

DOI: <https://doi.org/10.31392/NPU-nc.series9.2021.22.08>.

UDC: 811.133.1'33

**Iryna V. Strashko**

National Pedagogical Dragomanov University,  
Kyiv, Ukraine



## THE USE OF SPOKEN CORPORA FOR LEARNING FRENCH AS A FOREIGN LANGUAGE (BASED ON ESLO, FLEURON, CLAPI, AND CFPP2000)

### **Bibliographic Description:**

Strashko, I. V. (2021). The Use of Spoken Corpora for Learning French as a Foreign Language (based on ESLO, FLEURON, CLAPI, and CFPP2000). *Scientific Journal of National Pedagogical Dragomanov University. Series 9. Current Trends in Language Development*, 22. 110–122. <https://doi.org/10.31392/NPU-nc.series9.2021.22.08>.

### **Abstract**

*The paper deals with the educational and linguistic potential in use of oral corpora for learning French as a foreign language. The paper is divided into three parts. The first part illustrates the contribution of studies based on oral corpora in foreign language acquisition, French in particular. It also analyses the integration of corpora in educational process. The second one gives a brief overview on four speech corpora (the ESLOs, FLEURON, CLAPI, and CFPP2000), shortly describes their specifics, thematic orientation, some peculiarities of working with them. It also gives examples of concrete educational applications for vocabulary learning. The third part looks at some aspects of the use of corpora for language study, highlighting its advantages, difficulties, and challenges of oral corpora employment. Here it is emphasized the use of speech corpora in practice of teaching and learning French language learning is an effective addition to traditional educational technologies. They can be used as supporting material, which allows the teacher to correlate the learning objectives and the teaching methods. The way of working with them requires rethinking the way of teaching and learning: this approach combines empirical adequacy, authenticity, and adaptation, and the possibility of independent learner's work. It also implies technical, theoretical, methodological, and educational support for both the teacher and the learner. The use of these corpora can be effective and interesting if it bases on a clear program and is gradual, consistent, and systematic.*

**Keywords:** spoken corpus, corpus linguistics, French as a Foreign Language, learning resources.

### ***1. Introduction.***

The oral language is the main form of language existence. In comparison with written language it remains, as stated A. Martinet, first phylogenetically and ontogenetically (Surcouf & Ausoni, 2018, p. 73). However, studies on it lag behind. It is explained in first of all, still insufficient volume of a representative array of oral texts, although their number is steadily increasing, as well as technical, ethical, logical and methodological difficulties that stand in the way of a systematic analysis of oral speech.

C. Blanche-Benveniste remarks that for the general public, the opposition between the written language and the spoken language has long been, “a matter of combat between the good and the evil: spoken language spontaneous, possibly picturesque but definitely faulty; written language policed, testifying, especially due to the orthography, the true grammar of the language. The very notion of spoken language is often linked to the negative aspects of language: fault, incompleteness, peculiarities of delinquent suburbs, etc.” (Blanche-Benveniste, 2000, p. 5). Whereas the writing, by its capacity to fix definitively a language production, makes it observable and analyzable at will. Thus it naturally becomes a privileged object for analysis for the linguist, the grammarian and the pedagogue (Surcouf & Ausoni, 2018, p. 74).

Numerous projects for the constitution of oral corpora have appeared in France, this rise has amplified, having made possible to take the measure of the scattering of accessible data at the level of physical locations, materials, formats (Cappeau, & Gadet, 2007). This development was accompanied by a reflection on the challenges and the specifics of their constitution, the heaviness of the task and the quality data produced. (Baude, 2006; Bilger, 2008).

The use of corpora for the language acquisition purposes were analyzed in a number of research (Дерябина, 2012; Нагель, 2008; Садовникова, 2013; Сысоев, 2010; André, 2018, 2020; De Cock & Tyne, 2014; Chambers, 2009; Di Vito, 2013; Ravazzolo, Étienne & Ursi, 2021; Surcouf & Ausoni, 2018). As for implementing corpus methods for teaching and learning French we can name such researchers as V. André, S.Vito (André, 2018; Di Vito 2013). In the context of this work the most fruitful is the latest André's study (André, 2020) that concentrates particularly on possibilities for teaching and learning languages through the oral corpora use.

The research carried out by the LLL (Le Laboratoire Ligérien de Linguistique), as well as that of ICAR team (Interactions, Corpus, Learning, Representations) is particularly interesting for our study. Theoretical and methodological principles of the ESLOs corpora creation, as well as technical aspects of their use, the structure description and data process were described in detail in a number of research (Bergounioux, Baraduc & Dumont, 1992; Abouda & Baude 2005, 2007, 2009; Baude & Dugua, 2011, 2016; Eshkol-Taravella, Baude et al., 2011; Eshkol, 2015). The ESLOs provide a rich field for the study of French spoken language. Research carried out on this material in the field of vocabulary, grammar, pragmatics, discourse has already yielded interesting results (Baude, 2015; Bergounioux, 2016, 2018; Rendulić, Kanaan-Caillol, 2016; Abouda, Skrovec, 2018). As for ICAR team, we can cite the studies of a number of research (Alberdi, Etienne et al., 2018; Ravazzolo, Etienne, 2019; Ravazzolo, Etienne & Ursi, 2021). This research introduces the corpora excerpts and collections available in CLAPI-FLE platform, develops the contribution of these resources for different learning purposes, provides the methodological approach which led to the constitution of the CLAPI-FLE platform, uses of educational resources from CLAPI-FLE, in a classroom setting for different educational objectives, etc. In terms of a particular corpus specifics, the work of a number of researchers (Branca-Rosoff, et al., 2012) is very relevant.

To date, the problem continues to be the most studied on the basis of English language, well studied on the basis of Russian, there are few studies devoted to the German language, while the French language still does not know such coverage, especially from Ukrainian researchers. Although in recent years they have shown a keen interest in corpus linguistics, corpus-based studies for learning French are still rare. Therefore, the relevance of research is determined both by its object as in contemporary Ukrainian linguistics there are no corpus studies of a similar type based on the French language learning, and by the experimental base used to reveal its potential of teaching and learning. Thus, in a certain way this paper contributes to further studies of corpus-based speech studies in general and to studies French as a foreign language, in particular.

## **2. Aim and Objectives.**

The paper *aims* to analyze the use of spoken corpora for learning French as a foreign language.

The paper attempts to achieve the following:

- give a summary of current trends in corpus-based language learning;
- integrate theories from corpus-based research, linguodidactic studies, and sociolinguistics to language data processing;
- reveal and describe the potential of corpus-based learning for French language acquisition;
- give examples of the use of spoken corpora.

## **3. Methods.**

The study is based on the following research methods and techniques:

- descriptive method (including observation, generalization, and interpretation);
- verification method based on the use of the corpora sound materials;
- comparative analysis.

## **4. Results and Discussion.**

The corpus approach in language learning is based on the possibility to access a large amount of linguistic data. In this sense, the vast oral corpora allow to observe, analyze and compare the characteristics proper to authentic speech. The language material from oral corpora, as point out some researchers (Ravazzolo, Étienne, & Ursi, 2021) presents a variability and variations, which can be observed in the very nature of the interactions, depending on their character, context, format, type of presence, of the register, the number of speakers and their mother tongue. However, emphasizes V. André their use for educational purposes is neither obvious nor easy (André, 2020). Therefore, the question about the possibilities and the limits of the corpus-based learning methodology, as well as its theoretical benchmarks needs to be analyzed first.

Some researchers consider that corpus linguistics is as an applicable methodology for different disciplines (McEnery & Wilson 2001), others think that corpus linguistics is an independent discipline, with its own object (Sinclair, 1991; Leech, 1992; Tognini-Bonelli, 2001). These two approaches are not totally contradictory and can be complementary (Плунгян, 2008; Williams, 2005). In line with the first approach T. Johns (Johns, 1991) proposes to use corpus linguistics as a practical laboratory for language learning. S. Bernardini (Bernardini, 2004, p. 17) states that the analysis of the concordances is an ideal instrument to observe and learn the associations between the forms and contents that are adopted in a linguistic community.

Therefore, from the beginning corpus linguistics has been considered not only as an instrument to observe and analyze linguistic phenomena, but also as a methodology to promote a corpus approach, which is focused on applied language learning. This approach is called the “Three I’s” (illustration – interaction – induction) (McEnery & Xiao, 2010). Illustration concerns real data that are shown to learners; interaction involves discussion and exchanging opinions; the induction allows learners to create their own view on the studied phenomena. Speaking about the use of corpora in language teaching and learning, we mean both corpus tools and corpus approach.

When analyzing the available oral corpora A. Chambers arises the questions about the types of corpus that would best correspond to the various situations of learners. She states that their needs are so diverse that it is almost impossible to conceive and create a corpus, which meets all of them (Chambers, 2009, p. 17). On the other hand, André argues that the data collected and processed by corpus linguistics can compensate for the lack of oral resources representative of ordinary communication situations or adapted to the objectives of the learners (André, 2018).

I. Deryabina brings attention to the fact that corpora use (and she speaks about text corpora) also includes the identification of shortcomings in existing materials for teaching foreign languages. Drawing on the work of other researchers, she states that there is every reason to believe that “there are huge differences between the data in the textbooks and the use of speech samples by the native speakers, and this is confirmed by the corpora” (Дерябина, 2012, p. 158). She comes to a conclusion that corpora should be used for the development of teaching materials, as textbooks may not cover the important aspects, they may highlight phrases that are rarely used, which results in misleading and reduce the effectiveness of language learning (ibid).

An alternative to the use of language textbooks seems to be oral corpora that can be used for educational purposes. Data in the oral corpora present the samples of real colloquial French, which greatly differs from “ordinary French”. Whereas audio recordings, which accompany texts or dialogues in textbooks, despite their certain closeness to natural speech, still do not contain such a range of characteristic features inherent to natural speech behavior. It is about disfluencies at the level of phrases and words, enunciation marks, phatic elements and fillers at the phrasal level, the connectors at the proposal level and the digressions at the level of speech (Bergounioux & Eshkol, 2016).

The other question, which is particularly acute is authenticity of the recourses: authenticity of the texts (oral or written) in general and more particularly the authenticity of the corpora themselves. V. Chambers points out that the definitions of authenticity in language teaching embraces the text itself and the learner’s objectives. That is, oral or written text produced to meet real needs for communication, which in turns implies the “context”. In this case the authenticity of the activity of the learner and the understanding of the context, the study of an individual text, written or oral, extracted from a corpus, well represents an authentic activity. (Chambers, 2009, p. 18-19). V. André discussing the question of the authentic document used in foreign language class with reference to Abé and Cuq recalls that the authentic document is “produced for purposes other than learning of a second language”, it is a “message developed by francophones for francophones for real communication purposes” (Cuq 2003, p. 29). The use of the authentic documents is linked to the wish to exhibit and confront the learners with target language. But, as states V. André, the authenticity is not all, it is important to know how to do appropriate activities based on the language so that the documents enter into a didactic process (André, 2020, p.39).

The understanding of the features of oral language is important in foreign language teachers training, as they have to explain and demonstrate how the language works.

Moreover, these facts of oral speech production, as states V. André, generally go against what traditional grammars and textbooks present or are just not mentioned (André, 2018, p. 75). As affirms V. André access to large corpora modifies learners' sociolinguistic competence, as they have the possibility to observe, compare, analyze the functioning of the language "in cotext and in context", which facilitate its appropriation (ibid, p.81). What is important that oral socio-language skills are generally less worked than those in writing, precisely because they are less described (André, ibid), whereas learners' ability to communicate in the contemporary multicultural world will largely depend on their level of sociocultural competence.

Although the use of corpora for educational purposes seems promising and is often seen as opening up new possibilities and new perspectives, it deserves a reflection on the pedagogical practices that it involves. As simple exposure to the real language, and an enormous amount of linguistic data, is not sufficient to ensure efficient language learning (André, 2020, p. 41). For that reason, it is essential to consider some issues of a methodological nature concerning the corpus-based learning applied to learning French. That is about training in corpus use for all participants (teachers and learners), about students to be familiar with the key concepts of corpus analysis, such as the concordance, the frequency of words or expressions. Also it is crucial to filter the corpus as a number of occurrences may discourage the learner. This is extremely important because referring to J. Sinclair's words, on the basis of corpus data it is as easy to obtain meaningless conclusions as well as profound ones (Садовникова, 2013).

Beyond theoretical reflection on the issues of the exploitation of spoken corpora, the authenticity of their recourses, it is necessary to give an overview of the target corpora as well as examine the criteria to analyze the linguistic data.

Among the oral and multimodal corpora that can be used in teaching practices for French language learning, we can cite ESLO (<http://eslo.huma-num.fr>), FLEURON (Français Langue Etrangère Universitaire (<https://fleuron.atilf.fr/>), CLAPI (<http://clapi.ish-lyon.cnrs.fr/>), CFPP2000 (<http://cfpp2000.univ-paris3.fr/>).

It worth to mention at least two technical platforms, which give access to a list of speech corpora, in particular for French language. The first is the COCOON platform (that stands for "Digital Oral Corpus Collections", in French – "Collections de COrpus Oraux Numériques") that supports oral resource producers in creating, structuring and archiving the corpora (<https://cocoon.huma-num.fr/exist/crdo?%3Flang=en&lang=fr>). The other one is the ORTOLANG (stands for "Open Resources and TOols for LANGuage", in French – "Outils et Ressources pour un Traitement Optimisé de la LANGue"). It is a network infrastructure that includes among others (ex. text corpora, lexicons, dictionaries etc.) a number of speech corpora, in particular for French language (<https://www.ortolang.fr/>).

In the educational perspective a particular interest presents the ESLOs. The Socio-Linguistic Surveys in Orléans (fr. Les Enquêtes sociolinguistiques à Orléans – ESLO) form a large speech corpus of around seven million words. The corpus has been compiled in two stages. The first survey (ESLO 1), a closed corpus, started at the end of the sixties, was initiated by British researchers with subsequent participation of French academics. The second (ESLO 2) corpus, still in progress, started at the beginning of 2000, started in 2008 by the LLL team. "Firmly anchored in the current of sociolinguistics and variationist linguistics, the ESLOs form the basis of studies on the French spoken language in Orléans in a perspective that puts data at the heart studies on the social nature of language" (Baude & Dugua, 2016, p.1). Quantitatively significant "local" language data, formed within an interval of 40 years apart, whose production is contextualized in the form of metadata, not only allow the ESLOs to occupy a unique place in the landscape of databases on oral French,

but also give them special importance in the emerging field of study of micro-diachrony (Abouda & Skrovec, 2018, p. 1). The concept of “sound portrait of the city”, which formed the basis of ESLO 1, resulted in a combination of representativeness of language material and variations within a particular community in a clearly defined geographical and socio-economic space (Bergounioux et al., 1992, p. 79).

Being collected, organized and processed, according to explicit linguistic and extralinguistic criteria, the ESLOs serve as a sample of a particular variant of oral French language of a certain period. The ESLOs offer different recordings with the speakers of various ages and socio-cultural levels, expressing themselves in and authentic interactions. Along with that as remarks V. André the use of such sociolinguistic resource in education requires grasping the multiple inter-influences between language productions and the context (André, 2018).

The corpus is searchable in different ways. To search in catalogs, it is necessary to select a catalog (for recordings, speakers and transcripts). This selection may be done in accordance with metadata. In terms of recordings, it is possible to indicate a title, date, spatial sound, acoustic quality (poor, fair, good, excellent, or not specified). When choosing the speaker(s) the choice may be grounded on reference id, sex, age, professional category and level of studies. When selecting the transcription, it possible not only to choose brute, reread or validated version, but also title, the transcriber (s), guide version and summary. The procedure for finding an occurrence in the corpus is almost the same: it is based on the selection of corpus (ESLO 1 or ESLO 2, or both), characteristics of speakers, recordings, transcriptions, but also on the choice of conditions (exact word, word/segment starts with, word/ segment ends with), with sound event tag. The appropriate interface allows consulting the metadata (sociological categorization of the speakers, specifics of the recordings), which is stored in a database and therefore accessible in the form of cross-requests (for more details see: Eshkol-Taravella, Baude et al., 2011).

Below we will examine possible ways to use the tools made available by corpus linguistics to promote learning French as a foreign language. Based on our own teaching practices and other similar examples, we can say that there is a whole range of applicable possibilities, for instance: the observation of real, actual verbal productions, their lexical, syntactic, semantic, pragmatic, interactional and situational specificities. For instance, they can check the use of words, expressions or tenses they are not sure of. Just a few of the most typical difficulties, encountered by Ukrainian- speaking learners: *parce que /puisque, donc/ alors, environnement/ atmosphere, futur simple/ futur proche*.

In agreement with V. André, we have to notice that the results of sociolinguistic analysis of verbal interactions do have consequences for the teaching and learning a foreign language. That is particularly the case of observing the communicative interactions in oral corpora as they allow the learners understanding the following:

- there is no congruence between a language activity and a practice of language;
- a number of different language practices can carry out the same language activity;
- the same language practice can accomplish several language activities;
- a language activity can be carried out with a single word;
- a language activity can be carried out jointly by several speakers (André, 2018, p. 75).

In terms of learning grammar or vocabulary or conducting linguistic research arouses interest a subcorpus of ESLO constituted by L. Abouda and Mr. Skrovec (Abouda & Skrovec, 2015) with a view to a micro-diachronic approach to the analysis linguistic data. For educational purposes useful is the subcorpus of cooking recipes, extracted from ESLO 1 as a part of general interview. The subcorpus is composed of data of type “Cooking

recipes” obtained from the question “Comment est-ce qu’on fait une omelette?” (“How do you make an omelette?”). With a total duration of approximately one hour and forty minutes, this corpus comprises 96 question-answer exchanges, varying from 16 seconds to just over 3 minutes (Abouda, 2012). As emphasizes L. Abouda this “sociologically and discursively situated corpus” with its “paradigmatic structuring of comparable data” in terms of having a 96 responses to a simple question, “produced in the same discursive conditions, constitutes a privileged ground for a linguistic study” (ibid, p. 4).

In addition to the name above advantages, this subcorpus can be also interesting for linguocultural studies. In general, the ESLO 1 presents a unique source for sociocultural studies in terms of exploring cultural values, lifestyles, ethical and moral norms of behavior, as well as models of social situations of ordinary city dwellers of the second half of the XX<sup>th</sup> century. The dialogues reflect phonetic, lexical, stylistic rules of communication and linguistic features of social strata in particular proper to this period and current states. They also allow analyzing discursive methods of communication.

As for the ESLO 2, it reflects the characteristics of contemporary French and serves as a “reservoir” for oral, spontaneous communication, which can include neutral, literary, colloquial, and even non-literary units.

It is worth emphasizing that sociolinguistic analysis of speakers’ verbal interactions based on the ESLO 1 and the ESLO 2 sheds light on a number of specific language practices, interesting for understanding social, psychological and linguistic specificities, which are notably linked to the temporal gap, measured in almost 40 years. It is particularly interesting for micro-diachronic corpus.

The unique characteristics of the ESLOs corpora, with regard to speakers, type of discourse, and communication situations, allow learners to have access to linguistic diversity that does not exist in textbooks. In order to observe and analyze language practices, learners can consult available metadata to examine which language practices are appropriate for the communication situation. Figures 1 and 2 are extracts from the list of ESLO diachronic subcorpus metadata. They present data about the same speaker and the information about the recordings in two different periods.

Fiche locuteur	
Identifiant locuteur : CF4	
Anonyme:	OUI
Année de naissance:	1943
Tranche d'âge:	25/35
Lieu de naissance:	Tilly le Peneux (Eure et Loire)
Sexe:	Homme
Niveau d'études:	CEP - CAP
Commentaire:	Enseignement : primaire mixte collège d'enseignement technique diplômés : C.E.P. - C.A.P.
Age de fin d'études:	17
Catégorie Professionnelle (INSEE):	Ouvriers
Profession en termes propres:	Ajusteur
Langue(s):	
Commentaire niveau langue:	
Situation de famille:	Célibataire
Année d'arrivée:	1969
Domicile:	Orléans
Nombre d'enfants:	
Information sur les enfants:	
Remarques diverses:	Témoïn ré-interviewé dans ESLO2 - ESLO2_DIA_1223, avec le code CF4_DIA
Fiche modifiée par:	lkanaan
Enregistrements et transcriptions:	<ul style="list-style-type: none"> <li>• Enregistrement ESLO1_ENT_048</li> <li>• Transcription ESLO1_ENT_048_A</li> <li>• Transcription ESLO1_ENT_048_B</li> <li>• Transcription ESLO1_ENT_048_C</li> </ul>

  

Fiche locuteur	
Identifiant locuteur : CF4_DIA	
Anonyme:	OUI
Année de naissance:	1943
Tranche d'âge:	55/65
Lieu de naissance:	Tilly le Peneux (Eure et Loire)
Sexe:	Homme
Niveau d'études:	CEP - CAP
Commentaire:	
Age de fin d'études:	17
Catégorie Professionnelle (INSEE):	Retraités
Profession en termes propres:	Etait ouvrier, ajusteur dans une usine
Langue(s):	
Commentaire niveau langue:	
Situation de famille:	Célibataire
Année d'arrivée:	1969
Domicile:	Orléans, quartier Bourgogne-Théâtre
Nombre d'enfants:	
Information sur les enfants:	
Remarques diverses:	Témoïn déjà interviewé dans ESLO1 - ESLO1_ENT_048, avec le code CF4
Fiche modifiée par:	lkanaan
Enregistrements et transcriptions:	<ul style="list-style-type: none"> <li>• Enregistrement ESLO2_DIA_1223</li> <li>• Transcription ESLO2_DIA_1223_A</li> <li>• Transcription ESLO2_DIA_1223_B</li> <li>• Transcription ESLO2_DIA_1223_C</li> </ul>

Figure 1. Metadata of the speaker (CF4/CF4\_DIA) in the ESLO 1 and the ESLO 2.

Fiche enregistrement		Fiche enregistrement	
Référence enregistrement: ESLO1_BVT_048		Référence enregistrement: ESLO2_DIA_1233	
<b>Fichier son:</b>	ESLO1_BVT_048.wav	<b>Fichier son:</b>	ESLO2_DIA_1233.wav
<b>Corpus:</b>	ESLO1	<b>Corpus:</b>	ESLO2
<b>Catégorie:</b>	Entretien	<b>Catégorie:</b>	Diachronie
<b>Précisions sur la catégorie:</b>	Discussion en face à face entre un chercheur et un locuteur témoin à partir d'un questionnaire « ouvert »	<b>Précisions sur la catégorie:</b>	Entretien avec un locuteur déjà enregistré dans un entretien ESLO1. Discussion en face à face entre un chercheur et un locuteur témoin à partir d'un questionnaire « ouvert »
<b>Sujet:</b>	(text_and_corpus_linguistics) Français (Ethnologue: fra)	<b>Sujet:</b>	(text_and_corpus_linguistics) Français (Ethnologue: fra)
<b>Sommaire:</b>	01 9 ans02 parents habitent commune, venu trouver du travail03 ou04 oui, aime mieux ville01 àjusteur dans usine de machines agricoles02 description travail - journée de 9h03 03 il y a des choses qui lui plaisent, peut pas dire lesquelles04 un autre métier mécanique05 normal06 RR syndicalisme occupé tous ses loisirs - explication comité d'entreprise - syndicalisme son devoir07 01 on n'est jamais satisfait08 RR revendications - conditions, durée de travail - problèmes de la journée continue09 à chaque fois on se fâche10 pour assumer rencontres11 01 RR 01, qui a décidé le grève 07 les travailleurs 07 consultation travailleurs - réunion, vote12 augmentations03 pour être des hommes - on les forme pour la consommation, pas en homme libre03 histoire, égalisation02 aucune liberté facilité pour apprendre, professeurs05 16 ans05 ou02 chez parents03 se repose chez parents04 lire01 journée dans son organisation01 satisfait personnellement02 ne comprend pas la question04 RR usine occupée - gouvernement avait position dure - étudiants ont provoqué l'évènement - rapprochement entre étudiants et travailleurs assés dur05 oui - travailleurs/patrons/cadres06 Place Questionnaire sociolinguistique	<b>Sommaire:</b>	
<b>Editeur:</b>	LLL Université d'Orléans	<b>Editeur:</b>	LLL Université d'Orléans
<b>Créateurs:</b>	LLL Université d'Orléans - ESLOs	<b>Créateurs:</b>	LLL Université d'Orléans - ESLOs
<b>Chercheurs:</b>	• Blogg, Patricia	<b>Chercheurs:</b>	• Chesneau, Annie
<b>Chercheurs locuteurs:</b>	• Blogg, Patricia	<b>Chercheurs locuteurs:</b>	• Chesneau, Annie
<b>Participants:</b>	Témoin parlé très lentement, avec beaucoup de longues hésitations.	<b>Participants:</b>	
<b>Description des participants:</b>		<b>Description des participants:</b>	
<b>Descriptions annexes:</b>	Cet enregistrement fait partie du module "Diachronie". Le locuteur CF4 a été ré-intervié en 2007 avec le code locuteur CF4_DIA_of_ l'enregistrement ESLO1_BVT_048	<b>Descriptions annexes:</b>	
<b>Remarques:</b>		<b>Remarques:</b>	Cet enregistrement fait partie du module "Diachronie". Le locuteur CF4_DIA a déjà été interviewé en 1989 avec le code locuteur CF4_of_ l'enregistrement ESLO1_BVT_048
<b>Fiche modifiée par:</b>	gabuzo	<b>Fiche modifiée par:</b>	kanan
<b>Date d'enregistrement:</b>	04/04/1989	<b>Date d'enregistrement:</b>	13/03/2007
<b>Droits:</b>	Copyright (c) 2012 Université d'Orléans/LLL/Free available for non-commercial use. This file is licensed under a Creative Commons License.	<b>Droits:</b>	Copyright (c) 2012 Université d'Orléans/LLL/Free available for non-commercial use. This file is licensed under a Creative Commons License.
<b>Format:</b>	(IANA MIME Media Type: audio/x-wav)	<b>Format:</b>	(IANA MIME Media Type: audio/x-wav)
<b>Durée:</b>	01:38:00	<b>Durée:</b>	01:19:32
<b>Acoustique:</b>	Bonne	<b>Acoustique:</b>	Bonne
<b>Précisions acoustiques:</b>	bonne - témoin clair, légers bruits	<b>Précisions acoustiques:</b>	
<b>Lieu spatial:</b>	Orléans	<b>Lieu spatial:</b>	Orléans
<b>Lieu TGN:</b>	7008337	<b>Lieu TGN:</b>	7008337
<b>Lieu Point:</b>	east=1.90; north=47.90	<b>Lieu Point:</b>	east=1.915; north=47.902
<b>Locuteurs:</b>	• CF4 • PB	<b>Locuteurs:</b>	• CF4_DIA • cf4_of7
<b>Transcriptions:</b>	• ESLO1_BVT_048_A • ESLO1_BVT_048_B • ESLO1_BVT_048_C	<b>Transcriptions:</b>	• ESLO2_DIA_1233_A • ESLO2_DIA_1233_B • ESLO2_DIA_1233_C

Figure 2. Metadata of the records in the ESLO 1 and the ESLO 2.

When consulting the recording corresponding to this metadata the learners are able to see that the language practices observed are produced by the same person, a man (aged 26 and 64 respectively). They can analyze and compare changes (if it is) in his education level, social status, his accent, statements' sequences, vocabulary etc.

Another spoken corpus, which can be very useful for teaching and learning French as a second language is Fleuron. Contrary to the ESLOs, it is specialized only in the academic field. The Fleuron corpus offers authentic multimedia resources, classified by categories, which illustrate a set of communication situations in students' life on and off campus in France. Each resource is described by its title and by a short summary, and accompanied subtitles and transcript, see below:

The screenshot shows the Fleuron website interface. At the top, there is a navigation bar with 'Accueil', 'Comment utiliser Fleuron?', 'Ressources multimédias', 'Liens utiles', 'Profil', 'A propos', and 'Déconnexion'. Below this, there are two main panels. The left panel, titled 'Ressources multimédias', allows users to choose a category from a grid including 'Démarches administratives', 'Aides sociales', 'Explications du système universitaire', 'Témoignages', 'Vie en dehors du campus', 'Démarches à la préfecture', 'Questions pédagogiques', 'Utiliser les transports', 'Culture et lieux de culture', 'Etudiants en Doctorat', 'Santé', and 'La vie sur le campus'. The right panel shows a search bar with 'Mot recherché' and 'Rechercher', and a list of search results. One result is highlighted, showing a video titled 'Acheter au food truck du campus (2)'. The video player shows a student talking to a food vendor. Below the video, there is a transcript and subtitles in French. The transcript includes a search bar and a 'Rechercher' button. The subtitles are: 'A: dans toutes les questions d'ordre administratif et pédagogique', 'E: d'accord donc c'est vous qui allez est-ce que c'est vous qui lui parlez de la scolarité', 'E: qui lui expliquez un peu comment ça fonctionne les cours les emplois du temps et tout ça', 'A: avant son arrivée l'étudiant donc euh A: a déjà complété un contrat d'étude hein c'est le contrat pédagogique donc avec le choix des cours', 'A: euh mais il est vrai que lorsqu'il arrive il faut vérifier', 'A: euh que ses cours bah sont possibles', 'A: donc il y a un catalogue des cours qui a été réalisé', 'A: euh puis euh effectivement il y a aussi des réunions de pré-rentrée', 'A: euh qui sont données euh à chaque début de semestre', 'E: d'accord et ces réunions c'est vraiment...'

Figure 3. An example of multimedia resource from the FLEURON site with transcription and subtitles.

In the difference from the ESLOs where additional software is needed to be installed, in FLEURON all the resources are also searchable by a simple concordancer, created specifically for the site. It allows comparison of different usages of the same word, key words analysis and frequency analysis of words. Learners may be asked to discover a grammar rule or the use of a word in various contexts according to their level and learning objectives.

Searching for the occurrences in concordancer may help to better understand a confusing word. To illustrate the use of the corpus and the concordance in this concrete case, we will cite a work in which learners had been asked to search for “démarche” in the concordancer of FLEURON. The choice of this word is explained by the fact that Ukrainian speaking learners experience difficulties in understanding the meaning and the use of the word “démarche” in French, particularly used in educational context. It is explained by its meaning in Ukrainian language, where it signifies “The actions or measures of the state, diplomatically aimed at warning another state against certain actions, to achieve any concessions” (<http://sum.in.ua/s/demarsh>). The search here provides learners with a sample of lexical possibilities to use the target word.

The screenshot shows the FLEURON website interface. At the top, there is a navigation bar with links: Accueil, Comment utiliser Fleuron ?, Ressources multimédias, Liens utiles, Profil, A propos, Déconnexion. Below this is a search bar with the text 'Concordancier démarche' and a 'Rechercher' button. The search results are displayed in a table with columns: Concoardancier (13), Ressources (8), Mes notes (0), and CNRTL. The table has four main columns: #, A gauche du mot, Mot recherché, and A droite du mot. The search results show 13 occurrences of the word 'démarche' in various contexts, such as 'L2 : bonne question parce que pour dans la', 'A : voilà donc euh ensuite une fois que vous avez fait cette', and 'E : là spécialement pour ça pour lui expliquer la la et notre projet c'est de développer un peu la et le jour où les docteurs accepteront de participer à cette'.

#	A gauche du mot	Mot recherché	A droite du mot
1	L2 : bonne question parce que pour dans la	démarche	de demande d'aide L1 : d'allocation L2 : location pour il y a
2	A : voilà donc euh ensuite une fois que vous avez fait cette	démarche	d'inscription
3	A : la	démarche	quand vous voulez partir quelque part c'est de
4	E : je l'ai reçu que maintenant A : hum hum E : et je sais pas quelle	démarche	il faut suivre euh
5	E : là spécialement pour ça pour lui expliquer la la	démarche	OK A : voilà oui c'est leur travail
6	et notre projet c'est de développer un peu la	démarche	d'alumni d'anciens
7	et le jour où les docteurs accepteront de participer à cette	démarche	d'alumni
8	donc par exemple euh il y a la	démarche	d'inscription
9	pour euh ouais cette euh enfin le le tout était pour nous de rendre cette	démarche	-là très accessible
10	donc euh il faut être très renseigné sur euh ch- quelle	démarche	faire et quand la faire
11	donc c'est très important de se renseigner au préalable sur toute la	démarche	de soutenance et euh et tout ce qu'elle comporte
12	donc comment on fait pour candidater pour un doctorat à l'Université de Lorraine quelle est la	démarche	
13	dans le cadre de notre	démarche	qualité cet appel est susceptible d'être enregistré

Figure 4. Searching results for “démarche” in FLEURON.

Moreover, an option to view and to listen to the resources corresponding to the occurrences highly facilitates the learning process. Thus, due to the presentation in a list of occurrences and to the access to the corpus (transcription and sound), learners succeed in understanding the meaning and the pragmatic value in interaction. It is from the study of these two dimensions that learners have access to multimodality and metadata, and therefore they can better understand the regularities in the target word’s use.

Also, by choosing a subcorpus, learners can orient their own learning, according to their objectives and/or educational needs. For instance, students who have an intention to study in France may consult a subcorpus on concrete theme (how to register in an administrative service, get official university documents, discuss with a teacher about an educational path or a curricular etc.).

Another oral corpus that may be helpful in teaching and learning French is CLAPI (fr.Corpus de Langue Parlée en Interaction). It is a multimedia database of around forty oral

corpora recorded in real situations of social interactions, in various contexts (professional, private, institutional, commercial, medical, etc.) and a set of query tools. It offers a set of consultation and query tools with a multimedia display of the results: metadata, transcriptions (<http://clapi.ish-lyon.cnrs.fr/>). CLAPI-FLE is a platform, which offers “authentic, ready-to-use, free access and downloadable educational resources”, based on the results of research in interaction analysis and oral corpora of social interactions. They have been particularly designed for teachers and learners of French as a Foreign Language and French linguistics. Possible educational uses of the resources are clearly defined on the home page (see below):



Figure 5. Screenshot of the CLAPI homepage

The CLAPI-FLE platform includes around 40 extracts with an average duration of 2 minutes. This platform differentiates from others above cited that it offers a new “In Practice” resource, ready to be used directly in French lessons. It includes: downloadable audio- video recordings; a worksheet corresponding to a level of language (A1 / C2) with different objectives (linguistic, intercultural, conversation); recordings divided by level, with the main phonetic, lexical or syntactic difficulties; the transcription and the vocabulary list; splitting into language actions (to facilitate comprehension and interactional activities and to reveal the structure of the interaction).

In addition to these extracts, there are several educational files that document frequent oral expressions or functions, and explain their different pragmatic sense and specifics of their use or reveal speaking strategies and techniques. All these language features are analyzed and explained from different aspects:

- lexical, through the use of fixed expressions and discourse markers;
- syntactic, by identifying structures or the unexpected use of tenses (ex. the use of present tense instead of grammatically correct future or past tense);
- prosodic, in particular changes in intonation;
- multimodal, which analyses non-verbal means of communication;
- intercultural that accentuates on the specificity of cultural meanings and used linguistic forms.

In terms of vocabulary learning, apart from diverse themes (conversations during meals or gaming sessions, shopping, work meetings, medical consultations, phone calls, etc.), a quantitative report is given for each interaction, showing the number of (different) words, of speaking turns, number of overlaps and pauses, their length, etc. Also it is possible

to consult the vocabulary of the extract and the infrequent words used in it. Being analyzed in the light of metadata of the corpus (i.e. the elements of the communicative situation), the data obtained help understanding their values and pragmatic use.

Furthermore, sociolinguistic analysis of speakers' verbal interactions based on these materials helps understand on a number of specific language practices, for instance the meanings, functions, the pragmatic role and situational variation of discourse markers (ex. *bon, ben, alors, en fait, donc, quoi*), which are usually consider to be the most difficult to understand and to the appropriate use.

The phonetic learning perspectives within this platform worth mentioning too. Here we cite a few of oral phenomena to display and analyze: non-standard elision, unfinished word, rising or descending intonation, *h* aspiration, *h* expiration, immediate linking, short or long pause, etc. Also, while using this corpus resources learners are able to identify typical structures for oral discourse (dislocations, syntactic constructions) and interactional phenomena (repairs, repetitions).

CFPP2000 (<http://cfpp2000.univ-paris3.fr/>), unlike the ESLOs, which form the “sound portrait” of Orléans, makes up the sound portrait of Paris, presenting a set of interviews on the districts of Paris and the inner suburbs. CFPP2000 allows the study of colloquial French. As in the ESLOs, this corpus presents the interviews with the inhabitants of a big city. The difference is that they are based on the questionnaire on the concrete city. Therefore, “it constitutes an interesting and coherent testimony on the representations of Paris and its inner suburbs”. The questionnaire includes sections on the reasons for living in the district and on apprehension of the relationship to other districts of Paris or the suburbs, the reactions of residents to the changes (population, town planning, shops, etc.), it also concerns the way of population mixtures, the situation of other languages, or its impact on culinary and clothing styles, etc. (<http://cfpp2000.univ-paris3.fr/>).

There are plenty of possibilities to work and learn on this corpus, to name a few: the adopted transcription system allows to study on discourse, on sociolinguistic, on the syntax of spoken French, on phonetics, intonation or interactions, as the transcriptions have been aligned with the sound.

The learners interested in textometric explorations will benefit from CFPP2000 textometric base, which includes 32 interviews, with 3 layers of annotations (form, lemma, category). Here we can find both a breakdown in turn of words and 3 systems of parts (by quarter, by transcription and by speaker). To use it the iTrameur and Le Trameur software is needed.

### **5. Conclusions.**

The particular conclusions of the study may be reduced to the following:

The use of speech corpora in practice of teaching and learning French as a second language learning is an effective addition to traditional educational technologies. The authentic corpus resources can be used as supporting material (level B1-B2), which allows the teacher to correlate the learning objectives and the teaching methods. This approach combines empirical adequacy, authenticity, and adaptation, and the possibility of independent learner's work. The analyzed oral corpora give access to real, authentic language data, thus provide an alternative to textbooks.

Accompanied with transcriptions of authentic conversations, spoken corpora constitute a source for discourse analysis, language acquisition, phonetics, or sociolinguistics. Learner's exposure to these resources presents a possibility to familiarization with the cultural values, realities, norms and rules of behavior of native speakers of the target language; has an impact on development of listening and pronunciation skills, on mastering intonation, phonetic features; improves phonemic hearing; gives the

possibility to get acquainted with peculiarities of colloquial French at lexical, syntactic level; strengthens the understanding and of non-verbal means of communication; helps improving learners' level of intercultural communication.

The use of corpora in the language training can be realized in different ways, apart from concordancers, other corpus tools can be used. Moreover, the type and the level of difficulty of the corpus-based activities have to be varied and adapted to learners' competence, needs and objectives.

However, this use can be effective and interesting if it bases on clear methods and is gradual, consistent and systematic.

### References

- Abouda L., Baude O. (2007). Constituer et exploiter un grand corpus oral: choix et enjeux théoriques. Le cas des ESLO. *Corpus en Lettres et Sciences sociales, Des documents numériques à l'interprétation*, 161-168. [In French].
- Abouda, L. (2012). Le temps des omelettes. Une nouvelle valeur pour le présent?. *Revue de Sémantique et Pragmatique*, 71-94. [In French].
- Abouda, L., & Baude, O. (2005). Du français fondamental aux Eslo. In *Grand corpus de français parlé, Bilan historique et perspectives de recherche* (Vol. 33, No. 2, pp. 131-146). [In French].
- Abouda, L., & Skrovec, M. (2018). Pour une micro-diachronie de l'oral: le corpus ESLO-MD. In *SHS Web of Conferences* (Vol. 46, p. 11004). EDP Sciences. [In French].
- 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. [In French].
- 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). [In French].
- André, V. (2020). Faire de la linguistique de corpus avec des apprenants de français langue étrangère. [In French].
- Arnold, A., & Candea, M. (2015). Comment étudier l'influence des stéréotypes de genre et de race sur la perception de la parole? *Langage et société*, (2), 75-96. [In French].
- Baude, O., & Dugua, C. (2011). (Re) faire le corpus d'Orléans quarante ans après: quoi de neuf, linguiste? *Corpus*, (10), 99-118. [In French].
- Baude, O., & Dugua, C. (2016). Les ESLO, du portrait sonore au paysage digital. *Corpus*, (15). [In French].
- Bergounioux, G., & Eshkol, I. (2016). Quand faire, c'est dire: l'exemple de la recette. [In French].
- Bergounioux, G., Baraduc, J., & Dumont, C. (1992). L'étude socio-linguistique sur Orléans(1966-1991)–25 ans d'histoire d'un corpus –. *Langue française*, (93), 74-93. [In French].
- Bernardini, S. (2002). Exploring new directions for discovery learning. In *Teaching and learning by doing corpus analysis* (pp. 165-182). Brill Rodopi. [In English].
- Branca-Rosoff, S., Fleury, S., Lefeuvre, F., & Pires, M. (2012). Discours sur la ville. Présentation du Corpus de Français Parlé Parisien des années 2000 (CFPP2000). *article en ligne*, <http://cfpp2000.univparis3.fr/Articles.html>. [In French].
- Blanche-Benveniste, C. (2000). Approches de la langue parlée en français. Coll. L'essentiel français. Ophrys. 164 p. [In French].
- Cappeau, P., & Gadet, F. (2007). L'exploitation sociolinguistique des grands corpus. *Revue française de linguistique appliquée*, 12(1), 99-110. [In French].
- Chambers, A. (2009). Les corpus oraux en français langue étrangère: authenticité et pédagogie. *Mélanges Crapel*, 31, 15-33. [In French].
- Chambers, A. (2009). Les corpus oraux en français langue étrangère: authenticité et pédagogie. *Mélanges Crapel*, 31, 15-33. [In French].
- De Cock, S., & Tyne, H. (2014). Corpus d'apprenants et acquisition des langues. *Recherches en didactique des langues et des cultures. Les cahiers de l'Acedle*, 11(11-1). [In French].
- Di Vito, S. (2013). L'utilisation des corpus dans l'analyse linguistique et dans l'apprentissage du FLE. *Linx. Revue des linguistes de l'université Paris X Nanterre*, (68-69), 159-176. [In French].
- Eshkol, I. (2015). Interpréter le contexte dans un corpus oral: fonctions et limites du traitement automatique des données linguistiques. [In French].
- Eshkol-Taravella, I., Baude, O., Maurel, D., Hriba, L., Dugua, C., & Tellier, I. (2011). Un grand corpus oral «disponible»: le corpus d'Orléans 1 1968-2012. *Traitement automatique des langues*, 53(2), 17-46. [In French].
- Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. In T. Johns & P. King (ed), *Classroom Concordancing, English Language Research Journal*, 4, p. 1-16. [In English].

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). [In French].

Ravazzolo, E., Étienne, C., & Ursi, B. (2021). Apprendre les interactions en classe de français: enjeux et pratiques. *Recherches en didactique des langues et des cultures. Les cahiers de l'Acedle*, 18. [In French].

Rendulić, N., & Kanaan-Caillol, L. (2016). Je crois que, je pense que: valeurs et variation dans un corpus oral diachronique. *In SHS Web of Conferences* (Vol. 27, p. 02014). EDP Sciences. [In French].

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. [In French].

Дерябина, И. В. (2012). Характеристики педагогической работы с английским национальным корпусом [Characteristics of teaching work with the English National Corpus]. *Вестник Тамбовского университета. Серия: Гуманитарные науки*, (9). [In Russian].

Нагель, О. В. (2008). Корпусная лингвистика и ее использование в компьютеризированном языковом обучении [Corpus linguistics and its use in computerized language teaching]. *Язык и культура*, (4). [In Russian].

Плунгян, В. А. (2008). Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики [The corpus as a tool and as an ideology: on some lessons of contemporary corpus linguistics]. *Русский язык в научном освещении*, (2), 7-20. [In Russian].

Садовникова, О. Э. (2013). Прямое и косвенное использование корпусов в зарубежной лингводидактике. *Magister Dixit*, (2).

Сысоев, П. В. (2010). Лингвистический корпус в методике обучения иностранным языкам [Linguistic Corpus in the Methodology of Teaching Foreign Languages]. *Язык и культура*, (1 (9)). [In Russian].

### *Sources of Illustrative Material*

CFPP2000: [http://cfpp2000.univ-paris3.fr/\[en ligne\]](http://cfpp2000.univ-paris3.fr/[en ligne]).

CLAPI: [http://clapi.ish-lyon.cnrs.fr/\[en ligne\]](http://clapi.ish-lyon.cnrs.fr/[en ligne]).

CLAPI-FLE: <http://clapi.ish-lyon.cnrs.fr/FLE/accueil.php> [en ligne].

COCOON: <https://cocoon.huma-num.fr/exist/crdo?%3Flang=en&lang=fr> [en ligne].

ESLO: <http://eslo.huma-num.fr> [en ligne].

FLEURON: <https://fleuron.atilf.fr/> [en ligne].

ORTOLANG: <https://www.ortolang.fr/> [en ligne].

### **Бібліографічний опис:**

Страшко, І. В. Використання корпусів усного мовлення у вивченні французької мови як іноземної (на прикладі ESLO, FLEURON, CLAPI і CFPP 2000). *Науковий часопис Національного педагогічного університету імені М. П. Драгоманова. Серія 9. Сучасні тенденції розвитку мов*, 22. 110–122. <https://doi.org/10.31392/NPU-nc.series9.2021.22.08>.

### **Анотація**

У статті розглядаються освітні можливості використання усних корпусів для вивчення французької мови як іноземної. Аналізується застосування корпусів усного мовлення у навчальному процесі. Наведено огляд декількох корпусів усного мовлення, коротко описується їхня специфіка, тематична спрямованість, особливості роботи з ними. Даються також приклади конкретних освітніх застосувань. Аналізуються переваги, труднощі та проблеми використання корпусів усного мовлення. Наголошується, що їх використання у практиці викладання та вивчення французької мови є ефективним доповненням до традиційних освітніх технологій. Вони можуть бути використані як допоміжний матеріал, що дозволяє вчителю співвідносити цілі та методи навчання. Спосіб роботи з ними вимагає переосмислення способу викладання та навчання: цей підхід поєднує емпіричну адекватність, автентичність, адаптацію та можливість самостійної роботи учня. Він також передбачає необхідність теоретичної методичної та освітньої підтримки як для вчителя, так і для студента. Використання цих корпусних ресурсів може бути ефективним і цікавим, якщо воно ґрунтується на чіткій програмі та є поступовим, послідовним і систематичним.

**Ключові слова:** корпус усного мовлення, корпусна лінгвістика, французька мова як іноземна, навчальні ресурси.