*. Paris: Didier, Collection CREDIF
- Sinclair, J. (2004). *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins Publishing Company
- Tran, T. T. H. & Falaise, A. (2016). Un dictionnaire basé sur corpus pour une aide à la rédaction universitaire, *Lidil*, 58. doi: 10.4000/lidil.5378
- Van der Maren, J.-M. (1996). *Méthodes de recherche pour l'éducation*. Bruxelles: de Boeck.

ANNEXES

Liste indicative bilingue d'adverbiaux modalisateurs de l'assertion français-grec

1.	à coup sûr	στα σίγουρα	sta sigoura
2.	apparemment	προφανώς	profanos
3.	assurément	ασφαλώς	asfalos
4.	bien entendu	φυσικά	fisika
5.	bien sûr	βέβαια	vevaia
6.	certainement	βεβαίως	veveos
7.	certes	βεβαίως	veveos
8.	effectivement	όντως	ontos
9.	évidemment	φυσικά	fisika
10.	éventuellement	ενδεχομένως	endechomenos
11.	fatalement	μοιραία	mirea
12.	forcément	υποχρεωτικά	ipochreotika
13.	immanquablement	αναπόφευκτα	anapofefkta
14.	incontestablement	ομολογουμένως	omologoumenos
15.	indéniablement	αναμφισβήτητα	anamfisvitita
16.	indiscutablement	ασυζητητί	asizititi
17.	indispensablement	απαραίτητα	aparetita
18.	indubitablement	αναμφίβολα	anamfivola
19.	inévitablement	αναπόφευκτα	anapofefkta
20.	infailliblement	αναπόφευκτα	anapofefkta
21.	irréfutablement	αποδεδειγμένα	apodedigmaena
22.	manifestement	ολοφάνερα	olofanera
23.	naturellement	φυσικά	fisika
24.	nécessairement ⁸	αναγκαστικά	anagastika
25.	obligatoirement	υποχρεωτικά	ipochreotika
26.	peut-être	ίσως	isos
27.	plausiblement	πιθανώς	pithanos
28.	possiblement	ενδεχομένως	endechomenos
29.	présumablement	πιθανόν	pithanon
30.	probablement	μάλλον	mallon
31.	sans aucun doute	χωρίς αμφιβολία	choris amfivolia
32.	sans conteste	αναμφισβήτητα	anamfisvitita
33.	sans doute	αναμφίβολα	anamfivola
34.	sans l'ombre d'un doute	πέραν πάσας αμφιβολίας	peran pasas amfivoliass
35.	sans nul doute	οπωσδήποτε	oposdipote
36.	sûrement	σίγουρα	sigoura
37.	visiblement	ορατά	orata
38.	vraisemblablement	φαινομενικά	fenomenika

Liste des adverbiaux modalisateurs de l'assertion français et grecs distribuée aux étudiants

⁸ *Nécessairement* ne se trouve qu'une seule fois. Recours à *C'est nécessaire*. La réponse est donc **nécessairement** européenne. Η απάντηση είναι λοιπόν **αναγκαστικά** ευρωπαϊκή.

Introduire des constructions verbales en classe de français langue étrangère du niveau A1: une pratique sur corpus

Hasti NOGHRECHI

Université de Neuchâtel

Institut de langue et civilisation françaises (ILCF)

Faubourg de l'Hôpital 61-63, 2000 Neuchâtel, Suisse

hasti.noghrechi@unine.ch

This article aims to explore ways for teaching and learning the verb constructions in FFL (French as foreign language) class for beginner learners (levels A1.1 and A1). Starting from the observation of their incomplete representation in the manuals and the reference frames, I proceed, by the use of the tools of the corpus linguistics, to a selection of the verbs to present in relation to the fields of action of the reference frames for these levels (frequency and affinity criteria, among others) to better meet the needs of learners. I then describe their use in written texts, particularly in terms of complementation. Finally, I offer didactic activities adapted to the skills of learners, based on an inductive approach related to the use of corpora (*data-driven learning*).

Keywords:

verb constructions, corpus linguistics, data-driven learning, FFL, beginner learners, learning and teaching, technical material.

Mots-clés:

constructions verbales, linguistique de corpus, data-driven learning, FLE, apprenants débutants, enseignement/apprentissage, moyens d'enseignement.

1. Introduction

L'enseignement/apprentissage des verbes en classe de français langue étrangère (FLE) occupe une place importante dans les méthodes et les manuels de grammaire. Cependant l'analyse de leur contenu démontre que la quasi-totalité des chapitres traitant des verbes met l'accent sur leur aspect morphologique au détriment de leurs autres propriétés (syntaxiques, sémantiques, etc.). Le constat est identique pour les recherches universitaires; nombreuses sont les recherches qui s'intéressent à l'enseignement/apprentissage de la morphologie verbale aux apprenants allophones (voir notamment Ågren 2008, 2014; Granget 2005, 2009; Gerolimich & Stabarin 2007, 2014; Granfeldt & Schlyter 2005; Herschensohn 2001; Howard 2006, 2013; Kemps et al. 2010; Michot 2014; Michot & Pierrard 2014; Nadasdi 2001; Prévost 2009; Prévost & White 2000; Schlyter 2006; Sergeeva & Chevrot 2008), cependant les recherches sur l'enseignement/apprentissage des constructions verbales se font plus rares (Bruley-Meszaros 2010; Thomas 2009) et cela malgré le fait que la dépendance syntaxique ainsi que les relations prédicatives soient abordées dans un nombre considérable d'études (voir notamment

Feuillet 1978, 1998; Lazard 1994; Tesnière 1959) aux approches variées (la théorie de la valence verbale, l'approche fonctionnaliste, etc.).

Au manque d'intérêt porté à la complémentation verbale dans les moyens d'enseignement en FLE s'ajoute le fait qu'ils n'optent que très rarement pour un enseignement explicite de ce sujet aux apprenants débutants. Pour ces raisons, la présente contribution s'intéresse à la façon dont un enseignant peut introduire ce sujet dans une classe de français langue étrangère de niveaux débutants (A1.1 et A1) en répondant aux questions suivantes: que faut-il enseigner pour introduire des constructions verbales en classe de FLE? Et comment l'enseigner?

2. Constructions verbales dans les moyens d'enseignement

En s'appuyant sur la consultation des moyens d'enseignement de FLE pour les apprenants débutants (niveaux A1.1 et A1), cette étude s'intéressera particulièrement à la place que les constructions verbales (la morphologie et la syntaxe de verbes) y occupent et à la façon dont elles sont enseignées. Pour ce faire, le *Cadre européen commun de références pour les langues* (désormais *CECR*), les référentiels (A1.1 et A1) ainsi que quelques manuels de FLE seront examinés.

2.1 Ouvrages de référence

Avant d'étudier les constructions verbales dans les manuels de FLE, cette contribution étudiera brièvement des ouvrages de référence tels que *CECR* (2001), le *Volume complémentaire avec de nouveaux descripteurs* (2018) ainsi que les référentiels des niveaux A1.1 (Beacco et al. 2005) et A1 (Beacco & Porquier 2007).

2.1.1 CECR et volume complémentaire

Même si le *CECR*, en se référant à l'enquête du Fonds national suisse de recherche scientifique (1994-1995), précise que pour certains publics comme les "jeunes apprenants", il serait utile d'ajuster les compétences demandées au niveau A1 (Conseil de l'Europe 2001: 30), il ne propose pas pour autant de niveau en dessous de A1. En conséquence, dans les descriptions des compétences, il n'y a rien de précis pour ces apprenants. Le *Volume complémentaire*, quant à lui, est plus explicite à ce sujet et distingue clairement un niveau "Pré A1" qu'il considère comme "un jalon à mi-chemin vers le niveau A1" (Conseil de l'Europe 2018: 48). Selon cet ouvrage, de nombreux projets ont montré qu'"il était important de produire une description plus étoffée de la compétence des apprenants en A1 et d'inclure un niveau en dessous de A1" (*Ibid.*). Or, la description des compétences langagières des apprenants de ce niveau reste très générale: "l'apprenant n'a pas encore acquis une capacité de s'exprimer de manière autonome, mais s'appuie sur le répertoire de mots et d'expressions toutes faites" (*Ibid.*).

Les descriptifs des compétences du niveau A1 selon le *CECR* sont également vagues; selon lui un apprenant de ce niveau a "un contrôle limité de structures syntaxiques et de formes grammaticales simples appartenant à un répertoire mémorisé" (Conseil de l'Europe 2001: 90), mais il ne donne pas plus de détails. En ce qui concerne les compétences des enseignants ou des concepteurs de manuels, les explications ne sont pas d'avantage approfondies:

- les utilisateurs du *Cadre de référence* envisageront et expliciteront selon le cas
- sur quelle théorie grammaticale ils ont fondé leur travail
- les éléments, catégories, classes, structures, opérations et relations que les apprenants devront manipuler ou dont ils devront être outillés pour le faire (Conseil de l'Europe 2001: 90).

Comme Beacco et al., on peut déplorer le manque de précision dans les descripteurs, et le fait qu'ils "ne sont pas explicités [...]" (2005: 102).

2.1.2 Référentiels (A1.1 et A1)

Confronté à l'absence d'une réponse satisfaisante de la part du *CECR* à "la question de la sélection des caractéristiques morphologiques et syntaxiques considérées comme devant être connues des apprenants" (Beacco et al. 2005: 103), le référentiel du niveau A1.1 tente à son tour de remplir cette lacune en proposant un inventaire des notions que les apprenants de ce niveau doivent connaître. Pour ce faire, il tient essentiellement compte des éléments suivants:

- l'analyse des contenus retenus dans les manuels de langues destinés à des débutants;
- les connaissances disponibles sur les processus d'acquisition et les premières formes des interlangues des apprenants (Beacco et al. 2005: 103).

Plusieurs points sont communs aux référentiels niveaux A1.1 et A1. En ce qui concerne les connaissances morphosyntaxiques, l'apprenant doit prendre connaissance de l'existence des "mots variables et invariables", connaître "l'ordre des mots" et le fait que les mots sont "articulé[s] autour du verbe conjugué" (Beacco et al. 2005: 108 ; Beacco & Porquier 2007: 97). Il doit également être capable de reconnaître "des formes verbales à l'infinitif" ou "certaines formes de la conjugaison" de quelques verbes (Beacco et al. 2005: 108; Beacco & Porquier 2007: 97).

Un enseignement/apprentissage progressif de la morphologie et de la construction verbale est proposé dans les référentiels; par exemple ils mettent en avant, sans expliciter la raison de ces choix, l'apprentissage des "1^{er} 2^e et/ou 5^e personnes" (Beacco et al. 2005: 108), ou "des verbes en -er d'usage courant" (Beacco & Porquier 2007: 97), ou encore certains usages des verbes très fréquents avec des constructions variées:

être: + adj. (il est grand), + participe passé (à valeur d'adjectif) (je suis fatigué) de, ville/pays ou année (il est de Londres, elle est de 1996), on est/ nous sommes le + date/jour, + adv. (je suis ici), les expressions avec c'est...

avoir: + âge + ans (j'ai trente ans), j'ai trois enfants, j'ai mal/faim...

faire: il fait beau, faire la cuisine, faire de la musique, faire mal...

devoir, vouloir et pouvoir qui comportent des infinitifs objets (Beacco & Porquier 2007: 93).

2.2 Méthodes et manuels de grammaire en FLE

Pour comprendre la manière dont les constructions verbales sont présentées concrètement aux apprenants débutants, cette recherche passe brièvement en revue le contenu des méthodes et des manuels de grammaire pour les niveaux A1.1 et A1, en prenant en compte cinq titres pour chaque type de moyen d'enseignement¹. Ce choix porte sur des ouvrages variés selon les auteurs, les maisons d'édition et les années de parution.

Dans un premier temps, on constate que les verbes occupent une place importante dans les moyens d'enseignement analysés: en moyenne 57% du contenu des manuels de grammaire et 63% des leçons dans les méthodes contiennent des explications et des activités sur les verbes. Cependant, l'étude de ces manuels, comme ceux choisis par Bruley-Meszaros² montre "que les auteurs privilégient nettement l'aspect morphologique du verbe et plus précisément ce que l'on nomme traditionnellement la conjugaison" (2010: 56). La plupart des manuels consultés suivent le même schéma: ils commencent l'enseignement des verbes avec *être* et *avoir* au présent. Certaines méthodes introduisent quelques formes conjuguées du verbe *s'appeler* au même moment. Ensuite, la conjugaison des verbes en *-er* au présent occupe une place importante, et ces manuels continuent souvent avec les verbes qui se terminent en *-ir* et en *-re*, toujours au présent. Certains manuels enseignent les verbes *aller*, *venir* et *faire* séparément. La seule différence significative entre les deux types d'ouvrages est la façon dont ils enseignent la morphologie verbale, les

¹ Méthodes: Reflets 1 (1999); Alors? Méthode de français fondée sur l'approche par compétence niveau A1 CECR (2007); Écho 1 (2008); Vite et bien 1 (2009); Alter Ego + A1(2012). Manuels de grammaire: Exercices de grammaire en contexte: niveau débutant (2000); Grammaire progressive du français: niveau débutant (2004); Grammaire en dialogue: niveau débutant (2005); 100% FLE Grammaire essentielle du français: niveaux A1/A2 (2015); Pratique grammaire: 640 exercices niveaux A1/A2 (2019).

² Méthodes: Accord, Alter Ego, Café crème, Campus, Champion, Connexions, Escales, Festival, Forum, Initial, Métro Saint-Michel, Panorama, Reflets, Rond-point, Studio 100, Studio 60, Taxi!, Tempo, Tout va bien! Manuels de Grammaire: Atelier FLE, Cours de grammaire française, Exercices communicatifs de la Grammaire progressive du français, Exercices de grammaire (500), Exercices de grammaire en contexte, Exercisier, Français au présent, Français de A à Z, Grammaire des premiers temps, Grammaire en dialogues, Grammaire expliquée du français, Grammaire expliquée du français – exercices, Grammaire française, Grammaire pour l'enseignement/apprentissage du FLE, Grammaire pratique du français en 80 fiches, Grammaire progressive du français, Grammaire utile du français, Grammaire vivante du français, Je pratique – exercices de grammaire, Nouvel Entraînez-vous, Nouvelle Grammaire du Français (Sorbonne), Point par point – exercices de grammaire française.

manuels de grammaire optant dans la majorité des cas pour une méthode plus normative et explicite: une première partie avec la conjugaison du/des verbe(s) et ensuite la deuxième partie avec des exercices portant sur le(s) même(s) verbe(s). La plupart des méthodes consultées, quant à elles, choisissent la méthode inductive: ce sont les apprenants qui doivent trouver les formes conjuguées dans les petits textes ou les phrases présentés plus tôt dans la leçon, et à partir de là ces méthodes suggèrent également quelques exercices. En ce qui concerne les exercices, seule la quantité différencie les méthodes des manuels de grammaire: la majorité des méthodes contient moins d'exercices. Dans les deux cas, les exercices restent assez "classiques", comme compléter des phrases à trous ou écrire des phrases avec les verbes imposés.

Dans un deuxième temps, les concepteurs des manuels s'intéressent aux temps verbaux. Ils commencent tous avec les verbes au présent de l'indicatif, et dans la majorité des cas, continuent avec le passé composé. Ensuite viennent les temps du futur et les autres temps du passé. Dans de rares cas, l'impératif présent fait partie du contenu enseigné.

La seule méthode qui se distingue clairement des autres par rapport à l'enseignement des verbes est "*Alors? Méthode de français fondée sur l'approche par compétence niveau A1 CECR*" (2007). Cette méthode respecte plus fidèlement les propositions des ouvrages de référence. Toutes les leçons comportent un bref point grammatical sur les verbes. Les apprenants font parfois connaissance avec seulement certaines formes conjuguées des verbes (par exemple 1^{re}, 2^e et 5^e personnes). Cette méthode ne suit pas le schéma classique des manuels qui enseignent d'abord différents temps du mode indicatif, puis les autres modes; par exemple, l'impératif présent de certains verbes est introduit dès le début du manuel (p. 34 unité 2) et aussi presque à la fin du manuel (p. 148 unité 10), le verbe *devoir* au présent et au passé composé et le verbe *vivre* au présent y sont enseignés (p. 154 unité 10). Des constructions impersonnelles avec le verbe *falloir* (p. 47, unité 3) et *GN – V – Vinf*, comme: "Il sort acheter le journal." (p. 154, unité 10), souvent négligées dans les autres manuels, sont également présentées dans cette méthode. Ces démarches, même si elles sont à saluer, ne résolvent pas les difficultés des apprenants face aux constructions verbales, car comme dans tous les autres ouvrages consultés, l'enseignement/apprentissage explicite de ce sujet est totalement absent.

3. Choix des verbes

Cette étude ayant une finalité didactique du français langue étrangère, la première démarche dans le choix des verbes vise à répondre à de vrais besoins langagiers (usage quotidien ou professionnel) du public cible. La question qui se pose est donc de savoir comment choisir ces verbes, selon quels critères?

Dans un premier temps, la réponse repose sur deux facteurs primordiaux, à savoir l'affinité (cooccurrence) et la fréquence.

Pour l'affinité, ce sont les référentiels des niveaux A1.1 et A1 qui ont été consultés, car ils proposent un inventaire plus au moins complet de tout ce que les apprenants de ces niveaux doivent connaître. Plus de 140 verbes sont rassemblés dans ces deux ouvrages de référence. Il s'avère impossible pour les apprenants débutants de connaître les constructions d'une telle quantité de verbes dans tous leurs détails, et cela ne fait d'ailleurs pas partie des objectifs des concepteurs de ces ouvrages.

Pour diminuer le nombre de verbes à apprendre et pour sélectionner ceux qui sont les plus pertinents, il est donc nécessaire de retenir les verbes les plus fréquents. Pour ce faire, *A frequency dictionary of French* (Lonsdale & Le Bras 2009)³ ainsi que *Corpus français Leipzig*⁴ ont été consultés. Sachant que ce corpus n'est pas lemmatisé, pour avoir une idée plus précise de la fréquence des verbes, ils ont été examinés sous six formes qui semblaient être les plus représentatives⁵ (infinitif, première, deuxième et troisième personnes du singulier au présent, troisième personne du pluriel de l'imparfait et participe passé masculin singulier). La moyenne de "la classe de fréquence" de ces formes a été considérée comme "la classe de fréquence" des verbes. Les verbes très fréquents selon ces deux ressources ont été retenus; leur nombre se limite à 65 verbes. Sans grande surprise, les deux verbes les plus fréquents sont *être* et *avoir*. Même si la fréquence est un facteur nécessaire pour le choix des verbes, on peut raisonnablement se demander si elle est suffisante pour opérer un choix. Autrement dit, est-ce qu'on doit commencer l'enseignement/apprentissage des verbes et leurs constructions avec *être* et *avoir* comme c'est le cas dans de nombreux moyens d'enseignement?

Pour répondre à cette question, pour chacun de ces verbes, 200 phrases sélectionnées au hasard dans le corpus *Emolex*⁶ ont été examinées.

³ *A frequency dictionary of French* fournit une liste des 5000 mots les plus fréquemment utilisés en français. Basé sur un corpus de 23 millions de mots de français qui comprend des documents écrits et parlés de France et d'outre-mer, ce dictionnaire fournit à l'utilisateur des informations détaillées pour chacune des 5000 entrées, y compris les équivalents anglais, une phrase type, sa traduction anglaise, des statistiques d'utilisation et une indication de la variation du registre.

⁴ https://corpora.uni-leipzig.de/fr?corpusId=fra_mixed_2012, ce corpus fait partie de Leipzig Corpora Collection qui contient 233 collections monolingues pour 219 langues différentes. C'est un corpus évolutif et rassemble environ 75 millions de phrases, soit plus de 1,4 milliard de mots sélectionnés à partir des textes sur les web sites.

⁵ Les raisons de ce choix ont été abordées dans Kamber & Dubois (2014).

⁶ <http://persan.rom.uni-koeln.de/emolex/emoBase/index.php?disconnect>, le corpus Emolex rassemble des corpus dans cinq langues: le français, l'allemand, l'anglais, l'espagnol et le russe. Il comprend deux types de corpus: des corpus comparables d'un côté et des corpus parallèles (ou bilingues) de l'autre. Les données du corpus ont été étiquetées et annotées syntaxiquement. Dans cette recherche, le corpus comparable en français qui comprend environ 140 millions de

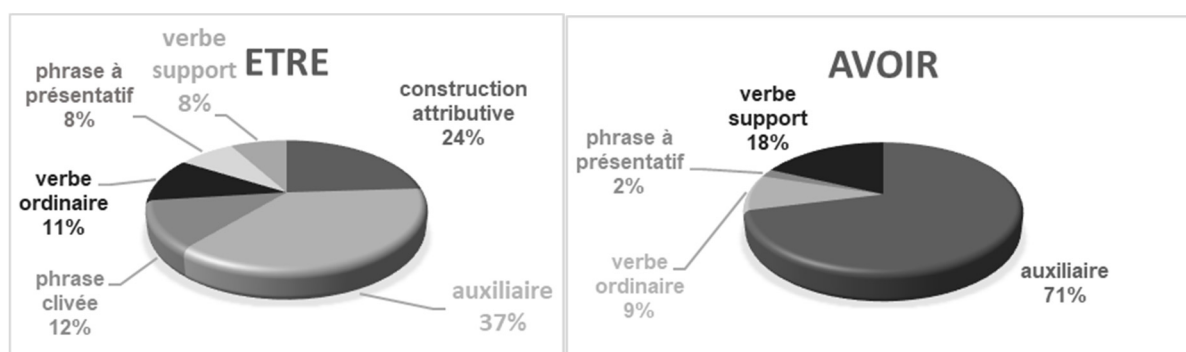


Fig. 1: Comparaison des complémentations – Être, Avoir

Comme on peut le voir dans la Fig. 1, les verbes *être* et *avoir* possèdent des constructions très variées: ils peuvent notamment être auxiliaires, verbes supports et verbes ordinaires.

Les verbes *être* et *avoir* en tant que verbes supports sont suivis dans la majorité des cas d'un nom avec lequel ils construisent des locutions à verbe support (*avoir besoin de*, *avoir lieu*, *être au chômage*, *être d'accord pour* – *Vinf*, etc.). Le verbe *être* peut également être un verbe support dans des locutions adverbiales: *être pour*, *être contre*, etc.

Le verbe *avoir* peut s'utiliser avec un complément direct qui est souvent un nom d'objet (J'ai un vélo.) tandis que le complément du verbe *être* est indirect et introduit par la préposition *à* généralement accompagnée d'un nom de lieu (Ils sont à Marseille.).

Les présentatifs qui "servent à présenter un groupe nominal" (Riegel et al. 2009: 757) sont employés avec les structures *C'est* – *GN* (*C'est la cuisine.*) et *Il y a* – *GN* (*Il y a un match ce soir.*).

Comme on peut le constater, *être* et *avoir* possèdent des structures très variées et cette pluralité de complémentations rend l'apprentissage de ces verbes particulièrement difficile pour les apprenants débutants (voir Noghrechi 2018). Il sera donc plus pertinent de choisir des verbes présentant des régularités morphosyntaxiques plus marquées pour introduire les constructions verbales en classe.

Parmi différents groupes de verbes, ceux qui se terminent en *-er* possèdent plus de régularités morphologiques. Cela permet aux apprenants de les employer plus facilement. Dans notre base de données, 37 verbes remplissent toutes les conditions susmentionnées: affinité, fréquence et se terminant en *-er*.

En se référant aux référentiels des niveaux A1.1 et A1, on constate que les apprenants, en plus des constructions impersonnelles (avec *falloir*), les phrases clivées, détachées et tronquées, doivent connaître et employer les structures

mots (environ 120 millions de mots de textes journalistiques ainsi qu'environ 20 millions de mots de textes littéraires) a été conçu.

suivantes: *GN – V*, *GN – V – GN*, *GN – Vêtre – Adj*, et quelques constructions avec *GN – V – Vinf* (Beacco et al. 2005: 109; Beacco & Porquier 2007: 98). Les apprenants doivent également comprendre dans certains cas les structures avec *GN – V – que P* (complétives) (*Ibid.*) mais ils ne sont pas encore en mesure de les employer eux-mêmes. Pour présenter des constructions verbales en classe, il est donc pertinent de ne pas enseigner dans un premier temps des verbes qui ont une attirance significative pour cette dernière structure. Pour ce faire, les voisins de droite et les cooccurrences significatifs des 37 verbes⁷ sous les six formes susmentionnées ont été étudiés dans *Corpus français Leipzig*. Les cinq verbes qui ont une attirance significative pour les complétives (*décider, expliquer, penser, raconter, répéter*) ont été éliminés.

De nombreux verbes, surtout les plus fréquents, possèdent des complémentations variées. Pour faciliter la tâche des enseignants et des apprenants, les verbes qui ont la/les structure(s) la/les plus régulière(s) seront sélectionnés. Pour cela les six formes susmentionnées des verbes restants ont été étudiées dans *Corpus français Leipzig*. Dans cette démarche, on s'intéresse particulièrement aux voisins de droite ainsi qu'aux cooccurrences significatives de chaque verbe. Dix verbes sont employés avec des complémentations moins variées: *aimer, écouter, envoyer, essayer, fermer, garder, manger, regarder, retirer* et *visiter*.

Le dernier critère de choix est purement sémantique. Parmi les quatorze séries de verbes proposées par Mathieu-Colas (2006), seules deux ont été choisies: "actions et activité" et les "verbes relatifs au corps". Pour enseigner ces verbes, l'enseignant peut avoir recours à plusieurs outils comme des images, des dessins ou il peut même mimer les verbes. Seuls quatre verbes ont rempli toutes ces conditions, les verbes *écouter, fermer, manger* et *regarder*. C'est donc avec ces quatre verbes que l'enseignement des constructions verbales aux apprenants débutants peut commencer.

4. Analyse

Pour pouvoir enseigner les constructions verbales en classe, l'enseignant a besoin de connaître celles pour lesquelles les verbes ont une attirance particulière. Dans le cas de cette étude, les constructions des quatre verbes susmentionnés ont été étudiées dans 200 phrases sélectionnées au hasard dans le corpus *Emolex*.

⁷ Les 37 verbes sont les suivants: acheter, aimer, arriver, changer, chanter, compter, continuer, décider, demander, écouter, envoyer, essayer, étudier, expliquer, fermer, gagner, garder, habiter, inviter, jouer, manger, marcher, parler, payer, penser, (se) préparer, prononcer, raconter, regarder, répéter, réserver, retirer, toucher, tourner, travailler, visiter, voter.

4.1 Écouter

Dans les phrases analysées, les deux structures pertinentes pour le verbe *écouter* sont les suivantes: *N0 – V – N1* (78%) et *N0 – V* (12%). Dans 10% des cas, les résultats ne sont pas pertinents:

- (1) "Mais entre *écouter* et agir, il y a une différence", a déclaré Farid, 22 ans, après avoir exposé ses difficultés à François Bayrou.

Comme nous pouvons le voir dans la Fig. 2, le verbe *écouter* a une attirance significative pour les noms humains en tant que sujet. Ce verbe est employé plus fréquemment avec un complément direct, dans la plupart des cas les noms humains, les éléments sonores ou la musique (abstrait). Dans ces phrases, l'emploi de modificateurs du type "modification d'un procès-verbal" (Riegel et al. 2009: 651) (*attentivement, gratuitement, religieusement*) est constaté.

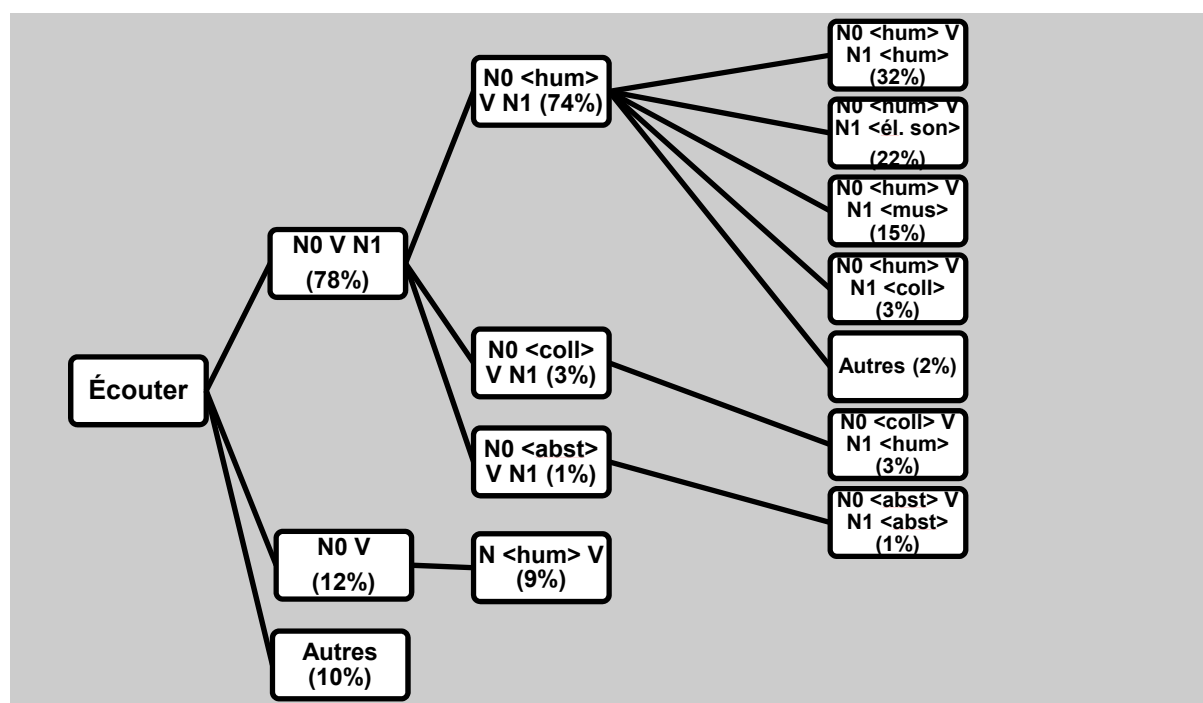


Fig. 2: Bilan de l'analyse – Ecouter

En tenant compte des éléments ci-dessus, les exemples suivants peuvent être pris en considération:

- (2) Dans ces locaux simples qu'on veut bien lui prêter, **Mark Spinks, le fondateur de l'association**, *écoute les hommes de sa communauté*, touchés, plus que d'autres, par le chômage ou l'alcoolisme.
- (3) Il *écoute le débat* sur la lecture: "Ceux qui sont pour la méthode globale se trompent, ceux qui sont pour la méthode syllabique aussi."
- (4) "On est de grands passionnés *de musique*, on *en écoute* partout", confie François Floret.

4.2 Fermer

Comme on peut le voir dans la Fig. 3, le verbe *fermer* a également deux structures courantes avec (70%) ou sans complément direct (22%). Dans les

structures sans complément, les noms de lieu sont les sujets les plus fréquents. Et, dans les structures avec complément les noms humains sont les sujets les plus courants suivis par les noms collectifs et les noms de lieu. Les compléments les plus fréquents sont les noms de lieu, les objets, ou les parties du corps. Même si certaines phrases comportent des modifieurs, le plus souvent du type "modification d'une expression quantifiée" (Riegel et al. 2009: 651) (*temporairement, définitivement*), l'usage des modifieurs avec ce verbe reste restreint.

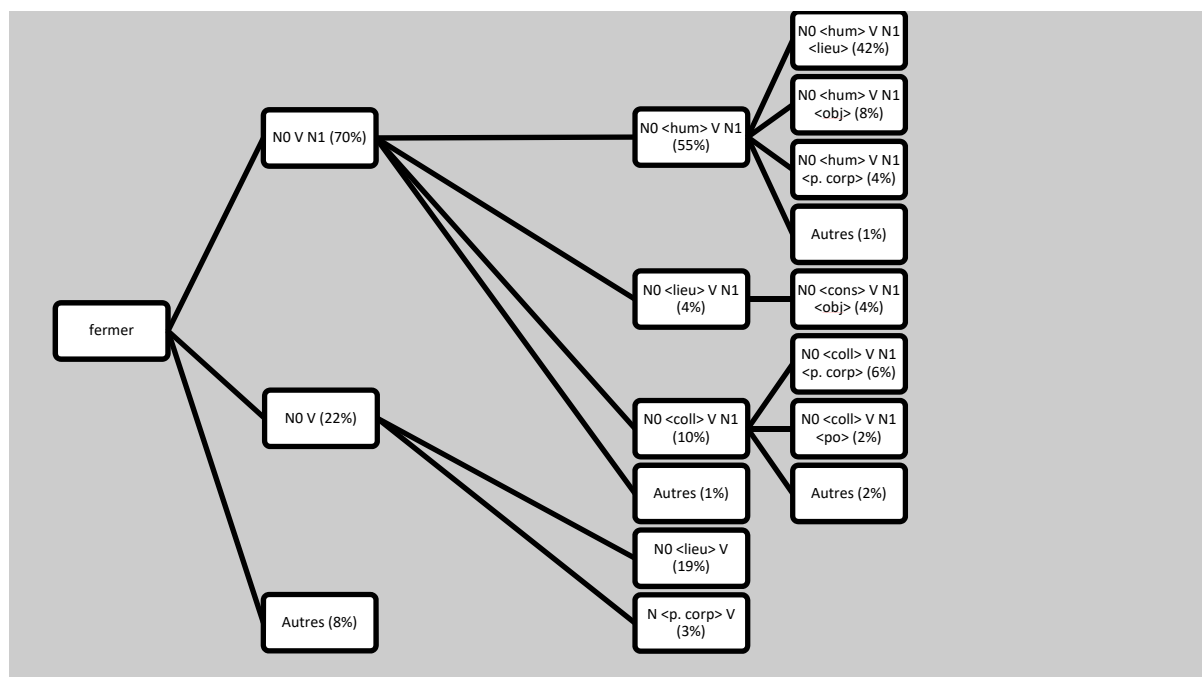


Fig. 3: Bilan de l'analyse – Fermer

En tenant compte de ces éléments, les exemples prototypiques pour ce verbe sont les suivants:

- (5) À Dakar (Sénégal), **des commerçants** ont *fermé* leur magasin à plusieurs reprises depuis 2002 pour protester contre "l'invasion massive des Chinois dans le secteur du commerce sénégalais", dénonçant l'arrivée de marchandises à bas prix et le non-respect de la fiscalité par leur concurrent asiatique.
- (6) **La police** a *fermé* les ponts et ordonné à la population de quitter les rues.
- (7) **La banque de l'hôpital Saint-Louis**, qui *avait fermé* en 2003 faute de crédits, devrait rouvrir au printemps 2007.

4.3 Manger

Le verbe *manger* n'a pas d'attirance particulière pour un complément direct, il peut autant être employé avec (41%) que sans complément direct (39%). Il est très régulier au niveau de la nature des noms qu'il attire comme sujet; tous les noms relevés dans les phrases analysées sont les noms animés: humains, collectifs ou animaux. Dans la plupart des cas, les aliments remplissent le rôle

de complément direct, cependant dans 8% des cas, les objets occupent cette place. L'usage des modifieurs avec ce verbe, surtout dans les constructions sans complément, est plus courant que chez les deux verbes précédents. "La modification d'une expression quantifiée", "l'indication du degré" et "la modification d'un rapport de caractérisation" (Riegel et al. 2009: 651-653) sont les sous-catégories les plus employées: *beaucoup*, *compulsivement*, *correctement*, *davantage*, *dehors*, *ensemble*, *lentement*, *normalement*, *proprement*, *sainement*, *suffisamment*, *tranquillement*, *trop*.

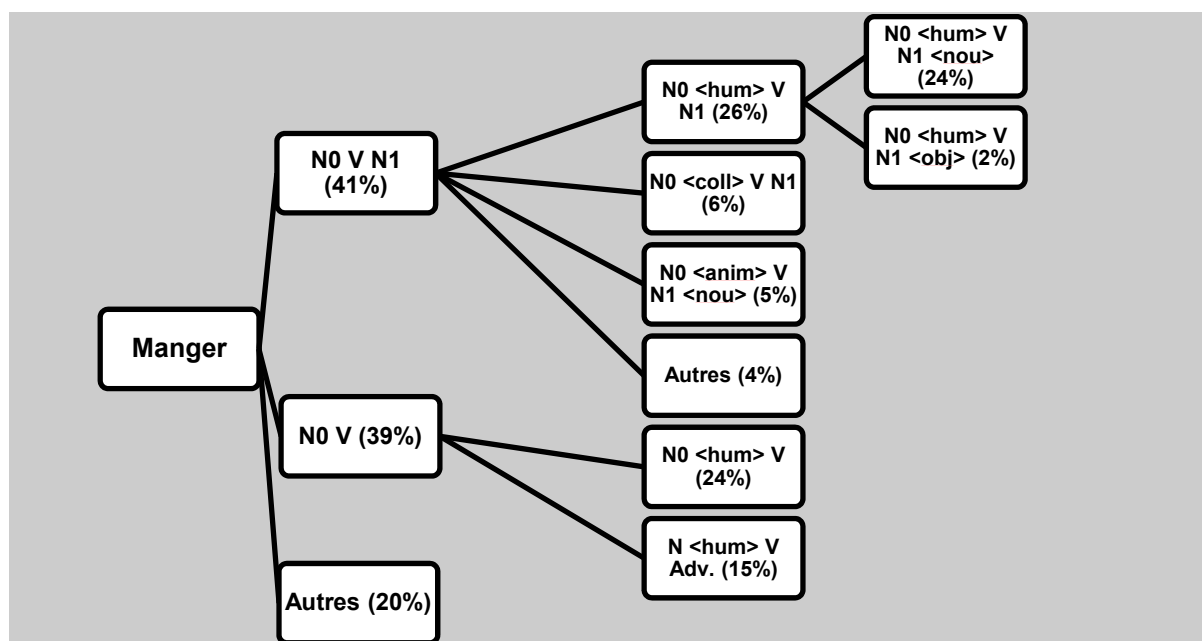


Fig. 4: Bilan de l'analyse – Manger

Les exemples suivants représentent les cas les plus constatés:

(8) À Lyon, un tiers des élèves ne mangent plus le plat principal, jugé impur.

Souvent dans les phrases avec le complément direct, on peut constater l'usage d'un adverbial:

(9) Je jongle avec les factures et je mange beaucoup de pâtes.

Les verbes à l'impératif sont très courants:

(10) Pratiquez une activité physique régulière, mangez au moins cinq fruits et légumes par jour.

(11) Quand je passe à table, je mange.

(12) Il ne fumait pas, se couchait tôt, mangeait sain et pratiquait le yoga.

4.4 Regarder

Regarder a une attirance significative pour les constructions avec un complément direct (73%). La plus grande partie des sujets sont les noms humains (62%), ensuite viennent avec un écart important les noms collectifs (9%) et abstraits (2%). Le complément peut être les noms humains, mais aussi les noms abstraits et les émissions télévisées ou les films. 40% des phrases

analysées comportent un modifieur du type "modification d'un procès verbal" (Riegel et al. 2009: 651-653) (*attentivement, bizarrement, gratuitement, tranquillement*).

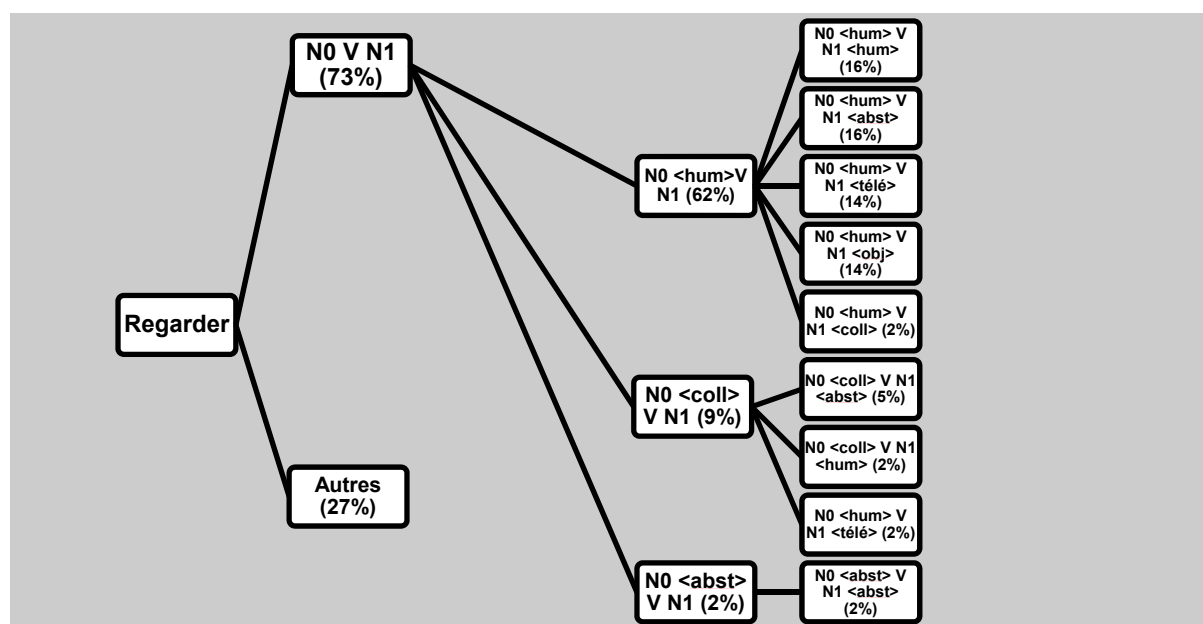


Fig. 5: Bilan de l'analyse – Regarder

Les exemples prototypiques de ce verbe peuvent être les suivants:

- (13) Une bonne raison pour que le ministre de la Santé, Xavier Bertrand, décide la création d'un groupe de travail sur l'image du corps: "Il est important, au moment où les **jeunes filles regardent les mannequins**, de s'interroger sur le véritable impact de ces images sur le comportement".
- (14) Si **vous regardez de près sa proposition** sur le système de couverture-maladie, il va bien plus loin que des démocrates.
- (15) **Les téléspectateurs regardent la télévision différemment**.

4.5 Résumé

Le Tab. 1 résume les complémentations les plus fréquentes de chacun de ces verbes. Les enseignants peuvent en tenir compte pour introduire les constructions verbales aux apprenants débutants.

Écouter	N0 <hum>	V	N1 <hum>, <él.son>, <mus>
Fermer	N0 <hum>	V	N1 <cons>, <obj>, <p. corp>
Manger	N0 <hum> N0 <hum>	V V	N1 <nou> (Adv.)
Regarder	N0 <hum>	V	N1 <hum>, <abst>, <télé>, <obj>

Tab. 1: Bilan de l'analyse – Ecouter, Fermer, Manger, Regarder

5. Propositions didactiques

Les précédentes parties de cet article ont répondu à la question: que faut-il enseigner pour introduire des constructions verbales en classe de FLE? Cette partie répondra à la question: comment l'enseigner?

L'enseignement des constructions verbales se fera en trois étapes. Dans un premier temps, les apprenants feront connaissance avec quelques exemples prototypiques pour les constructions de chaque verbe sélectionné. Pour choisir ces exemples, les histoires écrites en langage simplifié s'adressant aux apprenants débutants⁸ constitueront des ressources intéressantes grâce à leur lexique plus accessible. En classe, l'enseignant lit les premières phrases avec les apprenants et les guide en leur posant des questions sur la structure des phrases: quel est le sujet de cette phrase? quel est le verbe? les verbes, ont-ils des compléments? Pour que les apprenants deviennent autonomes dans leur apprentissage, il est nécessaire de les sensibiliser dès les premiers pas à la structure des phrases en français. Il faut préciser qu'un travail préparatoire est nécessaire pour que les apprenants comprennent les termes susmentionnés. Ensuite, il leur demande de travailler en binôme et de trouver les réponses ensemble. Enfin, ils discutent en classe pour comparer les réponses de chaque binôme.

Lisez les phrases. Remplissez le tableau.

- 1) Laura écoute la radio.
- 2) Jean-Baptiste écoute plein de gens de milieux très différents!
- 3) Élise ferme la fenêtre, prend son sac et sort.
- 4) Avant de regarder le deuxième indice, j'ai mangé un croque-monsieur.
- 5) Et je mange proprement!
- 6) L'enfant sursaute et regarde sa grand-mère.
- 7) Ils regardent les magasins, achètent des cafés, marchent lentement.

	Sujet	Verbe	Complément
1)	Laura	écoute	la radio
2)			
3)			
4)			
5)			
6)			
7)			

Image 1: Exemple d'exercice

⁸ Citons à titre d'exemples la collection LFF (édition Hachette) et la collection Mondes en VF (édition Didier).

Dès les premiers pas, l'enseignant peut encourager les apprenants à établir un lien entre leur langue première et le français. Pour ce faire, il peut sélectionner quelques phrases dans les deux langues dans des corpus parallèles comme *Corpus parallèle de l'Organisation des Nations Unies*⁹ ou des dictionnaires bilingues en ligne comme *Linguee*¹⁰, *Glosbe*¹¹ ou *Reverso*¹². Dans ce cas, les apprenants avec la même langue première peuvent travailler ensemble pour comparer les complémentations des verbes entre le français et leur L1.

Lisez les phrases en français et en russe. Complétez le tableau.

1) **Gardner** *mange* le sandwich¹³.

Гарднер *ест* бутерброд.

2) **Je** *mange* aussi vite que toi.

Я *ем* с такой же скоростью, как и ты.

3) Selon les estimations, **90% des hommes et 87% des femmes** *écoutent* la radio.

По оценкам, **90% мужчин и 87% женщин** *радио слушают* 90% мужчин и 87% женщин.

4) **Ils** *ferment* toutes les portes.

Они *закрывают* все двери.

5) **Ils** *ne le regardent pas* comme un enfant.

Они *не смотрят* на него так, будто ОН ребёнок.

6) **Tu** *regardes trop de* films américains.

Это все из-за американских фильмов, что **ты** *смотришь*.

	Français			Russe		
	Sujet	Verbe	Complément	Sujet	Verbe	Complément
1)	Gardner	mange	le sandwich	Гарднер	ест	бутерброд
2)						
3)						
4)						
5)						
6)						

Image 2: Exemple d'exercice

En comparant ces 6 phrases, les apprenants russophones peuvent trouver les ressemblances et les différences entre leur langue première et le français. Dans les deux langues, les verbes *écouter* (*слушать*) et *fermer* (*закрывать*) ont une attirance significative pour les compléments directs et le verbe *manger*

⁹ <https://conferences.unite.un.org/UNCORPUS/fr#test>

¹⁰ <https://www.linguee.fr/>

¹¹ <https://fr.glosbe.com/>

¹² <https://context.reverso.net/traduction/>

¹³ Les phrases sont sélectionnées dans *Reverso*.

(*есть*) peut être employé avec ou sans complément direct. De manière générale, la structure des phrases affirmatives en français est plus figée (sujet – verbe – complément) qu'en russe (exemples 3 et 6). Et enfin, le verbe *regarder* en français précède un complément direct tandis que son équivalent en russe (*смотреть*) peut être employé avec un complément direct ou indirect (avec la préposition *На*) selon le contexte.

Dans un deuxième temps, pour permettre aux apprenants de pratiquer à la maison, l'enseignant propose des activités sur une plateforme en ligne, comme Symbaloo¹⁴. Cette plateforme permet de créer un parcours complet avec le nombre d'étapes envisagées par l'enseignant. Elle donne la possibilité aux enseignants de voir le parcours de chaque apprenant, le temps nécessaire pour répondre, le nombre de réponses correctes et erronées, ou même le nombre d'essais nécessaire pour répondre correctement à une question. Dans un premier temps, les apprenants pourront regarder une courte vidéo pour apprendre/réviser la conjugaison de chaque verbe. Ensuite vient une série d'exercice sur la conjugaison des verbes et les apprenants écriront la forme correcte de chaque verbe et puis grâce aux enregistrements sonores, ils entendront des formes conjuguées (p. ex. elle écoute) et les écriront, quand plusieurs réponses sont possibles, ils doivent les écrire tous (p. ex. il ferme, ils ferment). Ensuite, ils liront des phrases et choisiront le verbe parmi les trois réponses proposées. Puis, ils trouveront le complément de chaque verbe parmi les quatre images proposées. Et enfin, parmi les six réponses proposées, ils choisiront le(s) complément(s) adapté(s) pour chacun des verbes étudiés. L'enseignant met à disposition de ses apprenants la/les bonne(s) réponse(s) pour chaque question et, si cela lui paraît nécessaire, il peut mettre à leur disposition quelques indices pour certaines questions. L'analyse des réponses des apprenants permet à l'enseignant d'identifier les difficultés des apprenants et par conséquent, il peut adapter le contenu de la séance suivante par rapport aux besoins réels de son public.

Enfin, à nouveau en classe, l'enseignant rebondira sur les erreurs commises par les apprenants et donnera des explications sur les exercices qui ont particulièrement posé problème. Ensuite, il proposera des activités de production orale et écrite, pour que les apprenants puissent employer ces verbes dans un contexte.

6. Conclusion

Comme nous l'avons vu dans cet article, pour introduire des constructions verbales en classe de FLE de niveau débutant, il faut d'abord choisir les verbes appropriés. Grâce à leur accessibilité, les verbes d'action très fréquents en *-er* possédant des structures régulières se prêtent plus facilement à cet exercice.

¹⁴ <https://www.symbaloo.com/home/mix/13eOhK6OgG>

Dès les premiers pas, l'enseignant peut encourager les apprenants à comparer des constructions entre leur langue première et le français. Pour aider les apprenants en difficulté, il peut faire appel aux corpus parallèles ou aux dictionnaires bilingues. Ces ressources permettent aux apprenants de mieux comprendre les constructions et d'apprendre en autonomie. Pour approfondir ce sujet, l'enseignant peut également proposer des activités sur différentes plateformes en ligne permettant de voir les réponses correctes et erronées des apprenants. Cela lui permet de préparer ses cours en fonction des vrais besoins de ses apprenants.

BIBLIOGRAPHIE

- Ågren, M. (2008). *À la recherche de la morphologie silencieuse: Sur le développement du pluriel en français L2 écrit*. Thèse de doctorat. Université de Lund.
- Ågren, M. (2014). Production et compréhension de la morphologie verbale en FLE: les difficultés de l'accord sujet-verbe en nombre. *Cahiers de l'Association for French Language Studies*, 19(1), 1-22.
- Beacco, J.-C. & Porquier, R. (2007). *Niveau A1 pour le français. Un référentiel*. Paris: Didier.
- Beacco, J.-C., de Ferrari, M., Lhote, G. & Tagliante, C. (2005). *Niveau A1.1 pour le français. Référentiel et certification (DILF) pour les premiers acquis en français*. Paris: Didier.
- Bruley-Meszaros, C. (2010). Quel enseignement des verbes en didactique du français langue étrangère? *Synergies France*, 6, 51-59.
- Conseil de l'Europe (2001). *Cadre européen commun de référence pour les langues: apprendre, enseigner, évaluer*. Paris: Didier.
- Conseil de l'Europe (2018). *Cadre européen commun de référence pour les langues: apprendre, enseigner, évaluer. Volume complémentaire avec de nouveaux descripteurs*. <https://rm.coe.int/cecr-volume-complementaire-avec-de-nouveaux-descripteurs/16807875d5> (consulté le 10 janvier 2020).
- Feuillet, J. (1978). Étude fonctionnelle de la phrase I. *Cahiers d'allemand*, 13, 100-120.
- Feuillet, J. (éd.). (1998). *Actance et valence dans les langues de l'Europe*. Berlin/New York: Mouton de Gruyter.
- Gerolimich, S. & Stabarin, I. (2007). Pour une représentation fonctionnelle de la conjugaison française. In C. Taylor (éd.), *Aspetti della Didattica e dell'Apprendimento delle Lingue Straniere* (pp. 147-165). Trieste: EUT Edizioni Università di Trieste.
- Gerolimich, S. & Stabarin, I. (2014). Une systématisation de la conjugaison française: feed-back et remodèlements didactiques. In M.-N. Roubaud & J.-P. Sautot (éds.), *Le verbe en friche*, (pp. 111-131). Série GRAMM-R. Études de linguistique française, Vol. 20, Bruxelles: Peter Lang.
- Granfeldt, J. & Schlyter, S. (éds.) (2005). Acquisition et production de la morphologie flexionnelle. Actes du Festival de la morphologie. Lund, *Perles*, 20.
- Granget, C. (2005). Développement de l'accord verbal avec un sujet pluriel dans les récits écrits d'apprenants germanophones scolarisés du français, Acquisition et production de la morphologie flexionnelle. Lund, *Perles*, 20, 111-123.
- Granget, C. (2009). L'acquisition de la morphologie verbale. In D. Véronique (éd.), *L'acquisition de la grammaire du français, langue étrangère* (pp. 164-218). Paris: Didier.
- Herschensohn, J. (2001). Missing inflection in second language French: accidental infinitives and other verbal deficits. *Second Language Research*, 17, 273- 305.

- Howard, M. (2006). The expression of number and person through verb morphology in French interlanguage. *International Review of Applied Linguistics*, 44(1), 1-22.
- Howard, M. (2013). La complexification du système morphologique de l'apprenant avancé: Une étude longitudinale de la morphologie verbale. In U. Paprocka-Piotrowska, C. Martinot & S. Gerolimich (éds.), *La complexité en langue et son acquisition* (pp. 337-357). Lublin: Towarzystwo Naukowe KUL.
- Kemps, N., Housen, A. & Pierrard, M. (2010). Instruction grammaticale et stabilisation de la morphologie verbale. L'accord sujet-verbe à la 3^e personne. In D. V. Raemdonck (éd.), *Enseigner les structures langagiers en FLE* (pp. 251-270). Série GRAMM-R. Études de linguistique française, Vol. 4. Bruxelles: Peter Lang.
- Lazard, G. (1994). *L'actance*. Paris: PUF.
- Mathieu-Colas, M. (2006). Les classes de verbes: syntaxe et sémantique. *Le traitement du lexique. Catégorisation et actualisation*, 10-24. https://halshs.archives-ouvertes.fr/halshs-00768381/file/Classes_de_verbes.pdf (Consulté le 10 janvier 2020).
- Michot, M.-E. (éd.) (2014). L'acquisition de la morphologie verbale en FLE. *Cahiers de l'AFLS*, 19(1).
- Michot, M.-E. & Pierrard, M. (2014). L'enseignement/apprentissage de la morphologie verbale en FLE: savoir "à enseigner" et "savoir acquis". *Le français dans le monde: Recherches et applications*, 55, 57-75.
- Nadasdi, T. (2001). Agreeing to disagree: Variable Sujet-Verb Agreement in Immersion French. *Revue Canadienne de linguistique appliquée*, 4(1-2), 87-101.
- Noghrechi, H. (2018). Enseignement/apprentissage des verbes "être" et "avoir": éclairage des corpus d'apprenants et d'experts. *Études en Didactiques des Langues, Les corpus*, 31, 37-52.
- Prévost, P. (2009). *The acquisition of French. The development of inflectional morphology and syntax in L1 acquisition, bilingualism, and L2 acquisition*. Amsterdam: John Benjamins.
- Prévost, P. White, L. (2000). Missing surface inflection or Impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, 16, 103-133.
- Riegel, M., Pellat, J.-C., Rioul, R. (2009). *Grammaire fonctionnelle du français*. Paris: PUF.
- Schlyter, S. (2006). *Le développement de la morphologie verbale de personne-nombre chez différents types d'apprenants*. Lund Universitet.
- Sergeeva, E. & Chevrot, J.-P. (2008). Influence de la fréquence sur la production et l'acquisition des formes verbales du français: le cas des russophones adultes. In J. Durand, B. Habert & B. Laks (éds.), *Congrès Mondial de Linguistique Française* (pp. 1831-1844). Paris: Institut de Linguistique Française.
- Tesnière, L. (1959). *Éléments de syntaxe structurale*. Paris: Klincksieck.
- Thomas, A. (2009). *Les apprenants parlent-ils à l'infinitif? Influence de l'input sur la production des verbes par des apprenants adultes du français*. Thèse de doctorat. Université de Lund.

Corpus et exercices en ligne, complémentarité ou redondance? L'exemple des pronoms relatifs *qui* et *que*

Anita THOMAS

Université de Fribourg

Département de Plurilinguisme et Didactiques des Langues Étrangères

Avenue de l'Europe 20, 1700 Fribourg, Suisse

anita.thomas@unifr.ch

The aim of this article is to discuss the use of online corpora as a resource for foreign language teaching. Corpora allow access to the language in its use, which can be considered more authentic than that found in exercises. This study is based on an analysis of the French relative pronouns *qui* ('who') and *que* ('which/that') in their most basic use. An analysis of data from learners of L2 French indicates a preference for the use of *qui* with animate antecedents and *que* for inanimates, but there are differences according to the tasks and the learners' L1. Correct production of relative pronouns seems to be more difficult in written than in oral production and less correct with *que* than with *qui*. A comparison between the teaching material available on the *Point du FLE* website and a sample of the ScienQuest written corpus reveals a similar distribution of the two relative pronouns. The discussion is focussing on the contribution of the comparison of different data sources for the development of corpus-informed teaching materials.

Keywords:

corpus informed teaching, learner corpus research, French L2, relative clause, foreign language teaching.

Mots-clés:

corpus et didactique, corpus d'apprenants, FLE, pronoms relatifs, didactique des langues étrangères.

1. Introduction: les corpus comme input de qualité

Pendant longtemps l'enseignement des langues étrangères a été basé sur le matériel fourni dans le manuel de cours ainsi que sur celui transmis par l'enseignant·e. Les personnes qui souhaitent avoir plus d'input, sans se rendre dans une région où la langue est parlée, pouvaient essayer de lire des romans, des journaux ou écouter la radio. À l'ère digitale, toutes ces ressources supplémentaires sont disponibles en un clic. Cette situation est une aubaine pour les apprenant·e·s motivé·e·s qui souhaitent s'exposer le plus possible à la langue cible. Pour les enseignant·e·s, elle demande aussi des compétences pour évaluer la qualité des ressources à disposition et parfois aussi pour rendre ces ressources accessibles en classe de langue. Il est donc important de mener des études évaluant la qualité de celles-ci.

L'objectif de l'étude présentée dans cet article est d'analyser et de discuter la qualité de ressources disponibles en ligne, et plus particulièrement de l'apport des corpus de langue cible comparé à des exercices en libre accès. Pour cela, je me concentrerai sur l'enseignement-apprentissage d'un phénomène précis

en français langue étrangère, à savoir les pronoms relatifs *qui* et *que* dans leur utilisation la plus simple. L'objectif de la comparaison entre les deux types de ressource est de contribuer à la discussion sur l'utilisation des corpus pour l'enseignement, en termes de qualité d'input.

2. Considérations théoriques

2.1 Corpus et enseignement des langues étrangères

Le nombre de publications traitant de l'utilisation des corpus pour l'enseignement n'a cessé d'augmenter durant les dix dernières années, à tel point qu'on a vu en 2017 la parution d'une *timeline* et d'une méta-analyse (Boulton & Cobb 2017). Par corpus, on entend traditionnellement "a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language" (Sinclair 1996: 4). Alors que les corpus sont des outils utilisés depuis longtemps pour l'analyse des conversations (voir par exemple le corpus CLAPI¹) ou du développement langagier (voir par exemple la base de données CHILDES²), ce n'est qu'avec l'avènement d'internet que les corpus ont été rendus accessibles et donc exploitables par le public. L'accès à de larges collections de langue dans son usage a donné lieu à une forme de regain d'intérêt pour la description linguistique, cette fois à partir de données empiriques (Tyne et al. 2014; Williams 2005), renouvelant ainsi la traditionnelle opposition entre la langue comme système et la langue comme corpus (Mayaffre 2005). Comme l'expliquent Boulton & Tyne (2014: 42-45), la linguistique traditionnelle est plutôt basée sur une réflexion menant à une théorie qui reflète la langue comme système. Dans cette approche, l'introspection et l'intuition jouent un rôle important. La linguistique de corpus se base en revanche sur l'observation de données qui reflètent la langue dans son usage, ce qui implique un certain niveau de variation. Ces descriptions-ci sont par contre basées sur des données empiriques. D'une certaine manière, ces deux approches correspondent aussi aux deux sortes de ressources traitées dans la présente étude: les exercices sont souvent construits à partir d'une réflexion sur la grammaire de la langue cible, alors que les données de corpus correspondent à des données d'usage. Ainsi, dans une visée de didactique des langues étrangères, plutôt que de considérer ces deux approches de manière opposée, on pourrait les voir comme complémentaires comme le suggèrent Boulton & Tyne (2014: 42).

Dans sa *research timeline* sur l'utilisation des corpus pour l'enseignement, Boulton (2017) présente 46 articles publiés touchant à différents aspects, tels que les attitudes des apprenant·e·s par rapport aux corpus, la manière d'utiliser les corpus ainsi que les effets de ceux-ci sur l'apprentissage ou comme

¹ <http://clapi.ish-lyon.cnrs.fr> (page consultée le 10 mars 2021).

² <https://childes.talkbank.org/access/> (page consultée le 10 mars 2021).

ressource de référence. Les avantages présumés de l'utilisation des corpus sont liés au fait que "such contact [with language through corpora] provides the massive contextualised exposure needed for language learning, but in a more controlled way than purely haphazard exposure via regular reading or listening, thus promoting or enhancing noticing, language awareness, autonomy, and ultimately producing 'better learners'." (Boulton 2017: 483).

Globalement, les résultats s'avèrent plutôt positifs. La *timeline* révèle que les corpus sont surtout utilisés pour le développement de l'écrit (académique), dans le cadre de séquences dans lesquelles les apprenant·e·s sont invité·e·s à comparer leur production avec des textes de corpus. Elle révèle aussi que l'utilisation des corpus donne de bons résultats avec des apprenant·e·s débutant·e·s. La méta-analyse de Boulton & Cobb (2017) confirme les aspects positifs de l'utilisation des corpus pour l'enseignement, en particulier pour le développement du vocabulaire et de *patterns* lexico-grammaticaux. Par contre, les études ayant examiné l'effet de l'utilisation des corpus sur le développement de la grammaire sont rares (Boulton & Cobb 2017: 383). Néanmoins, cette méta-analyse permet aux auteurs de constater que: "... we reach the somewhat surprising and possibly encouraging conclusion that DDL [Data Driven Learning] works pretty well in almost any context where it has been extensively tried." (Boulton & Cobb 2017: 386).

D'autres études actuelles soulignent les avantages de l'utilisation des corpus pour l'enseignement des langues étrangères: ils permettent notamment d'avoir un contact avec la variation et la culture de la langue cible (André 2018; Gajo et al. 2019: 189ss) et on voit de plus en plus d'études traitant du développement de l'enseignement des langues étrangères sur la base de données d'usage de la langue (Larrivée & Lefevre 2020; Tyler et al. 2018).

Un autre avantage de l'utilisation des corpus est qu'ils permettent "l'observation et la description des usages authentiques de la langue." (Jacques 2005: 24 en gras dans le texte). L'aspect authentique des corpus est également un argument pour leur utilisation en didactique des langues étrangères (Boulton & Tyne 2014: 22; Gajo et al. 2019: 189ss; Gilmore 2004; Vialleton & Lewis 2014). De ce point de vue, les corpus permettent un contact direct avec une utilisation dite réelle de la langue, ce qui est considéré par beaucoup comme motivant pour les apprenant·e·s. Le contact avec la langue dans son utilisation permet aussi aux enseignant·e·s de vérifier leurs intuitions au niveau de la langue dans son usage. Finalement, l'utilisation des corpus pourrait être particulièrement pertinente pour l'enseignement de l'oral et de l'interaction; paradoxalement, c'est un aspect de la langue qui est souvent sous- ou mal représenté dans les manuels de langue (Ravazzolo & Etienne 2019).

Mais l'utilisation des corpus comme matériel authentique ne présente pas que des avantages. Une des caractéristiques des corpus est qu'ils contiennent beaucoup de "bruit", des données qui ressemblent à ce que l'on cherche mais

qui ne le sont pas. Ainsi le travail avec les corpus demande un effort de sélection des données ou ce que Cuq & Gruca appellent une analyse "prépédagogique" (Cuq & Gruca 2005: 433). Les utilisateurs sont aussi confrontés à des difficultés techniques et il est parfois difficile de trouver ce que l'on cherche. Il s'agit donc d'une approche qui peut s'avérer chronophage. Ainsi, malgré l'enthousiasme pour l'utilisation des corpus comme ressource didactique, il est nécessaire de voir dans quelle mesure ils sont complémentaires à des exercices en libre accès, voire meilleurs que ceux-ci. Cette étude a ainsi pour but de comparer ces deux sources d'input pour les apprenant·e·s. Pour cela, je prendrai l'exemple d'un phénomène grammatical précis, à savoir les pronoms *qui* et *que* dans leur utilisation la plus simple (voir plus bas).

2.2 Apprentissage d'une langue étrangère et simplification

Les recherches sur l'apprentissage des langues étrangères ont montré que les apprenant·e·s passent par différents stades de production de la L2 et que ces stades reflètent différentes manières de traiter la langue cible. Parmi celles-ci, on trouve des processus de simplification de la langue cible (Ellis et al. 2016; Ortega 2009: 116ss), probablement pour pallier la surcharge cognitive liée au traitement de la nouvelle langue à apprendre. Cette simplification donne lieu à l'assignation d'une forme par fonction, un processus connu en tant que *one-to-one principle* (Andersen 1984). Ce processus est basé sur la sensibilité à la fréquence des formes et aux biais disponibles dans l'input. Selon ce principe, les apprenant·e·s en début d'apprentissage simplifient la langue cible en reproduisant ces biais, ce qui mène à une sur-généralisation des formes. Ce n'est qu'avec une certaine expérience de la langue (Bybee 2008) que les apprenant·e·s réussiront à percevoir des distinctions plus fines qui mèneront à leur tour à des restructurations et ainsi à une progression vers une langue reflétant de plus en plus les usages de la langue cible.

Certaines études suggèrent que les apprenant·e·s ont parfois de la peine à quitter des réflexes de simplification et de sur-généralisation (cf. fossilisation Klein & Perdue 1997). En ce sens, la classe de langue est une chance, car le matériel didactique proposé peut être construit de manière à soutenir la progression en veillant à ce que les apprenant·e·s aient la possibilité de dépasser les simplifications et les sur-généralisations en leur présentant plus spécifiquement ce qu'ils ne semblent pas noter (Collins et al. 2009; Lyster & Rebuffot 2002).

2.3 Les pronoms relatifs simples *qui* et *que*

Cette étude se concentre sur l'utilisation des deux pronoms relatifs simples *qui* et *que* (Riegel et al. 1994: 794). Néanmoins, la présentation comprendra aussi quelques autres formes, en particulier dans le but de montrer les difficultés auxquelles les apprenant·e·s du français L2 doivent faire face.

La fonction des pronoms relatifs est d'introduire une proposition relative et, dans les cas les plus simples, le pronom relatif reprend un antécédent qui se trouve dans la phrase principale³. En français, le choix du pronom relatif est lié à sa fonction, de sujet, par exemple "les livres *qui* sont sur la table me dérangent" ou d'objet, par exemple "les étudiants *que* je connais sont très engagés", dans la phrase subordonnée. Le pronom relatif simple ne s'accorde pas en genre, en nombre ou en personne mais reprend néanmoins les accords qu'on aurait eu à faire avec l'antécédent (Riegel et al. 1994: 795), par exemple "les photos que j'ai prises...".

Utilisé seul, le pronom relatif *qui* est utilisé que l'antécédent soit animé/humain ou non (Riegel et al. 1994: 794). Par contre, dans sa fonction de pronom interrogatif (direct ou indirect), par exemple "*Qui* reprend du dessert ?", "je me demande *qui* je vais voir ce soir", ou lorsqu'il se trouve dans une construction prépositionnelle, par exemple "voici la collègue avec *qui* je ne parle jamais", *qui* représente toujours un humain (Riegel et al. 1994: 794) (ou un animal) quelle que soit la fonction du pronom (sujet ou objet).

Le mot *que* n'est pas seulement un pronom relatif, il est également utilisé comme conjonction de subordination (Riegel et al. 1994: 788, 824), par exemple "je pense *que* la vue est belle". Selon la liste des mots les plus fréquents publiée sur le site *Eduscol*⁴ du ministère de l'éducation nationale, *que* comme conjonction serait presque deux fois plus fréquent que *que* comme pronom (176'161 *que* conjonction versus 99'412 *que* pronom; *qui* pronom 148'392).

Les recherches sur l'apprentissage des relatives suggèrent que la production des pronoms *qui* et *que* comme relatifs apparaît assez tôt, au stade 2 selon l'itinéraire d'apprentissage du français L2 proposé par Bartning & Schlyter (2004). Selon Bartning & Kirchmeyer (2003: 15), l'itinéraire de l'apprentissage des relatives se fait en trois étapes principales. Dans la première, les apprenant·e·s utilisent une forme, de base ou passe-partout, qui sert pour toutes les fonctions. Pour le français L2, il s'agit de *qui* (alors qu'en italien c'est le *che*). Selon les auteures, "[c]ette surgénéralisation de *qui* peut être liée à la saillance de la forme, la plus fréquente dans l'input." Dans un deuxième temps, les apprenant·e·s utilisent une forme pour une fonction, donc le *one-to-one principle* décrit plus haut⁵ et, finalement, ils/elles réussissent à utiliser les formes dans leurs diverses fonctions. Cette progression est basée sur des

³ Voir Cappeau (2015) pour une discussion concernant les différentes manières de considérer les relatives, ainsi que sa proposition de les traiter comme des expansions du groupe nominal (tel un adjectif).

⁴ https://cache.media.eduscol.education.fr/file/ecole/20/6/liste-mots-par-frequence_115206.pdf (page consultée le 10 mars 2021).

⁵ A noter que Bartning & Kirchmeyer (2003) s'intéressent à des fonctions discursives dans leur étude.

apprenant·e·s suédophones du français, mais elle a été confirmée par des études sur des apprenant·e·s japonais·e·s (voir Trévisiol-Okamura 2015: 109).

Les études sur le français L1 ont montré que les relatives en *qui* ainsi que les clivées forment la majorité des relatives produites (Welcomme 2013). Roubaud (2015: 133) montre que les enfants des classes du primaire utilisent les relatives avec des antécédents humains dans près de 80% des cas. Elle signale que ces données sont néanmoins toutes basées sur des tâches dans lesquelles on demande de décrire des situations liées à des personnes. Selon elle, des tâches de type explicatif ou argumentatif pourraient mener à la production d'autres pronoms.

Welcomme (2013) a mené une étude sur les relatives à l'oral avec trois groupes d'apprenant·e·s néerlandophones, âgés entre 12 et 18 ans, ayant eu entre 2 et 7 ans de français à l'école. Les résultats de son étude, basés sur une tâche de narration orale (*Frog Story*), montrent que la production des relatives augmente avec le niveau scolaire. Les relatives introduites par *qui* sont de loin les plus fréquentes alors que celles introduites par *que* sont peu fréquentes. Le nombre de relatives en *qui* augmente nettement selon le niveau des groupes, alors que le nombre de relatives en *que* stagne, un résultat qui reflète la fréquence de ces formes en français.

En résumé, le pronom relatif *qui* apparaît comme le pronom le plus fréquent et peut être considéré comme le pronom relatif par excellence. On s'attend donc à une production plus fréquente et correcte de celui-ci.

Étonnamment, il n'existe pas, à ma connaissance, d'étude qui s'est penchée sur l'influence de l'antécédent pour le choix du pronom relatif par les apprenant·e·s. Cet élément est pourtant fortement mis en avant dans les grammaires pour apprenant·e·s du FLE (voir par exemple Delatour et al. 2004: 204), même si, ou justement parce que l'antécédent ne joue pas de rôle déterminant pour le choix des pronoms *qui* et *que* dans leur utilisation la plus simple. Pourtant, pour des raisons de construction syntaxique – l'antécédent précède le pronom – l'antécédent pourrait être saillant, surtout pour les apprenant·e·s ayant une L1 ou une L2 (par exemple l'anglais) dans laquelle le pronom prend différentes formes selon la nature de l'antécédent.

Ainsi, à partir des études antérieures sur l'apprentissage des pronoms relatifs, on pourrait s'attendre à ce que les apprenant·e·s commencent par utiliser le pronom *qui* comme forme par défaut. Ensuite, par extension de l'utilisation de *qui* comme interrogatif pour les humains et de l'utilisation de *que* pour "d'autres choses" (il est peu probable que les apprenant·e·s débutant·e·s réussissent à faire la différence entre le *que* relatif et le *que* complétif), notamment les non animés. Si tel est le cas, les inputs et exercices devraient fournir aux apprenant·e·s de nombreuses occasions de réaliser que cette distribution est

simplificatrice et attirer l'attention sur la fonction des pronoms dans la subordonnée.

Ainsi, dans le but de discuter du possible apport des corpus comme ressource pour l'enseignement des langues, cette étude traitera les trois questions suivantes:

1. Observe-t-on une distribution dans l'utilisation des pronoms *qui* et *que* selon leur antécédent dans les données d'apprenant·e·s ?
2. La distribution des formes présente dans du matériel didactique en libre accès permet-elle de dépasser les préférences/biais observés dans les données des apprenant·e·s ou au contraire les renforce-t-elle ?
3. Les biais observés dans le matériel didactique en libre accès correspondent-ils à la distribution que l'on trouve en français tel que représenté par le corpus écrit ScienQuest ou l'usage représenté dans ce corpus est-il différent ?

3. Données des corpus d'apprenants

3.1 Choix des corpus

Les données des apprenant·e·s du français L2 sont des données de production qui proviennent d'un corpus d'apprenant·e·s ayant le suédois comme L1, le corpus CEFLE, et d'un corpus d'apprenant·e·s ayant l'allemand comme langue de scolarisation et le suisse allemand comme langue du quotidien, le corpus HarMoS (voir prochaines sections).

Les langues en présence sont donc toutes des langues germaniques, mais la manière de produire le pronom relatif le plus simple dans celles-ci diffère. En suisse allemand, on peut utiliser le pronom relatif *wo* pour les fonctions correspondant au *qui* et *que* du français ("d'Frau *wo* wartet" 'la femme qui attend', "d'Frau *wo* ich gseh" 'la femme que je vois'). En allemand, l'antécédent détermine le genre du pronom relatif, mais la construction verbale de la subordonnée détermine le cas de celui-ci ("der Mann, *der* wartet" 'l'homme qui [der = nominatif] attend', "der Man, *den* ich sehe" 'l'homme que je vois' [den = accusatif/objet direct]). L'attention à la fonction du relatif dans la subordonnée est donc un phénomène syntaxique auquel les apprenant·e·s ayant une bonne maîtrise de l'allemand sont déjà sensibilisés. En suédois, le pronom relatif *som* est le plus fréquent et commun. Il est utilisé quel que soit l'antécédent et quelle que soit la fonction du relatif dans la subordonnée. Toutefois, lorsqu'il prend la fonction de sujet, il est obligatoirement exprimé, alors qu'il peut être omis dans sa fonction d'objet ("kvinnan *som* väntar" 'la femme qui attend', "kvinnan (*som*) jag ser" 'la femme que [optionnel] je vois'). Le suédois connaît un autre pronom relatif correspondant aux utilisations les plus simples de *qui* et de *que*, mais qui est stylistiquement plus formel (*vilken*, *vilket*, *vilka*). Ce pronom s'accorde en genre et en nombre avec l'antécédent; comme *qui* et *que*, ce pronom peut aussi

être utilisé pour introduire une phrase interrogative. Finalement, en anglais, la première langue étrangère des suédophones, le pronom relatif est choisi selon l'antécédent, l'humain étant repris avec un *who*, alors que *which/that* reprend les non animés et les phrases.

Ces corpus ont été choisis afin de permettre une certaine variation dans les données en termes de niveau de français, de modalité de production et de tâche. Ils seront présentés plus en détail dans les deux sections suivantes.

3.2. Le corpus CEFLE

Le corpus CEFLE est un corpus en libre accès en ligne⁶ (Ågren 2008) comprenant des textes écrits par des apprenant·e·s suédophones de français langue étrangère qui suivent des cours de français au secondaire II (16 à 19 ans) ainsi que par un groupe contrôle de natifs de Paris du même âge. Les données utilisées pour la présente étude sont celles du sous-corpus transversal pour lequel chaque apprenant·e a produit un texte à partir de la série de 16 images *Le voyage en Italie*. Il s'agit d'une tâche de narration parlant de deux femmes qui font un voyage en Italie et qui y rencontrent deux hommes italiens. Les apprenant·e·s n'avaient aucune aide à leur disposition et ont pris environ 30 à 40 minutes pour produire leur texte. Les textes ont été répartis selon les stades de développement du français proposés par Bartning & Schlyter (2004): 16 textes au stade 1 (débutant) et 30 textes pour chacun des stades 2 à 4, donc 106 textes au total. Le groupe contrôle comprend également 30 textes.

3.3 Le corpus HarmoS

Le corpus HarmoS a été constitué en Suisse en 2007, dans le cadre d'un projet de recherche dont l'objectif était d'évaluer le niveau de compétence en langue étrangère au terme du primaire (12 ans) et du secondaire I (15 ans), à l'écrit comme à l'oral (Rey 2020; Schneider et al. 2009). Les apprenant·e·s ont donc un niveau débutant à intermédiaire. Le corpus comprend des productions orales en interaction et en monologue de 30 élèves, basées sur différentes tâches ciblant des compétences actionnelles décrites dans le CECR. Il comprend en outre les textes de 90 élèves de 6^e et de 60 élèves de 9^e qui ont écrit un texte à partir d'une série de quatre images, *Lisa et son chien Pacco*, traitant d'une jeune fille, de son père et d'un chien, ainsi que les textes de 175 élèves de 6^e et de 112 élèves de 9^e, dans laquelle les élèves devaient donner leur avis sur les devoirs (avantages et inconvénients, devoirs idéaux).

3.4 Compilation des données des apprenant·e·s

Les données des apprenant·e·s (ainsi que du groupe contrôle) ont été compilées manuellement en cherchant les occurrences de *qui*, *que* et *qu'* dans leur fonction de pronom relatif dans les textes et les transcriptions. Elles ont ensuite été reportées dans un tableau et codées au niveau de l'antécédent –

⁶ <https://projekt.ht.lu.se/cefle/> (page consultée le 10 mars 2021).

humain ou non animé – de la fonction du pronom relatif – sujet ou objet – et finalement de la forme – correcte ou incorrecte. Les pronoms relatifs produits à la place d'un pronom ou d'un subordonnant différent (*où*, *ce que*, *ce qui*, *quand*, *auxquelles* ainsi que *dont*) n'entrent pas dans les calculs, ni les utilisations ambiguës ("elles boivent du café et parlent beaucoup de ce voyage *qui* se rendre" texte original non retouché). La production de l'élision avec *qui* ("qu'est jaune") a été comptée comme incorrecte même si cette forme est attestée en français parlé. Dans le cas de répétitions du pronom seul (hésitation à l'oral, données HarMoS), seule la dernière production a été codée. Par contre, si le pronom a été produit avec un verbe, il a été compté ("qu'on peut, qu'on peut acheter", 2 occurrences). Les utilisations de *que* dans un contexte pour *ce que*, mais pour lesquelles le *ce* a été omis, n'ont pas été prises en compte. Les occurrences sans verbe qui suit ont également été exclues, puisque c'est le verbe qui permet de vérifier la fonction prévue.

La production a été évaluée comme correcte lorsque le pronom attendu a été produit selon la fonction qu'il prend dans la phrase, donc *qui* sujet et *que* objet. Les occurrences incorrectes correspondent donc à une erreur au niveau du choix du pronom selon la fonction que celui devrait avoir. Par exemple, dans l'énoncé "Le garde de nationale, qui picko conetre" (CEFLE, Billy), le *qui* est utilisé pour un antécédent humain (le garde national), mais la fonction est objet (que Picko connaît).

3.5 Résultats des apprenant·e·s

3.5.1 Corpus écrit CEFLE – apprenant·e·s

Au total, les apprenant·e·s du corpus CEFLE ont produit 151 pronoms, dont 130 *qui*, produits correctement dans 95% des cas, et 21 *que*, produits correctement à 62%. On voit donc une dominance de l'utilisation du pronom *qui* et on voit aussi que le pronom *que* est non seulement rare mais aussi plus difficile à produire correctement.

Le Tab. 1 présente la distribution des pronoms selon leur antécédent. Même si la différence au niveau des occurrences est importante (voir Fig. 1), on observe la distribution suivante des pronoms: *qui* est préféré avec les antécédents humains et *que* avec les non animés. On observe également que l'utilisation de *qui* avec les antécédents humains est très correcte dès le début, alors que l'utilisation de *qui* avec les non animés est un peu moins correcte. Ces résultats suggèrent que *qui* pourrait servir de pronom relatif par défaut et confirment ainsi les études antérieures.

On ne peut pas vraiment se prononcer sur l'utilisation du pronom *que* avec les humains vu leur rareté, mais on peut voir que les apprenant·e·s utilisent *que* avec les non animés même lorsqu'ils ont la fonction de sujet et que cette difficulté perdure jusqu'à un niveau avancé de français (stade 4). L'utilisation du pronom avec les antécédents non animés semble généralement plus difficile.

Pronom produit Total	stade 1 (16 textes)	stade 2 (30 textes)	stade 3 (30 textes)	stade 4 (30 textes)
<i>qui</i> avec humain 98% (101/103)		94% (15/16)	97% (37/38)	100% (49/49)
<i>qui</i> avec non animés 79% (23/29)		70% (7/10)	83% (5/6)	85% (11/13)
<i>que</i> avec humain 50% (2/4)	0% (0/2)	100% (1/1)		100% (1/1)
<i>que</i> avec non animés 61% (11/18)		0% (0/1)	33% (1/3)	71% (10/14)

Tab. 1: Proportion de production correcte (et occurrences) des apprenant·e·s du corpus écrit CEFLE

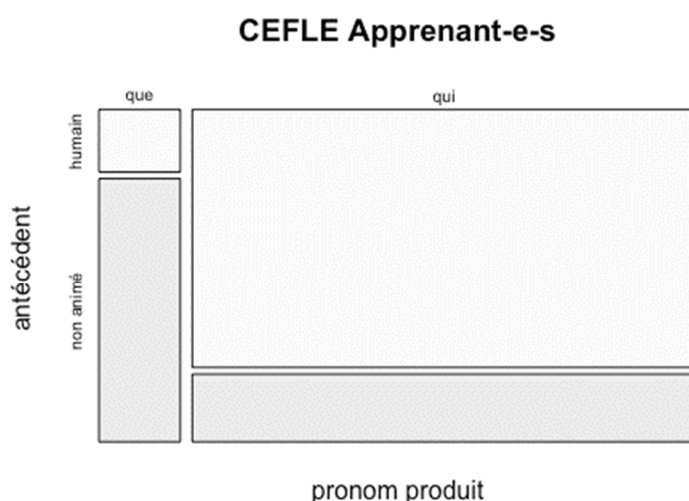


Fig. 1: Distribution des pronoms selon l'antécédent et le nombre d'occurrences chez les apprenant·e·s du Corpus écrit CEFLE

3.5.2 Corpus CEFLE – Groupe contrôle

La Fig. 2 présente la distribution des pronoms produits par le groupe contrôle (personnes natives du français). On voit que comme chez les apprenant·e·s le nombre d'utilisations du pronom *qui* (43 occurrences) est ici aussi nettement plus important que celles de *que* (6 occurrences), alors que le pronom *qui* est utilisé autant avec des antécédents humains (21 occurrences) que non animés (22 occurrences). La distribution entre les antécédents humains et non animés est donc plus équilibrée dans les données du groupe de contrôle que des

apprenant·e·s. On voit par contre que la proportion de *qui* est similaire à ce que l'on a vu chez les apprenant·e·s, ce qui suggère une influence de la tâche.

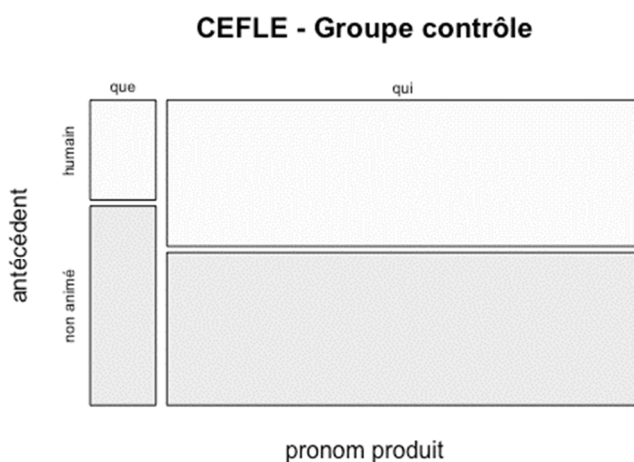


Fig. 2: Distribution des pronoms selon l'antécédent et le nombre d'occurrences chez les natifs du Corpus écrit CEFLE

3.5.3 Résultats des apprenant·e·s du corpus HarmoS

Les données du corpus HarmoS proviennent de trois tâches différentes (cf. Tab. 2). Au total, 177 pronoms relatifs ont été produits, dont 123 en production orale. Peu de pronoms relatifs ont été produits à l'écrit. Comme dans le corpus CEFLE, le pronom *qui* (130 occurrences) est plus fréquent que *que* (47 occurrences) et le pronom *qui* est aussi plus correctement utilisé (94%) que *que* (85%), même si la différence est plus faible que dans CEFLE. Les apprenant·e·s du corpus HarmoS ont moins de difficultés avec le pronom *que* que ceux du corpus CEFLE, ce qui pourrait s'expliquer par la L1 ou la langue de scolarité.

La Fig. 3, qui comprend les données des trois tâches, révèle une distribution des pronoms selon les antécédents, même si la différence n'est pas aussi marquée avec le pronom *qui* qu'avec le pronom *que*.

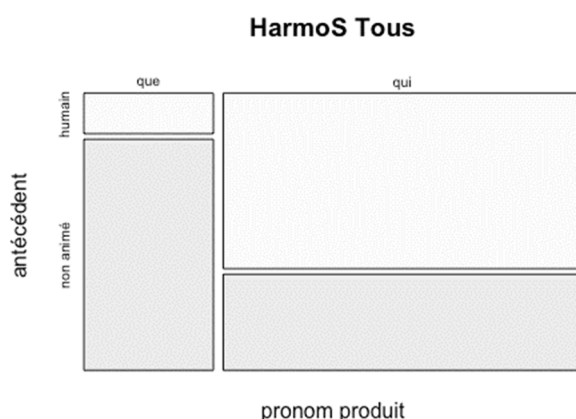


Fig. 3: Distribution des pronoms selon l'antécédent et le nombre d'occurrences dans les trois tâches du corpus HarmoS

Le Tab. 2 révèle qu'il existe des différences entre les tâches. À l'oral, les pronoms sont plutôt bien maîtrisés, sauf le *que* avec les humains, et il y a une distribution assez claire des pronoms selon les antécédents. La production des pronoms est un peu plus difficile à l'écrit. La Fig. 4 illustre les différentes proportions d'utilisation des pronoms: la tâche narrative, comprenant principalement des êtres animés, semble favoriser l'utilisation du pronom *qui*, alors que la tâche plus argumentative donne lieu à une production relativement équitable des deux pronoms. La tâche plus argumentative semble donc favoriser la production du pronom relatif *que* mais aussi l'utilisation des deux pronoms avec un antécédent non animé. On voit aussi que la production de *qui* est un peu plus difficile avec les antécédents humains dans cette tâche (mais il n'y a presque pas d'occurrences).

Même si les conclusions doivent rester modestes en raison du petit nombre de données, on peut malgré tout constater une distribution de l'utilisation des deux pronoms relatifs selon l'antécédent: l'utilisation de *que* avec un antécédent humain est rare et peu réussie et l'utilisation de *qui* avec un antécédent non animé est également plus rare et parfois moins bien réussie qu'avec un antécédent humain. On observe aussi des différences selon la tâche, ce qui pourrait confirmer la suggestion de Roubaud (2015).

Pronom produit Total	Oral (30 transcriptions)	Écrit Narratif (150 textes)	Écrit Opinion (287 textes)
<i>qui</i> avec humain 98% (82/84)	100% (72/72)	88% (7/8)	75% (3/4)
<i>qui</i> avec non animés 87% (40/46)	96% (26/27)	100% (3/3)	69% (11/16)
<i>que</i> avec humain 14% (1/7)	14% (1/7)	0	0
<i>que</i> avec non animés 98% (39/40)	100% (17/17)	75% (3/4)	100% (19/19)

Tab. 2: Proportion de production correcte (et occurrences) des apprenant·e·s du corpus HarmoS

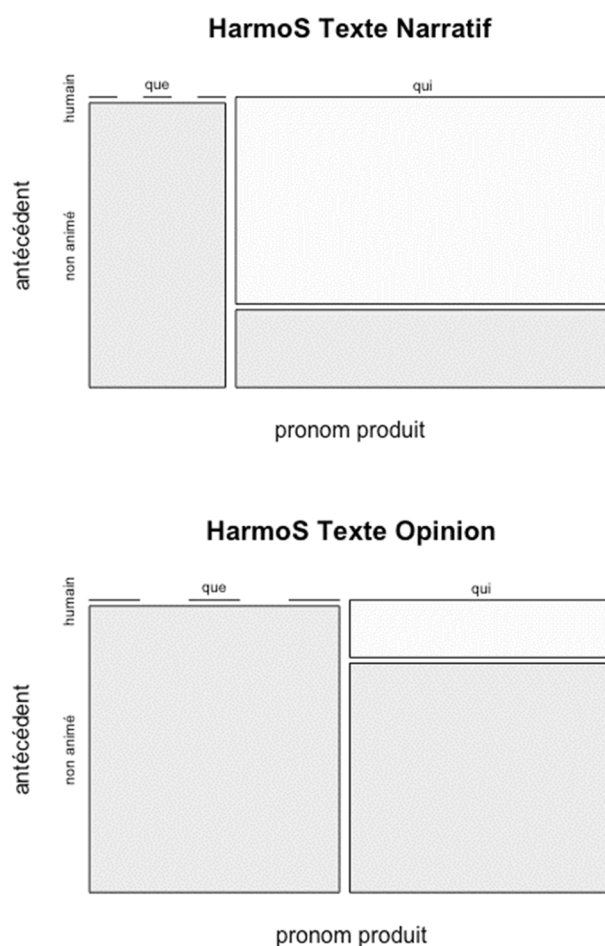


Fig. 4: Distribution des pronoms selon l'antécédent et le nombre d'occurrences dans les données HarmoS dans les tâches écrites

3.5.4 Synthèse des données d'apprenant·e·s

L'analyse des données des apprenant·e·s a globalement montré que le pronom *qui* pose peu de problèmes, en particulier lorsqu'il est utilisé avec un antécédent humain (ou un animal). Le pronom *que* est utilisé beaucoup plus rarement et de manière moins correcte que *qui*. Les données ont globalement confirmé l'hypothèse d'une préférence de *qui* avec les antécédents humains et de *que* avec les antécédents non animés. Mais on a également pu observer une différence selon la L1 des apprenant·e·s, la production étant généralement plus correcte chez les apprenant·e·s des écoles suisses que chez les suédophones, et surtout selon la tâche, une tâche argumentative favorisant l'utilisation du pronom relatif *que*. Alors que le pronom *qui* semble servir de forme par défaut chez les apprenant·e·s des stades 2 et 3 du corpus CEFLE, cela ne semble pas être le cas chez les apprenant·e·s du corpus HarMoS.

4. Données d'input: matériel didactique et corpus

Voyons maintenant quelle est la distribution de ces formes dans des données de français accessibles gratuitement en ligne. Pour cela, j'ai choisi les explications théoriques et les exercices disponibles de la base de données du site du *Point du FLE*⁷ et les données du corpus de français écrit ScienQuest⁸ (Falaise et al. 2011). Ces données peuvent être considérées comme des données d'input, dans la mesure où elles représentent différentes formes de français auxquelles les apprenant·e·s sont exposé·e·s.

4.1 Données du Point du FLE

4.1.1 Compilation des données

Les données du site du *Point du FLE* ont été compilées manuellement en janvier 2019. Tous les exercices et présentations théoriques ciblant notamment les pronoms relatifs *qui* et *que* aux niveaux débutants ont été récupérés, soit sous forme de pdf soit sous forme de capture d'écran. Au total, 26 documents de différentes longueurs ont ainsi été retenus. Les occurrences de *qui* et de *que* ont été compilées en calculant chaque occurrence dans les explications et introductions théoriques et chaque occurrence présentée ou attendue dans les exercices.

4.1.2 Résultats

Tout d'abord il est intéressant de constater que la question de l'antécédent est souvent thématifiée dans les explications et introductions théoriques. Dans l'exemple reproduit dans la Fig. 5, on voit qu'on s'est appliqué à montrer un exemple avec un antécédent humain et un avec un non animé pour chaque pronom. Par contre, le texte précédant les exemples est ambigu "QUI: reprend

⁷ <https://www.lepointdufle.net> (page consultée le 10 mars 2021).

⁸ <https://corpora.aiakide.net/scientext20/> (page consultée le 10 mars 2021).

le sujet du verbe qui suit", en effet *qui* "reprend" l'antécédent au niveau référentiel, mais prend la fonction de sujet à partir du verbe qui suit.

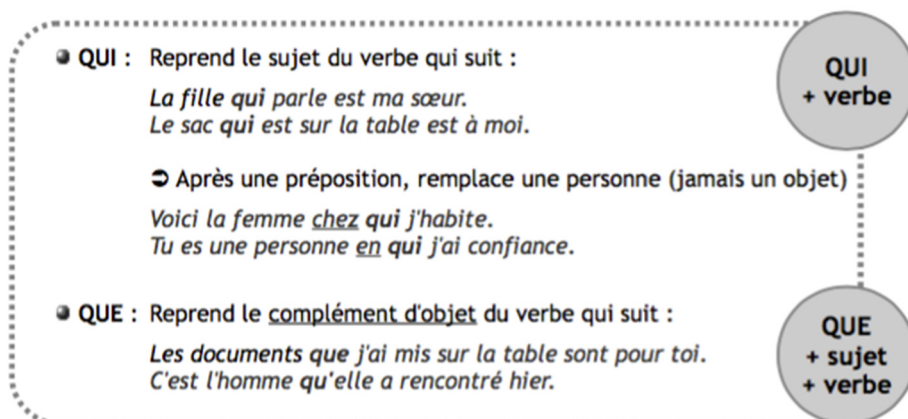


Fig. 5: Exemple de présentation théorique⁹

Dans l'exemple reproduit dans la Fig. 6, le lien vers l'antécédent est présenté de manière explicite, tout comme le fait que *qui* et *que* sont tous deux utilisés avec des animés et des non animés. Le seul problème est que ce n'est pas la fonction de l'antécédent qui compte, mais celle du pronom, et que c'est l'antécédent qui est animé ou non. En examinant les différentes propositions théoriques, on constate qu'il n'y en a en fait que très peu qui présentent le phénomène de manière non ambiguë.

antécédent	forme employée	exemple
sujet	animé et non animé : QUI	Le garçon <u>qui</u> joue là-bas est mon neveu. La chaise <u>qui</u> est là est cassée.
C.O.D.	animé et non animé : QUE	L'homme <u>que</u> j'aime s'appelle Jo. La pomme <u>que</u> je mange est délicieuse.

Fig. 6: Exemple de présentation théorique¹⁰

Au total, 392 pronoms ont été relevés dans les données du *Point du FLE*. Comme on le voit dans le Tab. 3, le nombre d'occurrences des deux pronoms est assez égal, mais le pronom *que* est légèrement plus fréquent (206 occurrences contre 186 pour *qui* au total). On observe aussi que les antécédents sont le plus souvent non animés (71% des données). Même si elle n'est pas très marquée (Fig. 7), on constate tout de même une plus haute proportion d'antécédents humains avec *qui* et surtout une très faible utilisation de *que* avec les antécédents humains, 21% (11/53 occurrences) dans les

⁹ https://flenantes.files.wordpress.com/2013/10/pronoms_relatifs_simples1.pdf (page consultée le 10 mars 2021).

¹⁰ https://www.lepointdufle.net/ressources_fle/a/pdf/pronoms_relatifs.pdf (page consultée le 10 mars 2021).

présentations théoriques et 16% (25/153 occurrences) dans les exercices. On ne retrouve donc pas la dominance de *qui* en termes d'occurrence et la distribution des pronoms selon les antécédents est moins marquée que chez les apprenant·e·s. Néanmoins, la très faible utilisation de *que* avec les humains, en particulier dans les exercices, ne vient pas contredire le système de simplification des apprenant·e·s.

	Théorie		Exercices		<i>Totaux</i>
	qui	que	qui	que	
humains	17	11	60	25	113
non animés	31	42	78	128	279
<i>Totaux</i>	48	53	138	153	392

Tab. 3: Nombre d'occurrences dans les données du *Point du FLE*

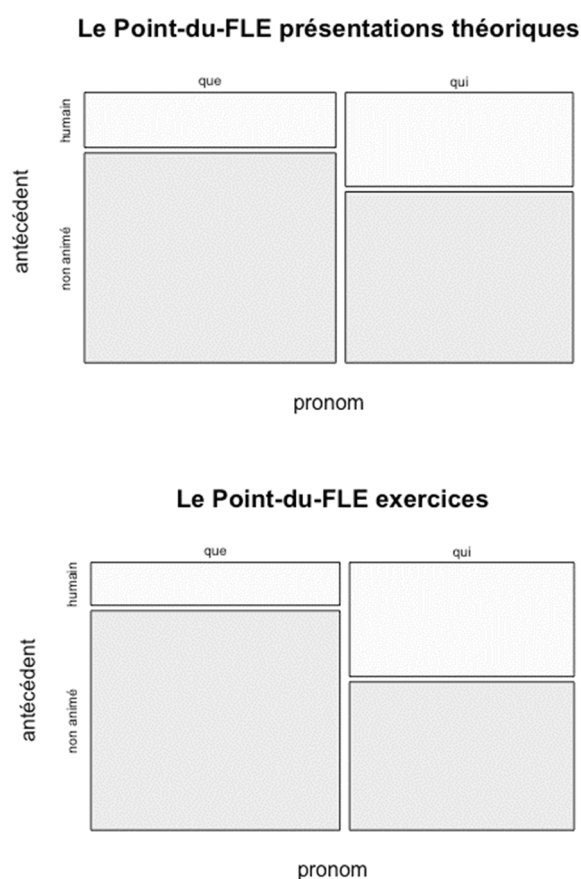


Fig. 7: Distribution des pronoms selon l'antécédent et le nombre d'occurrences dans les données du *Point du FLE* selon le type d'input

4.2 Données du corpus écrit de ScienQuest

J'ai choisi le corpus écrit *ScienQuest* (Falaise et al. 2011) comme élément de comparaison avec les exercices, car ce corpus est un corpus écrit facile d'accès et d'utilisation. Il pourrait donc être utilisé sans trop de difficultés pour l'enseignement du FLE (voir par exemple Yan et al. 2018). En outre, un travail didactique sur les pronoms relatifs se fait plus aisément sur des données écrites qu'orales.

4.2.1 Compilation des données

Le matériel pour cette analyse provient du sous-corpus *Démo*, qui comprend les données de cinq genres textuels (encyclopédie, littérature, presse, scientifiques, tourisme; voir la description du corpus sur leur site) pour un total de 903 textes et 6 millions de mots. La compilation des données a été faite au moyen de l'outil de recherche disponible sur le site. Ainsi 28'303 occurrences du pronom *qui* et 11'178 occurrences de *que* ont été trouvées. Les données pour la présente étude sont basées sur les 20 premières pages de résultats, correspondant à 350 lignes de résultats. Le codage a été effectué manuellement sur un *pdf* comprenant 57 pages pour chacun des deux pronoms. La Fig. 8 montre un extrait du codage, qui révèle la variété des utilisations du pronom *que*. Les antécédents des lignes 1 et 7 ont été codés comme *non animés*, ceux des lignes 3, 5 et 8 comme *humains*. L'antécédent "deux ans" a été codé comme *autre utilisation* (exclu des calculs).

Et avec le sourire , ainsi que le petit cadeau	qu'	il remet à chaque lecteur , en l' occurrence hier
Au lard et fromage " C' est la première fois	que	j' entends parler d' une opération aussi importan
il l' a été deux fois , ce saint éponyme	que	se sont choisi les boulangers pour patron . Car ho
faut mettre en valeur ce qui est bon , ce	qu'	on fait bien . Et c' est le cas du
us les ans d' environ 3.400 nouveaux marins ,	qu'	elle forme à une trentaine de métiers . Documen
Depuis deux ans	qu'	ils jouent ensemble , ils ont sillonné une bonne p
C' est ce travail	qu'	elle exercera à Cravanche tout en élevant ses tro
épousé , en deuxième noces , Julien Mettetal	qu'	elle avait eu le grand chagrin de perdre en 1974

Fig. 8: Extrait des données de ScienQuest

4.2.2 Résultats

Les résultats sont ainsi basés sur 447 utilisations simples des pronoms dans le corpus *ScienQuest* (Tab. 4). Le pronom *qui* est plus fréquent (66%) que *que* et les antécédents non animés sont plus fréquents (63%) que les humains. La distribution des antécédents est par contre plus claire que dans les données du *Point du FLE*, la proportion de *que* avec des antécédents humains ne s'élevant qu'à 17% (26/153 occurrences) et la proportion de *qui* avec les humains à 47%

(138/294 occurrences) (voir Fig. 9). Ces données confirment la fréquence du pronom *qui* ainsi que la distribution des pronoms selon les antécédents.

	qui	que	Totaux
humains	138	26	164
non animés	156	127	283
Totaux	294	153	447

Tab. 4: Occurrences compilées dans les données de ScienQuest

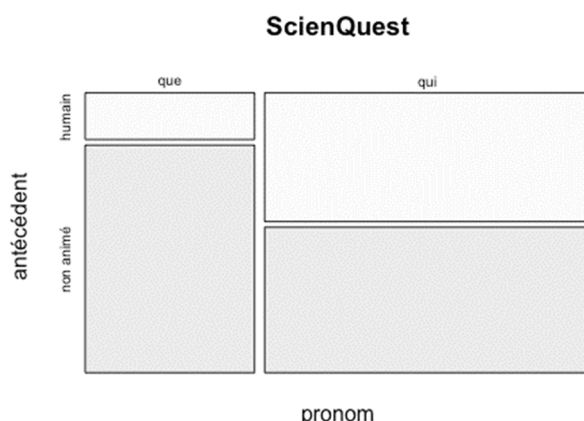


Fig. 9: Distribution des pronoms selon l'antécédent et le nombre d'occurrences dans les données du corpus écrit ScienQuest

4.3 Résumé des données d'input

L'analyse des données d'input a montré une distribution des pronoms selon les antécédents, même si l'association entre *qui* et *humain* était plus faible que chez les apprenant·e·s. En outre, la plus grande fréquence du pronom *qui* est apparue dans les données du corpus, mais pas dans les exercices.

5. Discussion

L'analyse de différentes données d'apprenant·e·s et d'input au niveau de la présence des pronoms *qui* et *que* et de leur distribution selon les antécédents *humain* et *non animé* a montré qu'une telle distribution était visible dans toutes les données, mais à différents degrés. Globalement, ces résultats confirment les approches basées sur l'usage, qui font l'hypothèse d'une relation assez étroite entre ce que l'on trouve dans l'input et la production des apprenant·e·s (Bybee 2008; Ellis et al. 2016).

Pour les apprenant·e·s, les résultats semblent varier selon la L1, le mode de production et surtout la tâche. Les tâches narratives, dans lesquelles des humains/animaux jouent un rôle central, donnent en effet lieu à une forte

production du pronom *qui*, qui semble parfois servir de pronom relatif par défaut. Les apprenant·e·s germanophones semblent aussi avoir plus de facilité à produire les pronoms relatifs correctement. Ces résultats montrent qu'il est nécessaire de conduire d'autres études, avec une variété de L1 et de tâches.

La fréquence de *qui* relevée dans les études antérieures (Bartning & Kirchmeyer 2003; Trévisiol-Okamura 2015) a été constatée dans l'échantillon des données corpus de français écrit ScienQuest, mais pas dans les données du *Point du FLE*. Cette différence pourrait refléter la volonté des personnes créant des exercices d'exposer les apprenant·e·s aux deux formes de pronoms relatifs.

Outre son apport empirique, cette étude permet de rappeler la pertinence de la mise à disposition de corpus de différents types, d'apprenant·e·s comme d'input, dans le but de les mettre en lien en vue d'une amélioration du matériel de didactique de langue étrangère. En même temps, elle révèle aussi la difficulté de manipuler ces données. Les outils de recherche automatique pour détecter certaines erreurs typiques des apprenant·e·s doivent encore être développés. Comme le suggère Granger (2009: 15), développer un enseignement informé par les corpus (*corpus informed teaching*) demande des connaissances dans plusieurs champs adjacents: linguistique de corpus, théorie linguistique, acquisition des langues étrangères/secondes, didactique des langues étrangères, recherche sur corpus d'apprenant·e·s. Une telle palette de connaissances risque de décourager plus d'un·e enseignant·e. Une des tâches de la linguistique appliquée à l'heure digitale, dans un effort conjoint de linguistes et de didacticien·ne·s, serait donc de préparer du matériel didactique sur base empirique, à l'instar de ce qui se fait par exemple sur le site de CLAPI-FLE¹¹.

Pour conclure, concernant l'étude présentée ici, on peut considérer que les données des exercices correspondent finalement bien à l'usage tel qu'il apparaît dans l'échantillon de ScienQuest. En ce sens, l'utilisation des corpus pour l'enseignement des pronoms relatifs *qui/que* semble avant tout redondante au niveau de ce qui est ciblé, sans compter que les données de corpus contiennent bien d'autres utilisations de ces relatifs. Alors que la richesse de telles données peut être fructueuse pour faire découvrir les usages multiples d'une même forme, elle exige un bon niveau de formation linguistique pour en faire le tri pour une application ciblée sur un point de grammaire précis (Thomas 2020).

Une autre piste pourrait être d'utiliser les corpus pour vérifier la présence de biais simplificateurs dans la langue cible et sur cette base construire des exercices qui permettent de les dépasser. Ainsi, plutôt que d'en faire des sources de données du français tel qu'il est vraiment utilisé, les corpus pourraient soutenir les didacticien·ne·s dans leur effort de créer des exercices

¹¹ <http://clapi.icar.cnrs.fr/FLE/> (page consultée le 10 mars 2021).

dans lesquels l'input ne vient pas soutenir les processus de simplification de la langue cible par les apprenant·e·s, comme le suggère VanPatten (2002 par exemple) dans ses travaux sur l'input.

BIBLIOGRAPHIE

- Ågren, M. (2008). *A la recherche de la morphologie silencieuse. Sur le développement du pluriel en français L2 écrit*. Thèse de doctorat, Université de Lund, Suède.
- Andersen, R. W. (1984). The One to One Principle of interlanguage construction. *Language Learning*, 34(4), 77-95.
- André, V. (2018). Nouvelles actions didactiques: Faire de la sociolinguistique de corpus pour enseigner et apprendre à interagir en français langue étrangère. *Action Didactique*, 1, 71-88.
- Bartning, I. & Kirchmeyer, N. (2003). Le développement de la compétence textuelle à travers les stades acquisitionnels en français L2. *Acquisition et interaction en langue étrangère*, 19, 9-39.
- Bartning, I. & Schlyter, S. (2004). Itinéraires acquisitionnels et stades de développement en français L2. *Journal of French Language Studies*, 14(3), 281-299.
- Boulton, A. (2017). Corpora in language teaching and learning. *Language Teaching*, 50(4), 483-506.
- Boulton, A. & Cobb, T. (2017). Corpus use in language learning: A Meta-Analysis. *Language Learning*, 67(2), 348-393.
- Boulton, A. & Tyne, H. (2014). *Des documents authentiques aux corpus: Démarches pour l'apprentissage des langues*. Paris: Didier.
- Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. C. Ellis (éds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 216-236). New-York: Routledge.
- Cappeau, P. (2015). La relative: Observation critique des deux présentations proposées par les manuels de FLE et les grammaires d'apprenants. In P. Trévisiol-Okamura & M. Kaheraoui (éds.), *Les subordonnées, corpus, acquisition et didactique* (pp. 21-34). Rennes: Presses universitaires de Rennes.
- Collins, L., Trofimovich, P., White, J., Cardoso, W. & Horst, M. (2009). Some input on the easy/difficult grammar question: An empirical study. *Modern Language Journal*, 93(3), 336-353.
- Cuq, J.-P. & Gruca, I. (2005). *Cours de didactique du français langue étrangère et seconde*. Grenoble: Presses universitaires de Grenoble.
- Delatour, Y., Jennepin, D., Léon-Dufour, M. & Teyssier, B. (2004). *Nouvelle Grammaire du Français. Cours de civilisation française de la Sorbonne*. Paris: Hachette.
- Ellis, N. C., Römer, U. & O'Donnell, M. B. (2016). *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of construction grammar*. Chichester: Wiley.
- Falaise, A., Tutin, A. & Kraif, O. (2011). Définition et conception d'une interface pour l'exploitation de corpus arborés pour non-informaticiens: La plateforme ScienQuest du projet Scientext. *Traitement Automatique des Langues*, 53(3), 103-128.
- Gajo, L., Luscher, J.-M., Racine, I. & Zay, F. (2019). *Variation, plurilinguisme et évaluation en français langue étrangère*. Bern: Peter Lang.
- Gilmore, A. (2004). A comparison of textbook and authentic interactions. *ELT Journal*, 58(4), 363-374.
- Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Aijmer (éd.), *Corpora and language teaching* (Vol. 33, pp. 13-32). Amsterdam/Philadelphia: John Benjamins.

- Jacques, M.-P. (2005). Pourquoi une linguistique de corpus? In G. Williams (éd.), *La linguistique de corpus* (pp. 21-30). Rennes: Presses Universitaires de Rennes.
- Klein, W. & Perdue, C. (1997). The Basic Variety (or: Couldn't natural languages be much simpler?). *Second Language Research*, 13(4), 301-347.
- Larrivée, P. & Lefevre, F. (éds.). (2020). *La didactisation du français vernaculaire*. Caen: Presses Universitaires de Caen.
- Lyster, R. & Rebuffot, J. (2002). Acquisition des pronoms d'allocution en classe de français immersif. *Acquisition et interaction en langue étrangère*, 17, 51-72.
- Mayaffre, D. (2005). Rôle et place du corpus en linguistique. Réflexions introductives. *Textol*, X(4), 5-17.
- Ortega, L. (2009). *Understanding second language acquisition*. London: Hodder Education.
- Ravazzolo, E. & Etienne, C. (2019). Nouvelles ressources pour le FLE à partir des études en interaction. *Linx. Revue des linguistes de l'université Paris X Nanterre*, 79. <https://doi.org/10.4000/linx.3454>
- Rey, K. (2020). *Acquisition des propositions subordonnées en français L2 par des apprenants germanophones. Résultats d'une étude transversale dans des productions écrites narratives et argumentatives* [Thèse de Master]. Université de Fribourg, Suisse.
- Riegel, M., Pellat, J. C. & Rioul, R. (1994). *Grammaire méthodique du français* (4^e éd.). Paris: Presses Universitaires de France.
- Roubaud, M.-N. (2015). Accéder aux compétences syntaxiques des élèves (8-11 ans): Le cas des relatives. In P. Trévisiol-Okamura & M. Kaheraoui (éds.), *Les subordonnées: Corpus, acquisition et didactique* (pp. 123-138). Rennes: Presses universitaires de Rennes.
- Schneider, G., Lenz, P. & Studer, T. (2009). *Langues étrangères. Rapport scientifique de synthèse et modèle de compétences. Version provisoire (avant adoption des standards de base)*. Bern: CDIP.
- Sinclair, J. (1996). Preliminary recommendations on corpus typology. *EAGLES Document*. <http://www.ilc.cnr.it/EAGLES/pub/eagles/corpora/corpusyp.ps.gz>
- Thomas, A. (2020). Didactisation du français ordinaire: Les défis de la variation dans l'enseignement des temps du passé en français langue étrangère. In P. Larrivée & F. Lefevre (éds.), *La didactisation du français vernaculaire* (pp. 153-186). Caen: Presses universitaires de Caen.
- Trévisiol-Okamura, P. (2015). L'acquisition et l'enseignement des relatives en FLE: regards croisés. In P. Trévisiol-Okamura & M. Kaheraoui (éds.), *Les subordonnées, corpus, acquisition et didactique* (pp. 103-120). Rennes: Presses universitaires de Rennes.
- Tyler, A. E., Ortega, L., Uno, M. & Park, H. I. (2018). *Usage-inspired L2 instruction: Researched pedagogy*. Amsterdam/Philadelphia: John Benjamins.
- Tyne, H., André, V., Benzitoun, C. & Greub, Y. (2014). *French through corpora: Ecological and data-driven perspectives in French language studies*. Newcastle: Cambridge Scholars.
- VanPatten, B. (2002). Processing Instruction: An update. *Language Learning*, 52(4), 755-803.
- Vialleton, É. & Lewis, T. (2014). Reconsidering the authenticity of speech in French language teaching: Theory, data, methodology, and practice. In H. Tyne, V. André, C. Benzitoun, A. Boulton, & Y. Greub (éds.), *French through corpora: Ecological and data-driven perspectives in French language studies* (pp. 293-316). Newcastle: Cambridge Scholars.
- Welcomme, A. (2013). Les subordonnées relatives à l'oral en français langue étrangère et en français langue maternelle. *Mémoires de la Société Néophilologique*, LXXXVI, 97-136.
- Williams, G. (2005). *La linguistique de corpus*. Rennes: Presses Universitaires de Rennes.
- Yan, R., Tutin, A. & Tran, T. T. H. (2018). Routines verbales pour les français langue étrangère: Des corpus d'experts aux corpus d'apprenants. *Lidil. Revue de linguistique et de didactique des langues*, 58. <https://doi.org/10.4000/lidil.5411>

Apprendre à interagir en classe de FLE: la situation d'invitation

Carmen ALBERDI

Université de Grenade

Département de Philologie Française

Campus de Cartuja s/n, Facultad de Filosofía y Letras, 18071 Grenade, Espagne

kalberdi@ugr.es

Carole ETIENNE

ENS de Lyon

Laboratoire ICAR

15 Parvis Descartes, ENS Lettres, 69007 Lyon, France

carole.etienne@ens-lyon.fr

While specialists in the teaching of a foreign language unanimously recommend immersion stays to have a better understanding of language practices, data collected and documented by specialists in the oral language are not widely used in classroom, despite the available on-line teaching resources. In collaboration with a network of language and learning specialists and teachers, the ICAR/LIS team has designed the CLAPI-FLE application, dedicated to the teaching of spoken French, which offers about forty extracts described, transcribed and didactically designed from everyday situations, as well as collections and explanatory sheets. Our paper will focus on how CLAPI-FLE makes learners feel immersed in French interactions inside their classroom, whatever their knowledge or their aims. We will detail and explain in context the pragmatic functions and processes used by speakers in a phone invitation and share our learning expertise and feedback on French interaction teaching.

Keywords:

Interaction, didactics, French as a Foreign Language, interactional competence, sociolinguistic competence, pragmatic awareness, spoken French, invitation.

Mots-clés:

Interaction, didactique, français langue étrangère, compétence interactionnelle, compétence sociolinguistique, conscience pragmatique, français oral, invitation.

1. Introduction

Le développement d'une compétence communicative en langue étrangère passe nécessairement par une appropriation des conditions d'emploi effectif de la langue en situation, aussi bien à l'oral qu'à l'écrit. Dans le cas de l'oral, un apprenant ne pourra pas répondre à une sollicitation d'un autre locuteur dans une situation donnée s'il n'en comprend pas les mécanismes. Au-delà de la connaissance des termes employés, il s'agit de comprendre l'organisation même de cette interaction pour suivre son déroulement *in situ*, identifier ses étapes et connaître ce qui est attendu avant d'en comprendre les détails et d'être capable d'intervenir. L'apprenant aura d'autant plus de facilité à appréhender une séquence qu'il en connaît la finalité; le cas échéant, il se contentera de projeter la même séquence dans sa langue avec le risque de

glisser vers une interprétation erronée si son déroulement n'est pas régi par des règles similaires. Ceci implique, pour l'enseignant, le choix d'un *input* aussi proche que possible de la réalité. Si un tel critère semble bien établi dans le choix de ressources écrites, l'oral paraît, lui, se réfugier encore dans une semi-authenticité préfabriquée qui prive l'apprenant de l'accès aux spécificités de l'interaction orale sous couvert d'une compréhension plus aisée. Ce critère exclut ainsi une variété de phénomènes pourtant observables dans toute interaction, formelle ou informelle: faux départs, répétitions, ambiguïtés, imprécisions, phrases inachevées, interruptions, chevauchements de parole, phatèmes, régulateurs... Autant de composantes de la parole en interaction que l'on a pu considérer comme des "scories" du discours oral (Riegel et al. 1999: 36), mais qui s'avèrent en réalité fonctionnelles et signifiantes (Kerbrat-Orecchioni 1990; Traverso 1999), puisqu'elles témoignent en première instance du caractère co-construit de l'interaction. Il en va de même des "petits mots" de l'oral (*voilà, enfin, tiens*, etc.), qui débordent très largement de la catégorie lexicale ou grammaticale à laquelle ils sont censés appartenir, pour se charger, en interaction, de nouvelles fonctions et valeurs pragmatiques qui permettent de les envisager ensemble, puisqu'"ils contribuent tous, en effet, à différents niveaux, à indiquer la continuité du discours" (Traverso 1999: 45).

Ce n'est certainement que dans l'usage naturel de la langue en interaction que l'apprenant sera exposé à un tel foisonnement de valeurs pragmatiques, accessible en immersion linguistique, mais aussi, comme il est de plus en plus mis en évidence par diverses études¹, à travers l'exploitation didactique de corpus d'interactions naturelles. Tel est l'objectif de la plateforme CLAPI-FLE, développée par l'équipe LIS au sein du laboratoire ICAR, volet pédagogique de la base de données Corpus de Langues Parlées en Interaction (CLAPI), dont des possibilités d'application ont été déjà avancées dans d'autres publications (Ravazzolo et al. 2015, Alberdi et al. 2018; Ravazzolo et al. [à paraître]; Alberdi et al. 2020).

Après une brève description des apports des recherches interactionnelles concernant les caractéristiques de l'interaction, en général, et un acte fréquent tel que l'invitation, en particulier, nous illustrerons les avantages de l'utilisation de corpus d'interactions naturelles dans le cadre de l'enseignement et apprentissage du FLE, où l'invitation est abordée au niveau A2 d'après les descriptions du *Cadre Européen Commun de Référence pour les Langues (CECRL)* (Conseil de l'Europe 2001 et 2018).

¹ Pour un aperçu de ces études centrées sur le FLE au niveau international, voir Guinamard et al. (2015).

2. L'analyse des interactions

2.1 *Les recherches en interaction*

Dans notre vie quotidienne, les interactions sociales ordinaires, qu'elles soient professionnelles ou personnelles, se caractérisent par leur omniprésence (Wierzbicka 1991) mais également par leur grande variété. En effet, elles peuvent prendre différents formats (réunion, repas, achat, consultation, appel, jeu, visite, conversation...), emprunter plusieurs canaux de communication (en face à face, au téléphone, en visioconférence, voire mixte quand seulement certains des locuteurs sont à distance), intervenir dans de nombreux contextes (amical, familial, associatif, entreprise, commercial, culturel...), se dérouler dans les lieux les plus divers (salle de réunion, appartement, salle de classe, commerce, cabinet médical, site historique...), à l'intérieur comme à l'extérieur, faire appel à différents registres (formel, informel, mais également mixte) et mettre en jeu un nombre variable de locuteurs dans des communications endolingues ou exolingues. Dans tous les cas, les locuteurs sont impliqués dans l'activité sociale elle-même, qui conditionne leurs productions verbales et multimodales, et dont le déroulement en retour se verra modifié par leurs contributions (Heritage 1984). Ainsi, l'interaction est par définition une pratique sociale située dont les différentes composantes ne peuvent être décontextualisées et ne sont pas réutilisables à l'identique dans une toute autre situation.

Par ailleurs, les locuteurs ne délivrent pas des productions indépendantes les unes des autres mais adaptent en permanence leur propos à la réception des autres locuteurs. "La temporalité de la parole en interaction est donc tributaire, non seulement du locuteur et de son travail de production, mais aussi du travail de réception et d'interprétation des interlocuteurs" (Traverso 2016: 17). Ainsi, l'échange est par essence collaboratif et se construit par l'intervention de l'ensemble des participants qui vont par différents procédés d'assertion, de complétion, d'ajout, de correction, de reformulation, de réparation ou d'abandon modifier son déroulement pour mener à bien leur objectif.

De plus, si l'étude des interactions s'intéresse aux productions verbales et à la manière dont elles s'articulent dans la séquence, elle interroge également les autres modalités comme les gestes, les regards, les mimiques... qui accompagnent le verbal ou viennent le remplacer, que ce soit par des haussements d'épaule, des hochements de tête, des regards adressés à certains des locuteurs pour les sélectionner ou bien des pointages vers un document, un objet ou un lieu pour désigner ou bien interroger.

Comme nous venons de le voir, le déroulement d'une interaction n'est pas figé suivant un schéma immuable reproductible à l'identique d'une situation à l'autre.

En termes métaphoriques (inspirés de la théorie des jeux): chaque intervention est comme un "coup" joué qui ouvre un paradigme plus ou moins large d'enchaînements possibles, mais qui ne sont pas tous équivalents en ce qui concerne leur degré de probabilité, ni leurs effets et conséquences sur la suite de l'échange (Kerbrat-Orecchioni 2001: 150).

Ainsi, une production initiale pourra susciter plusieurs réponses alternatives, mais parmi elles certaines seront davantage attendues, privilégiées par les locuteurs. On parle d'"enchaînements préférés", qui correspondront le plus souvent à un déroulement poli qui va dans le sens de la production initiale ou la minore par modestie, ou bien au contraire s'y oppose pour déjouer un conflit ou préserver la face du locuteur (Kerbrat-Orecchioni 2001: 77).

Par son adéquation au contexte et par ce procédé collaboratif d'ajustement permanent entre les locuteurs, l'interaction intéresse directement l'enseignement de la langue parlée à la faveur des approches didactiques qui montrent que

[...] la compétence ne peut plus être comprise en termes d'un savoir et d'un savoir-faire donnés une fois pour toutes, qui, une fois acquis, pourront être mobilisés tels quels dans des pratiques et des contextes les plus divers. À cette disponibilité universelle des compétences, la notion contextuelle, collective et contingente de la compétence oppose l'idée de la nature adaptative comme une caractéristique centrale de la compétence. (Pekarek-Doehler 2006: 40).

C'est dans cette perspective que l'équipe LIS (Langues, Interaction, Situations) du laboratoire ICAR a décidé de s'orienter vers la création de ressources pour l'enseignement (Etienne & Jouin 2019) à partir de ses recherches dans le domaine de l'interaction, depuis les travaux fondateurs de la discipline de Kerbrat-Orecchioni (1990, 1992, 1994, 2001) et Traverso (2016) portant notamment sur la conversation familière (Traverso 1996), les échanges en commerce (Kerbrat-Orecchioni & Traverso 2008), les réunions de travail (Bruxelles & Traverso 2006), les conversations en ligne (Traverso 2011), les visites guidées (Traverso & Ravazzolo 2016) jusqu'aux études plus récentes sur la participation des migrants dans les interactions plurilingues dans le contexte de l'immigration (Piccoli et al. 2019) ou l'analyse des situations de jeu (Baldauf-Quilliatre & Colón de Carvajal 2020).

Dans cet article, nous allons nous intéresser à une situation sociale ordinaire, avec un objectif à priori simple, l'invitation, et étudier de quelle manière son analyse interactionnelle peut bénéficier à son enseignement.

2.2 *L'invitation*

D'un point de vue sociolinguistique, l'invitation constitue un acte social fréquent dans nos cultures mais qui demeure néanmoins complexe dans la mesure où il met en jeu la face des locuteurs, et plus précisément de chacun des interlocuteurs, celui qui invite pouvant essuyer un refus mais empiétant par ailleurs sur la liberté d'action de l'invité qui peut se sentir ainsi "piégé" et qui rencontrera des difficultés pour refuser un acte censé être bénéfique pour lui et auquel son locuteur attache en principe de l'importance. Si la gestion des faces

dans les interactions a été définie comme un ensemble de procédés, verbaux et non verbaux, visant à assurer, tout au long de l'échange communicatif, l'harmonie interactionnelle (Kerbrat-Orecchioni 1996; Traverso 1996, 1999), elle n'en demeure pas moins très variable d'une culture à l'autre et pose donc problème à l'apprenant dans bien des cas.

De ces contraintes va découler l'organisation même de cette interaction, dont la structure va comporter une phase de pré-invitation qui, comme pour toutes les étapes préliminaires, "préparent un autre échange dont la réalisation dépendra de la réponse obtenue à l'échange préliminaire" (Traverso 2016: 89).

L'invitation va ainsi se dérouler en plusieurs étapes dans une structure incrémentale qui ne pourra pas se ramener à une invitation "directe" résolue rapidement en quelques tours de parole. Nous allons détailler ces étapes en omettant les séquences d'ouverture et de clôture, qui bien que fondamentales ne sont pas spécifiques à l'invitation mais à chaque interaction quel que soit son objet.

2.2.1 La pré-invitation

Dans une première étape, l'interlocuteur qui invite va donc procéder à une pré-invitation qui va venir amorcer l'invitation proprement dite. Si le locuteur natif reconnaît le plus souvent cette production et s'attend donc à une éventuelle invitation (Schegloff 2007: 29), la gestion de cet échange préliminaire s'avère délicate ne laissant pas beaucoup de marge à l'invité s'il n'y prête garde: "Pre-sequences are therefore one of the central elements that make the social activity of inviting a delicate issue for the speaker" (Traverso et al. 2018: 8).

Ces pré-invitations peuvent être orientées soit vers la volonté de l'invité ("*si ça te dit*", "*si tu veux*", "*si ça te tente*"), soit vers sa disponibilité ("*si tu es libre*", "*si tu peux*", "*si tu es dispo*"). Ce dernier cas se révèle particulièrement problématique pour un locuteur allophone qui n'anticipera pas l'invitation si suivant sa culture cette étape préliminaire n'existe pas; il pourra donc répondre par l'affirmative quant à sa disponibilité et aura alors bien des difficultés à refuser l'invitation qui suivra, alors qu'il vient de dire qu'il est libre.

2.2.2 L'invitation à proprement parler

Une fois cette pré-invitation mise en place, l'invitation à proprement parler intervient, mais là aussi il s'agit d'un acte complexe (Traverso et al. 2018: 10) qui va se co-construire entre les participants avec différents procédés: des marqueurs d'hésitation, des reformulations, des réparations, des abandons, des allongements, des pauses...

D'un point de vue pragmatique, le locuteur qui invite pourra faire appel à l'atténuateur "*petit*" pour minimiser sa proposition, son offre "*un petit verre*", "*un petit repas*" (Kerbrat-Orecchioni 2004), mais il pourra également laisser une ouverture à son invité avec une formulation ouverte comme "*se faire un truc*"

ou *"boire un verre et tout"*. Le locuteur emploiera souvent *"on"* pour engager l'invité *"on pourrait"* ou d'autres participants *"on se disait"*, *"on voulait te demander"*, en faisant appel au conditionnel ou au passé (imparfait, voire plus-que-parfait, *"on s'était dit que"*) qui atténuera là encore sa suggestion en évitant le présent trop souvent perçu comme injonctif.

2.2.3 Les réponses

La dernière phase de l'invitation concerne la réponse en elle-même. Comme nous l'avons vu dans le paragraphe précédent, un enchaînement préféré privilégiera dans ce contexte la politesse et, pour l'invitation, l'accord au refus. Si la réponse est positive, elle peut être directe mais s'accompagnera d'une certaine "emphase", par exemple avec des termes comme *"super"* ou *"génial"* et ne pourra se ramener à un simple *"oui"* qui décevrait le locuteur qui invite et s'attend en retour à une réponse enthousiaste. La réponse négative, quant à elle, devra être atténuée, soit en proposant une alternative concernant la date, le lieu ou le motif, soit en justifiant ce refus afin de préserver la face de son interlocuteur.

La réponse se poursuit ensuite par une phase de négociation sur l'heure ou le lieu de rendez-vous, le motif lui-même ou bien éventuellement ce que l'on peut apporter si on est invité chez quelqu'un; là aussi cette séquence correspondra à une réalisation collaborative de l'ensemble des locuteurs qui vont la co-construire pour atteindre leur objectif.

3. L'approche de l'invitation en FLE

Les analyses interactionnelles ont mis en évidence que la structuration de la conversation est généralement moins déterminée par la recherche de l'efficacité informationnelle que par le maintien de cette "harmonie interactionnelle" qui constitue, depuis les premières études de Brown & Levinson (1987), l'objet d'étude de la politesse linguistique. La structure de l'interaction étant influencée, en première instance, par sa nature intersubjective et co-gérée, co-construite au cours d'un processus dynamique et dialectique, les actes de langage fonctionnent *"à l'intérieur d'une séquence d'actes"* qui ne sont pas enchaînés au hasard" (Kerbrat-Orecchioni 2001: 53).

En tant que comportement social, ritualisé et codifié, l'interaction demeure par ailleurs, comme signalé précédemment, étroitement liée aux particularités de son cadre socioculturel de référence, ses composantes langagières et comportementales différant souvent d'une culture à une autre. Ce lien entre langue et culture constitue en fait l'un des principaux défis de la communication interculturelle, lieu potentiel de conflits et de malentendus porteurs de clichés et de préjugés, et définit l'une des compétences à développer dans l'acquisition d'une compétence communicative en langue étrangère: ce que le *Cadre*

Européen Commun de Référence pour les Langues nomme "compétence sociolinguistique".

Définie en termes d'adéquation, cette compétence devrait particulièrement éviter aux apprenants de langue étrangère de commettre des erreurs pragmatiques, ce que l'on connaît à la suite de Thomas (1983) comme "*pragmatic failure*". Le concept, développé dans les études interculturelles et appliqué à la didactique des langues (Kasper & Rose 2002), renvoie aux phénomènes de transfert ou d'interférence entre deux ou plusieurs langues. On en distingue deux: l'erreur pragmatolinguistique et l'erreur sociopragmatique, bien que la frontière entre les deux demeure floue (Kerbrat-Orecchioni 2001: 292-302). La première, placée au niveau paradigmatique, caractérise le transfert ou calque d'une expression appartenant à la langue source mais n'ayant pas la même signification – voire aucune signification – dans la langue cible. Ainsi, par exemple, les apprenants hispanophones ont tendance à utiliser l'expression "**passe, passe*" (calquée de l'espagnol) au lieu de "allez-y" pour accueillir et faire rentrer quelqu'un ou pour le laisser passer devant eux. L'erreur sociopragmatique, quant à elle, renvoie au degré d'adéquation et d'acceptabilité d'un énoncé dans le cadre des conventions socioculturelles de la langue cible. C'est souvent le cas rapporté par des enseignants français à propos d'étudiants universitaires espagnols qui s'adressent à eux par leur prénom en leur adressant un "*Salut!*" en guise de formule de salutation. Une erreur commise à ce niveau n'est pas perçue comme un déficit de compétence linguistique, mais comme un manque d'éducation ou de politesse (Piatti 2003), raison pour laquelle il importe de sensibiliser les apprenants, dès le début de l'apprentissage, aux variations interculturelles (Kasper & Rose 2002; Vinther & Dam Jensen 2008), afin de développer chez eux une conscience pragmatique – *pragmatic awareness* – (Meier 1997; Eslami-Rasekh 2005; Pizziconi & Locher 2015), qui assurera non seulement leur compétence en langue étrangère mais aussi leur compétence interculturelle.

Cet aspect s'avère déterminant dans la conception d'une démarche méthodologique. Dans les approches communicatives et actionnelles, il est d'usage d'adopter un schéma qui mène progressivement de l'exposition, ou mise en contact avec les actions langagières ciblées, à la mise en pratique d'une compétence partielle basée sur la reproduction d'un modèle. Entre les deux, des phases intermédiaires visent successivement à la découverte des contenus linguistiques concernés – phonétiques, morphosyntaxiques, lexico-sémantiques –, à la systématisation et à la synthèse. L'ensemble devrait aboutir à une appropriation permettant de dépasser la reproduction d'un modèle pour atteindre l'expression libre et le réinvestissement des compétences acquises pour l'exécution de tâches communicatives dans de nouvelles situations. C'est lors de cette conceptualisation qui sous-tend l'intériorisation des règles de fonctionnement de la langue, et donc l'émergence d'une conscience

linguistique, qu'il faut placer le développement de la conscience pragmatique, comme un processus de sensibilisation aux similitudes/différences entre langues/cultures sources et langues/cultures étrangères dans leurs conditions d'emploi.

Cette démarche peut être représentée comme suit:

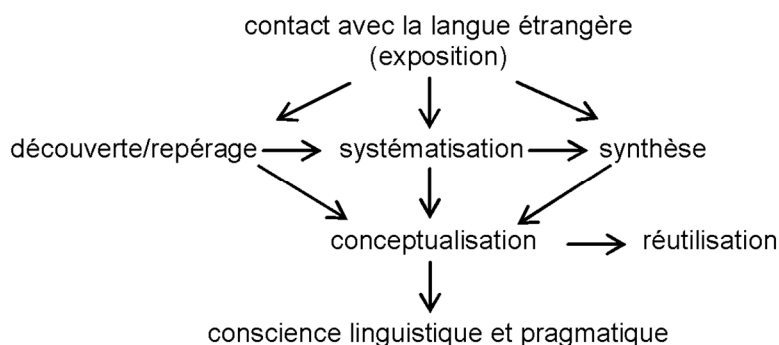


Fig. 1: Démarche méthodologique

Le développement de cette conscience requiert, en amont, un *input* aussi proche que possible de la réalité des usages langagiers, ce qui fait intervenir la question des ressources pédagogiques à utiliser, un choix souvent délicat dans un domaine d'enseignement devenu de plus en plus éclectique depuis l'irruption de l'"authentique" dans la salle de classe et exponentiellement complexifié en raison du nombre de ressources disponibles.

Nous aborderons cette problématique à travers cette action langagière relativement simple qu'est l'invitation, introduite au niveau débutant (A1-A2) et éventuellement reprise, pour l'élargir, au niveau intermédiaire. Pour ce faire, nous vérifierons comment elle est présentée sur trois types de supports: 1) un manuel, 2) un blog d'accompagnement à l'apprentissage, et 3) un site web de podcasts, avant d'explorer les possibilités offertes par un extrait de corpus oral en interaction.

3.1 L'exemple de deux manuels de FLE

Nous avons choisi, comme illustration, deux manuels spécifiquement conçus pour l'acquisition des deux volets de l'oral aux niveaux A1-A2 par Barféty & Beaujoin (2004a, b): *Compréhension orale. Niveau 1* et *Expression orale. Niveau 1*. L'approche proposée, dans les deux cas, comprend une première partie d'exposition à travers des micro-dialogues exemplifiant divers usages et formules d'invitation dont voici un exemple pour chacun des manuels.

(1) Compréhension orale

Claire: Paul, tu es libre samedi soir?

Paul: Ah non, désolé, je dois aller chez mes parents, je dîne chez eux.

Claire: C'est dommage (Barféty & Beaujoin 2004a).

(2) Expression orale

Un homme: Je vais au théâtre ce soir, vous voulez venir avec moi?

Une femme: Avec plaisir (Barféty & Beaujoin 2004b).

La vérification de la compréhension globale ("que se passe-t-il dans ces dialogues?") invite à l'identification de l'action langagière – bien que celle-ci soit déjà donnée dans le titre de l'unité – et propose ensuite le repérage des formules utilisées pour inviter quelqu'un et pour accepter ou refuser une invitation, la systématisation consistant, pour l'essentiel, dans la réutilisation des modèles. La phase de synthèse, enfin, est assurée par la proposition de tables, en guise d'aide-mémoire, regroupant les formules relevées dans les micro-dialogues selon la triade: inviter – accepter – refuser.

Un certain nombre de remarques s'impose, à commencer par le manque d'ancrage situationnel de ces micro-dialogues. En effet, hormis l'alternance du tutoiement/vouvoiement, aucune information contextuelle ne permet de les référer à une situation de communication concrète, ce qui empêche d'envisager les expressions utilisées en termes d'adéquation sociolinguistique. Le plus souvent, en dehors des formules de politesse conventionnelles de salutation ou de remerciement, les concepteurs de matériaux didactiques semblent encore confier ce qui relève de cette dimension à l'intuition de l'apprenant, qui tendra plutôt à calquer les modèles de sa culture d'origine. Deuxièmement, du point de vue de la structuration, les dialogues sont tronqués, aucune séquence d'ouverture ni de clôture ne venant les encadrer. En outre, si des pré-invitations sont repérables, rien n'est pourtant suggéré quant à leur importance interactionnelle et elles ne sont point reprises dans l'aide-mémoire. L'invitation semble ainsi pouvoir être introduite sans préalables et se résoudre en deux ou trois tours de parole, ce qui donne une image faussée de l'interaction. Il est important d'insister sur cet aspect, puisque le moment d'une révision venu – pour une épreuve, par exemple – l'apprenant aura tendance à se reporter à la seule fiche aide-mémoire plutôt que de revenir sur les dialogues introducteurs. Enfin, en ce qui concerne le contenu des formules proposées, il y a lieu de remarquer qu'elles semblent plutôt orientées vers la mise en valeur des verbes *vouloir*, *pouvoir*, *venir* et *devoir* que vers l'exploitation des formules susceptibles d'apparaître dans une vraie interaction, qui impliquerait des traits phonétiques, lexicaux, grammaticaux et syntaxiques propres à l'oral. Ces dialogues illustrent bien en somme la tendance aux dialogues "fabriqués" que Giroud & Surcouf (2016: 9) signalent comme prévalant encore dans la didactique du français langue étrangère.

3.2 Le blog d'accompagnement

Le blog d'accompagnement est une ressource utile et fréquemment utilisée afin de proposer des activités d'élargissement, de renforcement ou de révision. On y trouve, classées par niveau, des activités prêtes à être réutilisées dans les cours, des fiches de synthèse, des liens vers d'autres ressources, etc. Sur le

blog lebaobabbleu.com, par exemple, tenu à jour par une enseignante d'une École Officielle de Langues – établissement public voué à l'enseignement des langues en Espagne –, un menu permet de choisir parmi différentes catégories pragmatiques, grammaticales, lexicales...

Sous la rubrique "Inviter/Accepter/Refuser", des propositions sont faites aussi bien pour le niveau élémentaire (A1-A2) que pour le niveau intermédiaire (B1): des documents, dont on signale qu'ils ont été utilisés pendant les cours, présentés comme synthèse ou aide-mémoire et contenant le répertoire de base des formules employées pour faire une invitation, accepter et refuser. Voici quelques exemples pour chaque niveau:

- (3) Inviter (A1)
 - Tu viens? Vous venez?
 - Tu es disponible? Vous êtes disponible?
 - Est-ce que tu veux venir avec nous?
 - Je peux venir avec un(e) ami(e)? Je peux amener quelqu'un?
 - Ça te dit d'aller au théâtre ce soir?
 - Tu voudrais venir à une exposition avec moi?
- (4) Accepter (A2)
 - D'accord! Super! Génial! Cool!
 - Avec plaisir!
 - Je veux bien!
 - Bonne idée!
 - Oui, pourquoi pas?
 - Volontiers!
 - Parfait!
 - J'accepte
 - C'est parfait! C'est génial!
- (5) Refuser (B1)
 - Ça ne me dit (vraiment) rien
 - Désolé(e) mais...
 - Tu plaisantes!
 - Tu veux rire!
 - Jamais de la vie
 - Je refuse de voir...
 - Je n'ai pas envie de...

Dans ces fiches, l'invitation apparaît accompagnée d'autres actes qui rendent une image plus réelle des négociations survenant dans l'interaction, qui se déroule rarement sur le modèle dyadique invitation/acceptation ou invitation/refus. Outre les formules ci-dessus, sont ainsi mentionnées des expressions pour fixer le rendez-vous (A1), hésiter (A2) ou répondre par une autre proposition (B1). Conformément à une progression en spirale, du niveau A1 au B1, les mêmes contenus et actes sont donc repris en y ajoutant des élargissements. Certains marqueurs de l'oral en situation y apparaissent également tels que *ça te dit*, *on* à la place de *nous* et des exclamations pour l'acceptation.

Cependant si la pré-invitation figure comme vérification de la disponibilité ou de l'envie de participer, aucune indication n'est donnée qui permette de

comprendre son statut de "pré" ni son importance dans le déroulement de l'action du point de vue sociolinguistique. Dans cette même perspective, il y a lieu de noter la présence indifférenciée de formules de refus atténué –*désolé(e)* – et de refus catégorique – *tu plaisantes, jamais de la vie* –, alors que ce dernier est considéré comme un acte non-préférable, parce que potentiellement menaçant pour la face de l'interlocuteur. Enfin, la présentation visuelle sous forme de listes laisse difficilement imaginer le lien reliant interventions initiatrices et réactives.

3.3 Podcasts

De nombreux sites web offrent également des ressources prêtes à l'emploi, dont des archives audio ou vidéo permettant de travailler, en cours ou en autonomie, la compréhension orale. Sur le site "Français facile" de Vincent Durrenberger², nous trouvons par exemple de nombreux podcasts accompagnés d'activités de compréhension et parfois d'une vidéo explicative. Enseignant de français depuis une vingtaine d'années, il se présente lui-même comme ayant été membre du jury du DELF, concepteur de méthodes d'apprentissage et rédacteur des ressources disponibles sur son site.

Divers exemples d'invitation sont proposés, aussi bien pour un niveau débutant que pour un niveau intermédiaire, l'élargissement portant sur la façon de relancer une invitation après un refus. Examinons des extraits selon la transcription qui figure sur le site:

- (6) Niveau débutant. Invitation-acceptation
 Sylvain: Salut Chris, tu fais quoi demain?
 Christelle: Je rentre chez mes parents. Pourquoi?
 Sylvain: J'ai deux places pour le concert des Blacks, tu veux venir?
 Christelle: Génial, je suis trop contente. Ça se passe où? [...]
- (7) Niveau intermédiaire. Invitation-refus
 Franck: Allô? Margaux, tu fais quoi demain?
 Margaux: Rien, pourquoi?
 Franck: Ben, j'ai deux places pour le concert des Ramones.
 Margaux: C'est génial!
 Franck: Ouais, alors je me demandais si ça te dirait de venir avec moi...
 Margaux: Ben ça aurait été avec plaisir, mais je dois préparer un exposé pour mardi. [...]

Bien que lus et enregistrés en studio, les dialogues reprennent divers marqueurs de l'oralité – "petits mots" de l'oral (*ben, trop*) ou la structure interrogative (*tu fais quoi?*) – les rapprochant de l'usage réel de la langue. Si la structure présente certaines troncations – comme l'absence de salutations de Christelle en réponse à Sylvain (6) –, des dialogues complets sont néanmoins proposés avec des séquences encadrantes d'ouverture et de clôture, et ils comprennent également une pré-invitation accompagnée d'un acte de justification.

Or, après visionnage des vidéos explicatives, nous constatons que la pertinence de la pré-invitation – et donc de la structuration – n'est pas mise en

² <https://www.podcastfrancaisfacile.com/>

valeur, puisqu'elle se ramène à un traitement sémantique, le professeur proposant des formules apparentées pour la réalisation de la même action langagière:

et donc sylvain dit tu fais quoi demain↑ qu'est-ce que tu fais demain↑ t'as des projets pour demain↑ t'as prévu quelque chose pour demain↑ (.) tu fais quoi demain↑ christelle lui répond qu'elle rentre chez ses parents.

Le dialogue correspondant au niveau intermédiaire présente en outre une formulation indirecte de l'invitation, interprétée par le professeur comme une conséquence du caractère timide du locuteur (Franck), alors qu'il s'agit d'une atténuation doublement marquée par l'imparfait de politesse "je me demandais si" et l'usage du conditionnel "ça te dirait". Voici ce qui en est dit:

il dit ouais alors je me demandais si ça te dirait de venir avec moi↓ (.) bon c'est une formulation qui est un petit peu ((geste de la main en zig-zag)) pour (.) enfin il a l'air d'être un petit peu timide ce franck hein↑ il pourrait lui dire tu veux venir avec moi↑ ou bien c'est (.) enfin tu viens↑ on y va ensemble↑ mais non il dit je me demandais si ça te dirait de venir avec moi↓ c'est-à-dire qu'il se pose la question si elle aurait envie d'y aller (.) d'y aller avec lui (.) peut-être qu'il n'est pas tout à fait sûr de lui ici dans sa formulation↓ en tout cas avec cette formulation-là on a un peu cette impression-là (.) il est peut-être habitué à ce qu'elle lui refuse ou bien ou alors qu'il n'est pas du tout habitué à lui proposer ou c'est peut-être même la première fois qu'il l'invite.

Malgré la perception du caractère indirect de la formulation, l'explication demeure assez intuitive et risque de réduire à une particularité du locuteur ce qui relève du fonctionnement général de la politesse.

De ressource en ressource, nous nous rapprochons sans doute d'une reproduction plus fidèle des conditions de la parole en interaction, que ce soit du point de vue phonétique, lexico-sémantique ou syntaxique. Cependant, trois éléments sont encore nécessaires si l'objectif est de vraiment confronter l'apprenant à une situation proche de l'immersion: (1) la naturalité des interactions où des bruits, des interruptions et autres aléas situationnels viennent perturber la compréhension, contrairement au caractère préconstruit, récréé et aseptiquement enregistré des dialogues didactisés; (2) la prise en compte de l'importance de la structuration en termes de chaînes d'actions langagières, non dépourvues elles-mêmes d'interruptions, de reprises, de refus entraînant des négociations; et (3) un encadrement théorique permettant à l'enseignant de comprendre, puis de transmettre à son tour, les conditions d'emploi de ces formules et structures. Ces éléments sont accessibles pour l'enseignant dans les extraits et ressources offerts par CLAPI-FLE.

4. L'adaptation des recherches en interaction à l'enseignement

Si le *CECRL* accorde une place de plus en plus importante à l'interaction dans sa dernière version de 2018:

L'interaction, qui comprend deux partenaires ou plus co-construisant le discours, est centrale dans le schéma d'utilisation de la langue du CECR, résumé au début de ce document. On considère que l'interaction orale est à l'origine du langage, avec des fonctions interpersonnelles, collaboratives et transactionnelles (Conseil de l'Europe 2018: 84).

Il reste néanmoins assez vague dans la description des compétences, par exemple dans la formulation de la compréhension d'un interlocuteur au niveau B1: "Peut comprendre en détail ce qu'on lui dit en langue standard", ou au niveau A2: "Peut comprendre suffisamment pour gérer un échange simple et courant sans effort excessif." (Conseil de l'Europe 2018: 87).

Alors que les méthodes d'enseignement de FLE ou de FLS proposent des dialogues stéréotypés et normés ne rendant pas compte de la diversité des enchaînements des locuteurs ni de la variété des contextes, les chercheurs en interaction ont proposé des ressources à partir de leurs corpus et de leurs études pour rendre compte de cette variété et expliciter le fonctionnement des interactions. L'enseignant peut alors exposer l'apprenant à ces ressources de français parlé comme s'il était lui-même en immersion avec ces mêmes locuteurs (Alberdi et al. 2020), en fonction des objectifs pédagogiques et du niveau de langue, la variété des exemples donnant la possibilité d'ajuster ce niveau avec davantage de granularité dans des classes hétérogènes comme les classes intermédiaires (B1-B2 du *CECRL*) (Alberdi et al. 2018).

4.1 Un réseau pour concevoir de nouvelles ressources CLAPI-FLE

Dans l'objectif de concevoir des ressources adaptées aussi bien à la durée d'un cours et aux objectifs des enseignants qu'à leur manière d'aborder une thématique, l'équipe LIS du laboratoire ICAR s'est entourée d'un réseau de didacticiens et d'enseignants pour définir ensemble de nouvelles ressources, les mettre à l'épreuve en situation de classe et les faire évoluer en fonction des différents publics d'apprenants et de la variété des objectifs pédagogiques à l'université, en formation professionnelle ou dans les centres de formation pour migrants économiques ou politiques. C'est avec ce réseau que la plateforme CLAPI-FLE a été conçue et s'est enrichie en fonction des retours d'expérience (Ravazzolo & Etienne 2019).

4.2 Les extraits et les actions langagières

CLAPI-FLE comprend à ce jour une quarantaine d'extraits de différentes situations ordinaires du quotidien, privées ou professionnelles, décrits, transcrits et didactisés, suivant différents niveaux de granularité (simplifié, conseillé ou paramétrable) en fonction des besoins des enseignants. Les interactionnistes ont ensuite délivré une annotation des activités interactionnelles pour mettre en évidence la structure de chacun des extraits, mais également un deuxième jeu plus fin d'annotation de chacune des actions langagières. Ce second découpage apporte un éclairage pragmatique

particulièrement utile quand les productions verbales sont courtes et se composent de petits mots de l'oral bien souvent polysémiques.

4.3 Les fiches pédagogiques

A ces extraits, viennent s'ajouter des collections pour illustrer certains traits interactionnels de l'oral et des fiches pédagogiques qui expliquent de quelle manière une fonction pragmatique peut être accomplie, ou bien les différents emplois d'une expression fréquente à l'oral, dans une variété de contextes, en articulant les procédés lexicaux, prosodiques, syntaxiques, multimodaux et en soulignant les spécificités interculturelles les plus saillantes.

4.4 L'articulation des ressources

Suivant le profil des apprenants et ses objectifs d'apprentissage, l'enseignant pourra ainsi combiner dans une même session de cours l'étude d'une situation, comme l'invitation, à la réalisation de certaines fonctions pragmatiques comme, entre autres, la question, en comparant la formulation ("*tu fais quoi ce soir ↑*") de l'invitation à des attestations en contexte commercial ("*vous voulez un sac ↑*") ou à la forme négative dans une séquence argumentative en réunion professionnelle ("*il t'a pas dit pourquoi ↑*"), qui correspondent à des formulations attestées et fréquentes auxquelles l'apprenant sera confronté s'il interagit à son tour avec des locuteurs natifs.

5. Vers une approche interactionniste de l'invitation en FLE

Mettons-nous à la place d'un enseignant ayant choisi dans CLAPI-FLE un extrait pour travailler l'invitation. Après avoir cliqué sur l'extrait correspondant, il peut d'emblée prendre des décisions fondées tant sur le profil de ses apprenants que sur les objectifs d'apprentissage, grâce aux informations contenues sous les onglets "Contexte", "Exploitation" et "Autres extraits". Les fiches pré-pédagogiques permettent en effet de se faire rapidement une idée tant du contenu thématique en termes lexicaux et grammaticaux que du contenu pragmatique.

Considérons l'extrait "Invitation à une soirée" dans une transcription "simplifiée" où seules les intonations montantes/descendantes sont annotées – la même transcription dans le niveau de granularité "conseillé", où figurent les chevauchements de parole, les pauses, les allongements, les aspirations..., peut être abordée dans un second temps pour appréhender la temporalité de la séquence et apprendre à se familiariser avec les superpositions de parole (voir Fig. 6 en annexe):

1 Ami allô
 2 Invité allô oui
 3 Ami ouais salut ça va?
 4 Invité ça va bien et toi
 5 Ami ouais ça va super!! dis-moi euh tu fais quoi ce soir?
 6 Invité ce soir?
 7 Ami ouais parce que en fait on allait sortir moi Emmanuel et Manon!! on allait sortir pour euh boire un coup et tout en ville donc je sais pas si ça te dirait de venir
 8 Invité ben ouais! ouais ouais ouais
 9 Ami ouais d'accord
 10 Invité ouais ouais
 11 Ami et tu es où tu es où là tu es
 12 Invité ouais ouais
 13 Ami t'es chez toi?
 14 Invité euh ouais je suis euh non là en fait je suis pas chez moi je suis chez chez des potes mais je vais pas tarder à rentrer chez moi
 15 Ami ah ben ouais ben si jamais si tu veux tu peux venir avec tes potes hein!!
 16 Invité ouais je peux les ramener ça fait rien?
 17 Ami ouais ouais vas-y vas-y! euh dis moi vous serez prêts d'ici combien de temps?
 18 Invité ben un quart d'heure dix minutes!!
 19 Ami ben ok d'accord ben passez chez moi directement puis après on va descendre en ville ensemble d'accord?
 20 Invité ok ça marche très bien
 21 Ami on se dit d'ici euh d'ici une demi-heure trois quarts d'heure?
 22 Invité ouais ouais très bien ça marche
 23 Ami ok ok à plus tard allez vas-y ciao man
 24 Invité ok allez ciao

Fig. 2: Transcription: Invitation pour une soirée

Du point de vue phonétique et lexico-sémantique, l'extrait permet de mettre l'apprenant face à l'usage réel de la langue en interaction: les élisions du *e* muet et du pronom *tu*, le registre familier – *boire un coup*, *des potes*, *ciao man* –, des expressions – *ça te dirait*, *on se dit*, *ça marche* –, mais aussi les "petits mots" de l'oral (*ben*, *euh*, *et tout*) qui assument une fonction structurante. À cela s'ajoutent d'autres phénomènes généralement absents des dialogues fabriqués, tels que les chevauchements de parole, éliminés des supports didactiques en vue d'une parfaite intelligibilité et pourtant extrêmement fréquents dans l'interaction, *a fortiori* lorsque celle-ci se produit au téléphone. La construction syntaxique, quant à elle, permet d'approcher aussi bien la grammaire de l'oral à travers, par exemple, l'interrogation (*tu fais quoi*, *t'es où là*), sensiblement éloignée de ce que les apprenants ont intériorisé comme modèles, que les diverses valeurs des reprises, qu'il s'agisse d'auto-répétitions intra-tour (*ouais ouais ouais*, *vas-y vas-y*) ou d'hétéro-répétitions inter-tours (*ce soir* pour la demande de confirmation, *ok allez* à la fois pour la validation de l'accord et en tant que pré-clôture).

Contrairement aux dialogues fabriqués des manuels, ce type d'extrait permet en somme à l'apprenant d'expérimenter non seulement les spécificités de l'oral mais aussi de l'interaction, et de prendre conscience que ces spécificités ne constituent pas des erreurs de construction, comme on tend souvent à le croire, mais au contraire qu'elles réalisent des fonctions (réparations, reformulations,

questions, demandes de confirmation, etc.) et font donc partie de la communication quotidienne à laquelle ils seront confrontés au contact de cette langue étrangère.

Du point de vue de la construction, malgré la simplicité de l'action langagière et la brièveté de l'extrait, il est néanmoins possible de repérer la structure complète d'une invitation, avec des séquences d'ouverture et de clôture répondant au rituel interactionnel de la prise de contact (tours 1-5) et de la prise de congé (23-25), ainsi que le passage de la pré-invitation à l'invitation (tours 5-7), elle-même accompagnée de demandes d'information (11-14 et 17-18) et de propositions (15-17 et 19-20). Ces échanges rendent compte ainsi de l'enchaînement – et souvent l'emboîtement – des divers actes nécessaires à la réalisation complète d'une action langagière. C'est la connaissance et l'appropriation de ce fonctionnement discursif qui fait la différence entre l'acquisition d'une compétence à l'oral et d'une compétence à proprement parler interactionnelle.

Ces informations sont accessibles sous la rubrique "Actions langagières", qui reprend la transcription non plus seulement comme une suite linéaire de tours de parole mais découpée et réorganisée en fonction des actions accomplies – puisque divers actes peuvent être réalisés à l'intérieur d'un seul tour. Ceci permet de mettre en évidence l'organisation des interventions en échanges élémentaires (binaires, comme les paires adjacentes, ou ternaires), eux-mêmes regroupés en actions langagières constitutives d'une activité interactionnelle, donc d'une séquence. À partir de là, il est possible d'élargir les possibilités d'exploitation de cet extrait, applicable à un niveau débutant, selon la démarche méthodologique ci-dessus – notamment pour l'exposition, le repérage et la systématisation – et d'aller au-delà du contenu purement linguistique pour atteindre la dimension pragmatique.

Pour l'exposition, plutôt que d'annoncer le contenu pragmatique, il est intéressant de proposer à l'apprenant une première activité de compréhension globale dans laquelle il devra identifier, outre l'action langagière, le type de communication (émission radio, conversation téléphonique, interview, par exemple), le nombre de locuteurs et le lien qu'il suppose entre eux, tout comme les indices sur lesquels reposent ces hypothèses. Il s'agit en somme de le sensibiliser aux informations liées à la composante situationnelle. Une deuxième écoute, accompagnée d'un QCM, permet ensuite de repérer les particularités lexicales propres à l'oral dans cet extrait.

Le repérage, quant à lui, vise à élargir la recherche d'expressions habituelles pour faire ressortir leur rapport à la structuration. Par exemple, "*ça marche*" utilisée seule marque un accord qui dans l'invitation pourra porter soit sur la réponse soit sur le lieu et l'heure de rendez-vous, alors que "*ça te dit*" portera sur la pré-invitation. Partant du principe que l'intériorisation est encouragée par la découverte de similitudes et de différences, le repérage s'avèrera plus

enrichissant si le travail porte sur deux extraits, que ce soit un deuxième extrait incomplet ou tronqué tiré d'un manuel, ou un extrait présentant des variations situationnelles de registre, de relations, de canal...

Une fois ces expressions repérées, des ensembles d'étiquettes identifiant respectivement les actions langagières et les activités interactionnelles sont délivrées dans le désordre pour que l'apprenant les remette à leur place tout en délimitant les tours de parole qui en font partie. Dans l'exemple suivant, on peut apprécier que le tour numéro 5 se compose de deux interventions, dont l'une (5.1.) viendrait clore l'action précédente (saluer) et l'autre se poserait en pré-invitation (5.2.). En fonction des objectifs et du niveau des apprenants, on peut éventuellement approfondir les niveaux d'emboîtement, comme pour l'échange de confirmation 6.1-7.1.

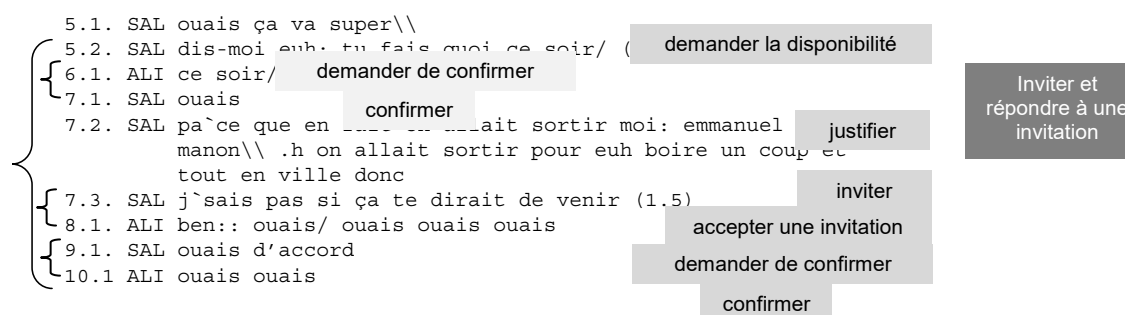


Fig. 3: Repérage de la structuration

L'activité de synthèse, enfin, sert à favoriser la conceptualisation. Plutôt que de proposer des fiches aide-mémoire toutes faites, qui semblent inviter l'apprenant à une mémorisation hors contexte, il est intéressant de faire établir aux apprenants leurs propres synthèses à partir de fiches à remplir qui reprennent aussi bien les expressions que les particularités de structuration observées. Cette activité peut être enrichie à travers l'usage d'un journal de bord d'apprentissage, où chaque apprenant peut en outre consigner des commentaires concernant ses difficultés, les différences par rapport à sa langue maternelle, etc.

6. Perspectives: vers des ressources prêtes à l'emploi

6.1 Les fiches "En pratique"

Pour répondre à la demande des enseignants, parfois déstabilisés par la richesse et la diversité des fonctions et des procédés mis en œuvre dans une interaction de courte durée, nous avons ajouté une rubrique "En Pratique" avec des ressources prêtes à l'emploi en situation de classe; celles-ci mettent l'accent sur les routines interactionnelles, précisent les objectifs linguistiques et socioculturels, incluent des activités sur les thèmes de l'invitation, de l'accueil chez soi, de la question, de la négociation et de la cuisine.

Organisation de la routine communicationnelle

1. SALUTATIONS : Phase ritualisée □
 2. PRE-INVITATION : Etape préalable à l'invitation pour vérifier
 - la disponibilité de l'invité.e → sa volonté de venir
 - ex : *si tu es libre / si tu peux* ex : *ça te dit / si tu veux*
 3. INVITATION → Structure complexe incrémentale avec des reformulations, questions, propositions, atténuations (ex : *un petit repas*) que l'invité.e peut compléter
 - Parfois vague pour rester ouverte
 - ex : *se faire un truc, boire un verre et tout*
 4. REPONSES POTENTIELLES :
 - a. Accepter une invitation : Retour positif avec des marques d'enthousiasme
 - 4.b. Proposer une alternative : Retour avec explication d'un obstacle, nouvelle proposition
 - 4.c. Refuser une invitation : Retour négatif avec justification et ouverture
ex : *on verra plus tard / une autre fois*
 5. NEGOCIATION DU RDV : Négociation et confirmation du lieu et de l'heure
 6. PRISE DE CONGE : Phase ritualisée
- Objectifs communicatifs : saluer, inviter, accepter, refuser, proposer, négocier lieu/horaire RDV, prendre congé
- Objectifs linguistiques et interactionnels : syntaxe, lexique, marqueurs, formules figées, prosodie
- Objectifs socioculturels : salutations, pré-invitation, expression « boire un verre et tout », événement festif (ex : Fête des lumières), emploi intergénérationnel du terme « génial »
- Exploitation pédagogique : niveaux A1-A2, niveaux B1-B2

Fig. 4: CLAPI-FLE. En Pratique: L'invitation

6.2 Des ressources directement utilisables par l'apprenant

Destinée à un public d'apprenants, la plateforme CORAIL (<http://clapi.icar.cnrs.fr/CORAIL>) documente une quinzaine de situations, les expressions figées de l'oral les plus courantes et certains procédés en réponse aux interrogations "Comment dire? Comment faire?" dans une application téléchargeable hors connexion internet et utilisable sur téléphone mobile. Conçue par un réseau de didacticiens, d'enseignants et d'apprenants, elle est actuellement en cours d'évaluation et sera enrichie par les différents retours d'expérience.

6.3 Une compétence interactionnelle plurielle

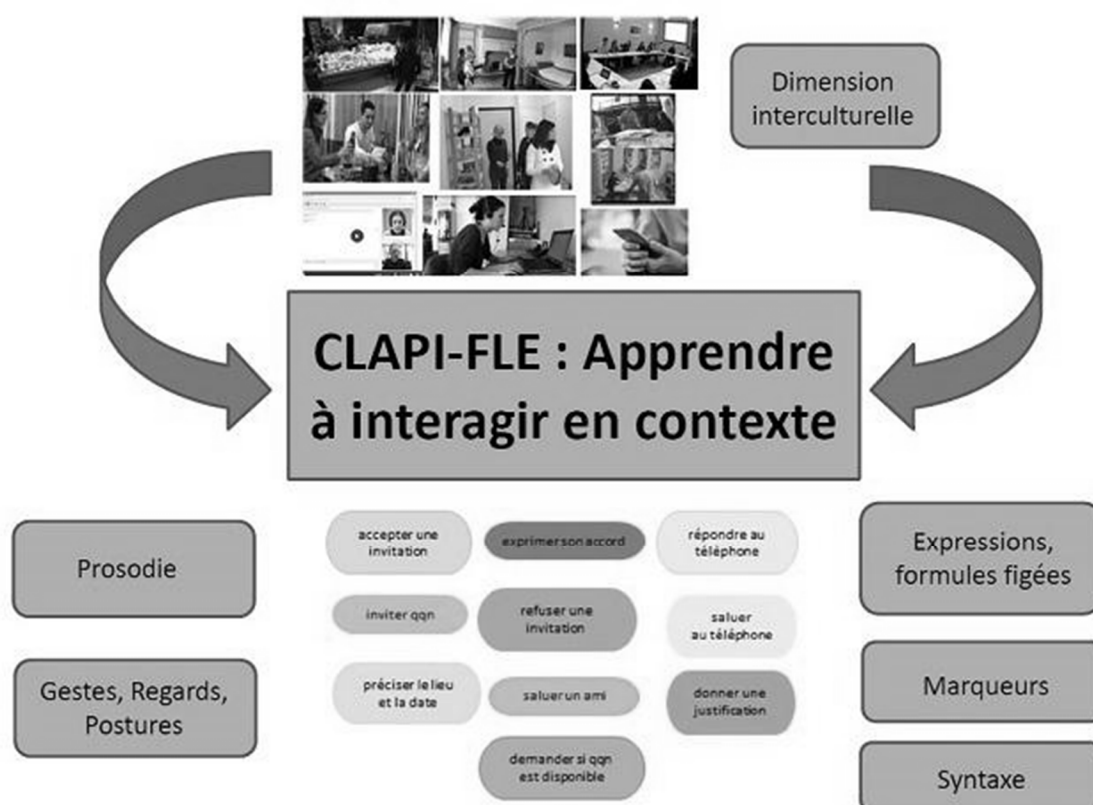


Fig. 5: CLAPI-FLE Apprendre à interagir en contexte

Comme nous venons de le montrer dans le cadre de l'invitation, la compétence interactionnelle est plurielle et son enseignement ne peut se satisfaire de quelques formulations ou règles de syntaxe qui lèveraient toutes les difficultés, à la manière d'une formule magique réutilisable dans n'importe quel contexte, mais demande au contraire à l'apprenant de développer des mécanismes d'écoute, de repérage et de compréhension, qu'il pourra éprouver dans d'autres situations interactionnelles pour acquérir, au fil de ses expositions à la langue orale dans sa naturalité, une perception satisfaisante du français parlé en interaction, lui permettant à son tour d'intervenir dans une interaction.

BIBLIOGRAPHIE

- Alberdi, C., Etienne, C. & Jouin-chardon, E. (2020). Comprendre les spécificités du français oral par l'immersion virtuelle: un défi possible pour les apprenants. In L. Meneses-Lerín, C. Molina & L. Montes (éds.), *Analyser, écrire, traduire et enseigner les langues à l'ère du numérique* (pp. 19-38). Grenade: Comares. halshs-03008804.
- Alberdi, C., Etienne, C. & Jouin-Chardon, E. (2018). Les apports des corpus d'interactions naturelles en situation de classe: enjeux et pratiques. *Action Didactique*, 1, 55-70. <http://www.univ-bejaia.dz/pdf/ad1/Alberdi-Etienne-JouinChardon.pdf>.
- Baldauf-Quilliatre, H. & Colón de Carvajal, I. (2020). Encouragement in videogame interactions. *Social Interaction. Video-Based Studies of Human Sociality*, 2(2).

- Barféty, M. & Beaujouin, P. (2004a). *Compréhension orale. Niveau 1*. Paris: Clé International.
- Barféty, M. & Beaujouin, P. (2004b). *Expression orale. Niveau 1*. Paris: Clé International.
- Brown, P. & Levinson, S. (1987). *Politeness: some universals in language usage*. Cambridge: Cambridge University Press.
- Bruxelles, S. & Traverso, V. (2006). Usages de la particule "voilà" dans une réunion de travail: analyse multimodale. In M. Drescher & B. Job (éds.), *Les marqueurs discursifs dans les langues romanes: approches théoriques et méthodologiques* (pp. 71-92). Berne: Peter Lang.
- Conseil de l'Europe (2001). *Cadre européen commun de référence pour les langues: apprendre, enseigner, évaluer*. Paris: Didier.
- Conseil de l'Europe (2018). *Cadre Européen Commun de Référence pour les Langues: apprendre, enseigner, évaluer. Volume complémentaire avec de nouveaux descripteurs* (<https://rm.coe.int/cecr-volume-complementaire-avec-de-nouveaux-descripteurs/16807875d5>).
- Eslami-Rasekh, Z. (2005). Raising the pragmatic awareness of language learners. *ELT Journal*, 59(3), 199-208.
- Etienne, C. & Jouin, E. (2019). Constituer des ressources pédagogiques pour enseigner le français oral à partir des recherches menées en interaction. In L. Gajo, J.-M. Luscher, I. Racine & F. Zay (éds.), *Variation, plurilinguisme et évaluation en français langue étrangère* (pp. 225-240). Berne: Peter Lang.
- Giroud, A. & Surcouf, C. (2016). De "Pierre, combien de membres avez-vous?" à "Nous nous appelons Marc et Christian": réflexion autour de l'authenticité dans les documents oraux des manuels de FLE pour débutants. *SHS Web of Conferences*, 27.
- Guinamard, I., Jouin-Chardon, É., Rispaill, M., Traverso, V. & Thai, T. (2015). *Langues parlées, interactions sociales. Une variété d'usages pour l'apprentissage du français*. Paris: L'Harmattan.
- Heritage, J. (1984). *Garfinkel and Ethnomethodology*. Cambridge: Polity Press.
- Kasper, G. & Rose, K. (2002). *Pragmatic development in a second language*. Oxford: Blackwell.
- Kerbrat-Orecchioni, C. & Traverso, V. (éds.) (2008). *Les interactions en site commercial. Invariants et variations*. Lyon: ENS Éditions.
- Kerbrat-Orecchioni, C. (1990, 1992, 1994). *Les interactions verbales* (tomes I-II-III). Paris: Armand Colin.
- Kerbrat-Orecchioni, C. (1996). *La conversation*. Paris: Seuil.
- Kerbrat-Orecchioni, C. (2001). *Les actes de langage dans le discours: Théories et fonctionnement*. Paris: Nathan.
- Kerbrat-Orecchioni, C. (2004). L'adjectif petit comme procédé d'atténuation en français. *Travaux et Documents 24*. Université Paris 8, Saint-Denis, 153-175.
- Meier, A.J. (1997). Teaching the universals of politeness. *ELT Journal*, 51(1), 21-28.
- Pekarek-Doehler, S. (2006). Compétence et langage en action. *Bulletin Suisse de Linguistique Appliquée*, 84, 9-45.
- Piatti, G. (2003). La cortesía: un contenido funcional para los programas de español como lengua extranjera. In D. Bravo (éd.), *La perspectiva no etnocentrista de la cortesía: identidad sociocultural de las comunidades hispanohablantes* (pp. 355-368). Stockholm: Université de Stockholm-EDICE.
- Piccoli, V., Ticca, A. C. & Traverso, V. (2019). "Go it's Internet here": Démarches administratives de personnes précaires ou en demande d'asile. *Langage et société*, 167, 81-110.
- Pizziconi, B. & Locher, M.A. (éds.) (2015). *Teaching and learning (im)politeness*. Berlin: Mouton de Gruyter.
- Ravazzolo, E. & Etienne, C. (2019). Nouvelles ressources pour le FLE à partir des études en interaction. *LINX*, 79 [<https://journals.openedition.org/linx/3454>].



- Ravazzolo, E., Etienne, C. & Ursi, B. (à paraître). Apprendre les interactions en classe de français: enjeux et pratiques, *Cahiers de l'Acedle*.
- Ravazzolo, E., Traverso, V., Jouin, É. & Vigner, G. (2015), *Interactions, dialogues, conversations: l'oral en français langue étrangère*. Paris: Hachette.
- Riegel, M., Pellat, J.-C. & Rioul, R. (1999). *Grammaire méthodique du français* (3e éd). Paris: PUF.
- Schegloff, E. (2007). *Sequence Organization in Interaction, vol. I: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.
- Thomas, J. (1983). Cross-Cultural Pragmatic Failure. *Applied Linguistics*, 4, 91-112.
- Traverso, V. & Ravazzolo, E. (2016). Définitions ostensives co-construites: le cas de la visite guidée. *Langages*, 204, 43-66.
- Traverso, V. (1996). *La conversation familière. Analyse pragmatique des interactions*. Lyon: Presses Universitaires de Lyon.
- Traverso, V. (1999). *L'analyse des conversations*. Paris: Armand Colin.
- Traverso, V. (2011). Chevauchements de parole, résolution et réparation dans la conversation en ligne. In C. Develotte, R. Kern & M.-N. Lamy (éds.), *Décrire la conversation en ligne. Le face à face distanciel* (pp. 117-145). Lyon: ENS Éditions.
- Traverso, V. (2016). *Décrire le français parlé en interaction*. Paris: Ophrys.
- Traverso, V., Ticca, A. C. & Ursi, B. (2018). Invitations in French: A complex and apparently delicate action. *Journal of Pragmatics*, 125, 164-179.
- Vinther, T. & Dam Jensen, E. (2008). El aprendizaje de la cortesía verbal a través de input audiovisual o escrito. In A. Briz, A. Hidalgo, M. Albelda, J. Contreras & N. Hernández (éds.), *Cortesía y conversación: de lo escrito a lo oral* (pp. 798-809). Valence: Université de Valence.
- Wierzbicka A., (1991). *Cross-Cultural Pragmatics. The semantics of Human Interaction*. Berlin/ New-York: Mouton de Gruyter.

Conventions de transcription

Les annotations des phénomènes spécifiques à l'interaction (chevauchement, pause, enchaînement des productions) et de la manière dont les mots sont prononcés (accentuation de syllabes, allongements, intonations, élisions non standard ou troncations) sont représentés dans la transcription par les signes suivants:

Notation de la parole

[]	chevauchements	:	allongement vocalique
(2.1)	pauses en secondes	exTRA	segment accentué
(.)	micro-pause	`	élision (<i>b`jour</i> ou <i>d`mande</i>)
↑ ↓	intonation montante - descendante	par-	troncation
((rire))	phénomènes non transcrits	.h	aspiration

Annexe

1 SAL allô
 2 ALI allô (0.9) oui (0.7)
 3 SAL ouais salut ça va↑
 4 ALI ça va bien et toi
 5 SAL ouais ça va super↓↓ (.) dis-moi euh: tu fais quoi ce soir↑ (1.6)
 6 ALI ce soir↑
 7 SAL ouais pa`ce que en fait on allait sortir moi: emmanuel et manon↓↓ .h on allait sortir pour euh boire un coup et tout en ville donc j`sais pas si ça te dirait de venir (1.5)
 8 ALI ben:: ouais↑ ouais ouais ouais
 9 SAL ouais d'accord
 10 ALI ouais ouais
 11 SAL [et t'es où] t'es où l- là t'es:
 12 ALI [ouais ouais]
 13 SAL t'es chez toi↑ (1.5)
 14 ALI euh: ouais j`suis euh: non là en fait j`suis pas chez moi j`suis chez des potes mais j`vais pas tarder à rentrer chez moi
 15 SAL ah ben ouais ben: si jamais: si tu veux tu peux venir avec tes potes hein↓↓ (1.3)
 16 ALI ouais j`peux les ramener ça fait rien↑
 17 SAL ouais ouais vas-y vas-y↑ (.) euh:: dis moi vous s`rez prêts: d'ici combien d` temps↑ (1.8)
 18 ALI be:n:: un quart d'heure dix minutes↓↓
 19 SAL be:n ok d'accord ben: passez chez moi directement: puis après: on va descendre en ville ensemble d'accord↑ .h (0.9)
 20 ALI ok ça [marche très bien]
 21 SAL [on s` dit d'ici::] euh: d'ici une demi-heure trois quarts d'heure↑
 22 ALI ouais ouais très bien [ça marche]
 23 SAL [ok] ok à plus tard (.)
 allez vas-y (0.3) ciao man
 24 ALI ok allez ciao
 25 SAL ciao↓↓

Fig. 6: CLAPI-FLE. Transcription conseillée

Corpus: exploration, médiation et autonomisation.

L'utilisation du concordancier de la plateforme FLEURON en cours de FLE

Biagio URSI

Aix-Marseille Université & Laboratoire Parole et Langage – UMR 7309 (CNRS, AMU)
Département de Lettres modernes, Aix-Marseille Université
29, Avenue Robert Schuman, 13621 Aix-en-Provence Cedex 01, France
biagio.ursi@univ-amu.fr

Virginie ANDRÉ

ATILF – UMR 7118 (Université de Lorraine et CNRS)
44, Avenue de la Libération BP 30687, 54063 Nancy Cedex, France
virginie.andre@univ-lorraine.fr

In this paper, using a specific didactic approach, data-driven learning, we study the use of the FLEURON platform (<https://fleuron.atilf.fr/>) and, in particular, its concordancer by FLE (French as a foreign language) teachers and allophone students of different levels. Learners rely on this tool to observe data, identify specific forms, propose categorizations and generalize the use of these forms based on real occurrences and sufficiently broad contexts. The concordancer serves as a resource for autonomous learning which, after initial trial and error, results in a real linguistic "investigation". Based on classroom recordings, we propose a multimodal interactional analysis of sequences where the use of certain words in context is illustrated through the mobilization of conversational data within the FLEURON platform. This will lead us to identify the emerging features of the learning process among users of the platform, as for learners' initiatives and their responses to teacher's solicitations.

Keywords:

Digital platforms, interactional corpora, French as a foreign language, concordancer, data-driven learning, multimodality, interaction analysis.

Mots-clés:

Plateformes numériques, corpus de la langue parlée en interaction, FLE, concordancier, apprentissage sur corpus, multimodalité, analyse des interactions.

1. Introduction

L'exploitation des corpus à des fins didactiques s'inscrit dans la suite et en complément de l'utilisation des documents authentiques, défendue depuis les années 1970 (Abé et al. 1979). Néanmoins, de cette exploitation, qui peut prendre différentes formes, émergent de nouvelles méthodologies dans le champ de l'enseignement et de l'apprentissage des langues (Boulton & Tyne 2014), notamment du français langue étrangère (FLE) et des compétences interactionnelles (André 2019). Parmi les différentes façons d'exploiter des corpus, nous nous intéresserons à la démarche associée au *data-driven learning* (Johns 1991; Aston 2001) qui propose à des apprenant·e·s de langue

d'interroger des corpus à l'aide d'un concordancier, puis d'observer les résultats, de les analyser, de les classer et d'induire les règles de fonctionnement des faits langagiers recherchés.

Le dispositif FLEURON (Français Langue Étrangère Universitaire: Ressources et Outils Numériques, <https://fleuron.atilf.fr/>) propose à des étudiant·e·s étranger·e·s, qui souhaitent se préparer à interagir dans les situations de communication qu'elles·ils vont rencontrer en France, un corpus d'interactions outillé d'un concordancier. Le corpus multimodal de FLEURON est constitué d'interactions universitaires et de situations vécues par des étudiant·e·s en dehors du campus (achats de billets en gare, de pain dans une boulangerie, explications d'itinéraires, etc.). Dans le cadre de cette étude, nous nous intéressons aux activités et aux processus analytiques mis en œuvre par des apprenant·e·s qui focalisent leur attention sur les résultats de leurs recherches dans le concordancier de FLEURON. Ces recherches, par chaînes de caractères, concernent des mots isolés, des blocs de mots ou des constructions spécifiques en français parlé. Nous proposons une analyse multimodale (Traverso 2016; Mondada 2017) des interactions qui prennent place lors de l'utilisation du dispositif FLEURON en cours de langue, entre les apprenant·e·s et entre les apprenant·e·s et l'enseignant·e qui les accompagne. Nous présenterons tout d'abord le cadre de cette étude et la mise en place des expérimentations avant d'exposer leurs analyses.

2. L'apprentissage sur corpus et le domaine de l'interaction

Le *data-driven learning*, ou en français l'apprentissage sur corpus (ASC) (Boulton & Tyne 2014), a fait ses preuves dans le domaine de l'enseignement et de l'apprentissage de l'anglais écrit (Breyer 2011; Boulton & Cobb 2017). Cependant, il n'existe que de rares travaux mettant en œuvre la démarche de l'ASC avec des corpus oraux (voir notamment Zorzi 2001 pour l'italien, Mauranen 2004 pour l'anglais ou Kerr 2013 pour le français) et encore plus rares avec des corpus multimodaux (Braun 2007). Les premières expérimentations de l'ASC pour acquérir des compétences interactionnelles en français s'inscrivent dans le cadre du projet FLEURON qui a expérimenté les principes de l'ASC avec son corpus multimodal et son concordancier (André 2017; André & Ciekanski 2018). L'acquisition de compétences interactionnelles permet à un·e apprenant·e de mobiliser les ressources (lexicales, syntaxiques, mais aussi multimodales) appropriées pour réussir à participer à un échange finalisé et situé avec un·e ou plusieurs locuteur·trice·s. Ces compétences reposent sur la triple capacité pour un·e locuteur·trice de s'adapter à la situation de communication et à tous les éléments qui la composent et qui influencent les interactions, de s'ajuster aux propos de son interlocuteur·trice pour produire des actions séquentiellement adéquates, et de construire collaborativement un discours commun.

En suivant la démarche de l'ASC, les apprenant·e·s sont amené·e·s à changer de paradigme, traditionnellement centré sur la transmission des connaissances, et à entrer dans un processus d'apprentissage de la langue qui prend appui sur trois étapes: observation – catégorisation – généralisation (Johns 1991). Ainsi, tout d'abord, l'apprenant·e est amené·e à observer les données, extraites par un concordancier et consultables sous la forme de listes de lignes de concordances. Elle·il analyse la distribution des occurrences d'un mot (grammatical ou lexical), c'est-à-dire qu'elle·il compare ses conditions d'apparition et ses usages. Elle·il procède ensuite à une catégorisation de ces données et réunit les occurrences qui ont le même sens, qui ont le même cotexte, ou encore qui s'intègrent dans des *chunks* ou des blocs de mots relevant d'une analyse syntaxique. Enfin, en s'appuyant sur cette catégorisation, l'apprenant·e induit les règles de fonctionnement des occurrences recherchées. Les activités cognitives et métacognitives développées lors de ces trois étapes semblent favoriser l'appropriation de la langue, de ses usages et de son fonctionnement en contexte. L'apprenant·e devient actif·ve dans son apprentissage en prenant le rôle d'un·e apprenti·e-linguiste ou d'un "Sherlock Holmes" (Johns 1997: 101). Gavioli (2001) et Bernardini (2004) parlent de *learner as researcher* pour évoquer les activités dans lesquelles s'investissent les apprenant·e·s en analysant des lignes de concordances. L'enseignant·e change également de rôle. Chambers (2010) explique ce changement en ces termes:

The role of the teacher also changes fundamentally, as s/he is no longer the sole source of knowledge about target language, but rather a facilitator of the learning process, helping the learners to interpret the data, and giving them advice on how best to search the corpus and analyse their results. (Chambers 2010: 355)

L'enseignant·e devient un·e guide, un·e conseiller·e ou un·e accompagnateur·trice afin d'impulser la démarche de l'ASC et les activités associées. En outre, en apprenant comment apprendre, l'apprenant·e s'autonomise au fur et à mesure de ses expériences. À ce propos, Aston souligne: "[p]erhaps the greatest attraction of corpora for language pedagogy is their potential for autonomous learning" (2001: 41).

La linguistique de corpus associée à la recherche en didactique des langues a impulsé l'ASC. La méthodologie d'analyse des données en corpus a été exploitée par des apprenant·e·s de langue pour entrer dans une démarche active de compréhension et d'analyse des données. Cette méthodologie se révèle particulièrement intéressante et pertinente pour des apprenant·e·s qui souhaitent développer leurs compétences socio-interactionnelles à l'oral, dans la mesure où elles·ils apprennent à communiquer de manière efficace en situations réelles. Tout d'abord, les manuels de langue ne présentent pas ou peu le français parlé en interaction tel qu'il est réellement produit (Giroud & Surcouf 2016). Ils prescrivent un enseignement de la norme, souvent en référence à l'écrit, et s'éloignent de l'exposition à une langue réelle et située.

Ensuite, la description dans les documents de référence du français parlé en interaction est encore parcellaire (Traverso 2016: 13-14) et les enseignant·e·s de FLE peuvent être légitimement démuni·e·s pour proposer à leurs apprenant·e·s des règles de fonctionnement. Ainsi, l'ASC permet aux apprenant·e·s ainsi qu'aux enseignant·e·s d'observer des interactions situées et de saisir des phénomènes langagiers en usage.

3. Les expérimentations de FLEURON

Dans le cadre du projet FLEURON, nous avons souhaité documenter les usages que des apprenant·e·s de langue étrangère font de cette plateforme et du dispositif d'apprentissage qu'elle propose¹. Notre attention s'est focalisée sur le concordancier multimédia (André & Ciekanski 2018). Cet outil permet non seulement de rassembler et de lister les occurrences d'une même forme linguistique (ou d'une même chaîne de caractères) présente dans différentes transcriptions, mais aussi de visionner les extraits vidéo des ressources correspondantes. En d'autres termes, toutes les occurrences présentées en liste dans les résultats du concordancier sont des hyperliens qui permettent d'afficher la ressource dans un nouvel onglet et d'avoir ainsi accès au signal et à la transcription de l'enregistrement dans lequel l'occurrence est prononcée. Plus précisément, la vidéo s'ouvre au moment où l'occurrence sur laquelle l'apprenant·e a cliqué est prononcée. Le titre de l'extrait vidéo ainsi qu'une description succincte de la situation de communication et des thèmes qui y sont abordés fournissent également des éléments de contextualisation précieux pour l'exploration de la ressource.

Nos expérimentations se sont déroulées avec des apprenant·e·s du Département de Français Langue Étrangère (DéFLE) au Centre de Langues Yves Chalon (CLYC) de l'Université de Lorraine avec la collaboration de Florence Poncet, enseignante de FLE. Les données sur lesquelles ce travail se fonde sont issues d'une campagne de collecte menée au CLYC en septembre 2019, pendant les cours intensifs de langue française destinés aux étudiant·e·s internationaux·ales, dans le cadre du programme ERASMUS. Nous avons enregistré les séances d'étudiant·e·s face à un ordinateur, seul·e·s et en binômes, accompagné·e·s d'un enseignant/tuteur qui a surtout assuré la médiation des fonctionnalités du site internet et des spécificités de l'interface numérique, auxquelles les apprenant·e·s avaient déjà été initié·e·s. Pour cet article, nous nous focaliserons uniquement sur une seule des configurations expérimentées, celle avec deux étudiant·e·s et un enseignant. Pendant ces expériences, les apprenant·e·s ont exploré les ressources de FLEURON en s'appuyant sur le concordancier; en particulier, elles·ils ont recherché des

¹ Cette étude expérimentale est soutenue par le projet OLKi (*Open Language and Knowledge for Citizens*) de Lorraine Université d'Excellence (LUE), ANR-15-IDEX-04-LUE.

formes linguistiques qu'elles·ils ont, dans la plupart des cas, elles·eux-mêmes choisies.

Concernant le dispositif de captation vidéo, nous avons utilisé deux caméras: l'une focalisée sur les participant·e·s, qui a offert une prise de vue latérale et générale sur les postures des corps et les orientations mutuelles des participant·e·s; l'autre focalisée sur l'écran. Nous avons opté pour cette dernière option, afin d'avoir accès aux détails des pointages et des gestes accomplis devant l'écran par les participant·e·s, plutôt que d'effectuer une captation depuis l'ordinateur (qui aurait permis de documenter exclusivement les mouvements du curseur et la transition entre les différentes fenêtres de la plateforme à l'écran). Par la suite, nous avons édité les données afin d'obtenir dans un même fichier les vidéos synchronisées des deux prises de vue, comme nous pouvons le voir dans la Fig. 1. Ici, des apprenant·e·s travaillent en binôme face à un ordinateur, en présence d'un enseignant; elles·ils parcourent une liste d'occurrences du concordancier FLEURON (qui apparaît dans la prise de vue incrustée en haut à droite, voir 4.1 pour des détails ultérieurs).

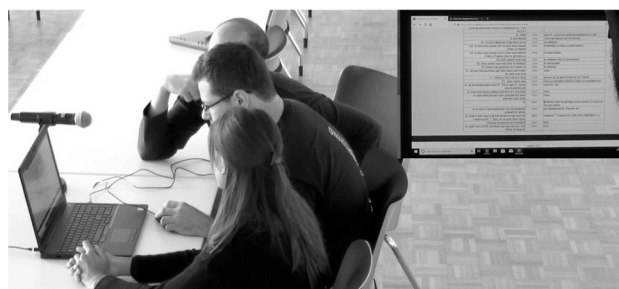


Fig. 1: Double captation vidéo

4. Analyse

L'analyse séquentielle et multimodale de trois extraits représentatifs nous permet de souligner des éléments de réflexion concernant les modalités d'exploration de la base FLEURON par le biais du concordancier et, en particulier, la découverte chez les apprenant·e·s des régularités linguistiques du français parlé à partir de corpus conversationnels. La nature de cette étude est donc principalement qualitative, mais nous avons observé des éléments récurrents qui se prêteraient à une appréciation quantitative. Notre regard analytique est focalisé sur le rôle des ressources multimodales mobilisées par les apprenant·e·s, dans le cadre d'une activité ciblée qui s'inspire des principes de l'apprentissage sur corpus et qui se fonde sur des documents conversationnels authentiques, en vue du développement des compétences linguistiques des apprenant·e·s. Comme nous le verrons, même l'initiative de l'apprenant·e à partir de problèmes locaux peut amener à un processus de découverte quant aux spécificités de la langue parlée en interaction et de

certaines formes linguistiques, notamment la préposition "depuis" (4.1) et la locution verbale "valoir mieux" (4.2).

4.1 Observation des formes linguistiques et problèmes émergents

Dans l'extrait suivant, trois participant·e·s sont présent·e·s. Deux étudiant·e·s Erasmus en troisième année de licence ayant un niveau de compétence avancé en français², LAU (Laura), une étudiante italienne, et MIR (Mircea), un étudiant roumain, se penchent sur la liste des occurrences de la préposition "depuis", qu'elles·ils ont elles·eux-mêmes choisi de rechercher; Georges (GEO) est l'enseignant de français. Cette liste est affichée dans le concordancier et elle est parcourue verbalement par les deux apprenant·e·s au début de l'extrait 1.

(1) CLYC_Groupe 1_Depuis_a (18:48)

```

01  LAU  cinq ans/ (0.3) oui (.) °ça ça indique toujours le [temps°
02  MIR                                     [depuis
03      cinq ans
04  GEO  depuis cinq ans/
05      (0.3)
06  MIR  x xx
07      (4.0)
08  MIR  hm la première année/
09      (0.9)
10  GEO  depuis la première année
11      (0.3)
12  MIR  dix minu:tes depuis quelque euh année/
13      (0.5)
14  MIR  *euh depuis: quelle année/
15  mir  *secouements de tête--->
16      (0.3)*(0.7)
17  mir  -->*
18  MIR  euh: depuis l'année dernière (0.4) de nouveau du temps
19      (0.5)
20  MIR  xxx
21      (2.5)*(0.3)
22  lau  *....>
23      *....>
24  LAU  et ici/*#1
25  lau  >...*parcourt du doigt une ligne du concordancier-->
26      >...*se penche vers l'écran

```

² B1/B2, d'après le test effectué par l'enseignante responsable du module de formation selon les critères du *Cadre européen commun de référence pour les langues*. La plateforme FLEURON a été précédemment présentée aux apprenant·e·s qui ont participé à nos expériences.

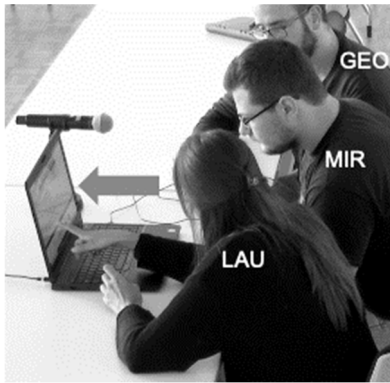


Image 1a

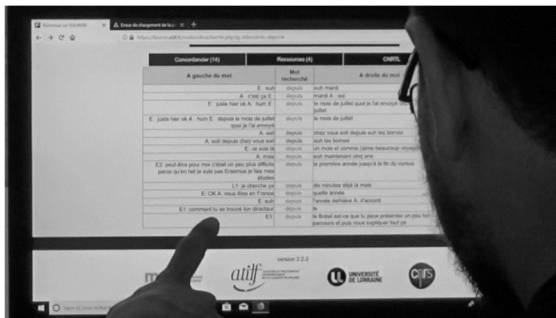


Image 1b

A gauche du mot	Mot recherché	A droite du mot
E : euh	depuis	euh mardi
A : c'est ça E :	depuis	mardi A : oui
E : juste hier ok A : hum E :	depuis	le mois de juillet quoi je l'ai envoyé depuis le mois de juillet
E : juste hier ok A : hum E : depuis le mois de juillet quoi je l'ai envoyé	depuis	le mois de juillet
A: soit	depuis	chez vous soit depuis euh les bornes
A: soit depuis chez vous soit	depuis	euh les bornes
E: je suis là	depuis	un mois et comme j'aime beaucoup voyager etc
A: mais	depuis	euh maintenant cinq ans
E2: peut-être pour moi c'était un peu plus difficile parce qu'en fait je suis pas Erasmus je fais mes études	depuis	la première année jusqu'à la fin du cursus
L1: je cherche ça	depuis	dix minutes déjà là mais
E: OK A: vous êtes en France	depuis	quelle année
E: euh	depuis	l'année dernière A: d'accord
E1: comment tu as trouvé ton directeur	depuis	le
E1:	depuis	le Brésil est-ce que tu peux présenter un peu ton parcours et puis nous expliquer tout ça

Image 1c

[illegible]

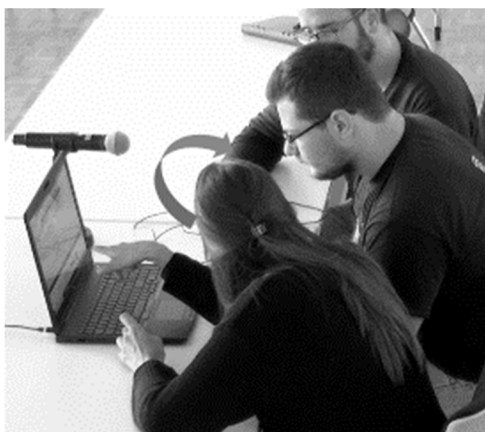


Image 2

```

24          (0.5)
25  GEO  alors on+ va VOIR*+ on peut* voir
    lau                                >,,*
                                -->+,,,,,,,,,+      *se redresse->
26          (0.8)*(0.6)
    lau      -->*
27  MIR  ((clique sur l'occurrence du concordancier))

```

Dès les premières lignes, les deux apprenant·e·s sont focalisé·e·s sur le contexte droit des quatorze occurrences de *depuis* listées dans le concordancier (image 1c). La lecture des différents éléments qui suivent cette forme linguistique ne s'inscrit pas dans une simple activité de restitution, mais dans une véritable démarche réflexive à partir de données linguistiques en contexte qui est réalisée de manière récurrente dans nos expériences. De ce point de vue, le tour ligne 1 est représentatif dans son articulation tripartite: la lecture du syntagme introduit par *depuis* ("cinq ans") est suivie d'une pause et, surtout, d'une particule ("oui") qui signale l'enregistrement de cette unité par rapport à sa simple mention. Une généralisation est alors produite de manière située, en guise de confirmation d'une hypothèse déjà formulée ("ça ça indique toujours le temps"). La structure de cette généralisation permet de relier la singularité de l'occurrence à une constatation qui est corroborée par la pluralité de données à disposition de l'apprenante-enquêtrice, par le biais d'une dislocation à gauche et d'une caractérisation fréquentielle ("toujours"), respectivement. Cette généralisation est prononcée à voix basse et, pour cette raison, ne fait pas l'objet d'une validation de la part de l'enseignant.

De son côté, Mircea est engagé dans une activité spécifique: il parcourt les occurrences affichées dans le concordancier. Il lit parfois les mots qui suivent la forme cible (par ex. "la première année", l. 08) parfois la totalité du syntagme prépositionnel (par ex. "depuis cinq ans", l. 02-03). Les mentions des lignes du concordancier sont hétéro-répétées par Georges, qui alimente ainsi la progression de l'activité par la reprise de formes étendues; ces ressources jouent donc un rôle structurant pour l'activité en cours (cf. Berger 2008). Lignes

01-14, l'intonation énumérative finale des interventions permet d'inscrire ces tours de parole dans une même trajectoire actionnelle.

Au sein de l'activité de lecture, des éléments problématiques émergent. Mircea revient sur la production de la ligne 12 "depuis quelque euh année", où un coup de glotte est perceptible entre les deux mots qui suivent la forme cible, et la répare par sa forme correcte "euh depuis: quelle année" (l. 14). Cette auto-réparation est accompagnée d'un secouement de tête évident chez l'apprenant, comme le montre la transcription multimodale. Comme Laura, Mircea ajoute lui aussi une caractérisation qui permet de spécifier rétrospectivement les mots situés à droite de la forme cible, qui font l'objet de sa lecture, et verbalise une généralisation en train de se faire ("de nouveau du temps", l. 16). Ligne 20, un autre problème est signalé par Laura, ce qui va déclencher une longue séquence de réparation (Schegloff et al. 1977; Schegloff 2007) présentant des réarrangements corporels notables. Elle se penche vers l'écran (image 1a) et indique la zone occupée par la transcription problématique (image 1b), à l'aide d'un pointage dynamique. En effet, elle parcourt du doigt la ligne du concordancier et puis s'arrête pour lire ce qui suit le *depuis* ("depuis le: (0.6)"), ou, pour mieux dire, afin de signaler un manque notable ("°après° c'est fini", l. 23), le pointage est alors figé (image 2). Pendant la lecture, elle réoriente son corps se penchant davantage en avant et tournant sa tête vers sa droite pour chercher un contact visuel avec l'enseignant, qui est ainsi interpellé: une reconfiguration du cadre participatif (Goodwin & Goodwin 2004) a lieu par rapport à la position initiale des participant·e·s, où leurs têtes étaient alignées (comme le montre la Fig. 1) et leurs regards focalisés sur l'écran.

Ligne 25, Georges réagit en formulant une instruction indirecte, qui signale le caractère interactif de l'interface du concordancier et évoque la possibilité de visionner l'enregistrement correspondant à la ligne problématique (une fonctionnalité proposée par le concordancier multimédia). Cette instruction est présentée comme envisageant une opération qui mobilise la totalité des participant·e·s (voir notamment l'utilisation du pronom sujet "on", dans les deux expressions verbales de ce tour), en tant qu'observateur·trice·s face aux données. C'est Mircea, la souris dans la main, qui accomplit l'action instruite, c'est-à-dire l'ouverture de la fenêtre présentant la vidéo obtenue par un clic sur l'occurrence désignée (l. 27). Ainsi, la recherche de l'occurrence problématique est réalisée dans le contexte élargi de la transcription intégrale de l'enregistrement (le deuxième extrait constitue la suite du premier).

(2) CLYC_Groupe 1_Depuis_b (18:48)

```

28   MIR   <(8.2) ((fait défiler la transcription vers le bas))>
29   MIR   ah non
30   <(7.0) ((fait défiler la transcription vers le haut))>
31   LAU   *[lâ]##3
32   MIR   *[hm]* ((petit rire))*
      lau   *....*pointe l'écran--*,,>

```

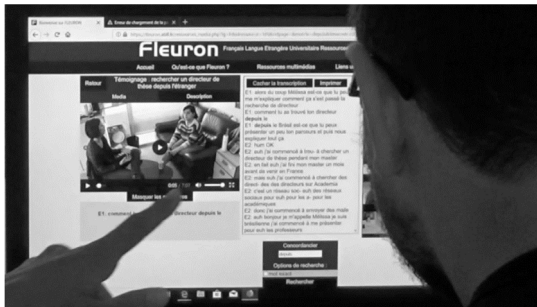


Image 3

33 (0.4)*(2.9)
 34 lau >,,,*
 35 MIR ((clique sur play))
 36 vid <((audio de la vidéo en ligne)) comment tu as trouvé ton
 37 directeur depuis le:: depuis le brésil est-ce que tu peux
 38 présenter un peu ton [parcours et puis [nous ex[pliquer&
 39 MIR [ah::
 40 LAU [ah o[kay
 41 MIR [c'est ça
 42 vid &tout ça hum okay\>
 43 MIR ((clique sur pause))
 44 GEO alors voilà\
 45 MIR depuis le brésil c- c'était ça
 46 (0.6)
 47 LAU mh mh
 48 (0.3)
 49 MIR mh mh
 50 (1.0)
 51 GEO donc c'é*tait un:*#4 quelque chose qui [arri-
 lau *.....*se penche en avant, se tourne vers GEO
 52 LAU [elle a *ré*pété*#5
 lau *...*-----*,,,,>

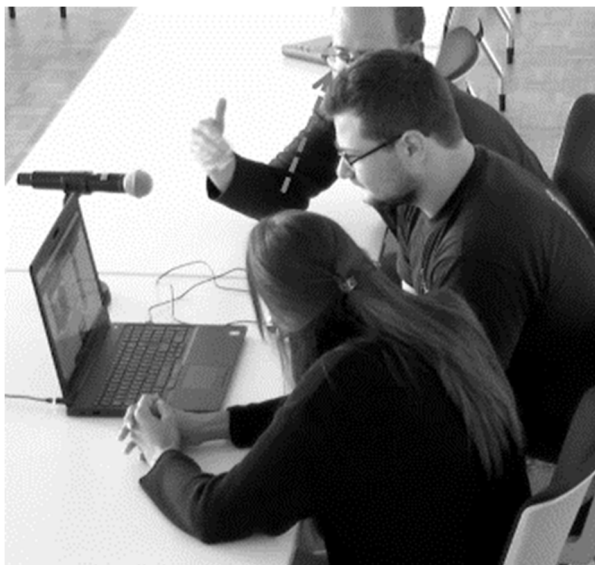


Image 4



Image 5

52 LAU e*lle a *con*tinué*#6 a*près
 lau >,* *...*-----*,,,,*(pointe l'écran)



Image 6

53 GEO voilà

Après l'affichage de la transcription de l'extrait intitulé "Témoignage: rechercher un directeur de thèse depuis l'étranger", la recherche de l'occurrence problématique prend un peu de temps (voir les longs silences lignes 28 et 30, dûs au défilement du texte de la transcription) et comporte également des erreurs (l. 29 Mircea s'arrête sur un tour de parole qui n'est pas celui recherché et s'exclame par un "ah non"³), mais cette activité est une activité collective, à laquelle les deux apprenant·e·s contribuent de manière active. Mircea manipule la souris et fait défiler le texte de la transcription jusqu'au moment où Laura indique l'expression recherchée de manière multimodale, par un pointage (image 3) et un déictique spatial ("là", l. 31). En même temps, l'étudiant prononce un "hm" et un petit rire qui signalent également l'aboutissement de sa recherche, il n'a pas encore posé le curseur sur le tour verbal ciblé mais il a repéré visuellement l'occurrence problématique, simultanément à Laura. Mircea lance alors la vidéo. Les lignes 35, 36, 37 et 41 reproduisent la transcription du tour de parole prononcé par la locutrice qui interviewe la doctorante brésilienne sur la recherche de son directeur de thèse en France.

Suite au visionnage et, en particulier, à l'écoute des données, Mircea et Laura reconnaissent finalement l'emplacement séquentiel de la production inachevée qui a posé problème ("depuis le:", l. 23, ex. 1): il s'agit d'une auto-réparation avec un premier syntagme non finalisé ("depuis le::"), suivi de sa réalisation complète ("depuis le Brésil", l. 36). La compréhension des deux apprenant·e·s est affichée par des *change-of-state tokens* (Heritage 1984), exprimant l'enregistrement d'une nouvelle information, et des marques de réception ("ah okay", "ah:: c'est ça", l. 38-40).

Mircea arrête l'enregistrement (l. 42), l'enseignant projette la clôture de cette séquence ("alors voilà", l. 43) et l'étudiant verbalise le syntagme complet, qui

³ Ce tour de parole verbalise ainsi une auto-réparation.

représente l'auto-réparation de l'occurrence problématique dans la transcription ("depuis le brésil c- c'était ça", l. 44). Les apprenant·e·s s'orientent donc vers une prise en compte de cette résolution (l. 46-48), mais c'est Georges qui essaie de proposer une explication au phénomène observé et la relie à ce qui vient d'être dit ("donc c'était un: quelque chose qui arri-", l. 50). C'est à ce moment-là que Laura se penche en avant, encore une fois pour établir un contact visuel avec l'enseignant (images 4 et 5), et donne une explication de ce qui a été remarqué ("elle a répété elle a continué après", l. 51-52). Son discours est rétrospectivement lié à l'observation effectuée sur l'écran, que l'apprenante pointe à deux reprises pendant son intervention (image 6). Enfin, l'enseignant valide cette explication et produit le marqueur de clôture "voilà" (l. 53) qui permet de fermer la séquence initiée par Laura à la ligne 20 de l'extrait 1.

Dans cette section, nous nous sommes focalisé·e·s sur le caractère émergent de l'observation de données à partir d'un concordancier multimédia. Dans la première partie de l'extrait 1, l'activité de lecture est accomplie de manière collaborative par les participant·e·s et est ponctuée d'observations généralisantes. Le caractère indéterminé d'une occurrence, observé par une apprenante, a donné lieu à une séquence de réparation étendue (Schegloff 2007). À l'aide de deux extraits représentatifs, nous avons pu montrer que les initiatives des étudiant·e·s peuvent porter sur des phénomènes locaux et très courants à l'oral tels qu'une auto-réparation (ce qui est réalisé par l'intervieweuse dans l'enregistrement FLEURON visionné par les apprenant·e·s, ex. 2). Ces initiatives sont caractérisées par une reconfiguration du cadre participatif (Goodwin & Goodwin 2004) et une réorientation du corps de l'apprenant·e, qui se trouve dans une position contrainte face à l'écran de l'ordinateur. En particulier, la recherche d'un contact visuel avec l'enseignant marque la mise en évidence du problème et sa résolution par l'apprenant·e, qui, dans ce dernier cas, sollicite la validation du participant "expert". La mobilisation des ressources multimodales (notamment, gestes de pointage dynamiques ou figés et reconfigurations corporelles) joue un rôle primordial dans les étapes saillantes (repérage et explication) portant sur le tour problématique. Son caractère problématique est, en définitive, associé à la transcription non suffisamment contextualisée, telle qu'elle apparaît dans la colonne droite du concordancier. Par ailleurs, la transcription d'un phénomène courant en français parlé tel qu'une auto-réparation et sa visibilité dans le signal correspondant permettent aux apprenant·e·s de se familiariser avec ce type de phénomènes, de les produire et de les interpréter comme des faits langagiers qu'il est possible de mobiliser dans la langue parlée en interaction.

4.2 La focalisation sur les formes linguistiques

L'extrait suivant montre un cas récurrent dans l'observation des occurrences d'une même forme linguistique, toujours sélectionnée par les apprenant·e·s, où la focalisation sur des éléments spécifiques est guidée par l'enseignant. Dans

l'extrait suivant, deux étudiantes Erasmus italiennes⁴, Maria et Elena, ont choisi de focaliser leur attention sur la forme *mieux* et elles remarquent plusieurs attestations de la collocation *vaut mieux*. Comme pour la situation précédente, cette activité se déroule en présence de Georges, l'enseignant.

(3) CLYC_Groupe 2_Vaut mieux (14:39)

01 GEO et donc euh et là qu'est-ce que vous >par exemple<
02 si on se focalise#7 sur VAUT MIEUX



Image 7

03 *(.) qu'est-ce que vous remarquez*#8
mar *penche sa tête en avant-----*



Image 8

04 (0.7)
05 GEO au niveau du pronom
06 (3.7)
07 GEO >donc< f'c'est-à-fdire mieuxf qui estf précédé par vaut/f#9
geo fbuste en arrière-----f
f.....fmain/table f.....fm -> g

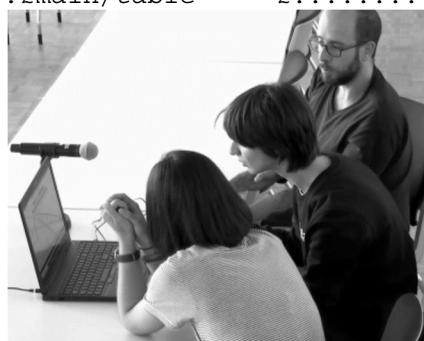


Image 9

⁴ D'après l'enseignante responsable du module de formation, les apprenant·e·s ont en français le même niveau de compétence que les apprenants de l'extrait 1.

08 (0.1)£(0.6)£ et vaut/£ (0.2) il est précédé par un pronom/
 geo m/t £,,,,,£ fbuste en avant-->

09 (0.4)

10 **ELE** euh£: [oui]

11 **MAR** [oui]

 geo ->£(buste en avant)

12 **GEO** £c'est £[il vaut oui]#10

13 **ELE** [il vaut mieux] oui [c'est IM]personnel

 geo £.....ftête en avant

14 **GEO** [°il vaut°]

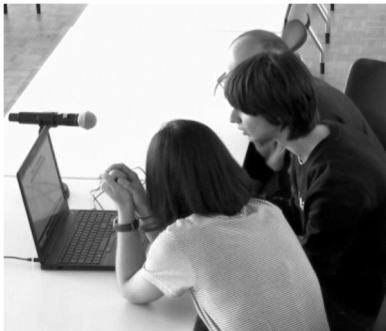


Image 10

15 **MAR** il vaut mieux\
 16 <(2.0) ((MAR pose le curseur sur une occurrence, le déplace))>
 17 **GEO** tou*jours/ (.) on va voir fon* va:£ vous pouvez cli[quer/
 mar *.....*redresse sa tête
 geo £mvt tête£

18 **MAR** [((clique
 19 sur une occurrence))

Au début de cet extrait, Georges sollicite les participantes pour qu'elles prêtent attention aux attestations affichées dans le concordancier et, en particulier, aux occurrences de *vaut mieux* (cette expression linguistique est prononcée de manière saillante l. 02). L'enseignant répare son tour initiatif qui est ainsi articulé en deux moments: (i) une *si*-proposition qui présente l'objet attentionnel sur lequel les participantes sont invitées à se focaliser ("si on se focalise sur VAUT MIEUX", l. 02); (ii) une interrogative partielle qui thématise la tâche observationnelle de l'activité ("qu'est-ce que vous remarquez", l. 03). De manière intéressante, les ressources grammaticales retracent la bipartition sur le plan pragmatique: si dans la première partie de ce tour multi-unités (Sacks et al. 1974) Georges utilise une marque pronominale inclusive ou générique ("on"), dans la deuxième partie, celle qui présente la forme d'une interrogative et projette donc une réponse, il s'adresse explicitement aux apprenantes par un pronom de deuxième personne pluriel ("vous"). Deux types de réponse sont fournies en réaction à ces deux unités.

La réponse des deux apprenantes à la première unité se manifeste de manière évidente sur le plan multimodal. Suite à la sollicitation attentionnelle produite par l'enseignant (via la *si*-proposition), Maria et Elena se penchent vers l'écran (le mouvement est particulièrement visible chez Maria, si on regarde l'image 8 et on la compare à l'image 7). Cette réponse corporelle, qui incarne la

focalisation attentionnelle précédemment sollicitée par l'enseignant, est un phénomène récurrent dans les expérimentations que nous avons pu enregistrer.

Ligne 5, Georges produit un incrément (Schegloff 1996; Ford et al. 2002) à sa question précédente, ce qui permet de mieux la spécifier. Cette extension mentionne la catégorie grammaticale que vise sa question ("au niveau du pronom"). Ensuite, il cherche à contextualiser sa question en retraçant l'historique des observations qui jusque-là ont été faites par les apprenantes ("c'est-à-dire mieux qui est précédé par vaut", l. 07) et il formule une véritable question portant sur la forme verbale *vaut*, qui est disloquée ("et vaut (0.2) il est précédé par un pronom", l. 08). Cette évocation du passé est accompagnée d'un réagencement multimodal: Georges éloigne son corps de l'écran (image 9). Les deux étudiantes produisent une réponse chorale ("oui" en chevauchement, l. 10-11) à la question de l'enseignant. Ce dernier se situe dans une position ambivalente: il répond lui aussi à sa propre question après avoir lu une occurrence du concordancier ("il vaut oui", l. 12), en se penchant vers l'écran (image 10).

Ligne 13, de son côté, Elena fournit une réponse plus complète: après avoir lu une occurrence ("il vaut mieux"), elle reproduit sa réponse positive ("oui") et donne une caractérisation morphosyntaxique précise du pronom en question ("c'est IMpersonnel"). Maria participe à cette activité suivant une temporalité distincte, après l'intervention de sa camarade, elle se limite à lire l'occurrence ("il vaut mieux", l. 15). Georges relance les participantes en questionnant la réalisation systématique du pronom ("toujours/", l. 17) et, pour inviter au visionnage de l'extrait, il mobilise la même structure que nous avons rencontrée dans le premier extrait analysé en 4.1 ("on va voir on va:", l. 17). Ensuite, il donne une instruction plus directe et s'adresse explicitement à Maria ("vous pouvez cliquer", l. 17): à ce moment-là, elle a positionné le curseur sur une occurrence qui n'est pas précédée d'un pronom. L'action instruite qui suit, comme dans l'extrait 1, consiste en un clic sur l'occurrence pour ouvrir la fenêtre qui permettra de visionner la vidéo correspondante (l. 18-19).

Dans les deux premiers extraits, la focalisation sur une ligne du concordancier émerge au cours de la lecture des occurrences et, en particulier, de ce qui se trouve "à droite du mot" recherché. Cette focalisation est opérée par l'un·e des apprenant·e·s. En revanche, dans ce troisième extrait, nous nous sommes penché·e·s sur une focalisation sollicitée par l'enseignant, qui essaie de diriger l'attention des apprenantes sur les occurrences affichées dans le concordancier et, plus précisément, sur ce qui se trouve ou ne se trouve pas "à gauche du mot" recherché (en l'occurrence, le pronom sujet de la forme verbale "vaut mieux").

5. Conclusion

Dans cette contribution, nous avons proposé une étude séquentielle et multimodale de trois extraits issus de deux séances de travail avec la plateforme FLEURON. Les analyses présentées dans la section 4 ont mis en évidence l'importance des ressources multimodales mobilisées chez les apprenant·e·s pendant ces séances. En particulier, le pointage et la réorientation du corps permettent d'initier des séquences de réparation qui visent à résoudre des problèmes émergents (section 4.1). Dans les extraits 1 et 2, le problème concerne un mode de production typique de la langue parlée, celui de l'auto-réparation. Par exemple, le regard de Laura se pose sur une ligne qui présente un problème pour elle: elle n'arrive pas à trouver le mot à la droite de la forme cible recherchée et remarque une absence notable dans ce contexte ("après c'est fini", l. 23, ex. 1), ce qui est démenti par le visionnage de la vidéo. Ce sont les va-et-vient entre la transcription ponctuelle telle qu'affichée dans le concordancier, la transcription complète et la vidéo de l'enregistrement correspondant qui peuvent fournir des éléments de contextualisation et, donc, d'explication des occurrences observées. La navigation entre les différents outils de la plateforme (concordancier, ressources, description du contexte) est possible grâce à la médiation de l'enseignant. Les marques linguistiques, les ressources multimodales et les manipulations au sein de la plateforme s'inscrivent dans une activité menée collaborativement par les binômes d'apprenant·e·s. Même si l'un des deux reste plutôt silencieux (en l'occurrence, Mircea dans l'extrait 1), il intervient avec les manipulations de la souris et avec des productions verbales ponctuelles, signalant sa compréhension effective de ce qu'il se passe dans l'interaction enregistrée.

Le court extrait de la section 4.2 nous a permis de présenter d'autres éléments récurrents dans notre corpus. Si, dans les deux premiers exemples, l'élément qui fait l'objet d'une enquête linguistique est repéré par une apprenante, dans le troisième c'est l'enseignant qui canalise les regards des deux participantes. En particulier, la sollicitation attentionnelle demandée par l'enseignant passe, sur le plan pragmatique, par une lamination actionnelle et, sur le plan de la structuration de la parole, par la mobilisation de ressources linguistiques s'inscrivant dans un tour multi-unités. La focalisation est accomplie de manière incarnée par les apprenantes⁵: elles se penchent vers l'écran et commencent la phase d'observation de données proprement dite, suite aux indications fournies par l'enseignant/médiateur. Dans l'extrait 3, l'initiative de l'apprenante est toujours présente, même si elle est silencieuse (en l'occurrence, Maria pose le curseur sur une attestation qui pose problème par rapport à la marque pronominale signalée par Georges et caractérisée par Elena). L'enseignant joue par moments le rôle de médiateur et sollicite la navigation entre les

⁵ Voir Mondada (2015) pour une analyse séquentielle détaillée dans des interactions en face-à-face, sans ordinateur.

ressources disponibles sur la plateforme, tout en conservant l'esprit collaboratif de l'activité en cours.

Les extraits analysés ici ont également fourni des exemples concernant la structuration de tours qui présentent des catégorisations spécifiant les traits sémantiques ("ça ça indique toujours le temps" pour *depuis*) ou grammaticaux ("c'est impersonnel" pour *il vaut mieux*) des formes cibles, ce qui illustre davantage l'usage acquisitionnel du concordancier. En particulier, les expériences que nous avons documentées attestent d'une structure tripartite composée des éléments suivants: (i) une occurrence et/ou son contexte, (ii) une particule qui est rétrospectivement liée à l'opération de lecture effectuée ("oui"), (iii) la catégorisation proposée (voir notamment le tour de Laura, l. 01, extrait 1; le tour de Elena, l. 13, extrait 3). La réalisation de la particule en position médiane n'est pas systématique. Quoi qu'il en soit, la catégorisation proposée par l'apprenant·e est le plus souvent ancrée dans une attestation (et s'inscrit donc dans un tour multi-unités), très rarement isolée.

Ce travail est exploratoire. Néanmoins, l'analyse multimodale des interactions accompagnant l'utilisation du dispositif FLEURON permet de rendre compte des activités cognitives et métacognitives réalisées par les apprenant·e·s, ainsi que du processus d'acquisition impulsé par ces activités. Parmi les ressources multimodales, les changements de postures par rapport à l'écran, l'orientation des regards ainsi que les pointages réalisés par les apprenant·e·s représentent des indices utiles pour le repérage de phases de "découverte", qui se distinguent de l'observation et de la simple lecture des lignes du concordancier. Les ressources linguistiques utilisées par l'enseignant, notamment les reprises (ex. 1) et les tours multi-unités (ex. 3), permettent de stabiliser les informations acquises pendant l'observation et de naviguer entre différents outils et onglets dans la plateforme numérique.

Nous avons mené des expérimentations avec des apprenant·e·s de FLE et, dans cette contribution, nous avons considéré une configuration participationnelle spécifique (celle d'un binôme d'étudiant·e·s universitaires ayant le même niveau de compétence linguistique, accompagné par un enseignant/tuteur) et nous nous sommes appuyé·e·s sur un dispositif d'enregistrement spécifique. Des études ultérieures pourront illustrer davantage le caractère émergent des observations et des commentaires (méta)linguistiques que les apprenant·e·s proposent pour les phénomènes langagiers accessibles depuis le concordancier multimédia. Une attention particulière a été accordée à la mobilisation, ici médiée par l'enseignant, des hyperliens donnant accès aux transcriptions complètes et aux enregistrements. Dans nos données, cette navigation s'avère centrale pour l'acquisition de certaines caractérisations proposées pour des phénomènes spécifiques, qui peuvent apparaître comme des récurrences ou des formes remarquables, voire problématiques, en parcourant les lignes du concordancier.

Par ailleurs, d'autres expérimentations sont en cours avec des apprenant·e·s adultes ayant différents profils en termes de niveau de scolarisation, niveau de langue, type de formation, compétences métalinguistiques, etc. dans différents contextes de formation (en autonomie, en classe). Elles permettront d'analyser également les effets de l'ASC sur le long terme et son efficacité quant à la rétention du fonctionnement de la langue en situation.

BIBLIOGRAPHIE

- Abé, D., Carton, F., Cembalo, S. M. & Régent, O. (1979). Didactique et authentique: du document à la pédagogie. *Mélanges pédagogiques*, 10, 1-14.
- André V. (2017). Un corpus multimédia pour apprendre à interagir en situations universitaires en France. In *Actes du troisième colloque international de l'ATPF "Enseigner le français: s'engager et innover"*, 292-315.*
- André, V. (2019). Des corpus oraux et multimodaux authentiques pour acquérir des compétences sociolangagières. In L. Gajo, J.-M. Luscher, I. Racine & F. Zay (éds.), *Variation, plurilinguisme et évaluation en FLE* (pp. 209-233). Berne: Peter Lang.
- André, V. & Ciekanski, M. (2018). Apprendre à interagir à l'oral à partir d'un concordancier multimodal: effets sur le développement de la conscience langagière et sur l'autonomie de l'apprenant dans le dispositif FLEURON. In C. Dejean, F. Mangenot, E. Nissen & T. Soubrié (éds.), *Actes du colloque EPAL 2018 (Échanger pour apprendre en ligne)*, Université Grenoble Alpes, 7-9 juin 2018.*
- Aston, G. (2001). Learning with corpora: an overview. In G. Aston (éd.), *Learning with corpora* (pp. 7-46). Houston: Athelstan.
- Berger, E. (2008). La reprise comme ressource interactionnelle en langue seconde. *Travaux neuchâtelois de linguistique*, 48, 43-61.
- Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In J. Sinclair (éd.), *How to use corpora in language teaching* (pp. 15-36). Amsterdam/Philadelphia: John Benjamins.
- Boulton, A. & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348-393.
- Boulton, A. & Tyne, H. (2014). *Des documents authentiques aux corpus. Démarches pour l'apprentissage des langues*. Paris: Didier.
- Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL*, 19(3), 307-328.
- Breyer, Y. A. (2011). *Corpora in language teaching and learning. Potential, evaluation, challenges*. Frankfurt am Main: Peter Lang.
- Chambers, A. (2010). What is data-driven learning? In A. O'Keeffe & M. McCarthy (éds.), *The Routledge handbook of corpus linguistics* (pp. 345-358). New York: Routledge.
- Ford, C. E., Fox, B. A. & Thompson, S. A. (2002). Constituency and the grammar of turn increments. In C. E. Ford, B. A. Fox & S. A. Thompson (éds.), *The language of turn and sequence* (pp. 14-38). Oxford/New York: Oxford University Press.
- Gavioli, L. (2001). The learner as researcher: Introducing corpus concordancing in the classroom. In G. Aston (éd.), *Learning with corpora* (pp. 108-137). Houston: Athelstan.

- Giroud, A. & Surcouf, C. (2016). De "Pierre, combien de membres avez-vous ?" à "Nous nous appelons Marc et Christian": réflexions autour de l'authenticité dans les documents oraux des manuels de FLE pour débutants. *SHS Web of Conferences* 27 - CMLF 2016.*
- Goodwin, C. & Goodwin, M. H. (2004). Participation. In A. Duranti (éd.), *A companion to linguistic anthropology* (pp. 222-244). Malden, MA: Blackwell.
- Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J. M. Maxwell Atkinson & J. Heritage (éds.), *Structures of social action. Studies in conversation analysis* (pp. 299-345). Cambridge: Cambridge University Press.
- Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *English Language Research Journal*, 4, 1-16.
- Johns, T. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (éds.), *Teaching and language corpora* (pp. 100-115). London: Longman.
- Kerr, B. (2013). Grammatical description and classroom application. Theory and practice in data-driven learning. *Bulletin suisse de linguistique appliquée*, 97, 17-39.
- Mauranen, A. (2004). Spoken corpus for an ordinary learner. In J. Sinclair (éd.), *How to use corpora in language teaching* (pp. 89-105). Amsterdam/Philadelphia: John Benjamins.
- Mondada, L. (2015). Introduction de nouveaux référents et gestion de l'attention conjointe: les apports de l'analyse interactionnelle multimodale. In X. Gradoux, J. Jacquin & F. Merminod (éds.), *Agir dans la diversité des langues. Mélanges en l'honneur d'Anne-Claude Berthoud* (pp. 121-137). Bruxelles: De Boeck-Duculot.
- Mondada, L. (2017). Le défi de la multimodalité en interaction. *Revue française de linguistique appliquée*, XXII, 71-87.
- Sacks, H., Schegloff, E. A. & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696-735.
- Schegloff, E. A. (1996). Turn organization: One intersection of grammar and interaction. In E. Ochs, E. A. Schegloff & S. A. Thompson (éds.), *Interaction and grammar* (pp. 52-133). Cambridge: Cambridge University Press.
- Schegloff, E. A. (2007). *Sequence organization. A primer in conversation analysis*. Cambridge: Cambridge University Press.
- Schegloff, E. A., Jefferson, G. & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53(2), 361-382.
- Traverso, V. (2016). *Décrire le français parlé en interaction*. Paris: Ophrys.
- Zorzi, D. (2001). The pedagogic use of spoken corpora: Learning discourse markers in Italian. In G. Aston (éd.), *Learning with corpora* (pp. 85-107). Houston: Athelstan.

* Ces références sont disponibles en ligne

Annexes

Conventions de transcription

Pour la notation des productions verbales, nous avons utilisé les conventions de transcription ICOR, dont une version développée est consultable à l'adresse <http://icar.cnrs.fr/projets/corinte/>:

[]	début et fin de segment en chevauchement	(.)	micro-pause
(2.1)	pause en secondes	xxx	segment inaudible
/ \	intonation montante/ descendante\	exTRA	segment accentué
((rire))	actions à l'écran, phénomènes divers	:	allongement vocalique
< >	délimitation des phénomènes entre (())	par-	troncation
&	continuation du tour de parole	=	enchaînement rapide
^	liaison	.h	aspiration

Pour la multimodalité, nous nous inspirons des conventions développées par Mondada et consultables sur son site personnel (<https://www.lorenzamondada.net/multimodal-transcription>):

Le verbal apparaît en gras pour le distinguer de la description de la multimodalité. Les gestes et les mouvements du corps sont repérés par rapport à une production verbale:

, £, +	indique le début et la fin d'un geste (, £) ou d'un regard (+) décrit à la ligne suivante
---	signale la durée du geste ou du regard
--->	indique que le geste ou le regard continue aux lignes suivantes jusqu'à la prochaine borne (*, £, +)
--->>	indique que le geste ou le regard continue au-delà de la fin de l'extrait
#1	dans le texte, le signe # suivi d'un chiffre indique l'emplacement et le numéro de l'image

Replication and extension of an empirical study about foreign language formulaic familiarity database: Reliability and emotionality¹

Yu KANAZAWA

Kwansei Gakuin University

École d'études internationales

1-155 Uegahara Ichiban-cho, Nishinomiya, Hyōgo 662-8501, Japon

yu.kanazawa.res@gmail.com; yu-kanazawa@kwansei.ac.jp

L'émotion, la familiarité et les séquences de formules (expressions préfabriquées / phrases à plusieurs mots) sont des aspects importants de la recherche en linguistique appliquée (acquisition d'une langue distante, par exemple: l'anglais pour les Japonais); mais ces aspects sont rarement liés les uns aux autres dans une étude quantitative. Les objectifs de la présente étude étaient les suivants: (n° 1) enrichir la base de données sur la familiarité avec des formules pour les étudiant-e-s des universités japonaises (Isobe et al. 2016) avec des données émotionnelles numériques pour poursuivre la recherche et créer des applications pédagogiques; et (n° 2) vérifier la fiabilité des données de familiarité en reproduisant une étude. Les étudiant-e-s d'anglais d'une université au Japon ont été invité-e-s à remplir une série de questionnaires en ligne dans lesquels elles et ils ont évalué chaque formule dans la liste PHRASE (Martinez et Schmitt 2012) en termes du sens positif ou négatif de sa signification (la valence émotionnelle - un facteur affectif digital / numérique; voir Kanazawa 2016) et la fréquence subjective / ressentie vue ou entendue (la familiarité - un facteur lexical important) sur une échelle de Likert à sept points. Les réponses ont été collectées avec succès et une liste de formules avec des données émotionnelles numériques a donc été produite. L'analyse corrélacionnelle a d'abord démontré que la valeur de valence émotionnelle n'était corrélée ni à la familiarité ni à la fréquence dans le corpus, indiquant son indépendance comme facteur. Il a également été indiqué que les éléments chargés d'émotion - qu'ils soient positifs ou négatifs - donnent des valeurs de familiarité plus élevées. Ensuite, l'étude de réplication a vérifié la fiabilité de la base de données en termes de familiarité des formules. Les résultats contribuent au développement de bases de données similaires pour les étudiants de langues étrangères.

Mots-clés:

Valence émotionnelle, séquence de formules, familiarité, fiabilité d'une base de données de formules, langues éloignées, apprenants japonais d'anglais, EmAL.

Keywords:

Emotional valence, formulaic sequence, familiarity, reliability of a formulaic database, distant languages, Japanese learners of English, EmAL.

1. Introduction

1.1 Lexical/Formulaic familiarity

Multiple lexical/formulaic attributes are known to play essential roles in foreign language vocabulary acquisition (FLVA). Namely, FLVA and processing are known to be influenced by such factors as the number of letters, the number of

¹ This paper is based partly on Kanazawa (2020a; 2020b).

syllables, phonetic characteristics, the (non)existence of homonyms, homophones, neighboring words, and *frequency* (Jiang 2018). Frequency can be defined as the number of occurrences of the target lexical/formulaic item in a given linguistic input (Conklin 2020). A plethora of empirical studies show that high frequency items are processed faster, remembered more easily, and are more likely to be acquired (Ellis 2002; Nation 2016). Methodologically, frequency is one of the inevitable variables that is to be controlled when conducting a quantitative lexical/formulaic study (Divjak 2019).

On the other hand, the prevailing overreliance on lexical frequency has recently been criticized (Hashimoto & Egbert 2019). Furthermore, it was theoretically proposed that lexical/formulaic frequency – an objective measure – and the actual degree of *familiarity* of the target item – a subjective measure – are similar but different constructs that should not be confused (Yokokawa 2006). Familiarity can be defined in terms of how often language users feel they see or hear each specific lexical/formulaic item (Amano et al. 2008). Lexical/formulaic familiarity is calculated by the mean scores of familiarity ratings, in which a large number of participants are asked to rate each word in a Likert scale ranging from "never seen or heard it" to "very frequently seen or heard it."

It is true that some applied linguists doubt the validity of the construct of subjective familiarity and propose that examining objective frequency suffices. This view, however, is rejected by many empirical studies. For example, Gernsbacher (1984) showed that subjective familiarity is diversely different among low-frequency words and it is familiarity that reliably affects word recognition, criticizing the confusion between familiarity and frequency. Kreuz's (1987) experimental study of homophonic word processing targeted to university students showed that rated familiarity could far better predict lexical decision task performance than word frequency. Connine et al.'s (1990) lexical access experimental results failed to find consistent frequency effects while showing solid familiarity effects for all the tasks: visual lexical decision, auditory lexical decision, cued naming. Lewellen et al. (1993) experimentally demonstrated that lexical familiarity affects processing efficiency in terms of naming, lexical decision, and semantic categorization. Cordier & Ny (2005) conducted a series of experiments, successfully achieving the goal of "disentangl[ing] familiarity from frequency (pp. 530-531)."

In relation to Japanese speakers, Amano et al. (1999) empirically showed that lexical recognition in the Japanese language is affected by lexical familiarity – broadening the scope of familiarity research in another language. There is even a large database of Japanese words solely dedicated to lexical familiarity, indicating the significance of familiarity (Amano et al. 2008). Based on an extensive survey, Yokokawa (2006; 2009) produced the English Lexical Familiarity Database for Japanese EFL (English as a foreign language) learners, which is available electronically and widely used in FLVA studies and

pedagogical practices. In sum, familiarity and frequency may well be separate entities not to be confused. Moreover, many studies show not only the discrepancy between the two but also the higher impact of familiarity effects compared to frequency effects on language processing.

Meanwhile, recent FLVA studies suggest that studying single word lexical items is insufficient, calling for more studies on formulaic sequences (Siyanova-Chanturia & Pellicer-Sánchez 2019; Webb 2020). Formulaic sequences (or formulae) can be defined as multi-word units perceived to have an identity as a single lexical unit (Wray 2019). Previous studies have shown that language recognition is affected by familiarity not only at the lexical level but also at the formulaic level (Moon 1998; Isobe 2011), validating the construct of formulaic familiarity – which is an underinvestigated topic despite its importance.

1.2 Formulaic database for English for distant language learners

In view of the lack of a formulaic familiarity database for distant language learners,² Isobe et al. (2016) conducted a large-scale study to produce the English formulaic familiarity database for Japanese learners of English. Participants were more than a thousand Japanese-speaking students learning English in 11 different universities in Japan. The target English formulae (n = 501) were selected from the PHRASE List tailored to EFL needs (Martinez & Schmitt 2012; 505 items) while deleting the overlapping items (n = 4) and subdividing the entire list into 5 subsets with 100-101 formulaic items. Online questionnaires created via SurveyMonkey were used to collect the data. Participants rated the subjective familiarity of each formula randomly presented on the screen on a seven-point Likert scale. After a pilot study, the data were successfully collected to produce the formulaic familiarity database. The exploratory analysis showed only weak correlation between formulaic familiarity ratings and the lexical familiarity ratings of the words composing the formulae, suggesting the discreteness of lexical familiarity and formulaic familiarity and the need to dissociate them (Fig. 1).

² A distant language can be defined as a language with linguistic distance far from your first language. Linguistic distance is an indicator of how similar or different the two languages are in terms of structural closeness and difficulty of learning (see Chiswick & Miller 2005). In comparison with learning a language with closer linguistic distance (e.g., English for French speakers), a distant language is usually more difficult to acquire and needs special effort – such as English for Japanese speakers and vice versa (Nishikawa-Van Eester 2016).

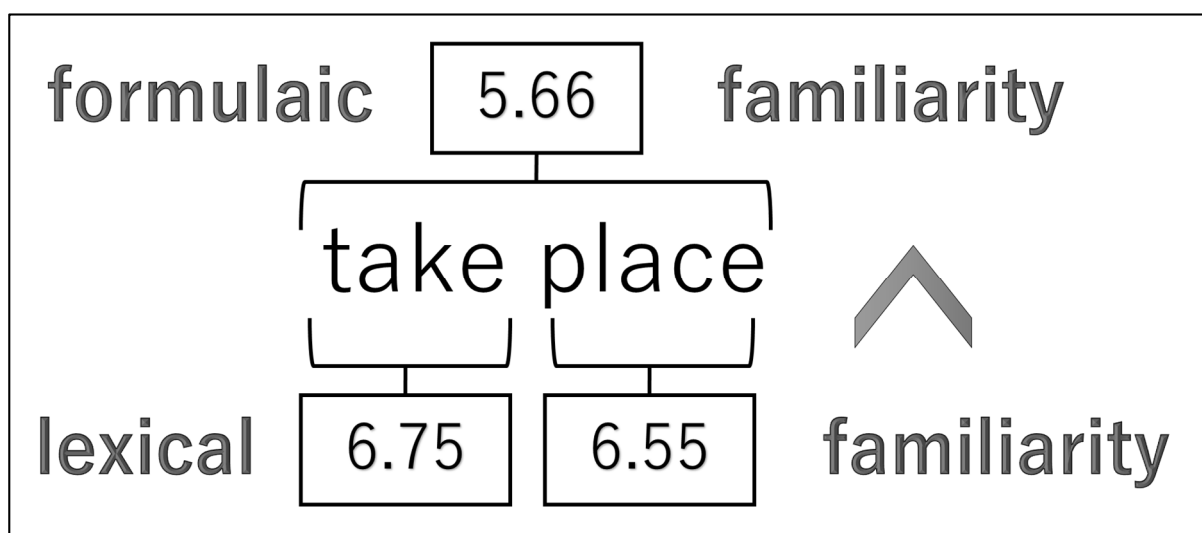


Fig. 1: Formulaic familiarity is dissociable from lexical familiarity – the familiarity of the formulaic unit (*take place*) is lower than that of individual composing words (*take* and *place*), reflecting their semantic inequivalence (cf. the sum is more than the whole of its parts).

Although Isobe et al.'s (2016) study paved the way for studying and developing pedagogical materials of formulae in a distant language, it has a limitation – each item in the list was rated only by one-fifth of the participants with potentially different individual profiles. If the aim of the research had been to compare various conditions allocated to different participant groups, the between-subject design would have been appropriate. In reality, however, the major aim was to formulate a database with no subsets as the final product. Since the participant groups were not controlled in terms of various individual factors (e.g., English proficiency, gender, age of acquisition), there remain possibilities of such factors interfering with the results and thus lowering the reliability of the database. Namely in Isobe et al. (2016), the whole list was subdivided into five sublists for the sake of expediency, and each participant rated only one of the sublists. For example, the item *as for* and the item *apart from* happened to belong to different sublists and were rated by different groups of participants. Therefore, the fact that these two items share the same familiarity score (4.8) can be explained not purely in terms of the equal level of familiarity but also owing potentially to the individual difference factors between the participants who rated *as for* and those who rated *apart from*. Since controlling all those between-subject factors is impossible to achieve, a better solution is to adopt a within-subject design – to have identical raters for all the items by having all the participants rate the whole list (see Charness et al. 2012 for benefits of within-subject design). Additionally, the study – although conducted on a large-scale – did not test the internal consistency (Cronbach α), leaving the inter-rater reliability susceptible to doubt.

The present study was designed to overcome these limitations by conducting a replication study with reliability testing in a within-subject design.

1.3 *Lexical/Formulaic emotionality*

Another pivotal subjective aspect that most previous FLVA studies have tended to fail to incorporate is emotion – another keyword in applied linguistics. Emotion plays fundamental roles in cognition (Barrett 2017); and its importance in foreign language acquisition (FLA) is increasingly being recognized (Dewaele 2015) not only as a social factor but also as a scientific factor in cognitive architecture (Sharwood Smith 2017). This *affective turn* (Pavlenko 2013) – the rise of EmAL (emotion in applied linguistics) – is also affecting the field of FLVA (Schütze 2017). For instance, Kanazawa (2017, 2021) empirically showed that processing which facilitates conscious emotional activation (Emotion-Involved Processing) enhances the retrieval of target words. A similar phenomenon has been reported under different lines of research and names such as emotionally enhanced memory (Talmi et al. 2007) and affective input enhancement (Truscott 2015). In order to study the effect of emotion at the micro-level of vocabulary, it is important to digitize and operationalize emotion as a numerical lexical/formulaic attribute. For example, Hashimoto and Egbert (2019) refer to affective sense as a variable related to word difficulty along with such important lexical/formulaic variables as orthographic similarity, length, polysemy/homonymy, and contextual distinctiveness. An established construct as a digitized emotional attribute in cognitive psychology is *emotional valence* – the subjective ratings about how positive or negative the stimulus is (Bradley & Lang 1999). Emotional valence for each stimulus is identified as the mean score of subjective ratings on a Likert scale collected from a large number of participants.

In the field of FLVA, however, emotional valence has rarely been taken into account as a lexical/formulaic attribute to be controlled and investigated. Kanazawa (2016) and Kanazawa & Lafleur (2019) were pioneering attempts to include emotional valence as a numerical vocabulary attribute by collecting data rated by foreign language learners. One of the limitations of these studies is that the targets were limited to decontextualized individual single words whereas multi-word formulae should also be considered. The insufficient attention to formulae has resulted in the lack of a formulaic list with numerical affective data and pedagogical adaptation. Through the study of lexical familiarity (i.e., the subjective version of word frequency; see above), it has been shown that linguistic databases that have been compiled by using native speakers' responses are not necessarily applicable to foreign language learners with different mental lexicon (Taylor 2012), indicating the need to develop a foreign language (L2, or LX; cf. Dewaele 2018) version of such a database, especially for subjective attributes.

1.4 Research questions of the present study

The present study was designed to produce an affective formulaic list for tertiary EFL learners while overcoming the limitations of Isobe et al. (2016). The research questions were as follows:

- [RQ N° 1]: Can Isobe et al.'s (2016) distant language formulaic database be reliably enriched with formulaic emotional valence data while demonstrating the independency of emotional valence as a formulaic attribute?
- [RQ N° 2]: Is Isobe et al.'s (2016) formulaic familiarity rating reliably replicable while overcoming its limitations?

2. Method

Participants were English learners at a university in Japan (N = 120). They were freshmen and sophomore students with varied proficiency levels of English (n = 10 [CEFR-A1]; n = 20 [CEFR-A2]; n = 67 [CEFR-B1]; n = 21 [CEFR-B2]; n = 2 [CEFR-C1]). The target formulae (501 items) were selected from the PHRASE List (Martinez & Schmitt 2012) since it is tailored to EFL pedagogical needs and it was utilized in Isobe et al. (2016). Participants were asked to complete web-based questionnaires which consisted of the following tasks:

(1) task explanation and consent form to abide by research ethics.³

(2) emotional valence judgment task (Warriner et al. 2013; Kanazawa 2016), in which they were asked to rate each formulaic item in terms of how positive or negative its meaning was on a seven-point Likert scale (1. Very negative ~ 7. Very positive).

(3) familiarity judgment task (Isobe et al. 2016), in which they were asked to rate each formulaic item in terms of how often they had seen or heard it on a seven-point Likert scale (1. Seldom ~ 7. Very frequent).

Fig. 2 and 3 are examples of the two rating tasks.

³ All the data collected and used in this research were from participants who agreed with all of the following statements (Original script was in Japanese): "I have read and understood the explanation." "I understand that the data will be processed and stored anonymously and that it does not affect the grading of the class what you answer in each question." "I agree that the data I provide will be used in further academic studies and material development."

<Valence Judgment Task>
(Translation)

- Please rate each formulaic item on a seven-point scale how positive its meaning is. (seven-point Likert scale: 1.Very negative ~ 7.Very positive)
- Please rate it not by how well you know its meaning but how often you see or hear it

それぞれの定型表現を、意味がポジティブかネガティブかについて7段階で判定しましょう。※1語1語の単語としてでなく、単語が組み合わされた定型表現全体として判定してください。

	1. とても ネガティブ ブ○○○	2. ネガテ イブ○○	3. ややネ ガティブ○	4. どちら でもない	5. ややポ ジティブ○	6. ポジテ イブ○○	7. とても ポジティ ブ○○○
go away	☐	☐	☐	☐	☐	☐	☐
such that	☐	☐	☐	☐	☐	☐	☐
once again	☐	☐	☐	☐	☐	☐	☐
switch on	☐	☐	☐	☐	☐	☐	☐
afford to	☐	☐	☐	☐	☐	☐	☐
just about	☐	☐	☐	☐	☐	☐	☐

Fig. 2: Emotional valence judgment task screen

<Familiarity Judgment Task>
(Translation)

- Please rate each formulaic item in terms of how often you see or hear it in seven-point.
- Please rate it not by how well you know its meaning but how often you see or hear it
- Please focus on the formula as a whole instead of its composing words

それぞれの定型表現を、見聞きする程度によって7段階で判定しましょう。※「意味を知っているかどうか」ではなく、「どの程度見聞きするか」で判定しましょう。※1語1語の単語としてでなく、単語が組み合わされた定型表現全体としての親密度を判定してください。

	1. まった く見聞き しない	2. 見聞き しない	3. あまり 見聞きし ない	4. どちら でもない	5. 少し見 聞かする	6. よく見 聞かする	7. とても よく見聞 かする
end up	☐	☐	☐	☐	☐	☐	☐
point of view	☐	☐	☐	☐	☐	☐	☐
but then	☐	☐	☐	☐	☐	☐	☐
in so far as	☐	☐	☐	☐	☐	☐	☐
hand over	☐	☐	☐	☐	☐	☐	☐
in the way	☐	☐	☐	☐	☐	☐	☐

Fig. 3: Familiarity judgment task screen

To decrease fatigue, the number of formulae for rating was limited to 50-51 items per week. Overall, it thus took ten weeks for students to finish rating all the items. The items were randomized using the random option on Google Forms. Participants accessed the questionnaire either via a smartphone or on a computer. It was ensured that participants rated all the formulae to overcome the limitations of Isobe et al. (2016), in which each participant rated only one-fifth of the whole list. In analysis, only the data with a high familiarity rating (5-7) were used while other data was omitted to ensure the valid semantic processing of each formulaic item. Statistical analyses were performed using SPSS software v.25.

3. Results

As a result, responses were successfully collected to produce a formulaic list with numerical emotional and familiarity attributes. The descriptive statistics are presented in Tab. 1. The number of component words in each formula means

how many words there are in each formula (e.g., *as well* = 2, *at the expense of* = 4). The number of syllables in each formula means how many syllables there are in each formula (e.g., *over time* = /óv·və·táim/ = 3, *on the other hand* = /ən·ðə·lð·ər·háend/ = 5). Formula frequency is based on British National Corpus and was calculated based on Martinez & Schmitt (2012). For formula familiarity ratings and formula valence ratings, see the previous section.

Description	N	Mean	SD
Mean <u>number of component words in each formula</u>	501	2.33	.56
Mean <u>number of syllables in each formula</u>	501	2.67	.79
Mean <u>formula frequency</u> (/100 M)	501	3,961.60	6,223.86
Mean <u>formula familiarity</u> ratings (Isobe et al., 2016)	501	4.78	.96
Mean <u>formula familiarity</u> ratings (this study)	501	4.97	.61
Mean <u>formula valence</u> ratings (this study)	501	4.28	.46

Tab. 1: The descriptive statistics

3.1 Formulaic emotional valence

The prerequisite for further inferential statistical analysis is to ensure the reliability of the collected data. The results of the reliability testing for subsets in terms of emotional valence are shown in Tab. 2. Because reliability (internal consistency) surpassing the acceptable level of .70-.80 was shown across all subsets (Larson-Hall 2010), it can legitimately be concluded that the data are nearly free from noise or measurement error, needing no data re-collection nor omission.

Reliability testing through Cronbach α			
Subset 1 (#1-50)	$\alpha=.982$	Subset 6 (#251-300)	$\alpha=.944$
Subset 2 (#51-100)	$\alpha=.973$	Subset 7 (#301-350)	$\alpha=.971$
Subset 3 (#101-150)	$\alpha=.952$	Subset 8 (#351-400)	$\alpha=.978$
Subset 4 (#151-200)	$\alpha=.981$	Subset 9 (#401-450)	$\alpha=.969$
Subset 5 (#201-250)	$\alpha=.966$	Subset 10 (#451-501)	$\alpha=.960$

Tab. 2: The internal consistency of emotional valence ratings

The correlational analysis demonstrated that emotional valence correlated with neither familiarity ($r = .11$; $p < .05$) nor frequency ($r = -.03$; $p = .48$). Although the correlation between emotional valence and familiarity was statistically significant, the fact that the Pearson r was .11 suggests there is no meaningful correlation (Patten & Newhart 2018: 231), as is also graphed in Fig. 4.

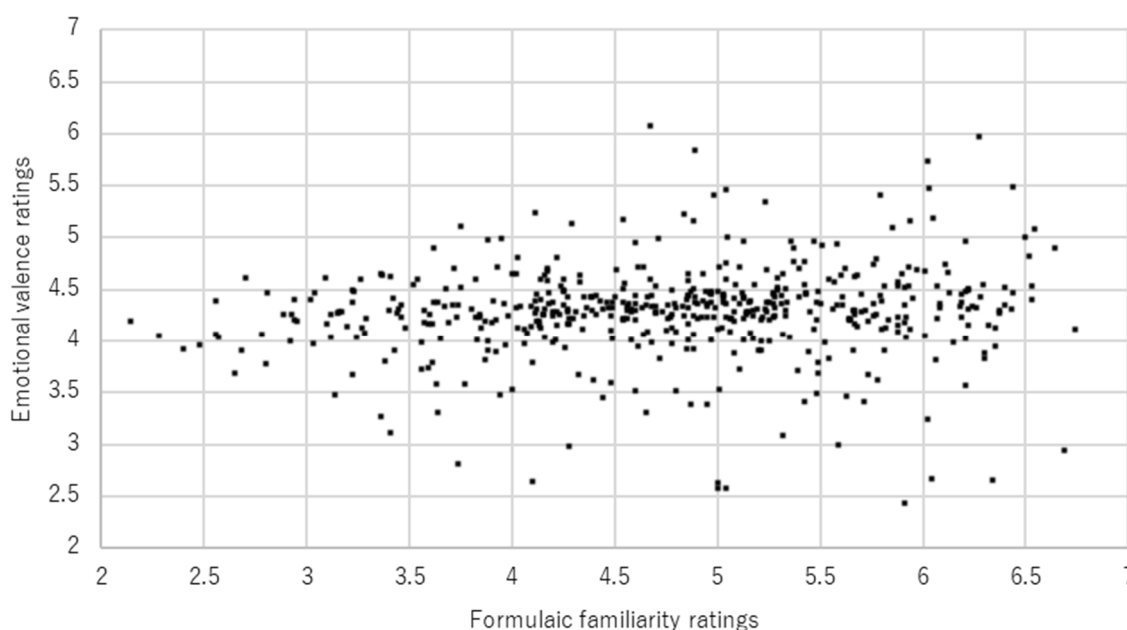


Fig. 4: Scatterplot showing the lack of correlation between emotional valence and familiarity

Although the correlation was not significant, Fig. 4 could also be interpreted to show the tendency that emotionally positive items and negative items are located more on the right side of the graph (familiarity > 4.5). To investigate further, the extremity of valence (the deviation from the midpoint of the valence scale; Rocklage et al. 2018) was calculated via the following equation: $\text{extremity} = |4 - (\text{valence score})|$. The higher the extremity of valence score is, the more emotional the item is – be it positively or negatively. It was found that the extremity of valence weakly yet significantly correlated with formulaic familiarity ($r = .21$; $p < .01$), which is graphed in Fig. 5. The result indicates that emotionally charged items – either positive or negative – yield higher familiarity values.

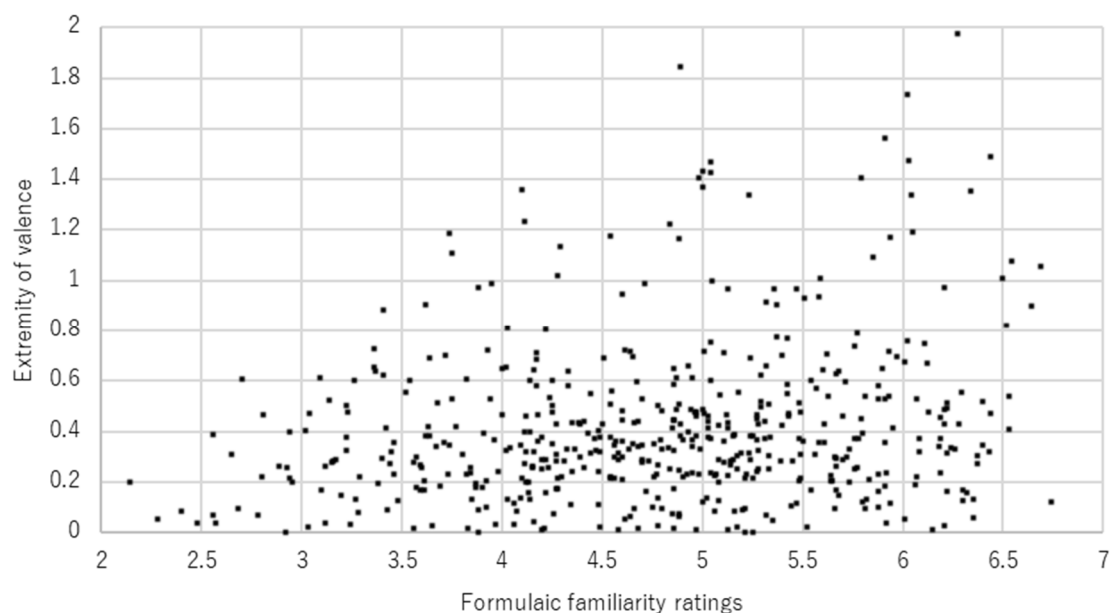


Fig. 5: Scatterplot showing the correlation between extremity of valence and familiarity

3.2 Reliability of formulaic familiarity scores

The reliability testing was implemented for subsets also in terms of familiarity (Tab. 3). High scale reliability (internal consistency) surpassing the acceptable level of .70-.80 was shown across all subsets), indicating that no data re-collection nor omission would be required.

Reliability testing through Cronbach α			
Subset 1 (#1-50)	$\alpha=.982$	Subset 6 (#251-300)	$\alpha=.976$
Subset 2 (#51-100)	$\alpha=.978$	Subset 7 (#301-350)	$\alpha=.965$
Subset 3 (#101-150)	$\alpha=.983$	Subset 8 (#351-400)	$\alpha=.961$
Subset 4 (#151-200)	$\alpha=.972$	Subset 9 (#401-450)	$\alpha=.970$
Subset 5 (#201-250)	$\alpha=.966$	Subset 10 (#451-501)	$\alpha=.974$

Tab. 3: The internal consistency of familiarity ratings

It was revealed that there is a statistically significant strong positive correlation ($r = .92$; $p = .00$) between the mean scores of formulaic familiarity produced in this study and the ones in Isobe et al. (2016). The result indicates that Isobe et al.'s (2016) formulaic familiarity rating is replicable, showing its solid reliability and generalizability (cf. Fig. 6).

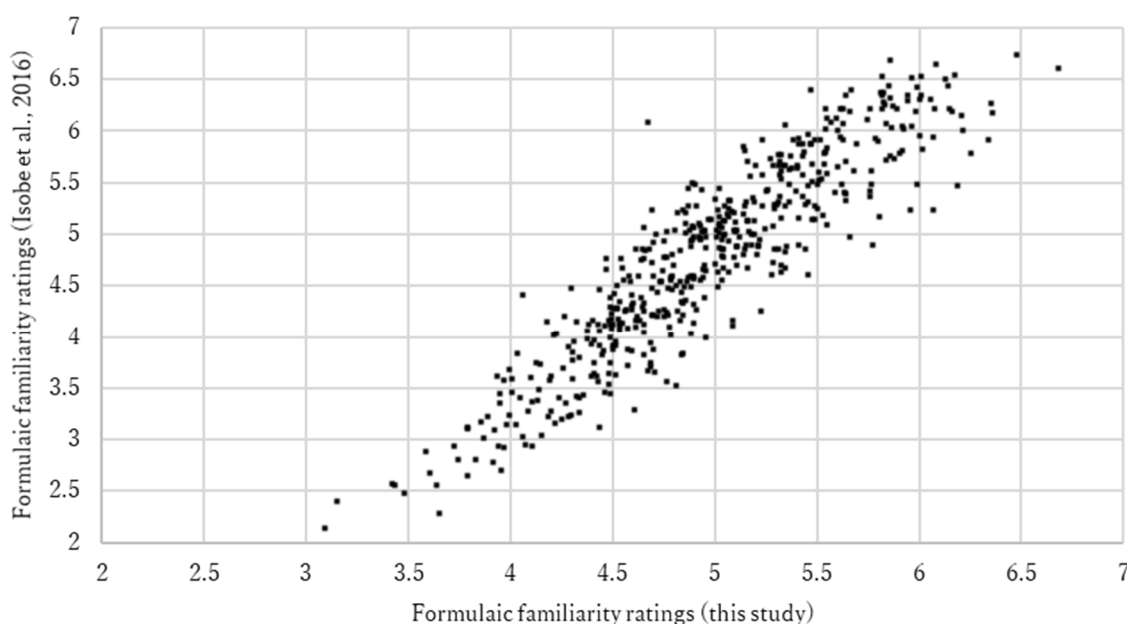


Fig. 6: Scatterplot showing the relationship between the familiarity ratings of English formulae in the present study (X-axis) and in Isobe et al. (2016) (Y-axis)

4. Discussion and Implication

The results in §3.1 indicated the independency of formulaic valence as a formulaic attribute, being discrete from frequency, the major formulaic attribute. It was also indicated that the extremity of emotion could be one of the underlying factors of the phenomenon of familiarity, calling for further investigations. Although there have been emotional databases of lexical items in both L1 (Bradley & Lang 1999; Warriner et al. 2013) and L2 (Kanazawa 2016), no database has extended the scope to emotionality of formulaic sequences. The present study paves the way for future quantitative studies of formulaic emotionality in a distant language and material development more mindful of the findings in the psychology of emotion, and EmAL. A promising next step forward would be to utilize the emotionality database to investigate or control the attribute for future formulaic studies. For instance, Kanazawa (2017) revealed that emotionality of lexical items and processing modes have a facilitatory impact on distant language vocabulary memorization under a more naturalistic (semi-)incidental learning condition. It is implied that such a cognitive advantage of making use of emotionality of the target stimuli would also be seen in learning formulaic language.

In addition, the strong positive correlation in §3.2 revealed that foreign language formulaic familiarity ratings – despite being a subjective measure – can be replicated with high reliability. The finding extended the robustness of lexical familiarity ratings (Amano et al. 2007) to formulaic familiarity ratings of a distant language. Isobe et al.'s (2016) large-scale database is now credited with higher

reliability as well as validity, making its use in future studies and pedagogical implementations more legitimate. As an example of pedagogical implications, formulaic items with disparity between familiarity ratings and objective frequency deserve more attention from teachers and learners. Teachers should encourage learners to learn and use formulaic items rated not familiar when they are actually frequent (e.g., *entitled to, on behalf of, prior to*); whereas learners should be aware that the items they tend to feel are familiar and which they use frequently can actually be relatively infrequent and should not be overused (e.g., *make up one's mind, look to, as good as*), a finding echoing with Hasselgren (1994).

Another important implication is in view of developing an L2/LX formulaic database for distant language learners. The present findings suggest the English database for Japanese speakers – an extreme case of learning a distant language – is reliably replicable. If the language distance was closer between the mother tongue and the target language (such as a French speaker learning German or vice versa), the result could also be reliable. Subjective ratings by a foreign language speaker – such as emotional valence and familiarity – may well be regarded as reliable data to be collected and used in multilingual applied linguistic investigations.

5. Limitations and Further Study

The present study successfully enriched Isobe et al.'s (2016) distant language formulaic database with formulaic emotional valence data and demonstrated its independency as a formulaic attribute. Furthermore, it replicated Isobe et al.'s (2016) familiarity study, successfully proving its robust reliability via strongly positive and significant correlation between their study and the present study. Although the aims of the study were achieved, it has several limitations.

First, the multifacetedness of formulae was not fully investigated. Classification of formulae is a controversial topic with different researchers defining and classifying formulae differently. For example, Lewis (1993) categorizes collocations, polywords, fixed expressions, and semi-fixed expressions, when Moon (1998) classifies different kinds of formulae such as compounds, phrasal verbs, and idioms. Moreover, Myles and Cordier (2017) distinguish learner-external formulaic sequences (linguistic clusters) from learner-internal formulaic sequences (processing units). As Wood (2020) succinctly notes, formulae are certainly not monolithic and many items defy strict and ready classification (Wood 2020: 42). This study took a detour to avoid this conundrum by utilizing an established formulaic list for foreign language learners (Martinez & Schmitt 2012). A future study could face this topic by investigating the variety of formulae separately.

Second, although the reliability of distant language formulaic familiarity was shown by the replicating part of the present study, reliability of the present

formulaic emotional valence list has not yet been investigated. Future research should replicate the survey to ensure the reliability of the emotional ratings.

Third, each formulaic item was displayed with no context in the present study whereas effects of context play an important role in vocabulary/formula acquisition (Webb & Nation 2017). In addition, contextual emotionality is known to affect acquisition (Kanazawa & Lafleur 2019; Kanazawa 2021). The present study was a preliminary study to pave the way for future investigations by providing a database, so the original aims were achieved. Future studies should investigate how different contexts in which the formulaic sequences are embedded affect differently how formulaic sequences are processed and learned.

Fourth, the formulaic list should be utilized with pedagogical applications. One ongoing attempt is a project to develop an electronic formulaic learning system for tertiary EFL learners, in which students utilise the Indirect Spaced Repetition Software (Lafleur 2016) – a special and original spaced repetition learning platform that enables learners to study the items repeatedly with optimal interval and task diversity. Moreover, educational activities utilising varied levels of processing (viz., perception as well as semantics) may well be adopted for foreign language formulaic learning (Kanazawa 2020c). In planning those, the foreign language formulaic database of the present study with enriched formulaic data and carefully selected items will cater for the educators' needs to plan such activities and to develop materials.

Fifth, the present attempt, which relied only on cognitive and psycholinguistic quantitative approaches, should be compensated with other approaches in multidisciplinary fields, including not only pragmatics and corpus-based approaches (Siyanova-Chanturia & Pellicer-Sánchez 2019) but also social and qualitative approaches that are popular in EmAL studies. Already in the domain of motivation studies – a big aspect of EmAL – researchers of formulae and motivation have collaborated, producing pioneering research investigating the role of motivation strategies in formulaic acquisition (Le-Thi et al. 2020). Such collaborative works with multiple research methodologies will produce fruitful outcomes beneficial to applied linguistics both theoretically and empirically.

Finally, the present findings should be extended and applied to other languages with different degrees of linguistic distance (Chiswick & Miller 2005) which can bring about a digital/numerical methodological advancement in emotion research in applied linguistics in different countries – and ultimately a better understanding and learning of foreign languages, including a distant language.

REFERENCES

- Amano, S., Kondo, T. & Kato, K. (1999). Familiarity effect on spoken word recognition in Japanese. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (eds.), *Proceedings of 14th International Congress of Phonetic Sciences* (pp. 873-876). San Francisco, CA: University of California.
- Amano, S., Kasahara, K. & Kondo, T. (2007). Reliability of familiarity rating of ordinary Japanese words for different years and places. *Behavior Research Methods*, 39(4), 1008-1011.
- Amano, S., Kasahara, K. & Kondo, T. (2008). *Nihongo goi tokusei: Dai 9 kan* [Characteristics of Japanese vocabulary: Volume 9]. Tokyo, Japan: Sanseido.
- Barrett, L. F. (2017). *How emotions are made*. Boston, MA: Houghton Mifflin Harcourt.
- Bradley, M. M. & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings*. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, Florida, FL.
- Charness, G., Gneezy, U. & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1), 1-8.
- Chiswick, B. R. & Miller, P. W. (2005). Linguistic distance: A quantitative measure of the distance between English and other languages. *Journal of Multilingual and Multicultural Development*, 26(1), 1-11.
- Connine, C. M., Mullennix, J., Shernoff, E. & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(6), 1084-1096.
- Conklin, K. (2020). Processing single-word and multiword items. In S. Webb (ed.), *The Routledge handbook of vocabulary studies* (pp. 174-188). London, England: Routledge.
- Cordier, F. & Ny, J.-F. L. (2005). Evidence for several components of word familiarity. *Behavior Research Methods*, 37(3), 528-237.
- Dewaele, J. -M. (2015). On emotions in foreign language learning and use. *The Language Teacher*, 39(3), 13-15.
- Dewaele, J. -M. (2018). Why the dichotomy 'L1 versus LX user' is better than 'native versus non-native speaker.' *Applied Linguistics*, 39(2), 236-240.
- Divjak, D. (2019). *Frequency in language: Memory, attention and learning*. Cambridge, England: Cambridge University Press.
- Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, 24(2), 143-188.
- Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of experimental psychology. General*, 113(2), 256-281.
- Hashimoto, B. J. & Egbert, J. (2019). More than frequency? Exploring predictors of word difficulty for second language learners. *Language Learning*, 69(4), 839-872.
- Hasselgren, A. (1994). Lexical teddy bears and advanced learners. *International Journal of Applied Linguistics*, 4(2), 237-258.
- Isobe, Y. (2011). Representation and processing of formulaic sequences in L2 mental lexicon. *JACET Kansai Journal*, 13, 38-49.
- Isobe, Y., Ikuma, Y., Izumi, E., Kadota, S., Kanazawa, Y., Satoi, H., Sugiura, K., Hirai A., Fujiwara, Y., Hori, T., Matsuda, N., Morishita, M. & Yabuuchi, S. (2016). *Nihonjin eigo gakushuusha no teikeiyougen shinmitsudo chousa* [Research on formulaic sequences of Japanese learners of English]. *Proceedings of the Japan Society of English Language Education, Saitama Japan*, 42, 84-85.



- Jiang, N. (2018). *Second language processing*. New York, NY: Routledge.
- Kanazawa, Y. (2016). Micro-level emotion as a factor of L2 vocabulary memory. *Language Education & Technology*, 53, 23-52.
- Kanazawa, Y. (2017). Emotion-Involved Semantic Processing and L2 vocabulary memory. *Vocabulary Learning and Instruction*, 6(2), 23-30.
- Kanazawa, Y. & Lafleur, L. (2019). *E-NAWL: Building of an enlarged, emotional, electronic NAWL*. Paper presented at the 52nd annual meeting of the British Association of Applied Linguistics, Manchester Metropolitan University, Manchester, England.
- Kanazawa, Y. (2020a). *Development of an affective formulaic list and its e-learning application in tertiary EFL education*. Paper presented at le Colloque Thématique de l'Association Suisse de Linguistique Appliquée, Université de Neuchâtel, Neuchâtel, Switzerland.
- Kanazawa, Y. (ed.) (2020b). *Formula to gaikokugo gakushu, kyouiku* [Formulaic sequences in foreign language learning and education]. Tokyo, Japan: Kurosio Publishers.
- Kanazawa, Y. (2020c). Micro-level emotion in shallow/perceptual processing: Testing the Deep Positivity Hypothesis on the valence-dependent difference for LX incidental lexical memory. *Language Education & Technology*, 57, 1-30.*
- Kanazawa, Y. (2021). *Lexical emotional valence and contextual emotional valence in foreign language vocabulary learning*. Paper presented at the 2021 Conference of the American Association for Applied Linguistics, Houston, TX.
- Kanazawa, Y. (2021). Do not (just) think, but (also) feel! *SAGE Open*, July-September 2021, 1-13.
- Kreuz, R. J. (1987). The subjective familiarity of English homophones. *Memory & Cognition*, 15(2), 154-168.
- Lafleur, L. (2016). Indirect Spaced Repetition Software (ISRS). *Vocab@Tokyo International Conference Journal*, 159-160.
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York, NY: Routledge.
- Lewellen, M. J., Goldinger, S. D., Pisoni, D. B. & Greene, B. G. (1993). Lexical familiarity and processing efficiency: Individual differences in naming, lexical decision, and semantic categorization. *Journal of Experimental Psychology: General*, 122(3), 316-330.
- Le-Thi, D., Dörnyei, Z. & Pellicer-Sánchez, A. (2020). Increasing the effectiveness of teaching L2 formulaic sequences through motivational strategies and mental imagery: A classroom experiment. *Language Teaching Research*. Advanced online publication.
- Lewis, M. (1993). *The lexical approach*. Hove, England: Language Teaching Publications.
- Martinez, R. & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33(3), 299-320.
- Moon, R. (1998). *Fixed expressions and idioms in English*. Oxford, England: Clarendon Press.
- Myles, F. & Cordier, C. (2017). Formulaic sequence (FS) cannot be an umbrella term in SLA. *Studies in Second language Acquisition*, 39, 3-28.
- Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. Amsterdam, The Netherlands: John Benjamins Publishing Company.
- Nishikawa-Van Eester, M. (2016). "Linguistic distance" and its influence on TESOL. *International Politics and Economics (Bulletin of Nishōgakusha University)*, 22, 21-32.
- Patten, M. L. & Newhart, M. (2018). *Understanding research methods*. New York, NY: Routledge.
- Pavlenko, A. (2013). The affective turn in SLA. In D. Gabryś-Barker & J. Bielska (eds.), *The affective dimension in second language acquisition* (pp. 3-28). Bristol, England: Multilingual Matters.
- Rocklage, M. D., Rucker, D. D. & Nordgren, L. F. (2018). The Evaluative Lexicon 2.0. *Behavior Research Methods*, 50, 1327-1344.

- Schütze, U. (2017). *Language learning and the brain*. Cambridge, England: Cambridge University Press.
- Sharwood Smith, M. (2017). *Introducing language and cognition*. Cambridge, England: Cambridge University Press.
- Siyanova-Chanturia, A. & Pellicer-Sánchez, A. (2019). *Understanding formulaic language: A second language acquisition perspective*. New York, NY: Routledge.
- Talmi, D., Schimack, U., Paterson, T. & Moscovitch, M. (2007). The role of attention and relatedness in Emotionally Enhanced Memory. *Emotion*, 7(1), 89-102.
- Taylor, J. R. (2012). *The mental corpus*. Oxford, England: Oxford University Press.
- Truscott, J. (2015). *Consciousness and second language learning*. Bristol, England: Multilingual Matters.
- Warriner, A. B., Kuperman, V. & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45, 1191-1207.
- Webb, S. (ed.) (2020). *The Routledge handbook of vocabulary studies*. New York, NY: Routledge.
- Webb, S. & Nation, P. (2017). *How vocabulary is learned*. Oxford, England: Oxford University Press.
- Wood, D. (2020). Classifying and identifying formulaic language. In S. Webb (ed.), *The Routledge handbook of vocabulary studies* (pp. 30-45). London, England: Routledge.
- Wray, A. (2019). Identifying formulaic language. In R. Corrigan, E. A. Moravcsik, H. Ouali, & K. M. Wheatley (eds.), *Formulaic language* (pp. 27-51). Amsterdam, The Netherlands: John Benjamins Publishing Company.
- Yokokawa, H. (ed.) (2006). *Nihonjin eigo gakushusya no eitango shinmitsudo moji hen: Kyoiku kenkyu no tameno daigengo database* [Familiarity of English words for Japanese EFL learners, written word version: The database for second language education and research]. Tokyo, Japan: Kurosio Publishers.
- Yokokawa, H. (ed.) (2009). *Nihonjin eigo gakushusya no eitango shinmitsudo onsei hen: Kyoiku kenkyu no tameno daigengo database* [Familiarity of English words for Japanese EFL learners, spoken word version: The database for second language education and research]. Tokyo, Japan: Kurosio Publishers.

*This reference is available online.

Acknowledgement

This paper is affiliated with the author's research project supported by the Japan Society for the Promotion of Science KAKENHI Grant-in-Aid for Scientific Research (C), Grant Number 19K00899 (<https://kaken.nii.ac.jp/en/grant/KAKENHI-PROJECT-19K00899/>) and with the collaborative research project conducted by the Japan Association for Language Education and Technology Kansai Chapter Fundamental Theory Special Interest Group (<https://let-kansai-fmt-sig.blogspot.com/>), of which the author is the coordinator and the project leader.

Le plurilinguisme dans les MOOCs: de nouvelles ressources pour la construction des savoirs

Mariana FONSECA & Laurent GAJO

Université de Genève

École de langue et de civilisation françaises (ELCF)

Rue de Candolle 5, 1211 Genève 4, Suisse

mariana.fonseca@unige.ch; laurent.gajo@unige.ch

MOOCs (Massive Open Online Courses) represent a new way of teaching and disseminating knowledge. They are based on digital technologies which make them not only a new medium, but a new didactic genre. By allowing a "massive" diffusion of knowledge, these courses re-examine the languages mobilized in the construction and the dissemination of knowledge as well as their cultural dimension (embedded knowledge). In this article, we try to show the tension between globalization and contextualization of knowledge that drives MOOCs, by focusing on several discursive functions (constitutive, vehicular and indexical functions). Through a multilingual perspective, we illustrate how these functions are found in MOOCs belonging to different disciplinary fields, and having different working and subtitle languages. In conclusion, we discuss how technology can work not only as a tool, but also as a revealer of the complex links between discourse and knowledge.

Keywords:

MOOC, plurilingualism, subtitles, knowledge construction, vehicular function, constitutive function, indexical function.

Mots-clés:

MOOC, plurilinguisme, sous-titrage, construction des savoirs, fonction véhiculaire, fonction constitutive, fonction indexicale.

1. Introduction

Dans le contexte actuel d'internationalisation des institutions d'enseignement supérieur, les MOOCs (*Massive Open Online Courses*) sont en plein essor. Du côté des universités, il s'agit de diffuser des connaissances produites localement auprès d'un vaste public et, ainsi, faire rayonner leur image au-delà des frontières nationales. Du côté des usagers, il s'agit de bénéficier d'enseignements dispensés par des spécialistes reconnus et, sous certaines conditions, d'obtenir un certificat délivré par une institution prestigieuse.

Comme son acronyme l'indique, la mise en place des cours "massifs", "ouverts" et "en ligne" est étroitement liée à l'outil numérique. Ces cours possèdent une organisation relativement constante du point de vue du scénario pédagogique¹: modules thématiques centrés sur des capsules vidéos, structurés autour d'un parcours hebdomadaire. Chaque semaine est validée par un quiz final dont la réussite conditionne l'accès au module suivant. Ces caractéristiques font de ce

¹ Nous nous référons ici au modèle pédagogique le plus répandu, appelé "xMOOCs".

dispositif "un 'genre' en voie de constitution" (Mangenot 2017: 164). Permettre l'accès au savoir au plus grand nombre à travers des cours disponibles sur des plateformes en ligne, comme le veulent les MOOCs, ouvre des questions originales aux chercheurs qui s'intéressent au rôle du discours et du plurilinguisme pour l'élaboration des connaissances.

On pourrait notamment se demander quelles sont les caractéristiques de ces discours didactiques qui s'adressent à un usager virtuel, qui ne partage pas nécessairement la même langue ni les mêmes références culturelles et qui n'a pas forcément de connaissances préalables sur le sujet du cours. On pourrait également se demander si une seule langue serait suffisante pour "parler à tous" ou si l'outil numérique pourrait proposer de nouvelles ressources qui permettraient non seulement une meilleure prise en compte de la diversité linguistique et culturelle des usagers (voir, par exemple, Kerr, Merciai & Eradze 2018; Stratton & Grace 2016; Phan 2018), mais également une nouvelle façon d'appréhender les savoirs.

Ces questionnements montrent la complexité sous-jacente à l'ambition d'universalité des MOOCs. Il s'agit, d'une part, d'interroger les caractéristiques didactiques du cours, ce qui présuppose une réflexion approfondie sur les savoirs à construire et sur la langue, à envisager comme un outil de problématisation (fonction constitutive). D'autre part, il faut répondre aux besoins d'une diffusion large, ce qui tend à l'utilisation d'une langue unique, souvent réduite à un simple support à la communication (fonction véhiculaire). S'ajoute à cela la "délocalisation" que permet l'outil numérique. Pourtant, les discours sont produits par des acteurs ancrés dans des contextes précis et parlant, ainsi, selon un certain angle de vue (fonction indexicale, voir ci-dessous). De la même manière, ils sont reçus par des usagers qui s'inscrivent, eux aussi, dans des contextes linguistiques et culturels spécifiques, et parfois très divers, et qui appréhendent les savoirs à partir d'un certain point de vue.

Dans cet article, nous nous intéressons aux trois fonctions du discours citées ci-dessus et montrerons comment elles se retrouvent dans différents MOOCs, dévoilant ainsi la tension entre contextualisation et globalisation qui anime ce dispositif. Après avoir exposé notre cadre théorique, nous illustrerons les trois fonctions du discours à partir d'un corpus varié, composé de trois éléments: a) extraits de MOOCs appartenant à divers champs disciplinaires, ayant différentes langues principales et de traduction; b) extraits d'entretiens réalisés avec les concepteurs; c) témoignages recueillis dans des questionnaires de recherche proposés aux usagers. Nous montrerons, en conclusion, comment la technologie peut fonctionner non seulement comme un outil, mais également comme un révélateur des liens complexes entre discours et savoirs.

2. Cadre théorique

Nos analyses se basent sur des outils linguistiques issus de quatre champs théoriques: l'analyse de l'interaction, l'analyse du discours, les études du bi-plurilinguisme et les théories de l'acquisition des langues secondes ou étrangères (voir Gajo et al. 2013). Nous tentons de comprendre, à travers une analyse fine des marques linguistiques, les processus de transmission et de construction des savoirs en situation plurilingue.

Dans notre perspective, le plurilinguisme est envisagé de deux manières, comme révélateur et comme déclencheur (voir Berthoud & Gajo 2016):

- en tant que révélateur, le plurilinguisme permet, par un processus de défamiliarisation, de mettre en évidence des mécanismes à l'œuvre dans toute communication, mais pourtant peu perceptibles en situation monolingue; il suffit de penser, par exemple, à la traduction (voir Cassin 2004), qui engendre un retour réflexif sur la manière dont la langue première fonctionne et encode les savoirs; c'est précisément ce qui se passe avec le processus de sous-titrage plurilingue des MOOCs, qui peut amener les concepteurs à revenir sur l'exposition de leur contenu en explicitant, en désambiguïsant ou en clarifiant des aspects restés invisibles dans une perspective monolingue;
- en tant que déclencheur, le plurilinguisme permet l'avènement de modalités spécifiques de travail pour la construction des savoirs; nous pouvons évoquer, par exemple, le processus de comparaison, qui opère sur des sources, des documents, des théories présentées dans une diversité de langues; le sous-titrage d'un MOOC peut, s'il est prévu en amont, favoriser l'utilisation d'un matériel plurilingue et une navigation éclairée entre plusieurs versions d'un contenu.

Analyser le discours plurilingue nous conduit ainsi à identifier des fonctions fondamentales de toute activité discursive. Dans le schéma ci-dessous, nous plaçons autour du discours les trois fonctions suivantes: la fonction véhiculaire, la fonction constitutive et la fonction indexicale (voir Berthoud & Gajo 2020). La fonction véhiculaire prédomine lorsque le discours apparaît avant tout dans son rôle de support à la communication. Il est considéré comme largement indépendant des objets de savoir qu'il transmet, doit viser une forme de transparence et, lorsque les ressources langagières font défaut, par exemple en langue étrangère, il activera des stratégies de résolution de problèmes de communication (*problem-solving*). La fonction constitutive prédomine quand le discours apparaît dans son caractère structurant pour les savoirs, qu'il façonne plus qu'il ne transporte. Le discours plurilingue fait apparaître les savoirs dans leur densité et sert à pointer des problèmes de conceptualisation (*problem-raising*) plutôt qu'à résoudre des problèmes de communication. La fonction indexicale, quant à elle, nous rappelle que les discours sont produits par des

voix ancrées dans des contextes, des trajectoires, des cultures. Ils expriment des angles de vue et peuvent aussi les remettre en question. Les discours ne sont jamais neutres, ils placent l'énonciateur face à son objet, ils orientent et sont orientés.

Dans notre schéma, le discours, en tant qu'activité, occupe la position centrale. Si on agit dans et par le discours, celui-ci entre dans des moules immédiatement reconnaissables et efficaces en situation. On pensera, en particulier, aux genres textuels ou discursifs qui organisent les activités dans un certain domaine. Le MOOC constitue un (macro-)genre² récent dans le domaine de l'éducation. S'approprier les règles de ce nouveau genre exige de la part des concepteurs une véritable formation. Par ailleurs, un MOOC décline et articule toute une série de (micro-)genres plus ou moins spécifiques et liés aux ressources numériques. La "capsule vidéo" pourrait constituer un élément de ces nouveaux genres.

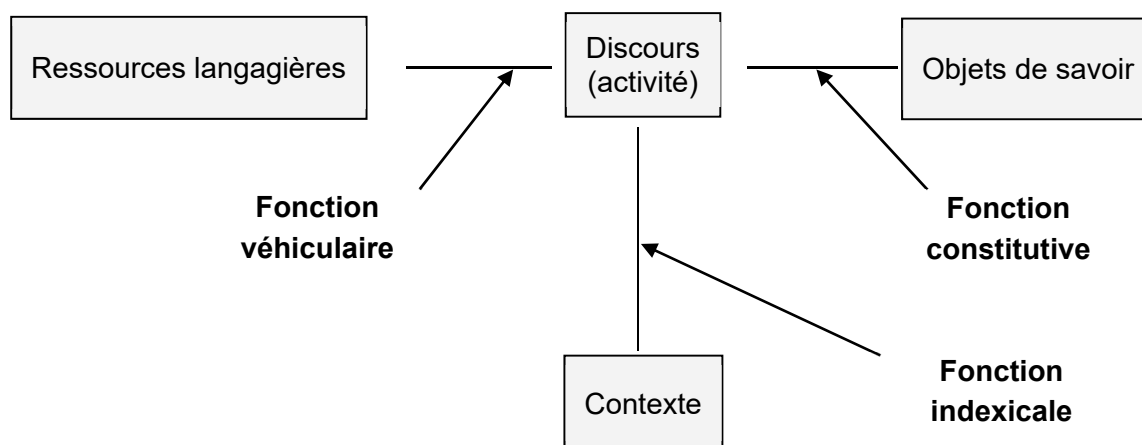


Fig.: 1 Les fonctions du discours

Ces différentes fonctions peuvent s'observer dans tout type de discours, mais demeurent le plus souvent implicites et non questionnées. De ce fait, il est intéressant d'analyser en quoi la communication plurilingue et numérique – dans sa mise en place didactique et dans son fonctionnement – problématise ces fonctions et rend plus ou moins saillant le recours à l'une ou l'autre d'entre elles.

3. Présentation de la recherche et du corpus

La réflexion que nous développons dans cet article fait partie d'une recherche plus vaste intitulée "plurilinguisme et construction des savoirs dans les MOOCs", qui s'intéresse à montrer comment le recours à différentes langues dans ces dispositifs – notamment à travers la traduction et le sous-titrage –

² Au stade actuel de la recherche, il n'est pas simple d'appliquer la notion de genre aux MOOCs, qui revêtent des formats divers. Il paraît toutefois raisonnable de postuler à la fois un air de famille et des spécificités communes aux différents MOOCs, d'où la notion heuristique de "macro-genre".

amène des changements dans les pratiques éducatives du côté des formateurs et des formés (pour une présentation détaillée, voir Fonseca & Gajo 2020). Pour cela, nous nous focalisons sur la dimension linguistique et culturelle des MOOCs et, en particulier, sur la place du plurilinguisme pour la construction et la contextualisation des savoirs, à travers la prise en compte de trois perspectives complémentaires:

- concepteurs (angles de vue)
- usagers (points de vue)
- savoirs (caractère plus ou moins universel et variablement problématisés)

Du point de vue méthodologique, chaque axe considéré a demandé la mise en œuvre d'un protocole spécifique: a) entretiens semi-directifs avec les concepteurs; b) questionnaires "Limesurvey" intégrés aux MOOCs étudiés portant sur le profil linguistique des participants et leur utilisation des sous-titres; c) analyse discursive des données des MOOCs, à partir notamment de la traduction et du sous-titrage effectué (voir plus bas).

Le corpus que nous allons analyser dans cet article se base notamment sur le contenu des capsules vidéos ainsi que sur les entretiens réalisés avec les concepteurs de trois MOOCs: "Introduction aux droits de l'homme", "Drugs, drug use, drug policy and health" et "Introduction au raisonnement clinique". Ces MOOCs appartiennent à différents champs disciplinaires et ont diverses langues principales et de traduction (voir Fig. 2). Il est important de préciser que la traduction et le sous-titrage vers d'autres langues constituent le point de départ de notre recherche et le point de convergence entre les MOOCs étudiés. Ce travail a été assuré par des membres de notre équipe ayant de connaissances très élevées en langue cible et en langue source et a fait souvent l'objet de riches échanges avec les concepteurs sur les choix de traduction effectués (voir Fig. 3, plus loin). Précisons encore que les MOOCs qui intègrent notre corpus ont été conçus par des chercheurs de l'Université de Genève (parfois en collaboration avec d'autres institutions) et qu'ils sont disponibles "à la demande" sur la plateforme Coursera³.

³ Ces MOOCs ont été lancés à des dates différentes et connaissent un nombre très inégal d'inscrits, comme voici : "Introduction aux droits de l'homme" (lancé en mars 2014, nombre d'inscrits 47338), "Drugs, drug use, drug policy and health" (lancé en juin 2018, nombre d'inscrits 10541) et "Introduction au raisonnement clinique" (lancé en janvier 2019, nombre d'inscrits 750). Données du 3 juillet 2020.

MOOC	Discipline de référence	Langue principale des capsules vidéos	Principales langues de traduction et de sous-titrage ⁴
Introduction aux droits de l'homme	Droit	Français	Anglais
Drugs, drug use, drug policy and health	Sciences médicales/ sciences sociales	Anglais	Espagnol Français Russe
Introduction au raisonnement clinique	Sciences médicales	Français	Anglais

Fig. 2: Vue d'ensemble du corpus

4. Analyses

4.1 Fonction véhiculaire

Notre recherche a permis de montrer que, malgré la conscience du rôle constitutif de la langue pour la construction des connaissances dont font preuve les concepteurs (voir rubrique 4.2), le plurilinguisme est rarement mobilisé dans la diffusion des savoirs. Ainsi, dans le MOOC "Introduction aux droits de l'homme", toutes les capsules vidéos (67 au total) ont le français comme langue principale. Si nous nous référons aux principaux types d'alternance (voir Gajo & Steffen 2015), nous pouvons affirmer que ce MOOC a été conçu selon un système de *macro*-alternance avec le français. Autrement dit, l'utilisation du français comme langue principale a été planifiée en amont lors de la conception du MOOC (voir plus loin). C'est notamment par son caractère programmable que la *macro*-alternance se distingue de la *micro*-alternance, également connue sous le nom de code-switching, qui peut être définie comme un changement spontané de langue dans le discours. Entre ces deux niveaux, il existe un niveau intermédiaire, appelé *méso*-alternance, qui consiste en un changement de langue entre deux activités ou tâches didactiques. Par exemple, lorsque dans un MOOC, on lit/projette l'extrait d'un texte dans une langue (souvent la langue originale) et on le commente dans une autre⁵. Du point de vue de l'alternance codique, la situation est similaire pour le MOOC "Drugs, drug use, drug policy and health", où, à part 5 vidéos où les intervenants s'expriment dans leur propre langue, dont 3 en espagnol et 2 en français, les autres 53 capsules vidéos du MOOC⁶ sont en anglais, souvent L2 des intervenants et inégalement maîtrisée par ces derniers. La structure linguistique de ce MOOC peut également être caractérisée en termes de *macro*-alternance avec l'anglais.

⁴ Ce tableau ne prend pas en compte les traductions et sous-titrages effectués par les bénévoles de la plateforme Coursera, mais seulement ceux qui ont fait l'objet d'une demande de la part des concepteurs.

⁵ Malgré l'intérêt didactique de la *méso*-alternance pour les MOOCs, nous ne la développerons pas dans le cadre de cet article.

⁶ Les chiffres annoncés ici se réfèrent aux vidéos obligatoires et ne prennent pas en compte les vidéos "bonus", dont le contenu ne fait pas l'objet d'une évaluation.

Les entretiens de recherche que nous avons menés avec les concepteurs contribuent à expliquer leur choix en ce qui concerne la langue principale du cours et la-les langue-s de traduction et montrent la tension qui existe entre, d'une part, la conscience de l'importance du discours et des langues pour la construction des connaissances et, d'autre part, le choix d'une langue unique pour leur diffusion. Le choix d'une seule langue pour la transmission des savoirs est notamment influencé par la discipline de référence, mais subit également le poids des contraintes pratiques ainsi que des représentations sociales. Le monolingisme observé dans la diffusion des savoirs nous rapproche de la fonction véhiculaire du discours, dans le sens où il y a une sous-exploitation des possibilités didactiques offertes par le numérique qui permettraient de donner voix à différents intervenants dans leurs langues et cultures éducatives respectives, à travers un recours réfléchi à la traduction et au sous-titrage. La diffusion plurilingue permettrait ainsi de construire un autre rapport au savoir, accordant plus de place à la contextualisation et à la problématisation. Comme l'observe Cerquiglini (2013: 16):

l'aller et retour entre les langues est une dimension même de l'activité scientifique; il est au cœur du transfert. La traduction, comme contrainte productive, est une source d'enrichissement de la pensée scientifique et de ses applications technologiques.

Dans le passage présenté ci-dessous⁷, extrait d'un entretien mené avec les concepteurs du MOOC "Introduction aux droits de l'homme", on remarque que la diffusion des savoirs a été pensée selon une perspective monolingue.

(1) Entretien avec les concepteurs du MOOC "Introduction aux droits de l'homme"

on s'est demandé si on allait le faire en anglais\ et euh là y avait un obstacle de taille c'est que j'aime beaucoup l'anglais mais je le je le parle extrêmement mal\
 [...] on s'est dit éventuellement peut-être que plus tard est-ce qu'une/ on s'est dit une traduction ou des sous-titres mais au début/ [...] c'est une question qu'on a même pas examinée parce qu'en réalité le rectorat voulait le le mooc rapidement\
 on a su l'espagnol c'était déjà fait français l'afrique mais après l'anglais s'impose quand même assez rapidement

Comme expliqué dans cet extrait, les concepteurs se sont demandé s'ils allaient faire le MOOC en anglais, mais apparemment ne se sont pas questionnés sur la possibilité de le faire dans différentes langues et de recourir ensuite à la traduction et au sous-titrage. Comme dévoilé par l'entretien, c'est là que les questions d'ordre pratique interviennent. Les concepteurs avaient une contrainte temporelle qui les obligeait à aller à "l'essentiel". Il faut dire que ce MOOC a été un des premiers à être conçus par l'Université de Genève, à un moment où la traduction et le sous-titrage ne faisaient pas partie des préoccupations des concepteurs. Remarquons, pour finir, la représentation selon laquelle l'anglais s'impose en tant que langue de diffusion des savoirs qui a fait que les concepteurs, quatre ans après le lancement du MOOC, ont voulu

⁷ Les conventions de transcription se trouvent dans la rubrique "annexe(s)".

le traduire prioritairement en anglais. Le choix de l'anglais comme langue de traduction ne semble pas encore avoir porté ses fruits, car d'après les données qui ressortent de notre questionnaire de recherche (proposé à toute personne qui suit ce MOOC au moins jusqu'au 3^{ème} module), les usagers qui visionnent ce cours sous-titré en anglais sont minoritaires⁸.

Le deuxième extrait que nous proposons pour illustrer le lien observé entre diffusion des savoirs et fonction véhiculaire du discours est issu d'un entretien avec les concepteurs du MOOC "Drugs, drug use, drug policy and health". Contrairement au MOOC précédent, pour lequel la traduction et le sous-titrage sont intervenus bien après son lancement, pour le MOOC en question, il avait été établi, dès la conception du projet, qu'il bénéficierait d'une traduction et d'un sous-titrage vers trois langues, ce qui rendait possible une diffusion multilingue. Néanmoins, comme explicité plus haut, cela n'a pas été le cas et c'est une perspective monolingue qui l'a emporté. Interrogés sur cet aspect, les concepteurs expliquent:

(2) Entretien avec les concepteurs du MOOC "Drugs, drug use, drug policy and health"

[...] il me semble que nous voulions un maximum d'anglais je je pousse un peu l'argument pour être certains de couvrir un maximum d'universalité\ euh et que o- o- au départ certaines des personnes que nous souhaitions que que pour lesquelles il était clair qu'elles s'exprimeraient que dans leur langue/ on le voyait un peu comme une exception\ maintenant cette discussion d'aujourd'hui et en vue du du du mooc à venir euh [...] y faut qu'on analyse bien les langues des des des apprenants [...] parce que je crois que encore une fois singulièrement pour tout ce qui touche aux aux aspects sociaux euh euhm euh peut-être au contraire faut-il qu'on force le trait [...] de la diversité des langues à dessein\ mais on ne l'a pas fait au départ hein

L'argument avancé ici selon lequel une diffusion en anglais permettrait plus d'universalité est devenu un lieu commun dans de nombreux domaines de la recherche scientifique, malgré les différentes mises en garde que l'emploi d'une langue unique (le plus souvent l'anglais) pour la diffusion des savoirs a pu susciter (voir, par exemple, Altbach 2014; Yanaprasart 2018). Ce qui est intéressant dans le cas de ce MOOC, c'est non seulement de voir une certaine prise de conscience, a posteriori, de la part des concepteurs, mais également de constater que la prétendue universalité de l'anglais n'est pas partagée par les usagers. Ainsi, les questionnaires de recherche que nous avons proposés aux participants du MOOC "Drugs, drug use, drug policy and health" ont montré que, loin de permettre une meilleure appropriation des savoirs, le choix d'une langue unique pour la diffusion des connaissances, souvent au détriment de la maîtrise effective de cette langue par les intervenants, devient une source de difficulté pour la compréhension du cours. Autrement dit, plusieurs usagers affirment avoir recours aux sous-titres, en anglais ou dans une autre langue,

⁸ Début juillet 2020, seulement 10% des participants ayant répondu à notre questionnaire de recherche affirmaient visionner ce MOOC sous-titré en anglais. A cette même date, le nombre total de réponses à notre questionnaire était de 335.

pour mieux comprendre le discours d'un intervenant qui s'exprimerait dans un anglais mal maîtrisé. On observe cela dans le témoignage ci-dessous:

- (3) [Subtitles] may also help understanding speakers whose English pronunciation is not very good (in such cases, it is best that they speak French, or whatever other language they speak best).

Ce type de témoignage est récurrent dans le corpus recueilli: environ 20% des usagers qui ont répondu à une question ouverte de notre questionnaire au sujet des apports des sous-titres dans le suivi d'un MOOC affirment y recourir lorsque les accents des intervenants posent des problèmes d'intelligibilité⁹. Par ailleurs, plusieurs répondants affirment apprécier être exposés à une autre langue dans le suivi d'un MOOC:

- (4) [...] The option of subtitles gave me the English translation-while at the same time, I was exposed to another language, which I deeply value.

Le point de vue des usagers, que nous n'allons pas développer dans cet article (voir Fonseca & Gajo 2020), mérite d'être pris en compte par les concepteurs des MOOCs. Les témoignages recueillis montrent qu'il est possible, et même souhaitable, d'introduire plus de diversité dans les MOOCs. L'enjeu serait non seulement d'envisager cette diversité en termes d'une plus grande accessibilité (ce qui renverrait à la fonction véhiculaire du discours), mais de la mobiliser comme un outil de problématisation (fonction constitutive).

4.2 Fonction constitutive

Pour illustrer la fonction constitutive du discours pour la construction des savoirs et montrer comment le plurilinguisme peut intervenir comme un outil de problématisation, nous allons tout d'abord analyser le passage d'ouverture du MOOC "Introduction aux droits de l'homme", que voici:

- (5) Une question préalable se pose cependant. Elle est d'ordre terminologique. Allons-nous parler des droits de l'homme ou des droits humains ? Dans d'autres langues, cette question ne se pose pas. En allemand, par exemple, le terme Menschenrechte désigne clairement tous les êtres humains. Il en est de même en russe: права человека, droits de tous les êtres humains. En anglais aujourd'hui, le terme Human Rights est clairement dominant. Cela n'a cependant pas toujours été le cas. Jusqu'à l'adoption de la Déclaration universelle des droits de l'homme, on utilisait la terminologie Rights of Man. Le changement de terminologie de Rights of Man vers Human Rights a été délibéré. Elle visait à souligner que les droits de l'homme ne sont pas l'apanage, le privilège du genre masculin, mais qu'ils reviennent bel et bien à tous les êtres humains, y compris les femmes. [...] Contrairement à la langue anglaise, « droits de l'homme » reste la terminologie officielle en français. (MOOC *Introduction aux droits de l'homme*, W1V1, 0.34 - 1.24)

Dans cet extrait, nous observons une réflexion sur la terminologie officielle utilisée en français pour parler des "droits de l'homme", principale notion traitée par ce MOOC. Ce qui est particulièrement intéressant, c'est que cette réflexion est élaborée à partir d'autres langues qui sont convoquées explicitement dans le discours ("en allemand", "en russe", "en anglais"), à travers l'alternance en

⁹ Le questionnaire proposé aux usagers de ce MOOC totalisait, début juillet 2020, 234 réponses, dont 151 pour la question évoquée.

mention des terminologies en vigueur dans ces langues ("Menschenrechte", "права человека" et "Human Rights", respectivement), dont le sens est, pour les deux premières, reformulé en français.

On remarque ainsi que, contrairement au français, l'allemand et le russe emploient une terminologie plus neutre, qui serait l'équivalent de "droits humains". Quant à l'anglais, on apprend qu'il y a eu un changement volontaire, passage de "Rights of Man" ("droit de l'homme") à "Human Rights" ("droits humains"), suite à un événement historique majeur, la *Déclaration universelle des droits de l'homme*. Le détour par ces différentes langues contribue à interroger la terminologie officielle en français et nous rappelle, dans un moment où l'écriture inclusive est un sujet de grande actualité, que les termes employés ont une incidence sur l'appréhension des notions. Ainsi, ne serait-il pas plus juste de parler de "droits humains" que de "droits de l'homme"¹⁰? Utiliser un langage précis est indispensable pour permettre une compréhension plus fine de la réalité tout comme pour la construction des savoirs plus étayés.

À ce propos, Lévy-Leblond (2013: 27) en se référant à ce qu'il considère comme de mauvais choix linguistiques pour désigner certains termes de la physique, tels la "relativité", parle du "mésusage des puissances de la langue". Selon lui, certaines terminologies gagneraient à être revues. Les entretiens que nous avons menés avec les concepteurs du MOOC "Introduction aux droits de l'homme" concourent à expliquer une telle introduction. Le témoignage ci-dessous (6) illustre la résistance que peuvent susciter des changements terminologiques ainsi que l'apport du plurilinguisme pour questionner la terminologie en vigueur dans une langue donnée (en l'occurrence le français) et, partant, apporter un nouveau regard sur la notion à laquelle elle renvoie. Comme le dit Lévy-Leblond (2013: 27), en parlant notamment des sciences de la nature et de la matière:

je ne peux pas me défendre d'entretenir l'idée, sans doute utopique, qu'il existe des fonds linguistiques multiples à exploiter pour nous permettre de retravailler de nombreux concepts scientifiques, afin de mieux les comprendre et les partager, quitte éventuellement à modifier des terminologies déjà admises.

(6) Entretien avec les concepteurs du MOOC "Introduction aux droits de l'homme"

la question terminologique elle est un peu une querelle interne aussi si vous voulez [...] j'ai un cœur très féministe aussi alors moi ça me gêne assez les les droits de l'homme en français [...] pour le professeur X mes droits humains il me dit chaque fois c'est une catastrophe\ alors en français ça veut absolument rien dire\ ça va ça va pas en français\ ça va dans toutes les langues/ ça va en anglais mais pas en français [...] et ok on s'est mis d'acc- j'ai dit pour moi **je peux accepter l'explication que ça reste la terminologie officielle**\ ça c'est vrai [...] ça a pas changé en français ce que je vois comme un retard de du français mais bon\ **mais je peux pas accepter cette première partie de dire alors ça va dans toutes les**

¹⁰ D'autant plus que dans le 1^{er} article de la *Déclaration universelle des droits de l'homme* ("tous les êtres humains naissent libres et égaux en dignité et en droits [...]") il n'est pas question d'un masculin pluriel "hommes" mais d'"êtres humains".

langues/ ça va pas en français\ et c'est pour ça j'ai mis cette introduction voilà en disant mais voyez en allm- en anglais aussi on a dû changer [...] et pourquoi le français pourrait pas faire ce virage/

Remarquons que lors de la traduction et du sous-titrage de ce MOOC en anglais, les concepteurs ont souhaité conserver la dimension plurilingue et comparatiste entre le français et l'anglais, proposée dans la réflexion d'ouverture, que le passage à l'anglais aurait pu ternir. L'image ci-dessous (Fig. 3) montre les corrections apportées par un des concepteurs à la version anglaise qui lui a été soumise:

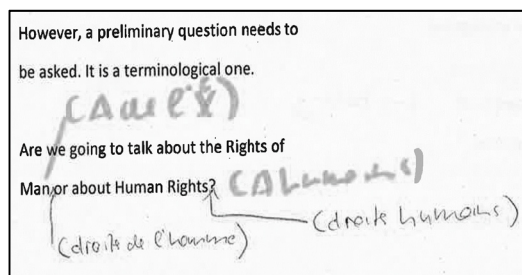


Fig. 3: Annotations à la traduction du MOOC "Introduction aux droits de l'homme"

On voit ainsi que, dans la traduction anglaise, l'expression "droits de l'homme" a été inscrite entre parenthèses juste après "Rights of Man". De la même manière, "droits humains" a été inséré à côté de "Human Rights". Loin d'être un détail, ce choix dévoile un point de vue analytique sur la langue, qui n'est pas sans incidences pratiques sur le visionnage du MOOC, car l'inscription des notions françaises a pour conséquence un dépassement du nombre de caractères conseillé par seconde pour le sous-titrage. Autrement dit, la synchronisation entre la parole et les sous-titres sera impactée et il est probable que pour que les usagers puissent lire l'intégralité de ce qui est écrit, ils doivent mettre la vidéo sur pause (voir la Fig. 4 pour observer le résultat final du travail de traduction et sous-titrage de cet extrait). Interrogé sur cet aspect lors d'un échange au sujet des corrections apportées, un des concepteurs explique:

- (7) Mon idée était de mettre les mentions françaises entre parenthèses. Je pense que l'ajout entre parenthèses se justifie, et **je préfère faire primer les raisons analytiques sur des considérations liées au confort d'utilisation.**



Fig. 4: Traduction et sous-titrage vers l'anglais du MOOC "Introduction aux droits de l'homme", avec des ajouts en français.

Dans l'introduction du MOOC "Drugs, drug use, drug policy and health", ci-dessous (extrait 8), on trouve également une problématisation de la principale notion traitée par le cours. Toutefois, contrairement au MOOC "Introduction aux droits de l'homme", où il y avait une référence explicite à d'autres langues, citées en mention, ici le plurilinguisme est présent implicitement. La réflexion proposée, même sans faire allusion à d'autres langues, est construite par rapport à d'autres langues. Autrement dit, si l'anglais a été choisi pour la transmission des savoirs, d'autres langues ont participé au processus de construction des connaissances.

- (8) **What do we mean by "drugs"?** The term "drug" is polysemous [...] **In this MOOC, we will assume the ambiguous meaning of the word "drug" in English** which perfectly reflects the fact that psychoactive substances can be both essential medications and potentially addictive and dangerous, both beneficial and harmful.
(MOOC "Drugs, drug use, drug policy and health", W1 V1, 0:31 et 2:01)



Fig. 5: Traduction et sous-titrage vers le français du MOOC "Drugs, drug use, drug policy and health"

On observe ainsi que le double sens du mot "drug" en anglais, qui signifie à la fois un médicament et une substance illicite, est à la base de la réflexion proposée dans ce MOOC. Ici, le choix de l'anglais se justifie dans la mesure où il apporte un autre regard à la notion et contribue à l'argumentation développée dans le cours. Le fait que les responsables du MOOC soient francophones corrobore à montrer que le plurilinguisme a été bien présent dans le processus de conception, même si le résultat final est monolingue (voir Nussbaum, Moore & Borràs 2013). A ce propos, le balisage "en anglais" n'est pas anodin et dévoile également une mise en relation implicite avec d'autres langues. Comme nous l'indique un des concepteurs de ce MOOC lors d'un entretien de recherche:

(9) Entretien avec les concepteurs du MOOC "Drugs, drug use, drug policy and health"

je crois que ce que nous aimons tous dans le mot *drugs* c'est son *ambiguïté en anglais* [...] puisque tout l'argument que nous développons euh c'est que l'héroïne et la diacétylmorphine c'est la même substance\ et que si vous le prenez dans le contexte de la prescription médicale c'est un médicament/ si vous le prenez dans la consommation de ri- rue/ c'est un cri- un crime au sens anglais du du du du terme\ un délit euh en français\ euh donc quelque part le mot *drug* en anglais aide [...] à la compréhension de cette *ambiguïté*\ en français/ médicament et drogue c'est totalement différent [...]

Les extraits que nous venons de présenter montrent que, loin d'être un simple "véhicule", la langue contribue à la construction des connaissances (même si la place des fonctions constitutives et véhiculaires du discours varie selon les

disciplines, voir Berthoud & Gajo 2016). Ce rôle constitutif est davantage mis en évidence dans des situations plurilingues: le passage par une autre langue fonctionne souvent comme une ressource supplémentaire pour l'élaboration des savoirs (Gajo 2013). Dans ce sens, la fonction constitutive du discours est à mettre en relation avec celle de re-médiation, où le préfixe "re" suggère une deuxième médiation proposée par le passage par une L2 (*id.*). Dans les exemples choisis, on a pu observer comment la fonction constitutive du discours intervient, en amont, dans la conception des MOOCs et est particulièrement visible dans les vidéos d'introduction, dans un souci de clarification des principales notions abordées par les cours.

4.3 Fonction indexicale

Comme nous l'avons vu plus haut, les activités de discours sont à la fois structurantes pour et structurées par le contexte. Le discours en tant que tel est moulé dans des genres et mis en circulation par des supports/canaux. Recourir à un MOOC renvoie déjà, en soi, à un genre particulier de discours formatif et à un support/canal appelant une certaine technologie. L'utilisateur d'un MOOC, par ses choix, s'oriente souvent vers certaines sources de savoirs et vers des voix autorisées, pouvant se faire entendre par-delà le contexte. Pourtant, les enseignants et les chercheurs adoptent toujours un point de vue et, si les effets de contextualisation n'affectent pas les savoirs de la même manière dans toutes les disciplines, ils interviennent toujours dans les stratégies d'exposition et, notamment, d'exemplification. Les concepteurs de MOOC en sont variablement conscients.

(10) Entretien avec les concepteurs du MOOC "Introduction au raisonnement clinique"

y a eu un autre mot qui nous a posé euh qui a eu des enjeux de ce type-là/ [...] dans l'entretien clinique c'est comment est-ce qu'on on parle du moment où le médecin pose des questions au patient\ et ce qui est très marrant c'est que en **france** y disent que c'est l'**interrogatoire**\ d'accord/ [...] c'est très connoté [...] donc nous c'est sûr qu'on allait pas utiliser ça/ nous **ici** on dit on fait l'**anamnèse**/ [...] mais en france y vont parler de l'interro- l'interrogatoire du patient\ nous ça nous fait a- alors là les québécois y SAUTENT au plafond [...] au **québec** y vont dire euh y vont dire euh **recueil de données par rapport au patient**/ un truc comme ça/ et pis bon alors nous on a pris anamnèse parce que finalement mais je peux je pense que y a une **connotation plus forte genevoise** parce que dans la manière dont le le mooc s'est constitué/

Cet extrait d'entretien avec un des concepteurs du MOOC "Introduction au raisonnement clinique" montre une conscience élevée de l'ancrage contextuel dans l'usage de la langue et de ses variétés. La manière de désigner l'entretien clinique avec le patient varie d'un point à l'autre de la francophonie et le format "MOOC" amène à interroger cette diversité, à identifier des connotations, directement liées au contexte. La diversité au sein de la langue (variation) opère un peu à la manière de la diversité entre les langues (plurilinguisme). Interroger cette diversité mobilise aussi la fonction constitutive, presque toujours impliquée par un travail sur la fonction indexicale. Cependant, les concepteurs/trices de

MOOCs finissent souvent par opérer un choix plutôt que de thématiser le problème. Ainsi, les fonctions indexicale et constitutive du discours ne transparaissent pas forcément dans le produit fini, mais l'ancrage contextuel demeure et pourrait déboucher sur un enseignement.

Dans ce même MOOC, le contexte occupe une place particulière, car il s'agit de former des praticiens en lien direct avec le terrain. Cela peut exercer une influence importante sur le contenu même de l'enseignement et, en particulier, sur le choix des exemples.

(11) Entretien avec les concepteurs du MOOC "Introduction au raisonnement clinique"

- 01 MCA: [...] on a parlé [avec les **collègues** de la faculté de médecine de **Marrakech**] de mon **mooc** comme une sorte d'outil de base qu'on peut utiliser pour une sorte de **langage commun** [...] et pis donc je leur montrais une démo/ des différentes méthodes dont on parlait tout à l'heure/ pis un moment donné **ils ont dit ouais mais y faut qu'on refasse nous des vidéos**\ j'ai dit parfait pourquoi/ ils ont dit mais parce que en fait ok **la faculté de médecine à marrakech c'est en français mais la plupart du temps avec les patients**
- 02 MF: c'est en arabe
- 03 MCA: c'est en arabe [...] et donc **y nous faudrait des vidéos en arabe parce que c'est pas la même chose**/ je leur ai dit ok je comprends bien mais pourquoi/ ils ont dit mais on parle pas pareil/ on dit pas les mêmes choses/ et **le patient y nous parle pas la même chose**\

La question de la variation se prolonge ici sous l'angle du plurilinguisme et de la pluriculturalité. Si le MOOC peut offrir un "outil de base", une sorte de socle commun, une partie de ses contenus ne peut être appliqué ni même compris dans certains contextes. Au-delà de la langue, l'interaction entre médecins et patients ne suit pas les mêmes routines au Maroc qu'en Suisse et il s'agit d'en tenir compte dans la formation. Un étudiant marocain remarquerait sans doute rapidement le caractère inapproprié de certains contenus, alors qu'un étudiant baignant dans la même culture que les concepteurs ne percevrait sans doute pas les effets de contextualisation, pourtant bien présents. Leur mise en visibilité peut intervenir dans le processus de didactisation des MOOCs.

La fonction indexicale intervient donc aussi bien dans la production que dans la réception. En prendre conscience demande toutefois un effort particulier, facilité par le travail en équipe internationale ou le recours à des modalités plurilingues. Cette prise de conscience peut admettre plusieurs degrés et répondre à diverses fonctions. Le premier degré concerne sans doute les contenus mêmes de l'enseignement et s'adosse à une fonction principalement didactique ou conceptuelle liée à la discipline en question, par exemple la médecine (voir extrait 11). Le deuxième degré concerne des choix terminologiques en lien avec la discipline ou le métier, comme dans l'extrait 9. Dans ce cas, la fonction est d'abord communicative et socioculturelle, dans la mesure où il s'agit d'être compris et, éventuellement, d'enseigner une certaine vigilance/adaptabilité métalinguistique pour des – futurs – praticiens appelés à parcourir le monde. Le troisième degré peine à devenir visible tant il est ancré dans des routines

sociocommunicatives et dans les pratiques linguistiques d'une communauté. Il concerne les images, les comparaisons, les expressions idiomatiques utilisées très largement et donc spontanément par les locuteurs (voir 10).

- (12) Les droits de l'homme, c'est comme le bon vin. C'est le grand cru, ce n'est pas le vin de table. (MOOC "Introduction aux droits de l'homme", W1V2, 12:35 – 12:41)
- (13) Au fond, sur le plan juridique, cette déclaration est un peu au monde des droits de l'homme ce qu'est la prise de la Bastille qui s'est déroulée quelque temps auparavant. (MOOC "Introduction aux droits de l'homme", W3V3, 4:30 – 4:41)

Dans ces extraits issus du MOOC "Introduction aux droits de l'homme", on recourt, à des fins de vulgarisation et d'explication, à des comparaisons. Dans une perspective pluriculturelle et internationale, le discours métaphorique peut opposer une certaine résistance à la compréhension et, dans tous les cas, fonctionne comme un marqueur contextuel. Ce marquage ne saurait être totalement gommé, mais il gagne à remonter à la conscience du locuteur, qui pourra davantage en maîtriser les effets rhétoriques et didactiques.

La mise en scène des registres, des variétés de langue et de la diversité des langues favorise une interrogation sur les relations entre discours et contextes. Elle permet aussi d'enseigner des savoirs tout en développant constamment une approche critique, n'omettant jamais l'effet des angles de vue sur les objets. Les fonctions constitutive et indexicale se présentent de manière complémentaire: la première se construit sur la dénotation, la seconde sur la connotation. Si les langues elles-mêmes nous placent face au monde (voir, par exemple, les valeurs différentes de "Mensch vs Mann" en allemand et de "homme" en français), le plurilinguisme nous aide à prendre conscience des effets de dénotation. Par la connotation, les locuteurs sont placés, se placent ou se déplacent dans des communautés et expriment des points de vue dont ils ne sont que partiellement les agents. Le débat au sein de la communauté large, linguistique (par exemple, la francophonie) ou professionnelle (par exemple, le monde médical), permet d'en comprendre les effets.

5. Discussion et conclusion

Le MOOC représente une modalité nouvelle d'enseignement et de diffusion du savoir. Il se base sur les technologies numériques qui n'en font pas qu'un nouveau support, mais un nouveau genre du discours didactique. Les mutations qu'il suppose doivent s'appuyer sur des expériences et des recherches émergentes. Dans notre projet, nous réfléchissons à un aspect particulier de ce nouveau genre, à savoir l'introduction d'un sous-titrage plurilingue dans les capsules vidéos.

Le plurilinguisme dans les MOOCs, via la mise à disposition de sous-titres dans plusieurs langues, interroge la sélection, l'exposition et l'appropriation des savoirs. La première question à se poser est de savoir si le sous-titrage plurilingue doit être au service d'un travail monolingue (rester dans sa langue

de confort) ou, au contraire, d'un travail plurilingue invitant les usagers à comparer plusieurs entrées sur le contenu et à percevoir la diversité des ancrages culturels du savoir. La prise de conscience des enjeux du plurilinguisme et de la pluriculturalité doit, dans un premier temps, atteindre les concepteurs de MOOC, variablement sensibles, suivant leurs disciplines, aux questions de langue et de culture. La sélection et l'exposition des savoirs peuvent s'en trouver ainsi rapidement impactées. Reste à travailler les modalités de fonctionnement des usagers pour qu'ils profitent pleinement de la perspective plurilingue qui s'ouvre à eux. C'est à cette condition qu'advientra une véritable didactisation du plurilinguisme et, en particulier, du sous-titrage dans les MOOCs. La technologie fonctionne alors non seulement comme outil, mais aussi comme révélateur des liens complexes entre discours et savoirs.

Comme nous l'avons vu à travers l'analyse de MOOCs sous-titrés et diffusés largement par l'Université de Genève à travers la plateforme Coursera, la question du plurilinguisme dans les MOOCs interpelle fondamentalement le projet d'un cours "massif", "ouvert" et "en ligne". Par définition, le MOOC enjambe les frontières institutionnelles, nationales, culturelles. Il prétend à une certaine "universalité" en s'adressant à un usager relativement abstrait, un "tout-usager". Or, la contextualisation reste importante, aussi bien dans la production que dans la réception des enseignements et des savoirs. Le plurilinguisme permet alors de valoriser une certaine polyphonie, de rétablir de la chair autour des voix, de comprendre d'où on parle et de mettre en comparaison des angles de vue.

BIBLIOGRAPHIE

- Altbach, P. (2014). MOOCs as neocolonialism: who controls knowledge? *International Higher Education*, 75, 5-7.
- Berthoud, A.-C. & Gajo, L. (2016). Le plurilinguisme comme révélateur de médiation et instrument de re-médiation. In L. Popesco & A. Neag Curea (éds.), *Discours en présence, Hommage à Liana Pop* (pp. 209-216). Presa Universitară Clujeană.
- Cassin, B. (éd.) (2004). *Vocabulaire européen des philosophies. Dictionnaire des intraduisibles*. Paris: Seuil/Le Robert.
- Cerquiglini, B. (2013). Le plurilinguisme en faveur de la science, *Synergies Europe*, 8, 11-17.
- Fonseca, M. & Gajo, L. (2020). Le plurilinguisme dans les MOOCs: profils d'usagers et fonctions du sous-titrage. *Apprentissage des langues et systèmes d'information et communication* [En ligne], 23(2). <https://doi.org/10.4000/alsic.4816>.
- Gajo, L. (2013). Le plurilinguisme dans et pour la science: enjeux d'un politique linguistique à l'université. *Synergies Europe*, 8, 97-109.
- Gajo, L., Grobet, A., Serra, C., Steffen, G., Müller, G. & Berthoud, A.-C. (2013). Plurilingualisms and knowledge construction in higher education. In A.-C. Berthoud, F. Grin, F. & G. Lüdi (éds.), *Exploring the dynamics of multilingualism. The DYLAN research project. Multilingualism and diversity management* (pp. 287-308). Amsterdam: John Benjamins.

- Gajo, L. & Steffen, G. (2015). Didactique du plurilinguisme et alternance de codes: le cas de l'enseignement bilingue précoce. *Canadian Modern Language Review / Revue canadienne des langues vivantes*, 71(4), 471-499.
- Berthoud, A.-C. & Gajo, L. (2020). *The multilingual challenge for the construction and transmission of scientific knowledge*. John Benjamins, Prestige Dylan Series.
- Kerr, R., Merciai, I. & Eradze, M. (2018). Addressing cultural and linguistic diversity in an online learning environment. *Educational Media International*, 55(4), 317-332.
- Lévy-Leblond, J.-M. (2013). La science au défi de la langue. *Synergies Europe*, 8, 19-28.
- Mangenot, F. (2017). *Formation en ligne et MOOC. Apprendre et se former avec le numérique*. Paris: Hachette.
- Nussbaum, L., Moore, E. & Borràs, E. (2013). Accomplish multilingualism through plurilingual activities. In A.-C. Berthoud, F. Grin, & G. Lüdi (éds.), *Exploring the dynamics of multilingualism. The DYLAN research project. Multilingualism and diversity management* (pp. 229-252). Amsterdam: John Benjamins.
- Phan, T. (2018). Instructional strategies that respond to global learners' needs in massive open online courses, *Online learning*, 22(2), 95-118.
- Stratton, C. & Grace, R. (2016). Exploring linguistic diversity of MOOCs: implications for International development. *Proceedings of the Association for Information Science and Technology*, [En ligne], 53(1). <https://asistdl.onlinelibrary.wiley.com/toc/23739231/2016/53/1>.
- Yanaprasart, P. (2018). Enjeux de la double médiation du plurilinguisme dans le milieu académique multilingue. Internationalisation ou diversité dans l'enseignement supérieur? *Recherches en didactique des langues et des cultures* [En ligne], 15(2). <http://journals.openedition.org/rdlc/2991>.

Annexe(s)

Conventions de transcription:

MCA, MF	abréviation des prénoms des concepteurs/enquêteur
/	intonation montante
\	intonation descendante
[...]	passages non transcrits
-	interruption
<i>drugs</i>	<i>micro</i> -alternance
SAUTENT	saillance perceptuelle

Overcoming language barriers and boundaries: Video-mediated eTandem

Tanya MCCARTHY

Kyoto University

Institute of Liberal Arts and Sciences

Yoshidanihonmatsucho, Sakyo-ku, Kyoto, Japan 606-8501

mccarthy.tanyamiranda.7n@kyoto-u.ac.jp

Matthew ARMSTRONG

Kyushu University

Faculty of Languages & Cultures

744 Motooka, Nishi-ku, Fukuoka, Japan 819-0395

armstrong@flc.kyushu-u.ac.jp

Ces dernières années, une initiative a été lancée dans les établissements d'enseignement supérieur du Japon pour faciliter les partenariats de recherche et la collaboration aux niveaux local et international. Les objectifs sont de fournir aux jeunes chercheurs une plate-forme pour perfectionner leurs compétences et promouvoir les échanges scientifiques. Cet article rend compte d'une initiative de recherche collaborative entre deux universités nationales au Japon qui ont utilisé une plate-forme *eTandem* à des fins d'échange scientifique. Les échanges en face à face et par vidéo ont permis aux étudiants de partager des idées de recherche, de travailler en réseau et de jeter les fondations possibles pour de futurs projets de collaboration. Les participants ($n = 26$) se sont rencontrés chaque semaine dans un environnement d'apprentissage informel. Des enregistrements audio de 13 dyades en tandem ont été collectés et analysés afin d'identifier les défis linguistiques et les échanges les plus significatifs. L'analyse des données a révélé que les étudiants étaient capables de donner et de recevoir des commentaires critiques sur divers aspects de leur apprentissage, en particulier dans les domaines suivants: la conception de posters, le contenu de la recherche, la présentation, la structure organisationnelle de la recherche scientifique et la terminologie hautement spécialisée. Nous avons noté que tout gain dans ces domaines variait de façon importante entre les individus et résultait peut-être davantage d'efforts multiples effectués au cours de la période plutôt que de l'expérience *eTandem* en tant que telle.

Mots clés:

Apprentissage en tandem, apprentissage collaboratif, autonomie de l'apprenant, anglais scientifique, environnement d'apprentissage informel.

Keywords:

Tandem language learning, collaborative learning, learner autonomy, SciTech English, informal learning environment.

1. Introduction

The Japan Society for the Promotion of Science (JSPS) has operated in Japan for the past 70 years as an independent institution that seeks to foster young researchers and promote scientific cooperation. In the 1990s, there was a significant increase in information exchange among young scientists (MEXT 1997) as global participation in joint research projects became part of the research culture. Since then, there have been greater calls for collaborative

partnerships to help young researchers to hone skills and promote scientific exchange. SciTech projects around the world, for example, have promoted scientific literacy (Oliveira et al. 2019), strong communicative skills in science (Scalice et al. 2019; Wang et al. 2018), and the creation of an international network of senior and junior researchers (Hastings et al. 2010). For these reasons, the researchers embarked on a joint research to facilitate a scientific Tandem Language Learning (TaLL) exchange program that would offer students a real-world opportunity to interact in an international setting with young researchers in various fields of study. This paper reports on this collaborative initiative.

TaLL has been defined as a collaborative, autonomous, and reciprocal method of language learning (Vassallo & Telles 2006). In its purest form, languages are kept separate and each interlocutor devotes equal amounts of time to learning the second language. Over the years however, TaLL has evolved with the digital transformation of second language education as well as the recognition of the necessity of encouraging more authentic language use in and/or outside of the formal learning environment. Email and video-based software have helped to build virtual bridges between international communities. Today, there is a plethora of online options available from various social networking sites, web conferencing tools, telecollaboration platforms and smartphones applications to create a virtual network which can expose second language learners to authentic language in wider contexts (O'Rourke 2007). Despite current and possibly future transformations of TaLL, it remains consistently grounded in the same basic principles: learner autonomy, language exchange and reciprocity.

1.1 Learner Autonomy

Learner autonomy, as one of the fundamental principles of the universities in which this research took place, was central to the TaLL program. "Self-teaching" or "self-learning" based on dialogue was promoted as the main teaching approach across all departments. Instructors were encouraged to act as guides to help students take a more proactive role in the global arena. Little's definition of learner autonomy as "a capacity – for detachment, critical reflection, decision-making and independent action" (1991: 4) meant for the tandem learner in this study, to show the ability to regulate learning, and become critically reflective on a metacognitive level.

1.2 Language Exchange

Traditional TaLL programs entail that learning partnerships exchange first languages. Tandem pairs function, in turns, as a model for their native language and then as a second language learner. As TaLL adapts to new and changing technologies, partnerships have now moved beyond second language exchange to international collaboration for professional training programs and for this study, scientific research exchange.

1.3 Reciprocity

Reciprocity, in essence, is the underlying tenet guiding the collaborative dimension of TaLL, giving it its strength. Tandem pairs must contribute equally to the relationship in terms of time, energy and interest. They must decide on learning objectives, tasks they would like to engage in, and how long, where and when they wish to meet. If objectives are not being met, tandem dyads will require the maturity to alter their learning approach for the partnership to remain mutually beneficial. Participants in this study were mostly master's and doctoral students, and as such, had the motivation, interest, and drive to participate in this mutual exchange.

With the central tenets decided, a TaLL model that would meet the requirements of a scientific exchange/intercultural program was designed based on an analysis of successful eTandem learning models.

2. eTandem learning models

eTandem learning models are widely used for educational purposes. However, its flexibility enables it to be used in wider pedagogical circles and contexts. A database created by *unicollaboration.org* showcases a wide range of cross-disciplinary telecollaboration and virtual exchange projects in various contexts which clearly demonstrates that eTandem is, and continues to be, a proven learning method that can be used for language learning purposes, increase intercultural awareness, and connect learners in distant geographical spheres. Aspects of the following programs informed decisions when designing the TaLL program.

Tian & Wang (2010) examined learner gains in language proficiency and intercultural understanding through the use of Skype. Their study was conducted with language learners in Australia and China, students taking on the role of "expert" and "learner" equally. The Soliya Connect Program (Elliott-Gower & Hill 2015) aimed to deepen intercultural understanding, but also widen perspectives on socio-political issues. This program was unique as shared knowledge of current issues was the focus rather than only language exchange. A separate teletandem model in Brazil (Leone & Telles 2016) similarly used VOIP tools (*Skype* and *Google Hangouts*) to engage students in out-of-class learning; however, what distinguished this program from others was the emphasis on inter-institutional integration and collaboration rather than as an individual endeavour. A new model appearing more and more these days, which especially intrigued the researchers, is the lingua franca model in which monolingual communication occurs between non-native speakers from various cultural backgrounds. This is becoming more commonplace in today's internationally oriented, social networking arenas, not only for intercultural exchanges, but also professional training programs focusing on content. It is

also the direction that this research took as the program got underway. Fig. 1 illustrates these current variations of eTandem in practice and the revisions made for the TaLL model used in this study.

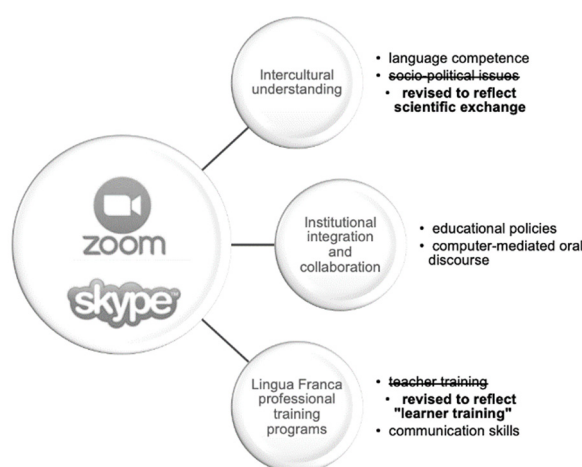


Fig. 1: eTandem Model variations which informed the study

3. Embedding the TaLL framework within the curriculum

The aim of this funded project¹ was to encourage Japanese and international students to participate in more out-of-class communication. Before undertaking such a large project, it was necessary to identify if there was a gap between current curricular goals and learner needs. Thus, a curriculum analysis was first carried out. It addressed three areas: policy, practice and process. Main weaknesses in the curriculum which became apparent were a lack of authenticity in language materials and transferability of academic skills for scientific research practices (McCarthy & Armstrong 2019). A learner needs analysis was then undertaken with the participants, to determine existing knowledge, abilities and research skills they wanted to improve. It was concluded from the analysis that the target learners had a high level of communicative competence to perform effectively in daily conversational and academic contexts. However, in research and professional environments, students considered their abilities to be inadequate. The gap, which the researchers thus aimed to fill, was to implement a program in which students could mutually benefit from sharing knowledge and experiences of research, networking, and conferencing. Accordingly, we set up the informal Learning Environment for Academics and Researchers through Networking (iLearn).

¹ Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan, Grant-in-Aid for Scientific Research (KAKENHI), project n° 19K00791/Facilitating Scientific Research and Intercultural Exchange through TaLL

3.1 iLearn

Four points emphasized in iLearn were as follows:

- linguistic competence
- knowledge-sharing for real-life contexts
- intercultural exchange
- connection to the university's educational philosophy and policies

Of the 26 participants in the program, 20% were Japanese. International students came from China, Vietnam, Cambodia, Korea, Bangladesh, Taiwan and Thailand. As such, students were able to improve linguistic competencies as well as increase awareness of different cultures through communication about research. The iLearn model followed current trends in educational practices of synergizing formal classroom learning and informal out-of-class learning (such as blended or self-access language learning), but with the language of scientific research being the focus. In this sense, the iLearn program is unique among typical TaLL projects conducted in other settings. Fig. 2 illustrates the organizational components of iLearn.

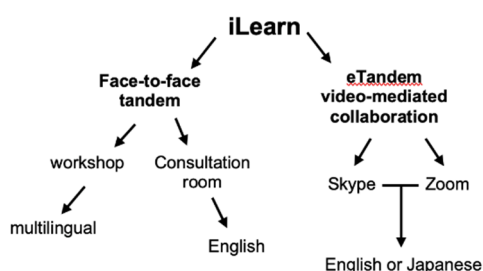


Fig. 2: iLearn eTandem model

A short, intensive 4-day, pre-tandem workshop was initially held to help students practice networking, discussion and research skills for poster and oral presentations. The workshop also helped students to bond and begin to form tandem pairs naturally. The morning session of the workshop introduced specific research skills students felt they lacked when in real-world situations. Students were paired off in face-to-face tandems in the afternoon sessions to connect theory with practice. The workshop culminated in a simulated poster session for an international conference, complete with a networking event. A consultation room was reserved for ten weeks to provide a space for face-to-face tandem partnerships for those who preferred this method to video meetings. Several computer-mediated communication studies have found that choosing interesting conversation topics over the duration of a tandem program is one of the challenges in maintaining successful tandem dyads (Beckett & Slater 2005; Okdie et al. 2011; Schwienhorst 2010). Thus, students were asked to pick weekly themes for their discussion on the basis of a list of research-

based topics². To ensure students regained their agency as learners, organizing a list of research-based topics in advance ensured that there was always a bank of subject matter of mutual interest. For most students, the preferred tandem option was video-mediated collaboration, as there were few chances to meet face-to-face due to conflicting schedules, lab experiments, report deadlines, and language classes. *Skype* or *Zoom* were used for synchronous video conferencing, as it allowed these students to interact in real time through screen sharing technology. Asynchronous communication via *LINE*³ was used by students to make quick decisions regarding scheduling and learning content for video conferencing.

4. Data collection and analysis

Data were collected by two methods. First, a descriptive case study design explored linguistic competence in using research language; and second, a grounded theory approach analysed meaningful discourse in dialogic exchanges.

4.1 A case-study of a tandem dyad

The tandem partnership between Learner A (a Japanese nurse-in-training) and Learner B (a Japanese doctoral student) who remained partners for the entire 10 weeks provided qualitative data. Tandem sessions lasted 45 minutes, during lunch time, when the teacher and students were available. Students met face-to-face four times and exchanges via *Skype* were held for the other six meetings. Learner B used the consultation room each week to set the audio recorder and the researcher collected the audio recording after the meetings. For this study, it was decided not to use video-based recordings, as it was seen as intrusive (Denzin & Lincoln 2011) and the researcher did not want the tandem pair to monitor their actions or language output. Direct observational research was also dismissed, as the researcher did not wish for students to interrupt their conversation and ask for assistance. It was of primary importance that the learners were given freedom and responsibility without interference from the researchers. An audio recorder was thus placed on a small table in the room or under the display (Fig. 3). In some cases, when students were working out how

² The following themes were negotiated between the students: 1. Reporting on current issues and needs; 2. Answering basic research questions for both general and specific audiences; 3. Describing the rationale for their chosen methodology; 4. Designing a research poster and analysing current posters for international conferences; 5. Describing data collection and analysis; 6. Describing eye-catching design elements and communication techniques for effective poster presentations; 7. Preparing for Q&A: specifically, challenging questions; 8. Practicing for presentation: Feedback and reflection-on-practice; 9. Job-hunting and interviewing for a professional position; 10. Publishing research papers and writing proposals. Themes were seen as guidelines which could be changed at any given moment.

³ LINE is a freeware application for instant communication on electronic devices. Users can exchange texts, images, video and audio, as well as conduct free video conferences.

to share documents and screens, jostling papers or moving around to practice a presentation, the quality of the recordings suffered. Therefore, short sections of these particular recordings were not transcribed.



Fig. 3: Tandem consultation rooms

4.1.1 Analysis of linguistic choices in three excerpts

Excerpt 1: Meeting 2

Learner A is applying for an international position and needs help with her CV.

Learner B is a doctoral student and between his master's and doctoral studies, he gained work experience as an IT engineer.

- A: I will do apply job for CDC.
 B: This is important.
 A: Yes, kind of. But, it will be failed, but I wanna challenge.
 B: Why do you say it will fail?
 A: Because lack of skill, especially in language.
 B: What do you need to do?
 A: Well, the program is on polio. I need...how do you say in English? [...silence... she looks up the word in a dictionary...] proficiency skills and doctor's or nurse's license is in....uhmm... public health? Uhm, certification? License is by practice in hospital. And additional language skill.
 B: Ah, you need to have a certificate but you don't have work experience?
 A: A little. I need communication skill and leadership skill
 B: So your, uhm, rirekisho⁴...
 A: CV
 B: CV?
 A: Yes, in English CV.
 B: What CV means?
 A: uhm.... [laughs] CV.
 B: [laughs]
 ---- [break in excerpt] ----
 B: Where is your CV?
 A: I can send it. I am thinking now of communication skill, like language
 B: What is your language skill?
 A: I have English, Japanese and a little bit Amharic.
 B: Amharic? What is Amharic?
 A: I worked in Ethiopia. It's communication for daily life. It's the Ethiopian focal language.
 B: ehhe? sugoi⁵

⁴ Rirekisho is the Japanese word for "Curriculum Vitae (CV)".

⁵ Sugoi is the Japanese word for "That's great!"

In the pre-TaLL workshop, Learners A and B demonstrated an ability to express themselves on a range of general topics, but once the conversation switched from familiar topics to specialized lexis, they both begin to struggle.

At the beginning of this excerpt, Learner A does not show a controlled usage of lexis or syntax whereas Learner B can provide immediate corrective feedback through usage of the correct linguistic form in question formation and verb tense.

A: I will do apply job for CDC.

B: This is important.

A: ...it will be failed, but I wanna challenge.

B: Why do you say it will fail?

As the conversation continues, Learner B is able to reconstruct the language exchange coherently and clarify meaning by summarizing what Learner A is trying to explain.

A: Well, the program is on polio. I need...how do you say in English? [...silence... she looks up the word in a dictionary...] proficiency skills and doctor's or nurse's license is in....uhmm... public health? Uhm, certification? License is by practice in hospital. And additional language skill.

B: Ah, you need to have a certificate but you don't have work experience?

What is notable in this dialogic exchange is that although Learner B does not explicitly tell Learner A her mistakes, he is able to give indirect feedback. Recast as corrective feedback is a frequently employed strategy used by teachers to show the correct form without explicit identification or explanation of the error (see, for example, Ellis & Sheen 2006). As the topic becomes even more specialized for a business English context, they both begin to struggle with finding the precise words for effective communication.

A: A little. I need communication skill and leadership skill

B: So your, uhm, rirekisho...

A: CV

B: CV?

A: Yes, in English, CV.

B: What CV means?

A: uhm.... [laughs] CV.

B: [laughs]

Learner B, for the first time, resorts to Japanese when he does not know the English vocabulary. Learner A responds this time with the direct translation, but she does not expand further. This became a meaningful learning moment; however, neither student considered using a dictionary. After listening to the recording, in the following week, the researcher was able to suggest that they record unknown words in a notebook to improve ability to express ideas accurately not only within their specific fields of study, but also for professional purposes. As the researchers' role was that of facilitator, any advice or feedback was given solely through suggestions of learning techniques, or specific actions that could help them to understand the connection to a real-life context.

At the end of each session, learners were encouraged to prepare for the next session by making a plan of action. At the end of meeting 2, they decide to exchange CVs to discuss further in meeting 3. As they wind down, Learner A mentions that one of her additional languages is Amharic.

B: Where is your CV?

A: I can send it. I am thinking now of communication skill, like language

B: What is your language skill?

A: I have English, Japanese and a little bit Amharic.

B: Amharic? What is Amharic?

A: I worked in Ethiopia. Its communication for daily life. It's the Ethiopian focal language.

"Amharic" is specialized vocabulary for even native speakers of English, but Learner A could easily use this word without hesitation. This implied that she was already familiar with using it in a particular context. For the researchers, this clearly indicated that learning specialized lexis for research and professional contexts is essential in how effectively learners are able to communicate in a real-world situation.

Excerpt 2: Meeting 3

Learner A and B have sent copies of their CVs to each other after their meeting, and in the following session they discuss it further.

A: This CV is so good. It have many experiences.

B: My background is in IT development, so maybe it is different from yours. I have many years of experience working in a company, writing programs, so I write about these experiences here.

A: Your CV have many experiences and networking with different companies.

B: My strength on my CV is my work experience. I am not working now. I worked before I started doctor's program. When I worked for some companies as an engineer, I just concentrate on working. I didn't, uhm, didn't focus so much on another thing like networking. At the time, I didn't want to have anything else. How about yours?

A: So, I have worked in Africa, in Ethiopia, training nurses in how to deal with HIV and sexual education programs. I worked in so many areas, in observing these, so I have some ideas about sexual education and health, but it is not public health. It is private activities. The job is public health.

B: But these experiences are connected to your communication skills?

A: Yes, I have consultation with patients. But I cannot write detail. Around medical area, doctors are so sensitive, the topic, so they can't show, uhm, can't say sensitive topics, but medical research want to show the beginning of the research, so they have to register the planning. Sometimes, doctors can't say what they are doing. They have to finish the research, be careful not to give ideas. Not so much sharing.

B: uhm, so you cannot write your work in detail, but you can write your experience. Your CV does not say about the type of participation experience. And you say last time language. You should write language skill because you cannot say communication part in detail.

A: Yes, I will finish now. Thank you for your CV. I have to submit tomorrow and then I will wait if I get email for interview.

B: Good luck!

A: Thank you. You went to conference last week? How was it?

There is a noticeable difference in the flow of conversation between Learners A and B in the following week. Whereas Excerpt 1 was filled with extended silences, and short sentences, in Excerpt 2, Learner A's lexical and grammatical choices are more accurate, the sentences are longer, and the rate of speech has increased which indicate a level of developmental progress. Learner B

continues to ask Learner A open-ended and follow-up questions which helps to elicit longer, more purposeful and meaningful responses.

B: ...How about yours?

B: But these experiences are connected to your communication skills?

This language-building technique is another strategy that Learner B uses effortlessly, which shows his ability as a skilled communicator. It is interesting that without teacher guidance, the learners are able to use effective conversational skills naturally. It is apparent from the exchanges that they have both read over each other's CVs carefully and have the necessary vocabulary to express themselves more fluently and spontaneously. For the researchers, this was seen as strong evidence of the need for learners to prepare in advance for discussions in academic, professional and/or research contexts, an essential component of tandem programs (Batardière & Jeanneau 2020; Brammerts 1996).

Excerpt 3: Meeting 10

Learner B talks about current research in his field and the conversation with Learner A turns into a discussion from the viewpoints of researchers in two very different fields.

B: Well, this company makes some robots which resemble creatures. This person, the creator develops the robot and kicks the robot and the robot tries to stand up again and again. So, I am writing language for AI programs and I am interested in learning how to write for this kind communication between human and robot. For me, there is a very close relation between human and artificial intelligence. That is a good point, however my research, 80% of my research is data analysis. My research needs more... I also interested in human, human, humanity or...

A: There was a humanity conference last week here, I couldn't attend, for medical and social research, but not include your area. They share information on newest trends. I think there should be collaborate with humans, mixed with AI...

B: Computer Science and AI and...

A: Yes, I feel some dangerous things. You have scope through artificial, and I have scope through human, but human already affected in good and bad communication way, actions changing, feelings changing, thinking changing, so AI or other artificial things, nothing to think... Do you understand? [laughs] Humanity. AI should understand more about this side. Nature, emotions.

B: Well in my research, more than 80% I focus on the computer science, but some researchers develop dialogue system and consider the communication between human and computers, so they must think about how humans behave when they hear some utterance uhm, speech system. So maybe we should collaborate between fields more.

Linguistically, the language in the final meeting is quite different from the earlier meetings particularly with turn-taking. In collaborative discourse, knowing when to speak without overlapping or interrupting, is essential for successful and meaningful conversational exchanges. This final excerpt provides an excellent example of Learners A and B's ability to discuss their research. Learner B has critically evaluated his current doctoral research and feels that it is lacking in an area in which he has little knowledge. Learner A is more specialized in this field and so he asks for her opinion. This conversation is the most challenging of all recorded sessions, as both students struggle throughout to switch lexis from a specific audience to a general audience.

Learner B opens with:

B: my research, 80% of my research is data analysis. My research needs more... I also interested in human, human, humanity or...

Immediately, he begins to struggle with lexis outside of his computing field of study. As his voice trails off twice, the conversation is picked up smoothly and naturally by Learner A whose expertise is in behavioural sciences:

A: There was a humanity conference last week here, I couldn't attend, for medical and social research, but not include your area. They share information on newest trends.

B: Computer Science and AI and...

A: ...I feel some dangerous things. You have scope through artificial, and I have scope through human, but human already affected in good and bad communication way, actions changing, feelings changing, thinking changing, so AI or other artificial things, nothing to think... Do you understand? [laughs]

As Learner A tries to connect human intelligence with that of AI, she also begins to falter as she also does not have the necessary lexis to discuss another field. She laughingly asks, "Do you understand?". Whereas in the first recordings, Learner A depended more on her dictionary, she now attempts to convey her meaning by using different words. She tries again to express her ideas using more general words that Learner B can connect with:

A: ...Humanity. AI should understand more about this side. Nature, emotions...

Learner B picks up on her cues and also tries to change his lexis. As can be seen from the example below, he adjusts his vocabulary from "dialogue system" to more comprehensible language, "utterance" and finally to "speech system"

... some researchers develop dialogue system and consider the communication between human and computers, so they must think about how humans behave when they hear some utterance uhm, speech system.

In the end, both Learner A and B seem to agree that there should be more interdisciplinary collaborative research.

B: So maybe we should collaborate between fields more.

A: I think there should be collaborate with humans, mixed with AI...

In the pre-tandem learner need's analysis, students had noted that speaking to professors about their research was not difficult because professors understood all the various complexities of their research, as well as the subject area in depth. However, when giving presentations to peers or lab members and more general audiences at international conferences, they found it extremely difficult to convey their message effectively. In this sense, TaLL helped these learners to not only improve conversational exchanges through natural turn-taking and adjusting lexis from specific to a general audience; but also, more importantly, they could understand the benefit of mutual collaboration or exchanges with researchers in other fields.

4.2 Meaningful exchanges

For the researchers, engaging in meaningful exchanges signified an exchange of knowledge or critical reflection on learning which could be of mutual benefit to students in a real-life research context (feedback or advice), rather than simply an exchange of information (greetings or sharing of personal information). Through qualitative coding (see Appendix 1), patterns in the data were found by identifying recurring themes (Urquhart 2013). The researcher used the categories to interpret the transcript data of the tandem partners. Initial coding categories and numerical density of each category are shown as Tab. 1.

Meaningful Exchange Category	Density
1. Design	18
2. Content	17
3. Delivery	12
4. Organizational structure	05
5. Lexis	02
TOTAL	54

Tab. 1: Coding categories and numerical density

4.2.1 Design

The majority of meaningful exchanges was connected to poster design, possibly as a result of the time spent on this area during the initial workshop. Students in post-graduate programs were expected to present their research at domestic and international conferences, and for some, the workshop was their first experience in designing scientific posters. As students could share screens and/or documents, they had more time to critically evaluate posters than other aspects of their research. Exchanges were focused mostly on poster layout and readability. The following are examples of meaningful exchanges in the design category:

I cannot read. This is too big and it is not useful. You do not need to put your questionnaire on the poster. This information is not helping anyone.

Your poster's heading says, "Research Question". This is not a question it is a statement. Sometimes in a poster session, the presenter is not there but the poster is up. They come to read the poster, so you cannot explain.

Key information should be big. This font size is too small for a poster. Bigger than 12.

4.2.2 Content

Having detailed knowledge of the specific content-area is essential for the researcher to clearly and confidently explain and expand upon the research topic, especially during Q&A sessions. This was the second area of density as students asked for clarification and expansion of specific points. Discussions were mostly concerned with presentation of data and students tended to be

especially critical of data collection methods and analysis. The following are examples of meaningful exchanges:

What is the number of subjects, uhm participants in your study? Do you think that this is enough? What is your idea?

There should be a comparison of results. I think this pie graph shows the rate, but I think you need another pie graph. So, one more pie graph. Young people's gender gap is this website, so if you want to discuss more the gender gap in another age, you need another age data. So, is there any difference? This information says, "the global gender gap report in 2018 shows that the index of gender gap better for fields economic education, health and politics". This index is better than the pie graph.

4.2.3 Delivery

There were fewer instances of feedback on delivery, possibly because of the limitations of video-mediated communication. In face-to-face situations, students are able to see a wider perspective, but tandem partners could only see faces. Delivering an effective virtual presentation was not a skill covered during the workshop. Learner exchanges focused mainly on time-management, speech rate, and attitude and less so with socially constructed rules of interaction (eye contact and facial or hand gestures). The following are examples of meaningful exchanges:

The method, result, discussion took too long. You are saying too much information. Only important points in the discussion.

"As you can see from above" is too casual. You should talk about the important data.

Your character should be so more bright. Make sure that you have eye contact and face your audience.

4.2.4 Organizational structure

The structure of scientific presentations is an integral part of how well research is received by an audience. There was not much feedback on structure as students had several years of experience giving academic presentations and writing academic essays in their undergraduate language and content-based courses. The following are examples of meaningful exchanges:

You didn't write an introduction on the poster. You need an introduction and also references since you have a source here. You do not need to put your research questionnaire on your poster. Where is your hypothesis?

The method is too long. Balance the presentation better.

4.2.5 Lexis

Before giving a presentation or writing a paper, it is important to understand the type of audience who will be attending the talk (whether general or specific) and use appropriate lexis accordingly. There were however, few instances in the recordings where students gave corrective feedback directly. This is possibly because the students communicated at a conversational level with their tandem partner for the most part using informal or academic level lexis instead of highly technical vocabulary. The following is an example of a meaningful exchange:

The word "variance" will show the gap...uhm, variance means "bunsan" in Japanese. Do you know? Variance is degree of distance with average, so if the variance is low number and average is individual and short. And if the variance is bigger, then average to individual score is low. So, I think that is a better word.

The feedback classification system constructed here, intended to illustrate meaningful exchanges found in the data. For future research in this area, more precise definitions of each category will be required to produce a clearer understanding and interpretation of the dialogic exchanges.

5. Discussion

This study was carried out in three distinct phases: 1) pre-TaLL; 2) TaLL implementation; and 3) post-TaLL. Each phase offered insights into how to improve the program. Each phase offered insights into how to improve the program.

5.1 pre-TaLL

Initially, research participants were asked about their motivations for joining a tandem program. Results of this open-ended question aimed to help tailor the program to suit learners' specific needs. Figure 4 illustrates that improving presentation and communication skills were the main motivators for participating in the project.

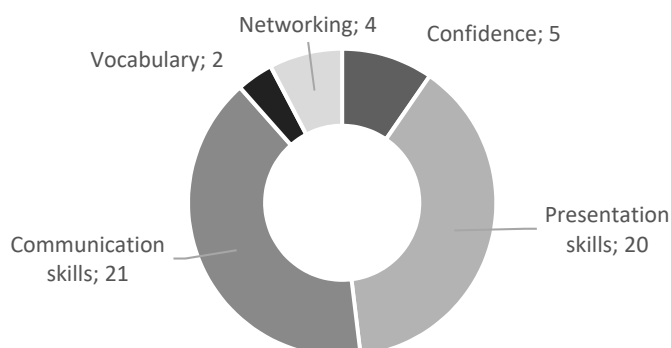


Fig. 4: What is your motivation for wanting to participate in a tandem program?

For many of these students, the intercultural aspect of communicating with international students was a big attraction. This has been a similar finding in several other studies over the years (Stickler & Emke 2011; Tardieu & Horgues 2019; Woodin 2001, 2003). A second survey (Tab. 2) sought to determine student's feelings about presentation skills, cultural awareness and confidence.

Self-evaluation	Pre-tandem	
	Yes	No
1. I understand the conventions of speaking in an academic context	33%	67%
2. I can express my ideas clearly	40%	60%
3. I can use appropriate vocabulary when discussing my research	33%	67%
4. I can plan and present a poster in my subject area	47%	53%
5. I can speak to a general audience using only notes or my poster as a guide	47%	53%
6. I can communicate with cultural sensitivity and awareness	71%	29%
7. Overall, I feel confident that in the future, I can give a poster presentation	33%	67%

(n=46)

Tab. 2: Pre-TaLL survey results

The results clearly indicate learners' lack of knowledge about conventions in academic and research contexts. Only in the area of intercultural awareness did they feel confident. This was mainly due to the fact that in the workshops, 80% of the participants were international students who were already comfortable conversing with students of various nationalities. Japanese students who participated in the workshop stated that in their research labs, there were typically one or two international students. The survey was given to students again at the end of the program to determine if there were any areas of improvement.

5.2 TaLL-in-Practice

During the program, students met directly in the consultation rooms or via video-based platforms. Poster design, research content, delivery, organizational structure of research presentation and specialized lexis were the five categories students focused on during their meetings. The lack of focus on syntax was initially surprising, as Japanese students typically tend to focus on form over content (Fujiwara 2018). However, on reflection, as the participants shared a similar status of researcher at one of Japan's top research institutions, it seemed more likely that they would focus their attention on the research content.

5.2.1 TaLL successes

The TaLL project can be regarded as somewhat successful in a few ways. First, iLearn attracted more students than expected, which meant that there was, without a doubt, a need for more support for an informal research environment where students could practice presentation, conference and networking skills with students across various disciplines. Having authentic communication opportunities was also a major benefit for the students as they could understand strengths and areas to improve, as well as develop relationships with students from different cultural backgrounds. Finally, the students could feel an increased sense of confidence as a result of being able to communicate comfortably in a real-life situation. Authenticity in tandem partnerships is an

essential component of tandem research projects (Brammerts 2003; Guth & Helm 2012) so that students can experience first-hand the intricate relationship connection between culture and language. Independent learning (Menezes De Oliveira 2011; Sasaki & Takeuchi 2011) and heightened intercultural awareness (Dodd 2001; Schwienhorst 2003) also played a role in helping to sustain student interest. These are some of the benefits that have been upheld as central to successful tandem partnerships.

5.2.2 TaLL challenges and strategies to overcome them

The first challenge, which is vital to the success of future programs, is to consider how to maintain tandem partnerships. Only one dyad stayed together continuously even though students had 'bonding' time during the initial workshop. The others met on and off as time and schedules allowed. As participation was voluntary, the students were free to make all decisions regarding participation. This disrupted the collection of audio recordings at times, which was the primary source of data. Solutions that proved effective for Learners A and B were setting achievable goals each week, actively preparing for follow-up sessions, and supporting their partner with their learning endeavours. A second challenge that students noted at the end of the program was that it was sometimes difficult to adjust language to suit a general audience when their oral presentations and written proposals were done mostly for their lab professors or specialist conferences. One solution the researchers considered was to pair lower proficiency students in similar fields at the start of a program. During the program, it was found that students tended to be able to converse more naturally and fluently with others in similar fields as they had deeper content knowledge and a wider range of lexis connected to similar research areas. More proficient learners were able to adapt with ease to an interdisciplinary situation. As such, the researchers will have to reconsider these aspects of the program design.

5.3 *Post-TaLL survey results*

In the final phase of iLearn, students were asked to once again complete the online survey that they had done in the pre-TaLL phase (Tab. 2). The purpose was to better understand student perceptions of the program in order to prepare for the next implementation. Tab. 3 shows the combined results of the survey.

Self-evaluation	Pre-tandem			Post-tandem		
	Yes	Just fine	No	Yes	Just fine	No
1. I understand the conventions of speaking in an academic context		33%	67%	93%	7%	
2. I can express my ideas clearly		40%	60%	60%	40%	
3. I can use appropriate vocabulary when discussing my research		33%	67%	20%	80%	
4. I can plan and present a poster in my subject area		47%	53%	93%	7%	
5. I can speak to a general audience using only notes or my poster as a guide		47%	53%	60%	40%	
6. I can communicate with cultural sensitivity and awareness		71%	29%	67%	33%	
7. Overall, I feel confident that in the future, I can give a poster presentation		33%	67%	100%		

(n=25)

Tab. 3: Comparison of Pre- and Post-TaLL survey results

Three salient points gleaned from the survey were:

1. Significant gains were documented in all areas as students became more familiar with the conventions of research presentations (especially, research organization, poster design, tone, and delivery).
2. Students' level of confidence about presenting in a real-life context in the future as well as interacting with others from a different culture was also considerably higher after 10 weeks of SciTech communication.
3. The ability for students to use appropriate lexis for general and specific audiences, although better, remained a challenge.

At the end of the program, an additional question was asked: What did you gain from your participation in the program? The two areas which students highlighted were:

1. An awareness of weaknesses in specific areas (such as English communication skill; ensuring accuracy in research content; lack of confidence and feelings of anxiety before giving a presentation).
2. Practical tips that they could use in a real-life research scenario (such as designing an eye-catching poster; enjoying the research experience; using presentation skills learned during the project in their labs; responding to research questions; having the opportunity to present research to people in different fields and countries; learning different points of view from international students; and considering collaborative projects with students in a different field of study).

6. Conclusion

The researchers implemented an eTandem program for research students as a result of spending many "off-the-clock" hours supporting them with submitting research proposals, writing abstracts, and preparing for conference

presentations. Without specific courses in the curriculum to meet this need, students were dependent on lab professors for guidance. For English presentation skills, it was not always easy for some Japanese professors to give adequate support due, at times, to language or time constraints. The TaLL proposal was, at first, met with scepticism, as it was thought that there would not be many participants, especially as it was launched during the summer vacation. However, as the enrolment numbers climbed, the researchers felt confident that the project would be successful. Post-tandem results showed that although there were challenges that needed to be overcome, the successes were significant enough to widen the scope of the project to a collaborative effort with another university. Student gains were evident in several areas over the 10 weeks, but it was also carefully noted that any increase in confidence and linguistic skills was subject to significant individual variability and came as a result of multiple efforts over the period rather than through the TaLL experience specifically.

eTandem is not without its set-backs and future programs will require a rethinking of tandem partnerships as well as other logistical matters; however, the results of this short study indicate that TaLL can contribute to confidence and enjoyment in SciTech language exchange, an awareness of the need to ensure accuracy when reporting research, an increase in intercultural awareness and most of all, the consideration of extending research networks not only to members of the immediate research community, but to labs within other faculties and even beyond the university walls.

REFERENCES

- Batardière, M. T. & Jeanneau, C. (2020). Towards developing tandem learning in formal language education [Formaliser les dispositifs d'apprentissage en tandem dans l'enseignement des langues]. *Recherche et pratiques pédagogiques en langues de spécialité. [Research and Teaching Languages for Specific Purposes]*, 39(1).
- Beckett, G. H. & Slater, T. (2005). The project framework: A tool for language, content, and skills integration. *ELT Journal*, 59(2), 108-116.
- Brammerts, H. (1996). Tandem language learning via the Internet and the international e-Mail tandem network. In D. Little & H. Brammerts (eds.), *A guide to language learning in tandem via the Internet* (pp. 9-22). Centre for Language and Communication Studies.
- Brammerts, H. (2003). Autonomous language learning in tandem: The development of a concept. In L. Lewis & L. Walker (eds.), *Autonomous Language Learning in Tandem* (pp. 27-36). Sheffield: Academic Electronic Publications.
- Denzin, N. K., & Lincoln, Y. S. (2011). *The SAGE handbook of qualitative research*. Thousand Oaks: Sage.
- Dodd, C. (2001). Working in tandem: An Anglo-French project. In M. Byram, A. Nichols & D. Stevens (eds.), *Developing intercultural competence in practice* (pp. 162-175). Clevedon, UK: Multilingual Matters.

- Elliott-Gower, S. & Hill, K.W. (2015). *The Soliya Connect Program: Two institutions' experience with virtual intercultural communication*. Gower Georgia College.
- Ellis, R. & Sheen, Y. (2006). Reexamining the role of recasts in second language acquisition. *Studies in Second Language Acquisition*, 28(4), 575-600.
- Fujiwara, Y. (2018). The role of grammar instruction in Japanese EFL context: Towards communicative language teaching. *JAS4QoL*, 4(4), 1-11.
- Guth, S. & Helm, F. (2012). Teaching culture through CALL. In A. Arnold & L. Ducate (eds.), *Present and future promises of CALL: From theory and research to new directions in language teaching*. San Marcos, TX: CALICO
- Hastings, M. G., Kontak, R., Holloway, T., Marin-Spiotta, E., Steiner, A. L., Wiedinmyer, C., Adams, A. S., de Boer, A. M., Staudt, A. C. & Fiore, A. M. (2010). Networking skills as a career development practice: Lessons from the Earth Science Women's Network (ESWN). *American Geophysical Union*, Fall Meeting 2010, abstract id. ED13A-0599.
- Leone, P. & Telles, J. A. (2016). The teletandem network. In R. O'Dowd & T. Lewis (eds.), *Online intercultural exchange: Policy, pedagogy, practice* (pp. 241-247). New York: Routledge Taylor & Francis Group.
- Little, D. (1991). *Learner Autonomy 1: Definitions, Issues and Problems*. Dublin: Authentik.
- McCarthy, T. M. & Armstrong, M. I. (2019). Developing an informal tandem learning scheme for young researchers and academics. *The Asian Conference on Education 2019: Official Conference Proceedings*. <https://papers.iafor.org/submission54082/>
- Menezes De Oliveira E Paiva, V. L. (2011). Identity, motivation and autonomy in second language acquisition from the perspective of complex adaptive systems. In G. Murray, X. (A.) Gao & T. Lamb (eds.), *Identity, motivation and autonomy in language learning* (pp. 57-72). Bristol: Multilingual Matters
- MEXT (1997). White Paper: Japanese government policies in education, science, sports and culture 1997. Scientific Research Chapter 2 Section 3.1. https://www.mext.go.jp/b_menu/hakusho/html/hpae199701/hpae199701_2_027.html
- Okdie, B.M., Guadagno, R. E., Bernieri, F. J., Geers, A. L. & Mclarney-Vesotski, A. R. (2011). Getting to know you: Face-to-face versus online interactions. *Computers in Human Behavior*, 27, 153-159.
- Oliveira, L. M., Bonatelli, M. L. & Pinto, T. C. (2019). *DivulgaMicro*: A Brazilian initiative to empower early-career scientists with science communication skills. *Journal of Microbiology & Biology Education* 20(1).
- O'Rourke, B. (2007). Models of telecollaboration (1): eTandem. In R. O'Dowd (ed.), *Online intercultural exchange: An introduction for foreign language teachers* (pp. 41-61). Bristol, UK: Multilingual Matters.
- Sasaki, A. & Takeuchi, O. (2011). EFL students' metalinguistic awareness in e-mail tandem. In M. Levy, F. Blin, C. B. Siskin, & O. Takeuchi, *WorldCALL: International perspectives on computer-assisted language learning* (pp. 55-69). New York, NY: Routledge.
- Scalice, D., Dolci, W., Brochu, L., Merriman, T., Davis, H., Billings, L. & Voytek, M. A. (2019). FameLab USA: Improving science communication skills for early career scientists. *Astrobiology* 19(4). doi.org/10.1089/ast.2017.1809
- Schwienhorst, K. (2003). Neither here nor there? Learner autonomy and intercultural factors in CALL environments. In D. Palfreyman & R. C. Smith (eds.), *Learner autonomy across cultures: Language education perspectives* (pp. 164-180). Basingstoke, UK: Palgrave Macmillan.
- Schwienhorst, K. (2010). Learner autonomy and tandem learning: Putting principles into practice in synchronous and asynchronous telecommunications environments. *Computer Assisted Language Learning* 16(5), 427-443.

- Stickler, U. & Emke, M. (2011). LITERALIA: Towards developing intercultural maturity online. *Language Learning & Technology*, 15(1), 147-168.
- Tardieu, C. & Horgues, C. (2019). *Redefining tandem language and culture learning in higher education*. New York: Routledge.
- Tian, J. & Wang, Y. (2010). Taking language learning outside the classroom: Learners' perspectives of eTandem learning via Skype. *Innovation in Language Learning and Teaching*, 4(3), 181-197.
- Unicollaboration (2020). Sample projects. <https://www.unicollaboration.org/index.php/sample-projects/>
- Urquhart, C. (2013). *Grounded theory for qualitative research: A practical guide*. London: Sage Publishing.
- Vassallo, M. L. & Telles, J. A. (2006). Foreign language learning in-tandem. Part I: A critical review of its theoretical and practical principles. *The ESPecialist, São Paulo*, 27(1), 83-118.
- Wang, J., Power, C., Kahler, C., Lyras, D., Young, P., Iredell, J. & Robins-Browne, R. (2018). Communication ambassadors—an Australian social media initiative to develop communication skills in early career scientists. *Journal of Microbiology and Biology Education* 19(1).
- Woodin, J. (2001). Tandem learning as an intercultural activity. In M. Byram & D. Stevens (eds.), *Developing intercultural competence in practice* (pp. 189-202). Bristol, UK: Multilingual Matters.
- Woodin, J. (2003). Encouraging intercultural competence in tandem learning. In T. Lewis & L. Walker (eds.), *Autonomous language learning in tandem* (pp. 71-78). Sheffield, UK: Academy

Appendix

Coding descriptors

Design	Clear articulation of research problem, purpose or focus, including appropriate-sized text and bold headings, as well as eye-catching visuals (e.g., photos, figures or tables)
Content	Having an appropriate content level for a scientific presentation, including critical evaluation of specific aspects of scientific content: research purpose, data, methodology, results, conclusion, and research questions.
Delivery	Evaluation of speech rate, time-management, pragmatics, and register.
Organizational Structure	Well-structured, detailed and logical progression of research ideas: Introduction, Method, Results, Implications, and References.
Lexis	Use of appropriate and precise vocabulary for a specific or general audience.

Digitalisation des outils de référence: enjeux et perspectives pour l'enseignement de la traduction vers le français L2

Sara COTELLI KURETH

Université de Neuchâtel
Centre de langues
Avenue du 1^{er}-Mars 26, 2000 Neuchâtel, Suisse
sara.cotelli@unine.ch

Alain KAMBER

Université de Neuchâtel
Institut de langue et civilisation françaises
Faubourg de l'Hôpital 61-63, 2000 Neuchâtel, Suisse
alain.kamber@unine.ch

Many universities offer translation courses in their modern language curricula as part of L2 teaching. Each one defines different modalities in terms of the resources allowed for these exercises, especially during exams. In recent years, however, a growing number of online tools – monolingual and bilingual dictionaries, bi-concordances, terminology databases, machine translators and corpora – have become available to learners. Such an explosion in the offer and accessibility of tools is likely to impact both teaching/learning and the way in which acquired skills are tested.

This article presents the first results of an action research on the progressive introduction of digital tools in two L2 French translation courses. Two series of semi-structured interviews conducted with our students will first allow us to establish what tools are spontaneously used by translation students today. Secondly, these results will be compared to the real practices of the candidates during the examination validating the semester.

Keywords:

translation into L2, dictionaries, digital resources, learners' practices and representations.

Mots-clés:

traduction en L2, dictionnaires, ressources numériques, pratiques et représentations des apprenant-e-s.

1. Introduction

Longtemps, les spécialistes en lexicologie ont porté leur attention principalement sur la conception des dictionnaires (organisation du contenu, structure des articles, sélection des informations, tri et choix des entrées, etc.). Dans un passé plus récent, les maisons d'édition se sont également efforcées de créer des versions électroniques de dictionnaires existants, essayant de les rendre plus adaptées aux attentes du public, même s'il faut bien admettre que l'expérience a montré que les versions numérisées de certains dictionnaires n'apportaient pas forcément toujours une véritable plus-value par rapport à leurs versions papier d'origine.



Mais, accaparés par le produit lui-même, les lexicologues ont largement négligé le destinataire de leur travail: l'utilisateur des dictionnaires, que Wiegand (1977: 59) qualifiait de "célèbre inconnu" (ein "bekannter Unbekannter"), et dont on ne savait que peu de chose. On savait qu'il existait, certes, mais on ignorait complètement quelles étaient ses attentes, comment il procédait, quels besoins linguistiques ou langagiers il avait et/ou quelles difficultés il rencontrait dans la consultation des dictionnaires.

Au cours des dernières années, la recherche sur l'utilisation des dictionnaires s'est attachée à combler ces lacunes. Logiquement, elle s'est concentrée sur l'utilisation des dictionnaires en ligne (voir notamment Domínguez Vázquez, Mollica & Nied Curcio 2014a, 2014b; Müller-Spitzer 2014; Tiberius & Müller-Spitzer 2015). Les études récentes montrent en effet que les utilisatrices¹ privilégient les dictionnaires en ligne (cf. Domínguez Vázquez, Mirazo Balsa & Vidal Pérez 2013; Domínguez Vázquez & Valcárcel Riveiro 2015) et les applications (cf. Nied Curcio 2014, 2015), au détriment des dictionnaires papier, qui ont pour ainsi dire disparu des salles de classe même à l'université (voir Mollica 2020).

Mais malgré un certain nombre d'études parues ces dix dernières années, beaucoup reste à faire, raison pour laquelle nous avons décidé d'entamer une recherche-action (Macaire 2011) qui mettrait au centre les propos et les pratiques concrètes d'utilisatrices, apprenantes du français à l'Université de Neuchâtel, non seulement face aux dictionnaires – papier ou numériques – et aux applications, mais aussi face à toutes les ressources numériques qu'elles connaissent ou découvrent dans leur apprentissage (concordanciers, traducteurs automatiques et corpus notamment).

Après une brève présentation des modalités de notre recherche-action, nous nous intéresserons aux représentations et aux pratiques déclarées des utilisatrices des ressources par l'intermédiaire d'entretiens semi-directifs effectués avec nos étudiantes. Nous présenterons ensuite quelques exemples concrets d'utilisation des outils recueillis durant l'examen validant le semestre. Nous tenterons enfin de dresser une typologie des utilisatrices et de tirer quelques conclusions encore très provisoires de nos observations.

2. Présentation du projet

Notre idée de départ était d'établir un état des lieux des pratiques dans un type de cours dans lequel les apprenantes d'une L2 sont *a priori* appelées à utiliser intensément des outils numériques: la traduction. Nous nous sommes donc focalisés sur deux cours de traduction (allemand-français et anglais-français) dispensés comme cours à option à l'Institut de langue et civilisation françaises

¹ Afin de faciliter la lecture de cet article, nous utilisons le féminin générique pour désigner l'ensemble de nos étudiant-e-s. Il désigne aussi bien les femmes que les hommes.

de l'université de Neuchâtel. Partant du constat que le travail de traduction a beaucoup évolué suite à la mise à disposition gratuite, sur internet, de toute une série d'outils, partant également de l'hypothèse que la plupart de ces outils sont connus de nos étudiantes, qui certainement y recourent pour faire leurs devoirs, nous avons estimé qu'il était temps d'envisager de thématiser leur utilisation dans notre enseignement.

2.1 Les étapes du projet

Pour dresser l'inventaire des pratiques effectives des personnes concernées et de leurs représentations sur leur activité de traduction, et donc de décider d'un éventuel changement dans nos méthodes d'enseignement et d'évaluation, nous avons imaginé une procédure en trois étapes, correspondant à trois semestres de cours:

Phase 1:

- aucun changement dans la façon de dispenser les cours de traduction allemand-français et anglais-français
- aucun changement dans les modalités d'examen
- enquête sur les outils de référence utilisés par les apprenantes, par l'intermédiaire de questionnaires et d'entretiens semi-dirigés

Phase 2:

- aucun changement dans la façon de dispenser les cours de traduction allemand-français et anglais-français
- changements dans les modalités d'examen: possibilité d'utiliser les ressources disponibles en ligne en laboratoire multimédia
- enquête sur les outils de référence utilisés par les apprenantes, par l'intermédiaire de questionnaires et d'entretiens semi-dirigés

Phase 3:

- en classe, introduction d'activités qui présentent et permettent d'apprendre à utiliser les principaux outils en ligne pour la traduction
- entretiens semi-dirigés et capture d'écran lors de la traduction d'un court texte pour documenter la façon dont les étudiantes travaillent.

Les résultats présentés ici portent sur les phases 1 et 2, la phase 2 ayant été répétée sur deux semestres en raison du nombre restreint de participantes aux cours (N total = 26) et la phase 3 devant encore être mise en œuvre.

2.2 Les données récoltées

2.2.1 Entretiens semi-directifs

Nous avons effectué des entretiens d'une vingtaine de minutes avec 26 étudiantes entre mai et décembre 2019. Les questions portaient sur leur utilisation des ressources numériques et, plus généralement, sur le cours de traduction. Nous souhaitons récolter à la fois des informations sur notre

enseignement (quelques questions générales sur l'apport du cours et l'intérêt de la traduction dans l'apprentissage des langues) et sur la connaissance et l'utilisation des différents outils par les étudiantes interrogées.

Nous avons ensuite effectué une transcription orthographique des entretiens selon les modalités décrites en annexe. Les transcriptions ont également été anonymisées. C'est sur cette base que nous avons effectué l'analyse du contenu de ces entretiens, en nous focalisant sur les représentations quant aux outils de traduction qui s'en dégagent.

2.2.2 Captures d'écran

Les cours sont validés par un examen écrit de deux heures pour lequel les étudiantes n'avaient traditionnellement à disposition que les versions papier d'un dictionnaire bilingue et du *Petit Robert*. Pour la phase 2, nous avons modifié les modalités d'examen et ajouté à ces deux dictionnaires la consultation de toutes les ressources en ligne souhaitées. En revanche, les candidates devaient rendre leur travail sous forme manuscrite (ceci pour éviter les simples copiés-collés des traducteurs automatiques) et n'avaient pas droit aux smartphones.

Les examens ont eu lieu au laboratoire multimédia (LMM) de l'institut. Un logiciel de capture d'écran, OBS-Studio², nous a permis d'enregistrer les écrans des candidates. Le système informatique du LMM permet de visionner tous les écrans en même temps ou d'agrandir l'un ou l'autre écran. Nous avons ainsi eu la possibilité d'observer un écran spécifique de manière ciblée lorsque nous remarquions qu'il s'y passait quelque chose de particulièrement intéressant.

3. Représentations des utilisatrices

Un point s'est révélé particulièrement frappant lors de l'analyse des entretiens: il semble en effet que les représentations des étudiantes s'appuient principalement sur une opposition entre dictionnaire papier et outils numériques. En d'autres termes, elles font une distinction entre les supports mais ne décèlent aucune différence dans le contenu des outils numériques. Les outils *online*, disponibles de façon digitale, sont évidemment très diversifiés, mais ils sont présents comme un bloc monolithique dans les représentations des utilisatrices. Autrement dit, elles pensent les outils en termes de **support** et non pas en termes de **contenu**. À part deux spécificités propres aux dictionnaires papier, toutes les représentations que nous avons pu extraire des données s'articulent autour de cet axe.

² <https://obsproject.com/fr>

3.1 La fiabilité

Le premier critère pertinent tourne autour de la notion de fiabilité (voir à ce propos Müller-Spitzer & Koplenig 2014). Pour les utilisatrices, les dictionnaires papier sont plus fiables, plus "reliable", comme l'exprime ET21:

(1) ET21EN, 9:14-9:20

le dictionnaire papier je crois parce qu'il faut toujours avoir parce qu'ils sont le le plus exact le plus **reliable**/

Cet avis est largement partagé par ET3:

(2) ET3EN, 9:45-10:24

si tu es dans la rue et tu veux chercher un mot ou traduire un mot spécifique et tu es avec tes amis et tu ne sais pas comment traduire un mot d'anglais en français tu peux utiliser les outils numériques mais si je dois faire une VRAIE traduction une traduction BIEN je voulais utiliser le dictionnaire pas le . (...) non je pense que **la qualité est mieux** dans le dictionnaire

Notons d'abord que pour nos participantes en général, "le dictionnaire" fait toujours référence au dictionnaire papier traditionnel, qu'il soit bilingue ou monolingue. En termes de contenu, ET3 fait ici allusion à l'utilisation du dictionnaire pour trouver la signification d'un mot. Elle établit une hiérarchie entre dictionnaire papier et dictionnaire numérique en leur attribuant des contextes d'utilisation, formels et informels. Non seulement les deux outils sont différents en termes de qualité mais ils induisent aussi une qualité dans l'output qu'ils permettent: la traduction.

ET23 pour sa part émet des doutes quant à la fiabilité de *Leo* et *Pons*³:

(3) ET23DE, 6:00-6:17

moi j'ai j'ai aucune expérience et je sais que pons et leo ce sont pas des des dictionnaires euh **très exacts spécifiques** alors peut-être c'est bon pour hum apprendre comment utiliser ces outils-là en classe

Ce dernier commentaire est intéressant dans la mesure où les doutes exprimés par l'étudiante (cf. Müller-Spitzer, Koplenig & Töpel 2011) appellent une réponse de la part des professeurs ainsi qu'un travail en classe sur le maniement des outils.

3.2 La vitesse de consultation

Le deuxième point saillant est celui de la vitesse de travail. Il est relevé par une très grande majorité des utilisatrices. Chercher dans le dictionnaire papier prendrait beaucoup plus de temps que de faire appel à des outils en ligne. L'une des étudiantes mentionne même qu'elle n'a plus ouvert un dictionnaire depuis l'école primaire:

³ Deux dictionnaires bilingues disponibles gratuitement en ligne: <https://dict.leo.org/allemand-fran%C3%A7ais/> et <https://fr.pons.com/traduction>

(4) ET25DE, 5:10-5:24

mais ça prend aussi **plus de temps** . De chercher des mots dans le pons au lieu de écrire **vite** les quatre lettres en ligne (rire)

(5) ET20EN, 5:10-5:23

sans l'internet je vais utiliser le dictionnaire . et i don't know in french but it would **take ages** . looking for the word

De même, il y a quasi-unanimité pour dire qu'utiliser les outils numériques permet de travailler plus vite:

(6) ET18DE-EN, 7:45-7:58

ben on quand on utilise les outils numériques je pense qu'on travaille **plus vite**

(7) ET19DE, 6:24-6:26

ET19: je n'ai pas encore pensé de ça euh oui . Je pense j'utilise plus les choses online que les dictionnaires

INT: les dictionnaires papier

ET19: les livres oui parce que **c'est plus rapide pour moi**

3.3 La difficulté de consultation

Le troisième point relevé est en lien avec la difficulté de consultation. Chercher dans un dictionnaire papier demande plus de travail: il faut connaître l'ordre alphabétique et surtout savoir comment orthographier le mot, comme le dit ET16, "dans les livres on doit chercher la lettre d'abord ensuite la suite de des lettres et tout et tout", cela requiert donc "plus de travail". Alors que comme le montre la citation (10), même si on orthographie mal en ligne, l'outil cherchera quand même et, malgré l'erreur, fournira souvent le mot recherché.

(8) ET16EN, 4:39-5:00

sur internet on cherche le mot on on a le mot on essaie de la chercher mais euh on fait pas le travail nous-même alors ben par contre sur euh dans les livres on doit chercher la lettre d'abord ensuite la suite de des lettres et tout et tout il y a **il y a plus de travail avec les livres** que l'outil sur internet

(9) ET15EN, 8:50-9:05

à mon avis les les outils en ligne ils sont plus faciles que les outils euh des les livres et tout (...) **ils sont plus faciles** si si vous voulez chercher un mot c'est plus facile à de trouver

(10) ET24DE, 0:58-1:22

parce qu'on a vraiment l'habitude avec le portable ou l'ordinateur . Et aussi si tu écris pas juste le mot parfois ça trouve le mot comme **ça c'est vraiment facile** de trouver la traduction je pense

3.4 Le plaisir de l'objet

Finalement, nous avons trouvé deux arguments qui concernent uniquement les dictionnaires papier. Certaines étudiantes ont un plaisir qui est directement associé à la matérialité de l'objet livre, comme dans la citation (12), avec cette

étudiante qui aime "toucher les pages", mais aussi partir à la découverte d'autres mots ou de citations.

(11) ET11DE, 5:00-5:40

(sur le Petit Robert) c'est très pratique parce que bon ça prend plus de temps évidemment de chercher les mots mais bon pour moi **c'est aussi un petit bonheur** de: ça me fait **plaisir** euh de chercher dans le Petit Ro- de travailler avec cet outil là et puis je trouve quand même c'est pratique (...) puis il y a les citations donc euh ça peut faire **plaisir** on voit un peu l'emploi du mot

(12) ET16EN, 3:54-4:24

moi j'aime bien déjà j'aime bien euh avoir un livre que: une tablette ou bien le net pour pour faire euh: mes devoirs euh XXX

INT: donc vous préférez le livre à la tablette ou au natel .. pourquoi/

ET16: moi j'aime bien les livres (toux) pardon parce que je touche les pages et tout et tout et regarder aussi des autres mots que j'ai pas besoin mais peut-être je vais les utiliser quelque euh . quelque part mais euh: effectivement ils sont hyper bien moi je je les trouve bien.

3.5 *Éloge de la lenteur*

Le second argument favorable au format papier fait référence aux avantages cognitifs de la consultation du dictionnaire traditionnel en raison de la lenteur du processus:

(13) ET25DE, 7:59-8:32

mais je pense si on peut regarder dans le dictionnaire en livre c'est ... **on: réfléchit un peu plus** pour ce mot on si on parce qu'on le cherche et on l'a dans la tête et on réfléchit un peu plus et peut-être **ça reste un peu plus dans la tête** pour le garder et si on le fait à l'ordinateur je pense d'une part ça va très rapide

3.6 *Des avantages (théoriques) évidents*

Pour conclure, selon les représentations des utilisatrices, il faudrait plutôt utiliser un dictionnaire papier: il est fiable et précis, il procure du plaisir et permet un apprentissage en profondeur. Son seul défaut est que sa consultation prend du temps et qu'il est peu pratique à transporter. Les outils numériques quant à eux ont l'avantage d'être efficaces et rapides mais ils sont peu fiables et peu précis. Malgré ce plébiscite du dictionnaire papier, une utilisatrice prédit sa disparition:

(14) ET10DE, 9:08-9:10

on a que les ordinateurs et les portables et **les livres vont disparaître**

Dès lors, la question se pose: qu'en est-il des pratiques? Les points forts supposés du dictionnaire papier poussent-ils les apprenties-traductrices à les utiliser en priorité? Comme le laisse entendre cette dernière citation, il n'en est rien.

4. Pratiques auto-déclarées des utilisatrices

Le tableau 1 présente l'ensemble des outils que les apprenantes ont déclaré connaître et/ou utiliser, au cours des entretiens semi-directifs, pour faire leurs devoirs:

Type d'outil	Dictionnaires bilingues					Bi-concordanciers		Traducteurs automatiques			Dictionnaires monolingues					Corpus	Autres	
Outil	Dictionnaire papier	Pons phone	Leo	Langenscheidt	wordreference	Linguee	Reverso context	DeepL	Google translate	Reverso traduction	Petit Robert (papier)	Grand Robert online	Larousse	www.cnrtl.fr	Reverso conjugaison	Wortschatz Leipzig	Google	BtB terminium cooc.
ET1EN																		
ET2EN																		
ET3EN																		
ET4EN																		
ET5EN																		
ET6EN																		
ET7DE																		
ET8DE																		
ET9EN																		
ET10DE																		
ET11DE																		
ET12EN																		
ET13DE																		
ET14DE																		
ET15EN																		
ET16EN																		
ET17EN																		
ET18DE-EN																		
ET19DE																		
ET20EN																		
ET21EN																		
ET22DE																		
ET23DE																		
ET24DE																		
ET25DE																		
ET26DE																		

Tab. 1: Ressources utilisées pour les devoirs à domicile (gris foncé: outil connu et utilisé; gris clair: outil connu mais pas utilisé)

Seules 5 des personnes testées utilisent le *Petit Robert* pour leurs devoirs, une seule un dictionnaire bilingue papier. Malgré les vertus qu'on leur prête dans le

discours, les dictionnaires papier ne sont donc utilisés que de façon très anecdotique. Ce sont les dictionnaires en ligne qui sont le plus utilisés, par 20 étudiantes sur 26. Il est intéressant de noter qu'en majorité les apprenantes utilisent plusieurs outils, dont un dictionnaire bilingue. La moyenne se situe à un peu plus de 2 outils par personne mais le nombre de ressources mobilisées oscille entre 1 et 5 selon les cas. Ce qui étonne particulièrement, à la vue de ce tableau, c'est le fait qu'individuellement les étudiantes utilisent et même connaissent en fin de compte peu d'outils.

Finalement, nous avons recensé les différentes affirmations dans les entretiens qui permettent de répondre aux deux questions suivantes, qui s'avèrent fondamentales: pour quels motifs les personnes testées élisent-elles ou au contraire rejettent-elles un outil donné?

Il apparaît que 4 raisons jouent un rôle dans le choix d'un outil. Tout d'abord, **l'efficacité** est présentée comme prépondérante, surtout en relation avec le type d'informations que l'outil fournit. Cet argument est souvent évoqué pour les bi-concordanciers, dont les utilisatrices apprécient les connaissances qu'ils leur apportent sur le contexte d'utilisation. La **familiarité** semble aussi jouer un rôle important: on utilise de préférence les ressources dont on a l'habitude, et peu d'étudiantes sont friandes de nouveauté. On peut y voir une preuve dans l'utilisation du corpus *Wortschatz Leipzig*⁴ par 4 participantes au cours de traduction allemand-français. L'enseignant a présenté ce corpus en classe et quelques participantes se sont aventurées à l'utiliser, mais pas très souvent selon leurs dires, et principalement pour la recherche de prépositions liées à des verbes. Cela nous amène à la raison suivante, la **connaissance de ces outils**. Si le tableau révèle une méconnaissance générale des ressources disponibles, l'idée revient à plusieurs reprises que les pairs se transmettent des informations concernant les sites internet et les applications. C'est le cas notamment pour *DeepL* et *Linguee*. Finalement, certaines mettent en avant le **côté plus pratique** des outils numériques. L'extrait suivant montre bien le processus pour l'application smartphone de *Pons*.

(15) ET26DE, 11:27-12:00

eum je pense que c'est vraiment à cause de application parce que l'application téléphone c'est une application qui marche sans internet alors je l'ai téléchargée une fois et depuis ce moment-là quand il y a les mots j'utilise seulement pons et sinon je je vais regarde sur internet et là c'est pas grave si c'est langenscheidt ou quelque chose d'autre je veux seulement trouver le mot

Au vu de ce témoignage et d'autres que nous avons récoltés, nous sommes amenés à préciser le constat fait plus haut sur l'importance du support des outils. Plutôt qu'un contraste entre dictionnaire papier et outils numériques, il faudrait ainsi insister sur la distinction entre livre et smartphone. Il semblerait en

⁴ https://corpora.uni-leipzig.de/fr?corpusId=fra_mixed_2012

effet que la plupart de nos étudiantes utilisent presque exclusivement les outils numériques sur leur téléphone. Cette précision est intéressante. En effet, les portables sont interdits dans les salles d'examen et l'habitude d'utiliser un smartphone semble avoir joué des tours à une étudiante qui a découvert lors de l'examen que l'application dont elle avait l'habitude ne fonctionnait pas de la même façon sur le téléphone et sur l'ordinateur.

A contrario, pourquoi les utilisatrices décident-elles de ne pas recourir à certains outils (voir les cases en gris clair dans le Tab. 1: outils connus mais non utilisés)? En premier lieu, certains outils ne sont **pas assez clairs et faciles d'utilisation** et, par conséquent, les apprenantes en privilégient d'autres. Par exemple, une étudiante dit détester *Linguee*, trouvant cette application trop "chaotique". La **fiabilité** de l'outil entre également en ligne de compte. On préfère éviter les outils qui n'ont pas une bonne réputation de fiabilité. L'outil le plus souvent incriminé est *Google Translate*, qui par ailleurs est utilisé et préféré par d'autres étudiantes. Finalement, il arrive que des **figures d'autorité** – notamment des enseignantes – mettent en garde contre certains outils. C'est apparemment le cas dans certaines universités, où on conseille par exemple aux apprenantes de privilégier l'application *Pons* et de ne pas utiliser *Leo*.

(16) ET8DE, 5:06-5:50

c'était à l'université à zurich qu'ils nous ont dit que pons c'est mieux que leo parce que . il y a plus des phrases ou ouais aussi le contexte ou je sais pas peut-être leo il y a plus seulement le vocabulaire/ aucune une idée mais ouais ils ont dit que c'est mieux d'utiliser pons que leo mais si on tape par exemple le verbe sur leo il il donne aussi le verbe fléchi et c'est pour ça que des fois j'utilise leo parce que si je veux voir par exemple la forme de passé simple on peut le voir là

On voit ici que la préférence personnelle de l'utilisatrice va à *Leo* qu'elle trouve pratique en termes de contenu. Elle consulte toutefois quand même *Pons* en priorité, car une figure d'autorité a décrété que *Pons* était de meilleure qualité.

5. Exemples concrets d'utilisation des outils

Dans un second temps, par le recueil de données en examen, nous avons cherché à savoir quelles étaient les pratiques et stratégies concrètes et réelles de nos apprenantes et, dans une certaine mesure, à quel point elles étaient conformes à leurs déclarations en entretien. Nous allons nous intéresser ci-dessous à l'utilisation des ressources numériques dans deux situations distinctes de l'examen de traduction allemand-français.

5.1 Des pratiques diverses pour des résultats contrastés

Le premier cas observé s'inscrit dans une démarche contrastive de phrases à corriger. Ces extraits ont été sélectionnés par l'enseignant dans les traductions rendues pour correction par les étudiantes durant le semestre et ont fait l'objet de discussions en classe. On peut dire pour résumer qu'il s'agit de phénomènes

du type "erreurs fréquentes ou difficultés récurrentes de germanophones en français".

2. Corrigez les phrases suivantes:

1. Il ne faisait pas du sens de se disputer.
2. Evelyn ne semblait pas d'avoir des intérêts.
3. C'était un de ces endroits où les repas portaient des noms originels et on offrait des demi-portions.
4. Sur la stupide petite table blanche sont des tulipes et des freesias dans un vase.
5. A moi, ça plaît.

Fig. 1: Extrait de l'examen allemand-français: phrases à corriger

La première phrase proposée (*Il ne faisait pas du sens de se disputer*) correspond au texte source *Es lohnte keinen Streit, denn im Grunde, wie gesagt, war es eine glückliche Ehe* (Max Frisch: *Stiller*) et pose le double problème de la traduction de "es" et de l'utilisation de locutions telles que "faire du sens", "faire sens" ou "avoir du sens". Des traductions possibles seraient par exemple: *Cela ne valait pas la peine de se disputer, [parce qu'au fond, comme je l'ai dit, c'était un mariage heureux]* / *Cela ne valait pas le coup de se disputer* / *Cela n'avait pas de sens de se disputer*.

Confrontée à cette tâche de correction, ET22 commence par essayer d'utiliser le corpus français de *Wortschatz Leipzig* et introduit "faire du sens" comme terme de recherche, une tentative vouée à l'échec puisque ce corpus n'accepte que des requêtes d'un seul mot-forme. Elle clique alors sur l'un des composants de la locution, le nom "sens", consulte les voisins de droite privilégiés de ce terme, puis... finit par revenir à *Pons*, un outil qui lui est familier et dans lequel elle introduit à nouveau l'expression complète. Le résultat de sa requête sera le suivant⁵:

1. ~~Il ne fait pas de sens de se disputer.~~

La correction proposée par l'étudiante, minimale, porte sur deux points: d'une part, elle met le verbe au présent, ce qui reste anecdotique, et d'autre part – et surtout – elle transforme le déterminant avec le passage de "du" à "de", une modification nécessaire eu égard à la présence d'une négation. En revanche, elle ne relève ni le problème lié au pronom impersonnel "il" ni celui lié à la locution elle-même.

Face à la même tâche, ET8 utilise, elle, *Linguee*, qui permet des recherches combinées. Elle y introduit "ne faisait pas de sens" et obtient comme première réponse la phrase *D'autre part, lors de l'expédition de la robe, le repassage ne*

⁵ Dans cet extrait de copie d'examen et les suivants, les mots uniquement biffés ou soulignés correspondent à la correction par l'enseignant.

*fait pas de sens car celle-ci est emballée dans un carton pour l'expédition*⁶. Elle y trouve donc en quelque sorte la confirmation de son hypothèse, qui était qu'il fallait modifier le déterminant à cause de la négation, et corrige l'extrait de la façon suivante:

1. ~~Il ne faisait pas du~~^{de} sens de se disputer.

Par contre, comme la candidate précédente, elle ne remet en cause ni le sujet "il" ni le choix du verbe "faire".

Enfin, une troisième étudiante, ET25, recourt à une stratégie très différente pour tenter de résoudre le problème. Elle décide d'utiliser le correcteur d'orthographe et de grammaire de *Reverso*⁷ et y introduit la phrase complète. Le correcteur ne décelant aucune erreur, elle change d'outil et passe à *Language Tool*⁸ (le fait qu'elle procède à une recherche d'un correcteur grammatical sur *Google* pour y arriver nous amène d'ailleurs à nous demander si elle avait connaissance de cet outil avant l'examen ou si elle l'a découvert par hasard). Après avoir à nouveau introduit la phrase entière, elle obtient une information intéressante, le logiciel indiquant que l'expression "faire du sens" est un anglicisme dérivé de "to make sense" et recommandant l'utilisation, en français de "avoir du sens". À la fin de cette recherche, elle nous livre le résultat suivant:

1. Ça n'avait aucune sens de se disputer

Il apparaît clairement dans ce cas que la variété des outils consultés et la répétition des recherches effectuées, couplées très probablement à une intuition de départ de l'étudiante, sont propres à permettre la production d'un énoncé adéquat.

5.2 L'utilisation intensive d'une ressource

Le deuxième cas que nous souhaitons aborder est un exemple d'usage intensif de *DeepL*⁹ par une étudiante, dans un contexte de traduction de texte inconnu cette fois. L'extrait est le suivant:

⁶ <https://www.linguee.fr/francais-allemand/search?source=auto&query=%CA%BAn+faitait+pas+de+sens%CA%BA+>

⁷ <https://www.reverso.net/orthographe/correcteur-francais/>

⁸ <https://languagetool.org/fr/>

⁹ <https://www.deepl.com/translator>

2. Peter Härtling: *Felix Guttman*

Ich traue ihm alles zu. *Er könnte wieder, wie vor Jahren, unerwartet im Treppenhaus stehen, nach den Kindern rufen, in der Küche nachsehen, was es zum Abendessen gibt, mich knapp grüßen, sich dann auf die Terrasse setzen, den abgenutzten Hut auf den Knien, und, kaum haben wir das Gespräch begonnen, wieder aufstehen und sich verabschieden.* Er geht durch den Garten, zur Pforte hinaus, den Mantel, wie immer, offen, den Hut wieder auf, klein, sehr klein, doch unverletzbar, merkwürdig dauerhaft.

Fig. 2: Extrait de l'examen allemand-français: texte à traduire (nos italiques)

L'élément saillant de ce texte est la syntaxe de la deuxième phrase, propre à poser des problèmes: un semi-auxiliaire de modalité *könnte* suivi d'une série de 7 infinitifs (*stehen, rufen, nachsehen, grüßen, sich setzen, aufstehen, sich verabschieden*), le tout entrecoupé d'éléments intercalés entre virgules (*wie vor Jahren, was es zum Abendessen gibt, den abgenutzten Hut auf den Knien, kaum haben wir das Gespräch begonnen*).

Nous allons illustrer par trois captures d'écran la stratégie de l'étudiante ET27 face à ce texte. Ces étapes sont représentatives d'un total de 10'27" de processus de traduction.

Étape numéro 1: on met tout dans *DeepL*

L'étudiante ET27 commence par introduire la phrase complète dans le traducteur automatique:

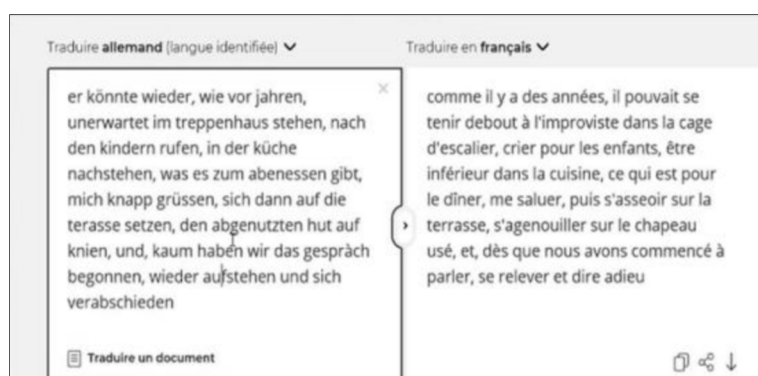


Fig. 3: Fenêtre *DeepL* après 1'38"

Comme on peut le voir, le logiciel propose une version française quelque peu approximative: *crier pour les enfants; être inférieur dans la cuisine, ce qui est pour le dîner; s'agenouiller sur le chapeau usé* en particulier s'avèrent très problématiques. En revanche, la structure verbale qui s'étend à travers toute la phrase, avec un semi-auxiliaire *pouvait* et 7 infinitifs, est correctement analysée et transcrite.

L'étudiante ne se contente toutefois pas de cette première suggestion. D'origine tchèque, elle va ainsi consulter par deux fois un dictionnaire bilingue, tchèque-

allemand d'abord (2'56") pour *zutrauen*, puis tchèque-français (03'08") pour *unerwartet*. Elle introduit ensuite (03'23") le même *unerwartet* dans *Linguee* allemand-français.

Étape numéro 2: on retranche

Ayant apparemment compris la structure générale de la phrase, ET27 met alors en place une stratégie de travail visant à en diminuer la complexité: elle commence à réduire systématiquement le contexte, en supprimant tour à tour les éléments situés en tête de phrase:

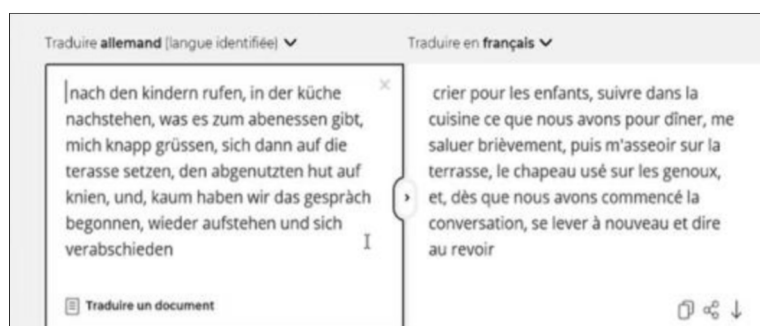


Fig. 4: Fenêtre *DeepL* après 4'45"

Cette stratégie de réduction du contexte se répétera en tout cinq fois durant le processus de traduction. À chaque fois, elle aura des implications sur le résultat obtenu: *être inférieur dans la cuisine, ce qui est pour le dîner* (Fig. 3) se transforme ainsi en *suivre dans la cuisine ce que nous avons pour dîner* (Fig. 4); *s'agenouiller sur le chapeau usé en le chapeau usé sur les genoux*, etc. Ces deux passages en particulier subiront des variations à chaque réduction. En revanche, les infinitifs sont maintenus tels quels.

Durant toute la procédure, des recours ponctuels au dictionnaire tchèque-français et à *Linguee* allemand-français interviendront.

Étape numéro 3: on réduit au maximum

Suite aux réductions successives, la dernière fenêtre *DeepL* enregistrée est la suivante:

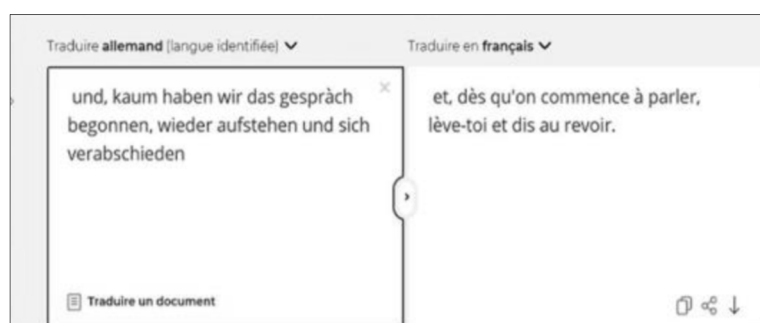
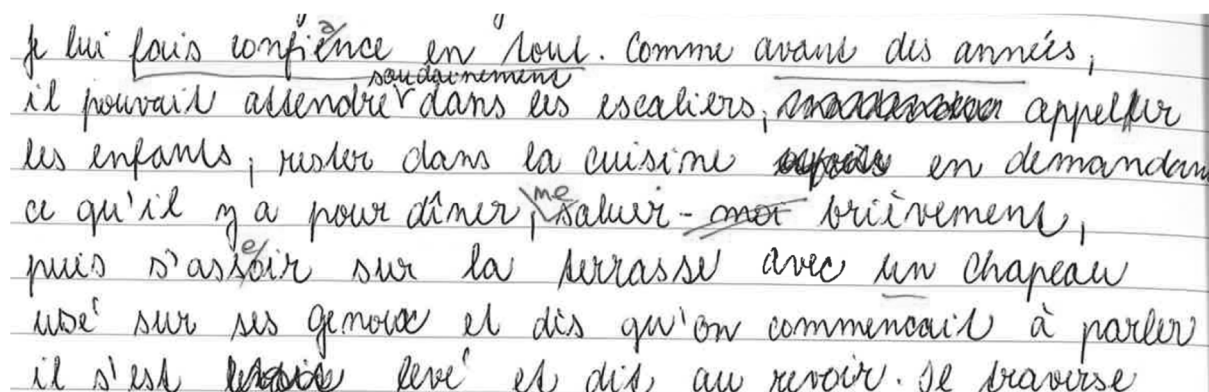


Fig. 5: Fenêtre *DeepL* après 8'34"

L'élément principal à relever est le passage des infinitifs, observés dans les phases précédentes, (*se (re)lever; dire*) à des impératifs (*lève-toi; dis*).

Cette dernière étape sera suivie de diverses vérifications dans le dictionnaire bilingue tchèque-français et un site de conjugaison des verbes français. De manière générale, on constate donc une utilisation particulièrement active du traducteur automatique. Loin d'un rôle de consommatrice, ET27 se révèle actrice et impliquée à chaque étape du processus, dont le résultat est le suivant:



Je lui fais confiance en tout. Comme avant des années,
il pouvait attendre ^{soudainement} dans les escaliers, ~~attendre~~ appeler
les enfants, rester dans la cuisine ~~après~~ en demandant
ce qu'il y a pour dîner, ~~saluer~~ ^{me} brièvement,
puis s'asseoir sur la terrasse avec un chapeau
posé sur ses genoux et dès qu'on commençait à parler
il s'est ~~levé~~ levé et dit, au revoir. Il traverse

On peut noter un certain nombre d'éléments dans la traduction proposée:

1. Malgré sa stratégie consistant à supprimer les éléments en tête de phrase déjà "traités" et les changements qui en ont résulté dans les propositions de traduction de *DeepL*, avec le passage aux divers impératifs (Fig. 5), l'étudiante a été capable, dans une large mesure, de maintenir la "tension" entre le semi-auxiliaire et les infinitifs; il n'y a que les deux derniers verbes qui sont conjugués – à l'indicatif – et qui rompent donc la syntaxe de la phrase d'origine.
2. La question du chapeau aussi a été réglée correctement: personne ne s'est agenouillé dessus, contrairement aux indications de *DeepL* (Fig. 4).
3. La vérification supplémentaire de *unerwartet*, un adjectif-adverbe en allemand, a porté ses fruits, puisqu'on trouve bien un adverbe (*soudainement*) dans la traduction, contrairement à d'autres copies d'examen récoltées qui contiennent à tort un adjectif.

5.3 Typologie des pratiques en examen

Les observations faites durant les examens permettent d'établir une typologie encore toute provisoire de l'attitude des étudiantes face aux ressources numériques, qui demandera à être affinée par des récoltes de données supplémentaires. On peut *grosso modo* distinguer les types suivants:

- Les "conservatrices", que l'on pourrait également qualifier de "prudentes": par principe, elles s'appuient sur ce qu'elles connaissent. Chez les germanophones, elles se distinguent notamment par le recours

systématique et quasi exclusif à *Pons*, un outil qui leur a été recommandé et dont la qualité est validée par une autorité.

- Les "éclectiques": elles essaient plusieurs outils, parfois aussi en concurrence, pour tenter d'atteindre le meilleur résultat possible. Elles prennent également des risques en s'essayant à des outils qu'elles ne maîtrisent qu'imparfaitement.
- Les "exhaustives": elles introduisent rigoureusement tout dans les traducteurs automatiques, la plupart du temps *DeepL* pour les anglophones et certains germanophones, mais parfois aussi *Reverso* pour d'autres germanophones. Ce faisant, elles ne se bornent toutefois pas forcément à reprendre mot pour mot les traductions proposées, qui font au contraire l'objet d'une analyse personnelle.

On notera enfin que, malgré leur attachement déclaré au format papier, gage de qualité supposée, de rigueur et parfois même de plaisir, les étudiantes n'ont à aucun moment consulté les dictionnaires à leur disposition durant l'examen.

6. Conclusion

Les deux premières phases de la recherche-action ont tout d'abord permis de mettre en lumière un fort décalage entre les représentations des étudiantes et leurs pratiques. En effet, si le dictionnaire papier, symbole de sérieux et de qualité, est plébiscité dans leurs propos, il s'avère pratiquement absent du travail à domicile et en classe. Un second critère qui est présenté comme important dans les entretiens est celui du support: alors que nous étions partis de l'idée d'une dichotomie entre papier et numérique, il s'avère que dans les pratiques courantes, c'est bien la distinction entre téléphone et ordinateur qui est souvent prépondérante, notamment pour les applications dont le fonctionnement diffère entre les deux versions proposées.

De manière générale, on constate que l'on se trouve, dans le domaine de l'enseignement/apprentissage des langues et des *digital tools*, dans une phase de transition qui s'avèrera probablement décisive. En effet, alors que certaines étudiantes se montrent déjà très à l'aise avec le maniement de ces outils et produisent des résultats intéressants (et probablement supérieurs à leurs compétences intrinsèques dans la L2), d'autres tâtonnent encore largement. Il semble dès lors indispensable non seulement de thématiser le recours aux ressources numériques, comme nous l'avons fait dans les phases 1 et 2 de notre recherche-action, mais de présenter et expliquer leur usage durant le cours et leurs potentialités respectives pour favoriser une égalité des chances en contexte d'apprentissage et plus généralement dans la vie civique.

La phase 3 devra ainsi impérativement être consacrée aux stratégies de consultation des diverses ressources. Il s'agira tout d'abord d'illustrer ce que les unes et les autres sont capables de faire ou non. Il s'agira ensuite de mettre

l'accent notamment sur les reformulations du texte source pour obtenir de meilleurs résultats dans la L2, de réfléchir ensemble aux avantages possibles de la rétrotraduction pour mettre le doigt sur des faiblesses de formulation, et enfin, dans le cas des traducteurs automatiques (sur la nécessité d'intégrer ceux-ci dans une démarche didactique, voir Nied-Curcio 2015: 461), de mettre en place des stratégies quant à la gestion des variantes proposées par le logiciel. De telles activités réflexives ne devraient en aucun cas contribuer à fragiliser la place de la traduction dans l'enseignement/apprentissage des langues mais seraient au contraire de nature à lui donner une valeur supplémentaire dans une démarche contrastive.

BIBLIOGRAPHIE

- Domínguez Vázquez, M. J., Mirazo Balsa, M. & Vidal Pérez, V. (2013). Wörterbuchbenutzung: Erwartungen und Bedürfnisse. Ergebnisse einer Umfrage bei Deutsch lernenden Hispanophonen. In M. J. Domínguez Vázquez (éd.), *Trends in der deutsch-spanischen Lexikographie* (pp. 135-172). Frankfurt a. M.: Peter Lang.
- Domínguez Vázquez, M. J., Mollica, F. & Nied Curcio, M. (2014a). Zweisprachige Wörterbücher, Didaktik und Translation: Einführung. In M. J. Domínguez Vázquez, F. Mollica & M. Nied Curcio (éds.), *Zweisprachige Lexikographie zwischen Translation und Didaktik* (pp. 1-12). Berlin, Boston: de Gruyter.
- Domínguez Vázquez, M. J., Mollica, F. & Nied Curcio, M. (2014b). Simplex-Verben im Italienischen und Spanischen vs. Präfix- und Partikelverben im Deutschen. Eine Untersuchung zum Gebrauch von Online-Wörterbüchern bei der Übersetzung. In M. J. Domínguez Vázquez, F. Mollica & M. Nied Curcio (éds.), *Zweisprachige Lexikographie zwischen Translation und Didaktik* (pp. 179-220). Berlin, Boston: de Gruyter.
- Domínguez Vázquez, M. J. & Valcárcel Riveiro, C. (2015): Hábitos de uso de los diccionarios entre los estudiantes universitarios europeos: ¿nuevas tendencias? In M. J. Domínguez Vázquez, X. Gómez Guinovart & C. Valcárcel Riveiro (éds.), *Lexicografía de las lenguas románicas II. Aproximaciones a la lexicografía contemporánea y contrastiva* (pp. 165-189). Berlin: de Gruyter.
- Macaire, D. (2011). Recherche-action en didactique des langues et des cultures: changer les pratiques et pratiquer le changement. In M. Molinié (éd.), *Démarches portfolio en didactique des langues et des culture* (pp. 113-124). Amiens: Encrage.
- Mollica, F. (2020). Funktionsverbgefüge in ein- und zweisprachigen Wörterbüchern (für das Sprachenpaar Deutsch-Italienisch) aus der Perspektive der DaF-Benutzer. In S. De Knop & M. Hermann (éds.), *Funktionsverbgefüge im Fokus. Theoretische, didaktische und kontrastive Perspektiven* (pp. 137-178). Berlin: de Gruyter.
- Müller-Spitzer, C. (éd.) (2014). *Using Online Dictionaries*. Berlin/Boston: de Gruyter.
- Müller-Spitzer, C., Koplenig, A. & Töpel, A. (2011). What Makes a Good Online Dictionary? - Empirical Insights front an Interdisciplinary Research. In I. Kosem & K. Kosem (éds.): *Electronic lexicography in the 21st century: New applications for new users. Proceedings of eLex 2011, Bled, 10-12 November 2011* (pp. 203-208). Ljubljana: Trojina, Institute for Applied Slovene Studies.
- Müller-Spitzer, C. & Koplenig, A. (2014). Online dictionaries: expectations and demands. In C. Müller-Spitzer (éd.), *Using Online Dictionaries* (pp. 143-188). Berlin/Boston: de Gruyter.

- Nied Curcio, M. (2014). Die Benutzung von Smartphones im Fremdsprachenerwerb und -unterricht. In A. Abel, C. Vettori & N. Ralli (éds.), *EURALEX International congress: The user in focus. 15-19 luglio 2015* (pp. 263-280). Bolzano: Eurac.
- Nied Curcio, M. (2015). Wörterbuchbenutzung und Wortschatzerwerb. Werden im Zeitalter des Smartphones überhaupt noch Vokabeln gelernt? *Info DaF 5/2015* (Themenreihe "Wörterbücher für Deutsch als Fremdsprache"), 445-468.
- Tiberius, C. & Müller-Spitzer, C. (éds.) (2015). *Research into dictionary use / Wörterbuchbenutzungsforschung. 5. Arbeitsbericht des wissenschaftlichen Netzwerks "Internetlexikografie"*. Mannheim: Institut für Deutsche Sprache (= OPAL 2/2015)
- Wiegand, H. E. (1977). Nachdenken über Wörterbücher. Aktuelle Probleme. In G. Drosdowski, H. Henne & H. E. Wiegand: *Nachdenken über Wörterbücher* (pp. 51-102). Mannheim, Wien, Zürich: Bibliographisches Institut.

ANNEXE: CONVENTIONS DE TRANSCRIPTIONS

. micro-pause ; .. pause moyenne ; ... pause longue

[] chevauchements

- troncation d'un mot (from- fromage)

: allongements

/ intonation montante

oui emphase, accentuation

OUI voix plus forte

°oui° voix moins forte

@oui@ voix riante

XXX segment incompréhensible

(rit) commentaires

(...) ellipse dans la transcription

Écouter la "voix" du texte: modélisation d'une approche de l'oral en L2 au sein d'un environnement numérique

Monica MASPERI

Enseignante-Chercheure en Études Italiennes
Laboratoire LIDILEM, Université Grenoble Alpes
Bâtiment Stendhal, CS40700, 38058 Grenoble cedex 9, France
monica.masper@univ-grenoble-alpes.fr

Triscia BIAGIOTTI

Maître de Langue
Institut d'Études Politiques de Grenoble
7 Place Bir Hakeim, 38000 Grenoble, France
triscia.biagiotti@iepg.fr

Sandra CANELAS-TREVISI

Enseignante-Chercheure HDR en Sciences du Langage
Membre du GRAFE, Université de Genève
24 rue du Général-Dufour, 1211 Genève 4, Suisse
canelas.trevi@orange.fr

All'interno del paesaggio di modelli didattici ampiamente riconosciuti, fondati su una logica di progressione graduale, fedele ai descrittori del QCER, lo sforzo di trasformazione didattica è il più delle volte delegato alla mediazione digitale e alle modalità di interazione virtuali. Ribaltando questa tendenza e schierandoci sul fronte di una "innovazione educativa sociale", proponiamo un approccio didattico in L2 fortemente interconnesso con la ricerca fondamentale in L1 e incentrato sul testo orale in quanto oggetto complesso: solidamente organizzato, culturalmente denso e socialmente ancorato. Tali riflessioni, condotte nell'ambito del progetto IDEFI Innovalangues si concretizzano nella realizzazione di percorsi digitali in italiano L2, elaborati a partire da risorse video selezionate in base a valori culturali rappresentativi della lingua-cultura target. In tale contesto, le tecnologie digitali vengono messe a profitto di una "lettura" pluridimensionale del testo, rivolta al linguaggio, al discorso e alla lingua. Attraverso questo approccio intendiamo interrogare non soltanto in maniera più approfondita l'accesso ai testi ma anche lo strumento digitale in quanto strumento di ricerca, messo a contributo della conoscenza linguistica, didattica e ingegneristica.

Parole chiave:

Testo, discorso, L2, comprensione orale, ambiente digitale.

Mots-clés:

Texte, discours, L2, compréhension de l'oral, environnement numérique.

1. Introduction

La compréhension orale d'une langue autre que la langue première est un des objectifs visés par les institutions d'enseignement. S'il existe un consensus absolu sur la nécessité de créer les conditions pour qu'un niveau de compréhension déterminé soit atteint à un moment donné du cursus

d'apprentissage, les points de vue sont nettement moins consensuels lorsqu'il s'agit de définir la compréhension orale et d'en cibler l'objet. Aussi, la compréhension de l'oral est-elle souvent pensée en relation étroite avec l'agir social en interaction et les apprentissages tendent à être évalués à partir de l'efficacité des conduites adoptées par les locuteurs. La conception du langage sous-jacente intègre les faits langagiers dans le fonctionnement social mais l'agir langagier est massivement conçu comme évaluable en termes de stratégies et de leur degré d'efficacité selon les attentes socialement partagées de manière dominante. Cette tendance s'est renforcée suite au lent affaiblissement institutionnel de l'enseignement de la littérature de la langue étudiée, qui peut être considéré comme un des révélateurs de l'orientation vers les langues de service au détriment des langues de culture (Rastier 2007).

Les choix didactiques présentés dans cette contribution sont fondés sur la conception théorique de l'activité langagière et des textes que développe J.-P. Bronckart, articulée avec la notion de "passage" émanant des travaux de J.-F. Rastier. Il nous faut préciser d'emblée que l'articulation entre la filiation théorique convoquée et la proposition didactique ici exposée fait l'économie d'une jonction explicite avec des travaux menés en compréhension de l'oral en L2. Ce choix n'entend aucunement éluder les apports que la didactique des langues nous livre depuis plus de vingt ans dans le domaine de la compréhension de l'oral (i.e. Cornaire et Germain 1998 ; Field 1998, 2000, 2008, 2015, 2019 ; Hilton 2000 ; de Pietro et al. 2017 ; Gagnon et al. 2017 ; Lopriore 2011 ; Nunan 2002 ; Roussel 2014 ; Vandergrift 1997, 2007). Toutefois si ce socle de connaissances représente une base incontournable de notre travail de didacticiennes, notre objectif est en l'occurrence de produire un effet de zoom sur un travail de conception didactique qui tente de réinvestir en L2 des orientations théoriques problématisées en L1 dans des branches des sciences du langage telles que la psychologie du langage ou la sémantique, qui se sont concrétisées entre autres dans des ouvrages francophones destinés aux élèves (notamment en Suisse et au Québec)

L'objet didactique que nous présentons – à savoir un parcours numérique "tout à l'oral" – est donc le produit de la transposition et de la reconfiguration didactique en L2 de nos choix théoriques.

Mis en place dans le cadre du projet Innovalangues (ANR-11-IDFI-0024, Masperi 2011) pour l'enseignement de la compréhension orale en L2, le dispositif que nous présentons ici place l'apprenant face à des vidéos et lui propose une série d'activités construites selon deux critères méthodologiques strictement articulés: une conception du texte fondée sur des apports théoriques déjà largement reconfigurés et mis à l'épreuve, en particulier dans l'enseignement de l'écrit (*cf. infra* 2); les possibilités offertes par quatre types d'opérations numériques: le chapitrage, le bornage guidé, le bornage libre, l'écoute active enrichie (*cf. infra* 4.2).

Dans la première partie, nous développerons notre conception de l'activité langagière en l'inscrivant dans le champ de référence de l'interactionnisme socio-discursif de Bronckart et de son équipe. Ce modèle théorique nous aide à cerner les choix individuels du locuteur lorsqu'il s'inscrit dans les conventions langagières en vigueur dans une culture donnée.

Dans la deuxième partie, nous présentons notre démarche didactique de sélection et de traitement textuel et numérique de notre corpus de vidéos. Les critères de sélection des vidéos font partie intégrante de la construction méthodologique et peuvent être regroupés selon trois axes. Le premier concerne l'ancrage explicite du texte dans la langue de culture; le deuxième a trait à la présence des traces nettement repérables du contexte de l'interaction; le troisième fait référence aux types de discours identifiables dans la linéarité du texte, produit des opérations psychologiques qui le sous-tendent.

Dans la troisième partie, sont illustrées les conduites des apprenants engagés dans le travail de compréhension des vidéos, guidés par les activités du dispositif numérique. Leurs conduites nous apportent un tout premier éclairage sur l'interaction entre la démarche didactique adoptée, le dispositif numérique et les vidéos proposées.

2. Le modèle théorique de référence

La construction didactique de nos parcours d'enseignement s'appuie sur la théorisation de l'activité langagière que mettent en place Bronckart et ses collaborateurs dans le cadre de l'interactionnisme socio-discursif. Cette théorisation se fonde sur le principe que tout individu se développe au contact d'entités physiques et d'activités collectives instrumentées, socialement formatées.

Ainsi, le fonctionnement du langage qui se déploie dans l'activité langagière est appréhendé comme le produit des pratiques effectives, mais également des pré-construits socio-historiques, dont les représentations collectives. Les formats des activités langagières se définissent dans l'interaction complexe entre l'individu et le fonctionnement collectif, sous l'effet de la réflexivité, dont témoignent aussi bien les actants directement impliqués dans l'activité que les observateurs externes (Bronckart 1997).

S'opposant à la position logique de Piaget et prenant appui sur la conception saussurienne des signes discrets, radicalement arbitraires et essentiellement sociaux, Bronckart défend le caractère social de l'entrée dans le langage chez l'enfant et s'engage à théoriser "les mécanismes d'interaction" entre activité langagière et organisation psychique présents pendant toute la durée de la vie humaine.

Les réalisations empiriques de l'interaction sont conceptualisées à l'aide de la notion de texte. Le terme désigne toute production verbale située (Bronckart

1997: 73), orale ou écrite, gérée par une seule instance énonciative ou par plusieurs énonciateurs. La diversité des textes, illimitée, dépend de la diversité des contextes dans lesquels ils sont ancrés. Toutefois, si chaque texte est singulier, les activités humaines auxquelles s'articulent les textes présentent des formes d'organisation identifiables, qui dépendent du fonctionnement des sociétés.

Ainsi, puisque le contexte de toute production langagière est inscrit dans l'organisation sociale, les textes empiriques adoptent et adaptent des formats qui se sont plus ou moins stabilisés au fil de l'histoire d'une société donnée. Ces manières de faire les textes, historiquement identifiées comme genres, sont d'abord codifiées dans la littérature et dans des pratiques sociales hautement valorisées comme celles du domaine religieux et juridique. Ensuite, progressivement, l'existence de ces sortes de modèles a été identifiée dans toutes les pratiques langagières, écrites et orales. Toutefois, si tout locuteur perçoit l'existence d'une multiplicité des genres, les critères de description et de classement des genres demeurent extrêmement difficiles à identifier.

Notre modèle théorique et méthodologique de référence s'appuie sur une double prise de position. La première est ancrée dans le travail que l'équipe de recherche dirigée par Bronckart mène depuis longtemps sur le classement des textes, notamment avec une perspective didactique (Bronckart et al. 1985). Des textes relevant de genres divers, produits par des adultes et des enfants, écrits pour la plupart, ont été soumis à un ensemble d'analyses. Dans chaque texte écrit, sont d'abord sélectionnés trois sous-ensembles d'observables: d'ordre sémantique, lexico-syntaxique, para-linguistique (Bronckart 1997: 84).

Ensuite, dans les mêmes textes empiriques, sont découpés des segments à partir de deux critères: le degré élevé de proximité entre les segments et les descriptions et analyses théoriques des types de discours émanant de la recherche scientifique; le degré de présence des trois sous-ensembles d'observables retenus. Pour éviter la circularité qui consisterait à isoler des segments de texte relevant d'un type de discours particulier à partir d'indices observés à la lecture et à confirmer ensuite le statut discursif du segment en se fondant sur la simple quantification des indices, les différents segments "font l'objet d'analyse quantitatives complexes qui portent notamment sur des unités linguistiques qui n'ont pas à priori la réputation de contribuer à la différenciation des types" (Bronckart 1997: 85).

Les résultats de cette recherche apportent des éléments de réponse à un questionnement largement partagé sur le classement des genres, aussi bien à l'écrit qu'à l'oral, et sur le traitement des segments de textes présentant des caractéristiques spécifiques. L'étiquetage terminologique de ces segments, de même que le terme qui les regroupe, est un des révélateurs de la difficulté à définir leur statut et leur forme, comme le soulignent, entre beaucoup d'autres, Kerbrat-Orecchioni et Traverso (2004) qui mentionnent les étiquettes genre 1

et genre 2, pour souligner l'ambiguïté de la différenciation, ainsi que Philippe et Adam (2015) pour qui les effets de textualité et de généricité dans l'analyse textuelle des discours ne peuvent être disjoints.

Les termes que nous utilisons, en assumant les questionnements encore ouverts de notre modèle théorique de référence, sont ceux de "genre de texte" et de "type de discours". Les recherches sur l'oral de l'interaction ont interrogé en profondeur les enjeux d'une organisation super-ordonnée de genres et de types sans pour autant la remettre radicalement en question, malgré l'extrême difficulté de parvenir à cerner des modèles prototypiques.

La seconde prise de position sur laquelle nous nous sommes appuyées nous amène à aborder les types et les genres à partir des notions de choix psychologiques du "monde discursif" et d'opérations psychologiques. Dans sa construction théorique, partant du principe que l'activité langagière, sémiotique par sa nature même, s'appuie sur la création de mondes virtuels, Bronckart s'engage à décrire les mondes et les opérations psychologiques qui les sous-tendent.

Ces mondes sont des systèmes de coordonnées formelles, qui d'une part sont radicalement "autres" que les systèmes de coordonnées des mondes représentés dans lesquels se déploient les actions d'agents humains, mais qui d'autre part doivent exhiber le type de rapport qu'ils entretiennent avec ces mondes de l'activité humaine. Par convention, nous qualifierons les mondes représentés des agents humains de mondes ordinaires et les mondes virtuels créés par l'activité langagière de **monde discursifs**. (Bronckart 1997: 153).

Deux ordres d'opérations rendent compte de la construction des mondes discursifs. Le premier ordre d'opération a trait aux rapports entre les coordonnées générales qui gèrent la construction du contenu thématique du texte et les coordonnées du monde ordinaire, dans lequel se déploie l'action langagière dont le texte est le produit. Le second ordre d'opérations explicite d'une part le rapport entre les instances intervenant dans le texte (personnages, institutions etc.) et leur environnement spatio-temporel tel qu'il est mis en texte et d'autre part les paramètres physiques de l'action langagière (producteur, interlocuteur éventuel et espace-temps de production) (ibidem: 154). Cette analyse binaire permet de partager les textes en deux grands ensembles en fonction de leur appartenance à l'un des deux mondes discursifs: le monde du RACONTER et le monde de l'EXPOSER (ibidem: 159)

Défini d'abord à partir de son appartenance au monde discursif, le texte demande ensuite à être appréhendé dans sa dimension linéaire, car lors de toute production textuelle les représentations et les connaissances relatives à un thème donné sont nécessairement réorganisées dans le successif.

Bronckart considère que tout texte est constitué de types de discours, formes d'organisation linguistique qui traduisent les mondes discursifs. Toutefois, bien que son infrastructure comporte des composants différents (Adam 2015: 24), tout texte est pensé comme une unité articulée à une situation d'action

langagière particulière et tendant à produire un effet de cohérence sur le destinataire.

Deux ordres de mécanismes assurent la cohérence du texte: les mécanismes de textualisation que sont la connexion et la cohésion nominale, la cohésion verbale; la prise en charge énonciative (ibidem: 153). Les mécanismes de prise en charge énonciative contribuent à faire apparaître d'une part les responsabilités énonciatives en jeu dans le texte et d'autre part les multiples jugements, commentaires et évaluations sur des aspects du contenu thématique qu'expriment les modalisations. Les marques de la modalité sont très hétérogènes et leur fonction n'est pas liée directement à la linéarité du texte.

3. La construction didactique

Les choix théoriques de Bronckart et de son équipe ont guidé notre travail didactique, mais la construction de nos "parcours numériques" nous a nécessairement amenées à repenser la conception des textes sélectionnés ainsi que leur infrastructure. Il s'agissait en effet pour nous de construire des activités d'apprentissage visant à faciliter l'accès au texte pour des étudiants travaillant en autonomie. Nos destinataires sont en l'occurrence les étudiants de l'Université française qui ont un niveau de compétence B2 en voie d'acquisition ou acquis.

L'enseignement d'une langue en milieu institutionnel s'appuie traditionnellement sur la notion de langue comme système, comportant des ensembles de règles phonologiques, lexicales et morpho-syntaxiques, relativement stables donc analysables. Dans notre construction didactique, l'objet de la compréhension est un texte oral concrétisé dans un document vidéo, diffusé par un média.

Ayant admis le principe que toute langue naturelle se matérialise dans des productions langagières inscrites dans des situations d'interaction particulières, nous avons fondé notre réflexion didactique sur l'articulation étroite du langage et de l'action socialement formatée. Les activités dans lesquelles nous engageons les étudiants les renvoient au monde socio-culturel dont le texte émane et aux choix énonciatifs qui le sous-tendent. Nous visons ainsi à les engager dans la démarche de compréhension en prenant conscience des enjeux du parcours proposé, car comme le souligne Rastier:

Si (...) le sens d'un texte est construit plutôt que donné, son objectivation n'est pas un processus unique, fixé une fois pour toutes. Elle est certes fondée sur l'objectivité matérielle du texte mais non fondée et garantie par elle. (Rastier 1989: 19).

Aussi, notre construction didactique s'appuie-t-elle sur l'aspect configurationnel des textes mais également sur les valeurs des mondes psychologiques que les textes expriment.

Dans la phase de construction de notre dispositif d'enseignement numérique, nous avons centré le travail didactique sur trois critères: a) la sélection de documents vidéos exhibant les traces d'un contexte institutionnel clairement identifiable. Plus exactement, doivent pouvoir être identifiées les relations entre les contenus de la vidéo et les domaines d'intérêt mobilisés à un moment donné de la vie sociale et culturelle d'un pays; b) le statut social et énonciatif des locuteurs; c) le caractère cohérent mais non spécialisé des échanges.

Dans la démarche adoptée, la première étape de l'accès au texte est une invitation à faire des hypothèses sur le genre à partir des indices fournis, en général les premiers moments de la vidéo. On commence ainsi à établir des relations encore sommaires entre le contexte de production et l'unité texte. Ensuite, il s'agit de cerner dans l'unité texte le mode d'organisation et la visée énonciative, que révèlent le type de discours dominant d'une part, la prise en charge énonciative d'autre part. Rappelons que l'action langagière implique la mise en interface des connaissances de l'agent avec les connaissances qu'expriment les genres de textes socialement disponibles. Cela entraîne la mise en confrontation des représentations personnelles et des représentations des autres dans un espace mental commun, une instance collective nécessairement impliquée dans les opérations ayant trait à la textualisation.

Lors de la production d'un texte, écrit ou oral, monologal ou dialogal, le locuteur fait appel à ses connaissances et les réorganise selon un plan textuel, qui se déploie dans la linéarité de l'écrit ou le successif de l'oral. Ensuite, la définition même du texte implique qu'il soit perçu comme une unité continue et cohérente. Les procédures de segmentation et les différents types de liage qui assurent la continuité des textes ont été analysés en profondeur notamment par Adam et nous avons fait référence à la notion de séquence développée par Adam et reprise par Bronckart (Adam 2011, 2014; Philippe et Adam 2015).

Notre travail didactique porte sur des textes oraux et les activités visant à en favoriser la compréhension nous ont amenées à segmenter les textes. Pour préserver l'impression de continuité d'une part et de cohérence d'autre part, nous avons procédé en essayant de maintenir l'équilibre entre deux points de vue: identifier et délimiter les types de segments qui composent le texte; préserver la cohérence sémantique et socio-discursive. Les plans des types de discours et l'organisation des séquences selon Adam ont constitué un premier ordre de repères, ensuite nous avons délimité à l'aide des faits intonatifs des ensembles d'énoncés qui, à l'intérieur d'un segment plus large, sont porteurs d'une valeur particulière du point de vue socio-discursif.

Mais avant d'aborder ce niveau de décision, nous avons dû affronter le problème que partage toute intervention didactique: celui de la création d'un milieu susceptible de favoriser l'engagement des étudiants. Nos étudiants travaillent en autonomie, face à leur écran. En tant qu'enseignant.e.s nous ne

sommes pas en interaction directe avec eux, aussi avons-nous créé un environnement susceptible de favoriser l'agir des étudiants en nous appuyant sur un double repérage, les traces langagières de la segmentation et nos propres parcours interprétatifs. C'est la notion de *passage* adoptée par Rastier qui nous a servi de repère¹.

Le choix didactique de segmenter le texte a été d'autant plus justifié à nos yeux que nos textes oraux ne présentent pas de traces de segmentation homogènes, comme les marques de connexion stabilisées dans l'écrit, mais des faits intonatifs parfois difficiles à saisir.

Lors de la conception de nos activités, la première étape a consisté à segmenter le texte avec un double objectif: faire ressortir les caractéristiques du type de discours dominant; préserver l'effet de cohérence que le texte produit dans la continuité. Les types de discours constituent des sortes de modèles disponibles dans la langue culture et jouent un rôle dans la production et la compréhension des textes empiriques.

Aussi, dans les activités proposées, attirons-nous prioritairement l'attention des étudiants sur les aspects typiques. Bien que dans les textes empiriques, les types de discours ne se réalisent pas forcément selon un plan entièrement prévisible, le repérage des types de discours demeure une procédure efficace pour comprendre les textes, car le choix du type de discours reflète les représentations du contenu et du destinataire.

Pour les textes de l'ordre de l'EXPOSER, la visée énonciative peut être l'intention de convaincre, d'expliquer ou simplement d'informer. Comme le souligne Bronckart (1997), si l'énonciateur considère que le contenu risque d'être problématique pour le destinataire, il peut adopter le type de discours explicatif. Inversement, s'il pense que le contenu peut paraître contestable, il choisira le type de discours argumentatif. Dans les textes de l'ordre de l'EXPOSER, l'architecture du type de discours nous permet de délimiter les segments qui présentent une intention énonciative particulière.

Pour ce qui est de l'ordre du RACONTER, les textes retenus possèdent un ensemble de caractéristiques qui les singularisent. En effet, les quatre unités constitutives du parcours sont construites à partir d'une seule vidéo représentative du genre, contrairement aux autres parcours où chacune des quatre unités a été sélectionnée comme représentative d'un genre différent, le lien entre elles étant créé par la présence du même type de discours nettement dominant.

¹ Un passage suppose une sélection et donc un point de vue. Etablies par décision de méthode, ces frontières sont donc toutes relatives et naturellement révisables. [...] Cette conception sélective du passage résulte directement de la problématique interprétative : il s'y définit tout à la fois comme un *lieu* du texte et comme un *moment* du parcours qui le choisit et l'isole. (Rastier 2015: 82)

4. Mise en œuvre des parcours

Les potentialités d'une approche didactique de l'apprentissage des langues par "genres de textes et types de discours" telles qu'elles ont été présentées ont commencé à être explorées dans notre projet en 2016. Des expérimentations ont été menées entre 2017 et 2020 en italien (langue pilote) dans des formations "En route vers B2" déployées dans deux contextes contrastés: en situation collective-présentielle-tutorée, en situation individuelle-distancielle (en autonomie).

Deux types de public ont été impliqués: des étudiants non spécialistes en langues (filière lansad) et des étudiants spécialistes d'italien (filière LLCER – Langues, Littératures, Civilisations Etrangères et Régionales). Au cours de la première phase de la mise à l'épreuve du terrain (collectif-présentiel), les activités prévues dans le parcours ont été présentées par l'expérimentateur au groupe et les réactions des étudiants ont été enregistrées. Cette dynamique a permis d'ajuster la transposition didactique, en conformité avec l'approche théorique adoptée.

En 2017-2018 le panel des contenus d'apprentissage proposés s'est notablement enrichi (dix modules en italien, répartis en quatre parcours) et la démarche expérimentale a intégré le "think aloud protocol". Bénéficiant de ces retours sur expériences, nous avons consolidé la démarche didactique en 2018-2019 (Masperi & alii 2019; Biagiotti & Quaranta 2020), finalisé le développement de l'outillage technique intégré dans la plateforme et mené de nouvelles expérimentations en 2019 (en italien et en anglais) et en 2020 (en italien).

4.1 Architecture

La formation "En route vers B2" (italien langue cible) est structurée en cinq "Parcours". Chaque parcours relève d'un type de discours et se compose d'un ou de plusieurs modules. Cinq types de discours ont été retenus pour le niveau ciblé: EXPLIQUER, RACONTER, INFORMER, ARGUMENTER, JOUER avec la LANGUE. Chaque module présente deux volets: les activités à réaliser (Attività) et un questionnaire d'opinion (Questionario) que les apprenants sont invités à compléter après avoir réalisé les activités (cf. annexe 2 et infra 5.2).

Les activités se répartissent en deux séries: "cerner le genre" et "cerner le type de discours". La série "cerner le genre" est finalisée à la découverte du genre de texte (document vidéo) proposé: i.e. conférence, interview, récit de vie. Les apprenants sont invités à ce stade à identifier des indices de type contextuel qui caractérisent le genre en question, comme le type de média, les caractéristiques visuelles et sonores, le public et les objectifs visés. La deuxième série d'activités – "cerner le type de discours" – oriente quant à elle la compréhension du discours, de manière de plus en plus fine.

Les activités sur le type de discours sont conçues de manière à faire ressortir les éléments langagiers qui sont directement liés au type de discours dominant. Elles guident l'apprenant dans la découverte du texte et sont focalisées en particulier sur la façon dont les énonciateurs construisent leur discours.

4.2 Les outils développés

Faire le choix d'un environnement de formation à distance n'est pas chose aisée, notamment lorsque le présupposé est de contribuer à générer par la réflexion didactique des innovations techniques et à les traduire en fonctionnalités, concrètement opérationnalisées dans une plateforme.

La nécessité de prendre part à la construction d'un Learning Management System orienté "langues" nous a incités dès 2013 à nous tourner vers une plateforme de toute dernière génération, dont le "noyau dur" était tout juste en phase de conception: Claroline Connect². Le panel de fonctionnalités de la plate-forme conçues spécifiquement pour travailler les ressources audio s'est progressivement enrichi à partir de 2018, ce qui nous a permis d'expérimenter nos parcours numériques en langues dans une modalité "tout à l'oral" (cf. annexe 1). Quatre outils de traitement de l'oral par l'oral sont actuellement disponibles pour assister les enseignants dans la didactisation des textes oraux et les étudiants durant le processus de compréhension: les outils "chapitrage" et "bornage audio guidé", intégrés dans un éditeur d'activités, ainsi que la ressource "écoute active enrichie", augmentée d'une fonctionnalité "bornage audio libre".

Nous allons ci-après brièvement les décrire et les illustrer.

- Le "chapitrage" permet d'écouter ou de visionner un extrait audio ou vidéo de manière segmentée, permettant d'isoler et de mettre en évidence des éléments présents dans l'extrait sonore ou visuel. Il peut se révéler particulièrement pertinent lorsqu'il s'agit d'amener l'étudiant à identifier un ou plusieurs segments ("chapitres") relevant d'un "type de discours" dominant. Il est possible de revenir sur un segment autant de fois que nécessaire et l'écouter avec un débit variable. Cet outil est associé à un protocole d'activité spécifique (QRM, QRU, association, classement, ordonnancement, tableau, etc.) en mesure de fournir un feedback.

² <https://claroline.net/>. Ce choix de départ vers cette solution LMS s'est trouvé conforté par les retours d'utilisateurs et d'experts ainsi que par la publication en 2018 du Guide "LMS & MOOC: comment choisir la plate-forme?" édité par le Forum des acteurs de la formation digitale (<http://www.fffod.org/>). Ce document comparatif met en avant de nombreux points de force de Claroline Connect et ceci est d'autant plus encourageant que nombre de ces atouts sont directement issus de notre recherche.

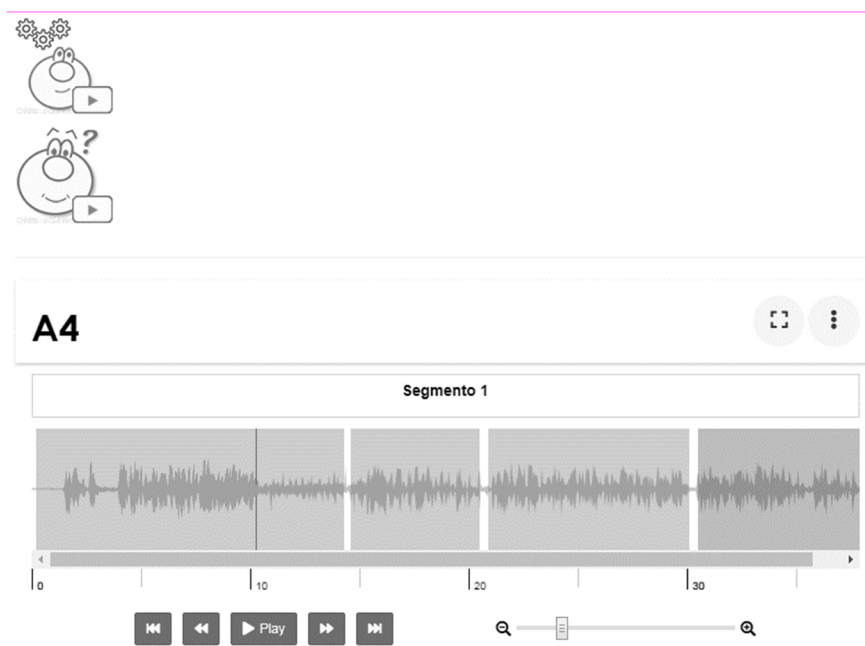


Fig. 1: Capture d'écran de l'outil "chaptirage"

- Le "Bornage Audio Guidé" est un protocole d'activité qui permet à l'étudiant de sélectionner une ou plusieurs zones à l'intérieur d'un extrait audio. La sélection se fait en mettant la ou les zones en surbrillance. Ces zones peuvent être ajustées en utilisant les curseurs et le zoom et réécoutées à volonté, en appliquant si nécessaire un débit variable. Le concepteur détermine la ou les zones de réponse correctes ou incorrectes et fixe une tolérance de dépassement de zone à droite et à gauche, correspondant au début et à la fin du segment. Il ajoute également le feedback à afficher. Les zones sélectionnées par l'étudiant sont analysées et comparées à celles déterminées par le concepteur en tenant compte de la tolérance indiquée dans les paramètres de l'activité.

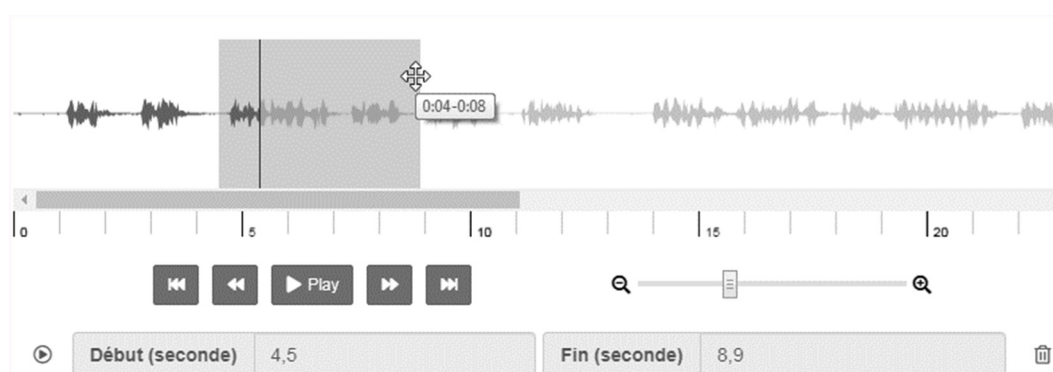


Fig. 2: Capture d'écran de l'outil "Bornage Audio Guidé"

- L'"Ecoute Active Enrichie" est une ressource dotée de fonctionnalités spécifiques d'assistance à la compréhension. Elle permet de mettre en évidence des segments auxquels sont associées des aides. Après avoir lancé l'écoute de l'audio, l'utilisateur peut revenir sur un segment autant de fois que nécessaire et l'écouter avec un débit variable. A chaque segment sont associés différents types d'aides: transcription, aides multimédia (images, vidéos, textes, liens externes, etc.). L'ajout d'un fichier audio complémentaire permet en outre de proposer de réécouter un extrait en jouant sur les variables segmentales et suprasegmentales et sur des paramètres acoustiques (variété standard vs. variété diatopiquement ou diastratiquement marquée, hypoarticulation vs. hyperarticulation; voix masculine vs. féminine; intensité faible vs. moyenne ou forte, etc.). La possibilité est également offerte à l'étudiant d'ajouter un commentaire portant sur l'activité qu'il est en train de réaliser.



Fig. 3: Capture d'écran de l'outil "Ecoute Active Enrichie"

- La fonctionnalité "Bornage Audio Libre" permet aux étudiants de participer activement à l'amélioration de la conception didactique: les segments de texte qui font obstacle à la compréhension peuvent être mis en surbrillance directement sur l'extrait audio (un seul mot; une phase toute entière; un fait intonatif, une variété de langue, etc.) et librement commentés, à l'écrit ou à l'oral en intégrant un fichier audio.

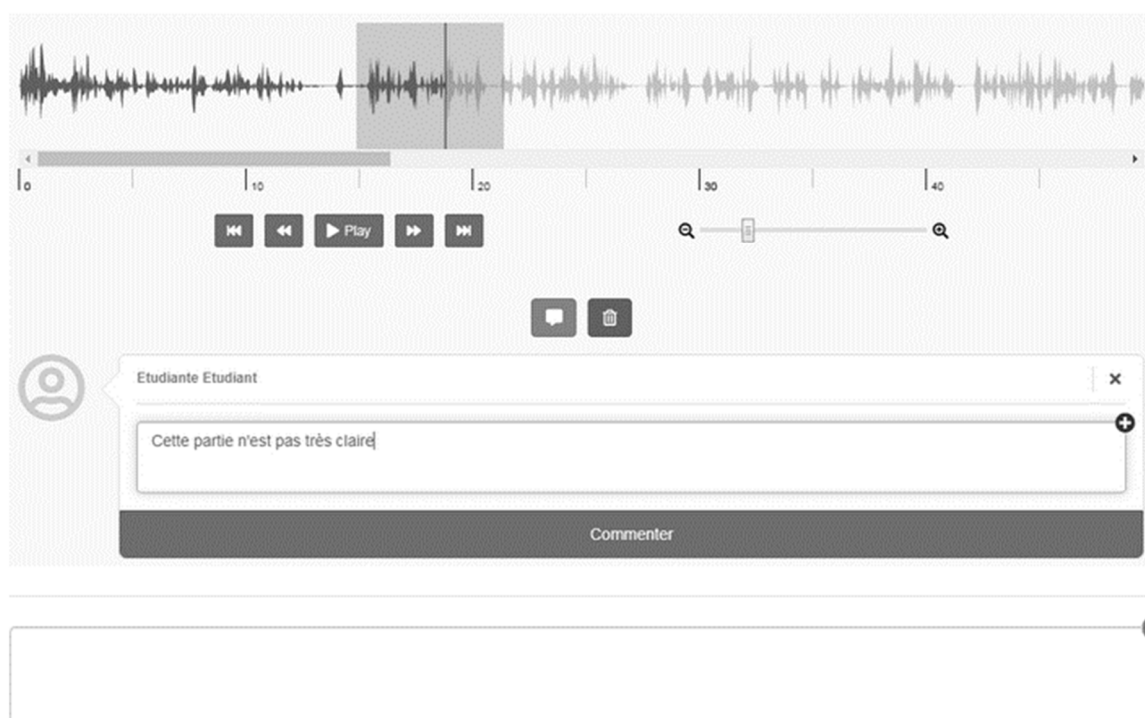


Fig. 4: Capture d'écran de l'outil "Bornage Audio Libre"

5. Mise à l'épreuve du terrain

5.1 Contexte

Au cours du deuxième semestre de l'année universitaire 2019-20, nous avons mis à l'épreuve du terrain deux unités du parcours "Expliquer" (niveau B2), conçues à partir de deux genres de texte différents: une conférence de vulgarisation scientifique "grand public" ayant pour objet le peintre italien Caravaggio³; une interview au réalisateur italien Michelangelo Frammartino au sujet de son film "Le quattro volte".

L'expérimentation a impliqué un échantillon de 53 étudiants ayant un niveau de compétence en italien évalué entre B1 et C1, inscrits dans trois cours d'italien de l'Université Grenoble Alpes relevant de trois filières différentes (spécialistes et non spécialistes de langues).

En l'occurrence, notre public se répartit comme suit:

- 20 étudiants spécialistes de la L2 LLCER Italien (Langues, Littératures, Civilisations Etrangères et Régionales);
- 13 étudiants inscrits au cours thématique 'L'italien à travers l'art' du Service des Langues de l'Université Grenoble Alpes (filière "Lansad");
- 20 étudiants de la L1 de l'Institut d'Etudes Politiques de Grenoble.

³ Animée par l'historien et critique d'art Vittorio Sgarbi dans le cadre du Festival "Incontrarsi a Salsomaggiore" en 2010

Les étudiants ont réalisé les activités individuellement, à distance, via la plateforme Claroline Connect, en totale autonomie. Préalablement à l'ouverture de l'espace de formation, une partie d'une séance en présentiel a été consacrée à l'illustration du contexte de l'expérimentation, de l'approche didactique adoptée ainsi que des principales fonctionnalités de la plateforme, avec une focale sur les nouveaux outils d'assistance et de support à la compréhension de l'oral. Une vidéo-tutorielle, intégrée dans la page d'accueil du parcours, ainsi que quelques consignes techniques à suivre complétaient la préparation au bon déroulement des activités.

Un questionnaire d'opinion (cf. annexe 2) a été soumis à nos étudiants testeurs. Il se compose de sept questions à réponse ouverte et de neuf QCM pour lesquels nous avons adopté une échelle d'évaluation de Lickert à cinq options (allant de l'accord total au désaccord total). Les avis des étudiants ont été sondés principalement autour de deux dimensions: l'adhésion à la démarche didactique adoptée; les retours sur les usages relatifs à l'ingénierie didactique, avec une attention particulière accordée aux éléments d'innovation technique.

Nous avons notamment sollicité des retours sur: les activités de mise en contexte; la pertinence du choix des extraits; le choix du "tout à l'oral"; la segmentation et la longueur des extraits; l'intelligibilité des consignes; l'explicitation du but de la consigne; l'adéquation entre le niveau du public testé et le matériel didactique proposé; l'intérêt pour les ressources vidéos sélectionnées; la possibilité de visionner la vidéo en entier; les points forts et faibles du dispositif; d'éventuels conseils et suggestions d'amélioration du dispositif.



Fig. 5: Image de l'outil 'Bornage Libre' tirée du questionnaire, assortie de la question qui lui est associée.

5.2 Les premières impressions

Bien qu'en situation de travail en autonomie non contraint, les publics testeurs sollicités ont pour la plupart mené à bien les tâches proposées⁴. Nous allons ici rendre compte de quelques éléments saillants qui se dégagent des réponses collectées dans l'unité 1 du parcours 'Expliquer' (Caravaggio), ainsi que dans le questionnaire associé à cette unité (cf. annexe 2). Ces éléments relèvent de la démarche didactique adoptée (points a, b, c, ci-après) et des retours sur les usages relatifs aux éléments d'innovation technique (point d).

5.2.1. Le "tout à l'oral"

Le principal objectif des parcours étant le travail sur la compréhension de l'oral, nous avons pris la décision de proposer tous les éléments qui composent les activités contenues dans chaque unité sous forme orale, sans médiation de l'écrit.

⁴ En référence à l'unité 1 du Parcours Expliquer, achevée à ce jour par nos publics, les taux des réponses aux activités et au questionnaire ont été respectivement de 69,8% et de 45%.

Ce choix assez radical a été apprécié par la majorité des sondés (66,6 %). Le 'tout à l'oral' semble constituer en effet un véritable atout des parcours. A titre d'exemple, voici quelques témoignages:

- Le fait que tout soit à l'oral permet d'être réellement concentré sur ce qui est dit, et sur le but de l'activité [A1019]
- Le point fort est clairement l'aspect tout en audio, qui nécessite une concentration lorsque l'on [n']est pas habitué à écouter de l'italien dans un débit de parole classique voire rapide. Le thème est très intéressant, et le fait d'avoir une évolution de la conférence tout au long des activités est super. Enfin, les différentes formes des activités permettent de travailler la compréhension sous plusieurs aspects et à différents niveaux de concentration. [C3011]
- Le tout à l'oral qui aide à la compréhension [C3015]

5.2.2. Premier contact avec le texte

Nous l'avons évoqué précédemment: les premières activités de l'unité sont censées faciliter l'identification du genre du texte à travers la mise en évidence d'indices de type contextuel (i.e. le statut du locuteur, les attentes envers les destinataires, le contexte physique et social de l'énonciation etc.).

En l'occurrence, l'extrait qui sert de support aux premières activités de l'unité introduit le lieu social de production de la conférence (le Festival "Incontrarsi a Salsomaggiore"). Il donne à voir le logo et le titre du Festival, le titre de la série des conférences programmées ("L'incontro come ultima spiaggia prima dello scontro") ainsi que le titre de la conférence en question ("Caravaggio, lectio magistralis di Vittorio Sgarbi").

Lors de ce premier contact avec le texte, il est demandé aux étudiants de formuler des hypothèses sur le genre de la vidéo qui leur est proposée. Un QCM (activité 1) propose à cet égard une série d'options parmi lesquelles les étudiants sont invités à choisir celles qui se rapprochent le plus de leurs hypothèses personnelles sur le genre. L'activité 2 se présente en revanche sous la forme d'une question ouverte: les étudiants sont ici invités à formuler librement leurs hypothèses sur le genre.

Nous observons à cet égard que les étudiants ont plutôt adhéré à cette démarche: ils ont en effet formulé, en italien⁵, leurs hypothèses, de manière plus ou moins poussée.

Certains d'entre eux se sont appuyés de manière explicite sur les indices contextuels présents dans l'extrait:

- La video puo' essere una conferenza, una lezione [A1013]
- Secondo me è una conferenza filmata [B2029]

alors que d'autres ont fait appel à leurs propres références culturelles:

- Mi sembra di essere un video abbastanza ufficiale, formale [D1012]

⁵ Ont été rapportées ici les productions originales des étudiants, d'où les nombreux écarts à la norme.

- Non ho idea come questa video potrebbe essere, non so chi Caravaggio è e non ho alcuni elementi per cominciare a riflettere [D1016]
- Penso che questo video presenterà la vita dell'artista e quindi il contesto in cui ha dipinto le sue opere perché trovo che sembra un inizio di reportage con la società di produzione all'inizio e una frase di richiamo poi [A1011]

D'autres encore ont poussé leurs hypothèses jusqu'à évoquer le type de discours dominant:

- Il genere sarà piuttosto spiegativo che argomentativo perché è una leziokne da un professionista. Il presentatore darà probabilmente anche delle informazioni e raconterà, ma il genere principale sarà la spiegazione [C3025]

Deux autres questions, intégrées dans le questionnaire, nous fournissent des compléments de réponses au sujet de ces activités. Il s'agit en l'occurrence d'une question fermée (avec cinq options de réponse) et d'une question ouverte⁶.

Les témoignages collectés dans les réponses ouvertes tendent à conforter l'intérêt de nos orientations didactiques:

- Je suis plutôt d'accord car cela permet de réfléchir à plusieurs possibilités, tout en écoutant beaucoup plus attentivement les vidéos [A1019]
- Oui car elles permettent de mieux rentrer dans le texte, à comprendre le contexte [C3010]
- Je trouve que c'est une bonne méthode pour améliorer les capacités de concentration lors de l'écoute [B2010]

5.2.3. Segmentation du texte

Comme nous l'avons souligné plus haut (cf. infra 3), nous avons fait le choix de segmenter la vidéo et d'établir des activités numériques visant à faciliter la compréhension d'un segment après l'autre. Les étudiants sont donc amenés à la découverte progressive du texte qui leur est présenté au fur et à mesure qu'ils effectuent les différentes activités. Les critères de la segmentation respectent toutefois la contrainte de préserver et de souligner l'effet de cohérence que le texte produit dans la continuité.

Cette prise de position didactique a été particulièrement appréciée par les étudiants, qui ont décidé de valoriser la segmentation textuelle dans leurs réponses à la question sur les points forts de l'unité. En voici quelques-unes:

- Le fait de couper la vidéo en plusieurs parties et de travailler pendant plusieurs questions sur une même partie permet d'organiser ce que l'on écoute et de bien comprendre et retenir ensuite [A1010]
- Une meilleure compréhension de l'oral grâce aux découpages pertinents de la vidéo [A1021]
- Je pense que le fait d'étudier par petit segment la vidéo est un bon moyen de tout comprendre et de rester focaliser petit à petit sur toute la vidéo et tout le long de l'activité. Avoir des activités

⁶ Il s'agit des questions n° 5 et 5bis du questionnaire d'opinion (cf. annexe 2).

différentes (des questions, le bornage, relier des segments) rend le tout très ludique [A1022]

- Ces activités m'ont permis de comprendre quels étaient les éléments essentiels à retenir des différents extraits alors qu'en général j'ai beaucoup de mal à discerner l'important du superflu [B2026]

5.2.4. Outillage technique

Un autre aspect important sur lequel nous avons voulu questionner les étudiants concerne leurs avis par rapport aux innovations techniques proposées dans l'unité en question, à savoir, le Chapitrage, le Bornage Guidé et le Bornage Libre.

L'utilisation du Chapitrage avait déjà fait l'objet de quelques retours sur expériences, avec des résultats globalement satisfaisants (Masperi & alii 2019). Cette tendance s'est à nouveau confirmée:

- Ce qui est selon moi positif est le découpage des extraits qui permet de faire des pauses pendant l'écoute afin de stimuler la concentration. Le chapitrage m'a surtout servi d'indicatif de temps. Je trouve que le système fonctionne plutôt bien [B2026]
- Le Chapitrage m'a permis de revenir plus facilement sur les parties qui me posaient problèmes, celles que je ne comprenais pas bien [A1014]
- C'était très pratique, et ça permettait de se concentrer sur des petits extraits plutôt que d'être "submergé" par un trop d'informations [C3025]

Par rapport au Bornage Guidé, les avis sont plus partagés.

D'une part, la possibilité de sélectionner en surbrillance le segment correspondant à la réponse attendue, plutôt que d'effectuer un choix parmi des segments prédéfinis, a été mise en valeur:

- Le Bornage était également simple d'utilisation, et très pratique. C'est un exercice qui permet de varier le type de réponses à donner [A1019]
- Le bornage était très amusant à faire et permettait de réécouter le passage en entier [C3010]
- Très bien, l'outil était facile à utiliser et permet d'individualiser les réponses plutôt que de proposer des réponses déjà faites [C3025]
- C'est pratique, on peut écouter plusieurs fois et faire des modifications [B2021]

D'autre part, une telle sélection se révèle toutefois à l'usage plus difficile à opérer:

- Je trouve que ce système fonctionne bien même si parfois j'avais du mal à comprendre les différentes propositions à cause de mon manque de vocabulaire. Sinon je trouve que c'est très intéressant de proposer un système de bornage, mais il m'est finalement assez difficile d'identifier exactement les passages demandés [B2026]

Par ailleurs, une première analyse des réponses a mis en évidence que les sélections des étudiants étaient parfois considérées comme incorrectes alors qu'elles ne s'écartaient que de quelques millisecondes de ce qui avait été défini comme étant correct en phase de conception. De tels feedbacks binaires

(correct vs. incorrect) peuvent être source de frustration et s'avérer préjudiciables en phase de découverte d'un nouvel outil. Ces remarques nous incitent à nous pencher sur des questions – importantes mais jusque-là éludées – d'évaluation formative ainsi que sur les moyens de rendre l'utilisation de notre outillage technique globalement plus performante.

Ensuite, à la fin de l'unité en question, nous avons intégré trois Bornages Libres sur trois différents extraits de la vidéo. Comme le montrent les réponses collectées dans le questionnaire qualitatif, l'utilisation de cet outil a eu de très bons retours. Les étudiants ont particulièrement aimé le fait d'être amenés à réfléchir et à s'exprimer sur les écueils qu'ils pouvaient rencontrer lors du visionnage de chaque extrait:

- Le bornage libre m'a permis de surligner rapidement les passages qui m'ont posé problème lors de l'écoute. Utilisation relativement facile [C3012]
- J'ai bien apprécié cette fonctionnalité, ça permettait de clarifier exactement pourquoi est ce qu'on ne comprenait pas quelque chose [C3025]
- Le bornage libre est une fonctionnalité intéressante, elle permet de se rendre compte de ce que l'on comprend et de ce que l'on ne parvient pas à saisir. Généralement, ce sont surtout les termes et la rapidité d'élocution qui posent problème [C3011]

6. Pour (ne pas) conclure

Le projet que nous portons a l'ambition de contribuer à développer des réflexions, analyses et actions didactiques de soutien aux acquisitions en L2, notamment dans la sphère de l'oralité. A ce titre, les démarches présentées sont envisagées comme une recherche de solutions susceptibles de mieux rencontrer les besoins qui émergent d'un groupe ou d'un environnement socio-éducatif donné.

Si la création de "nouveaux" outils numériques participe à l'élaboration du dispositif, la focale portée sur la visée sociale et culturelle de l'intervention didactique place clairement les moyens technologiques au service de la finalité du projet (Masperi & Quintin 2014). Ce positionnement sur l'échiquier de la didactique des langues est loin d'être trivial: une réponse qui se veut "innovante" à des attentes et à des questionnements soulevés dans un domaine aussi "brumeux" que le traitement de l'oral en L2 nécessite la convocation d'un outillage scientifique complexe – à la croisée de la psychologie du langage, de la sémantique, de l'analyse du discours, de la multimodalité – une solide reconfiguration didactique, une vision claire et concrète des potentialités que peut offrir le numérique et une mise à l'épreuve sur le terrain.

Si nous poursuivons cette voie aujourd'hui, c'est que nous croyons en la plus-value d'une démarche qui se démarque des approches techno-actionnelles de ces vingt dernières années. Sans minimiser ou déprécier l'impact des recherches adossées aux courants générés par le CECRL et ses modélisations,

nous soutenons toutefois la nécessité de penser à la didactique des langues d'une manière moins compartimentée, moins échantillonnée et à la fois plus trans-langagière et plus ouverte sur la culture de chaque langue-cible.

C'est la raison pour laquelle nous plaçons le texte, avec ses déclinaisons en genres et en discours, résolument au centre de notre action didactique. En cela, nous nous efforçons de répondre de manière complémentaire au développement de cultures langagières larges, allant du linguistique au textuel, tout en établissant des ponts explicites et visibles entre langue(s) première(s) et langue(s) seconde(s), ainsi qu'entre L2 entre elles.

Par ailleurs, nos dispositifs reposent sur des innovations didactiques et techniques qui se veulent ouvertes aussi bien vers l'extérieur (par une mise à disposition au sein de la communauté éducative) qu'à l'extérieur (par une ouverture à toute forme de collaboration, avec des chercheurs, praticiens et Institutions) (Ibidem 2014: 12). Notre volonté est en effet d'enrichir ces parcours par les contributions d'enseignants/chercheurs qui souhaiteraient s'inscrire dans cette démarche didactique et participer à la conception de nouveaux modules de formation.

En guise de perspectives, signalons enfin que nous œuvrons afin d'assortir le dispositif d'un corpus organisé par genres textuels et types de discours (et avec des préconisations d'usage par niveau) de vidéos illustratives et prototypiques, sous licence libre ou obtenues avec cession des droits d'exploitation. Au sujet de cette banque de données originales, nous avons bon espoir d'aboutir à une mutualisation inter-Établissements.

BIBLIOGRAPHIE

- Adam, J.-M. (2011). *Genres de récits. Narrativité et généricité des textes*. Louvain-la-Neuve: Editions Academia.
- Adam, J.-M. (2014). Le paradigme du texte: regard rétrospectif et perspectives pour les sciences des textes. In M. Monte & G. Philippe (éds.), *Genres & Textes* (pp. 297-323). Lyon: Presses universitaires de Lyon.
- Adam, J.-M. (2015). *Faire texte. Frontières textuelles et opérations de textualisation*. Besançon: Presses universitaires de Franche-Comté.
- Biagiotti, T. & Quaranta, G. (2020). Approcher les textes par les notions de "genres" et de "types de discours": présupposés didactiques et mise en œuvre au sein d'un environnement numérique. In R. Comotti, W. E. Ouedraogo, C. Wolfarth & C. Noûs (éds), *Qu'a changé le numérique ? Perspectives linguistiques et didactiques*. Colloque international des Etudiant.e.s chercheur.se.s en Didactique des langues et Linguistique, CEDIL'18, May 2018, Grenoble, France, 2020. <http://hal.univ-grenoble-alpes.fr/hal-02641560/document>
- Bronckart, J.-P., Bain, D., Schneuwly, B., Davaud, C. & Pasquier, A. (1985). *Le fonctionnement des discours*. Neuchâtel-Paris: Delachaux & Niestlé.
- Bronckart, J.-P. (1997). *Activité langagière, textes et discours - Pour un interactionnisme socio-discursif*. Paris: Delachaux & Niestlé.

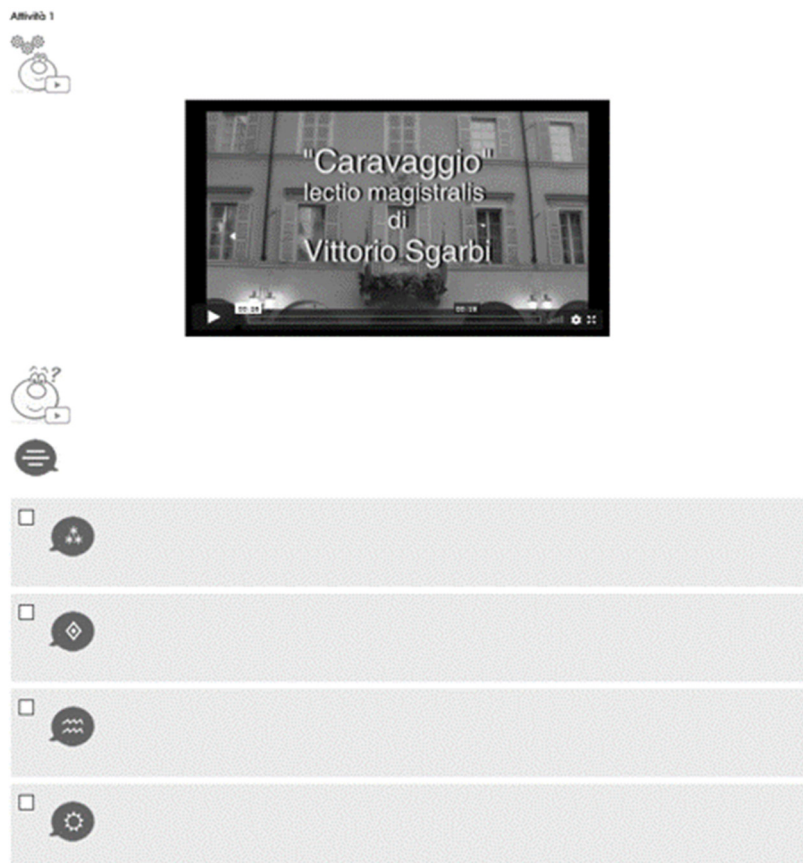
- Canelas-Trevi, S. (2009), *La grammaire enseignée en classe. Le sens des objets et des manipulations*. Berne: Peter Lang.
- Cornaire, C. & Germain, C. (1998). *La compréhension orale*. Paris: Clé International.
- De Pietro, J.-F., Fisher, C. & Gagnon, R. (2017). *L'oral aujourd'hui: perspectives didactiques*. Namur: Presses universitaires de Namur.
- Field, J. (1998). Skill and strategies: towards a new methodology for listening. *ELF Journal*, 52(2), 110-118.
- Field, J. (2000). Finding One's Way in the Fog: Listening Strategies and Second-language Learners. *Modern English Teacher*, 9(1), 29-34.
- Field, J. (2008). *Listening in the language classroom*. Cambridge: Cambridge University Press.
- Field, J. (2015). *Dans la tête de l'auditeur: le défi de la compréhension orale d'une langue étrangère*. Conférence donnée à l'UGA dans le cadre du projet Innovalangues. <https://innovalangues.univ-grenoble-alpes.fr/innovalangues/contributions-scientifiques/videos-et-podcasts/dans-la-tete-de-l-auditeur-le-defi-de-la-comprehension-orale-d-une-langue-etrangere-699872.kjsp?RH=4830846371796247>
- Field, J. (2019). *Rethinking the Second Language Listening Test: from theory to practice*. Sheffield: Equinox Publishing.
- Gagnon, R., De Pietro, J.-F. & Fisher, C. (2017). L'oral aujourd'hui: perspectives didactiques [introduction]. In J.-F. de Pietro, C. Fisher & R. Gagnon (éds.), *L'oral aujourd'hui: perspectives didactiques* (pp. 11-40). Namur: Presses universitaires de Namur. (Recherches en didactique du français 9).
- Hilton, H. (2000). La Didactique de la compréhension aurale: une approche stratégique. *Annales de l'Université de Savoie*, 28, 23-39.
- Kerbrat-Orecchioni, C. & Traverso, V. (2004). Types d'interactions et genres de l'oral. *Langages*, 153, 41-51.
- Lopriore, L. (2011). La comprensione orale dall'ascolto all'interazione: aspetti didattici e valutativi. In P. Leone & T. Mezzi (éds), *Didattica della comunicazione orale. Lingue seconde e italiano a scuola e all'università* (pp. 21-38). Milano: Franco Angeli Edizioni.
- Masperi, M. (2011). *Innovalangues: Innovation et transformation des pratiques de l'enseignement-apprentissage des langues dans l'enseignement supérieur*, en ligne sur HAL: <https://hal.archives-ouvertes.fr/hal-02004250>
- Masperi, M. & Quintin, J.-J., (2014). L'innovation selon Innovalangues, *LEND – Lingua E Nuova Didattica*, 1, 6-14. <https://hal.archives-ouvertes.fr/hal-01955544>
- Masperi, M., Biagiotti, T., Quaranta, G. & Barletta, M. (2019). Appréhender la diversité langagière et culturelle au fil des textes et des discours: une entrée par la notion de "genre". In: *Des genres en Méditerranée: pratiques, représentations et transfert*, ILCEA, 37. <https://journals.openedition.org/ilcea/7430>
- Nunan, D. (2002). Listening in Language Learning. In J. C. Richards & W. A. Renandya (éds), *Methodology in Language Teaching. An Anthology of Current Practice* (pp. 238-241). Cambridge: Cambridge University Press.
- Philippe, G. & Adam, J.-M. (2015). Continuité et textualité. In J.-M. Adam (éd), *Faire texte. Frontières textuelles et opérations de textualisation* (pp. 35-80). Besançon: Presses universitaires de Franche-Comté.
- Rastier, J.-F. (1989). *Sens et textualité*. Paris: Hachette.
- Rastier, F. (2007). Eloge paradoxal du plurilinguisme. In *Texto!*, 3(XII).
- Rastier, F. (2015). *Saussure au futur*. Paris: Edition Les belles lettres.

- Roussel, S. (2014). A la recherche du sens perdu: comprendre la compréhension de l'oral en langue seconde. *La Clé des Langues* (Lyon: ENS LYON/DGESCO).
- Thévenaz, T. & Canelas-Trevisi S. (2018). L'écriture et la lecture littéraire: une relation qui se transforme. In C. Ronveaux & B. Schneuwly (éds.). *Lire des textes réputés littéraires: disciplinaton et sédimentation*. Berne: Peter lang.
- Vandergrift, L. (1997). The Comprehension Startegies of Second Language (French) listeners: A Descriptive Study. *Foreign Language Annals*, 30, 387-409.
- Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language teaching* 56, 431-462.

Annexes

Annexe 1

Capture d'écran d'une activité du parcours 'tout à l'oral'



Légende:



Capacité



Consigne



Amorce de la question



Options de réponse (elles sont représentées par des images différentes afin de faciliter le repérage des contenus)

Annexe 2 – Questionnaire

Unité 'Caravaggio' Parcours 'Expliquer' Niveau B2

1. De quelle manière le Chapitrage vous a aidé lors de la compréhension de la vidéo ? Avez-vous apprécié les différentes fonctionnalités disponibles ? (Réponse ouverte)

2. Comment avez-vous utilisé le Bornage (protocole question-réponse) ? Avez-vous apprécié les différentes fonctionnalités disponibles ? (Réponse ouverte)

3. Comment avez-vous utilisé le Bornage Libre ? Avez-vous apprécié les différentes fonctionnalités disponibles ? (Réponse ouverte)

4. Avez-vous rencontré des problèmes techniques lors du déroulement des activités en ligne ? Laissez un commentaire pour nous aider à les résoudre. (Réponse ouverte)

5. Les hypothèses à faire au début de l'unité m'ont aidé à mieux rentrer dans le texte.

Accord total	Plutôt d'accord	Moyennement d'accord	Plutôt pas d'accord	Désaccord total
--------------	-----------------	----------------------	---------------------	-----------------

5bis. Si oui, pourquoi ? (Réponse ouverte)

6. La vidéo proposée est intéressante.

Accord total	Plutôt d'accord	Moyennement d'accord	Plutôt pas d'accord	Désaccord total
--------------	-----------------	----------------------	---------------------	-----------------

7. La vidéo me paraît en adéquation à mon niveau d'italien.

Accord total	Plutôt d'accord	Moyennement d'accord	Plutôt pas d'accord	Désaccord total
--------------	-----------------	----------------------	---------------------	-----------------

8. Je pense qu'il est utile d'avoir la possibilité de visionner la vidéo en entier.

Accord total	Plutôt d'accord	Moyennement d'accord	Plutôt pas d'accord	Désaccord total
--------------	-----------------	----------------------	---------------------	-----------------

9. La visée explicative de la vidéo émerge de manière claire.

Accord total	Plutôt d'accord	Moyennement d'accord	Plutôt pas d'accord	Désaccord total
--------------	-----------------	----------------------	---------------------	-----------------

10. Les extraits proposés ont une longueur raisonnable.

Accord total	Plutôt d'accord	Moyennement d'accord	Plutôt pas d'accord	Désaccord total
--------------	-----------------	----------------------	---------------------	-----------------

11. Je trouve les consignes claires.

Accord total	Plutôt d'accord	Moyennement d'accord	Plutôt pas d'accord	Désaccord total
--------------	-----------------	----------------------	---------------------	-----------------

12. La désignation de la "capacité" a été utile pour comprendre les activités.

Accord total	Plutôt d'accord	Moyennement d'accord	Plutôt pas d'accord	Désaccord total
--------------	-----------------	----------------------	---------------------	-----------------

13. Lors du déroulement des activités, le choix du "tout à l'oral" me semble utile.

Accord total	Plutôt d'accord	Moyennement d'accord	Plutôt pas d'accord	Désaccord total
--------------	-----------------	----------------------	---------------------	-----------------

14. Quel protocole d'activité avez-vous particulièrement apprécié et pourquoi ? (Réponse ouverte)

15. Quels sont pour vous les points forts des activités proposées ? (Réponse ouverte)

16. Pourriez-vous suggérer des améliorations ? Si oui, lesquelles ? (Réponse ouverte)