



2395-2636 (Print);2321-3108 (online)

Using the Corpus in FOS Teaching

YANG Tong

School of Foreign Languages, North China Electric Power University 102206, China

Corresponding Author Email: tongyang@ncepu.edu.cn

DOI: [10.33329/rjelal.13.4.193](https://doi.org/10.33329/rjelal.13.4.193)



Article info

Article Received: 09/10/2025
Article Accepted: 17/11/2025
Published online: 22/11/2025

Abstract

This paper begins by comparing three teaching models (FOS, FS, and FLP) from the perspective of contemporary specialized French language teaching, introducing the curriculum development of French for Specific Purposes. Secondly, given the challenges in data collection for FOS instruction, corpus linguistics can address this difficulty by providing multiple authentic examples through the impetus of Natural Language Processing (NLP) at two levels: automatic data collection and processing of data within the corpus. At the end of this article, we provide a concrete example of a corpus-based teaching exercise.

Keywords: French for Specific Purpose, Specialty French, French Professional Language, Corpus Linguistics.

1. Current Status and Trends in FOS Teaching Research

Unlike long-term, progressive foundational French language instruction, professional French teaching focuses on achieving specific objectives within specialized fields. Due to complex historical backgrounds and social environments, the terminology for professional French has varied across different developmental periods, such as Military French (*français militaire*) in the 1920s, Scientific and Technical French (*français scientifique et technique*) in the 1960s, French for Specific Purposes (*français de spécialité*) in the 1960s, Instrumental French (*français instrumental*) in the early 1970s, and Functional French (*français fonctionnel*) in the late 1970s. In the current trend of global integration, exchanges between

France and the world have grown increasingly close, and a favorable market economy environment has provided fertile ground for the development of French for Specific Purposes. Since the 1990s, professional French has existed in three main forms: *Français sur Objectif Spécifique* (FOS), *Français de spécialité* (FS), and *Français Langue Professionnelle* (FLP). The following discussion will primarily explore the similarities and differences among these three teaching models from the perspective of contemporary professional French instruction.

1.1 FS: course offering

As Cuq and Gruca pointed out, learners of FOS (*Français sur Objectif Spécifique*) are typically adults who wish to acquire or improve their French language skills for professional or academic purposes (Cuq & Gruca, 2005). FOS

instruction aims to provide French courses within a relatively short timeframe to meet the specific needs of particular groups, primarily targeting adult learners who require language enhancement for career advancement or academic pursuits. FOS focuses on case-specific studies. Consequently, the linguistic competencies to be acquired—including vocabulary, pronunciation, syntax, and other aspects—are derived from authentic materials and are closely tied to specialized fields (e.g., business, tourism, healthcare, academia). Français de spécialité (FS) emerged in the 1990s (distinct from the vocabulary-centered approach of the 1960s), building on professional training and international work experiences initiated by the French Ministry of Foreign Affairs' cultural department. This period saw a proliferation of teaching materials, such as *Business French*, *French for the Hotel and Restaurant Industry*, *French for Tourism*, and *Legal French*. Next, we delineate the distinctions between FS and FOS across five dimensions: clarity of training objectives, duration of training, diversity of topics and skills, lesson preparation methods, and teaching material development.

FS training objectives are generally broad and encompass a wide range of themes, aiming to cover as many communication scenarios related to a discipline as possible. As a result, FS programs tend to be longer, spanning several months or even years. Examples include courses like "Business French," "Legal French," "Tourism French," or "Scientific French" offered by universities or training institutions. In contrast, FOS training objectives are highly specific and targeted, focusing on particular situational needs or tasks. Since learners often have existing professional or academic commitments, FOS programs are typically condensed, lasting from a few days to several weeks. For instance, when the French government called for thousands of Spanish nurses to address a shortage of healthcare workers, these nurses underwent short-term FOS training to improve their

language skills and integrate into their new work environments in French hospitals.

In terms of lesson preparation, FS instructors can often rely on existing teaching materials and work independently to design courses. For FOS, however, instructors must engage with industry professionals and source authentic, practical materials to meet the precise learning objectives.

A comparative analysis reveals a fundamental difference in the "supply and demand" dynamic between FS and FOS: FS emphasizes course **supply** (offre), where courses are designed first, and students are recruited afterward. For example, an institution may offer training programs for potential audiences without targeting a specific niche but rather an entire field or profession. In contrast, FOS is driven by learner **demand** (demande), where courses are tailored to the specific needs of learners. For instance, a Chinese accountant seeking an internship at the Paris Chamber of Commerce and Industry (CCIP) would trigger the creation of a customized language course to meet their precise requirements.

1.2 FLP: Mastering Professional Competencies

French for Professional Objectives (FVP), which gained prominence in the early 2000s, was dedicated to teaching French by focusing on common professional skills across various specialized fields. These skills included tasks such as writing reports, letters, memos, and using various media (forms, meetings, telephone conversations). Driven by economic and political factors, **French for Professional Purposes (FLP)** gradually replaced FVP starting in 2006. According to Mourlhon-Dallies (2006), FLP is taught in French and is intended for learners who must practice their profession entirely in French. In this context, "even if part of the professional activities can be conducted in English or other languages (of colleagues), the entire professional framework is structured in French—including legal and institutional aspects, communication with colleagues and

hierarchical structures, and professional practices" (Mourlhon-Dallies, 2006).

Next, FOS and FLP are compared based on the following three aspects: French proficiency level, degree of specialization, and learning needs:

- **French Proficiency Level:** In FOS, the target audience primarily consists of non-native speakers (professionals, immigrants, and international students). FLP, however, can target both immigrants with insufficient French skills and native speakers facing communication challenges (particularly in writing). Examples include immigrants from French-speaking Maghreb countries in Africa and native speakers with literacy issues.
- **Degree of Specialization:** The second distinguishing feature between FOS and FLP lies in the professional specialization of the target audience. FLP is aimed at individuals in training or already working as professionals. In FOS, the focus is often on individuals who are already professionals (e.g., receptionists in foreign hotels, engineers working in their home countries, businesspeople). For instance, students seeking to continue their studies in French-speaking schools, or professionals already established in their home countries who wish to acquire specific skills in their field to stand out among their colleagues.
- **Learning Needs:** The third level of distinction between FOS and FLP relates to learning needs. In FOS, learning needs vary depending on the nature of the audience. In FLP, the audience's goal is typically to gain or retain employment.

Most educators fail to recognize or even overlook the specificity of the target audience when designing courses. The use of commercially available textbooks often proves ineffective, as these materials tend to be generalized and thematic in nature, failing to

directly address learners' actual needs. French publishers predominantly focus on disciplines most demanded by FOS audiences – particularly business, tourism, and law – while other specialized fields (such as agronomy, science and technology, computer science, etc.) receive little attention from textbook developers.

Certain FOS experts caution against systematically relying on textbooks to meet learners' needs. As noted by Mangiante & Parpette (2004: 7), "While specialized French textbooks exist (particularly in tourism, law, and notably business), which may provide course sequences or materials adaptable to FOS instruction in related fields, this approach remains uncommon. Consequently, instructors must develop their own courses and teaching materials."

Most teachers still do not adequately account for the distinct characteristics of their audiences when developing curricula. Since readily available textbooks often fail to align with learners' requirements, educators are compelled to create customized syllabi and pedagogical resources.

2. the new discipline of Corpus Linguistics

In linguistics, a corpus is "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). The discipline of Natural Language Processing (NLP) emerged around the same time as the advent of computers, with the objective of modeling and automating part of language analysis (Condamines, 2005). Since its inception, new tools and methodologies for analyzing textual data in digital formats have proliferated, and these tools have begun to be standardized using generic resources such as concordancers, conjugators, taggers, lemmatizers, transcribers, web crawlers, etc. (Antoniadis et al., 2006).

In the 1990s, corpus linguistics became an independent discipline (Sinclair, 1991; Leech, 1992; Habert *et al.*, 1997) with its own publications (among others, the *Journal of Corpus Linguistics*). Some specialists (including Kennedy, 1998; Condamines, 2005; Azzopardi, 2010) position corpus linguistics solely within the framework of NLP at two levels: automatic data collection and corpus processing. As Kennedy states, "Over the last three decades the compilation and analysis of corpora stored in computerized databases has led to a new scholarly enterprise known as corpus linguistics" (1998: 1).

Written corpora can be automatically compiled from the web using a crawler, which scans and indexes the web automatically to map its content. Platforms for such crawlers can include websites like *WebBootCat* (Baroni & Bernardini, 2004) or software such as *Gromoteur* (Gerdes, 2014). Spoken corpora are often compiled through sociolinguistic surveys, as seen in the first two spoken corpora used in teaching French as a foreign and native language: *Français Fondamental* (Gougenheim *et al.*, 1956) and *Les Orléanais ont la parole* (Biggs & Dalwood, 1976).

Once a corpus is constructed, query tools can facilitate analysis. Various tools provide valuable assistance in the linguistic description of languages, particularly in textometry, with examples such as *AntConc* (Anthony, 2005).

3. Using the Corpus for FOS

The primary contribution of corpora in language teaching is to replace introspection based on intuition or isolated examples provided by linguists with authentic language phenomena. Recognizing that intuition can sometimes be misleading (Sinclair, 2005), corpora serve as a "revealer of language" (Tognini-Bonelli, 2001) and provide access to the "intuitions of thousands of speakers" (Frankenberg-Garcia, 2005: 192).

3.1 Using corpora for language teaching

Corpus-based language teaching and learning gained momentum primarily through Johns' (1991) data-driven learning (DDL) approach, earning him the title of "founding father of pedagogical applications of corpus linguistics" (Boulton & Tyne, 2014: 75). The essence of the DDL approach is that learners can inductively formalize their own knowledge rather than deductively applying rules transmitted by the teacher (Landure & Boulton, 2010). Over the past thirty years, integrating corpora into language classrooms has generated significant interest across various fields of language teaching and learning: error correction (O'Sullivan & Chambers, 2006; Gaskell & Cobb, 2004), academic writing (Yoon & Hirvela, 2004; O'Sullivan, 2010), language acquisition (Colletta, 2004; Hoey, 2005; Tyne, 2009), and writing competence (Gaskell & Cobb, 2004; O'Sullivan & Chambers, 2006). Furthermore, corpus-based learning represents a "natural" activity where learners attempt to identify linguistic regularities, as opposed to the "artificial" transmission of knowledge (Gaskell & Cobb, 2004; Scott & Tribble, 2006; Boulton, 2009). Some view corpus use as a way for learners to engage directly with the target language rather than merely learning grammatical rules (Loewen *et al.*, 2009). Thus, corpora can provide "scaffolding" that enables learners to construct their own knowledge without disregarding any learning theories (Kirschner *et al.*, 2006; Mayer, 2004). Finally, corpus-based learning can "lead to greater autonomy, which will benefit learners in their future professional lives" (Landure & Boulton, 2010: 3). Consequently, using corpora in French as a Foreign Language (FFL) classrooms, as well as in our FOS project, is essential.

3.2 Example of a corpus-based teaching exercise

Here are an overview of corpus construction and its subsequent use in FOS teaching. Our corpus, named *Cuisitext* (Yang,

2016), includes both written and spoken components. For the written corpus, we used the *Gromoteur* web crawler to collect thousands of French recipes from culinary websites such as Marmiton, 750g, and Cuisine AZ. In professional culinary contexts, oral discourse is also frequent. Thus, our corpus includes three types of video-based spoken data. Transcribing spoken corpora is essential for pedagogical applications. Since our corpus was designed for teaching and learning culinary lexicon in FOS, the *CLAN* tool was used for orthographic transcription, enabling lexical, syntactic, and interactional analysis.

The selected keyword should be representative of the culinary domain. Among the analyzed and selected NAdj expressions for teaching, the lexeme "HACHÉ" (minced/chopped) was chosen as the keyword for this session. A2-level learners (according to the CEFR) struggle with this lexeme both syntactically and semantically, and ingredients described by this term can be easily categorized into several groups.

Once concordances are extracted, the teacher needs to clean the data. Students are grouped to discuss and exchange ideas about the meaning and constructions of this lexeme in context. Each group receives different concordance examples. After observation, analysis, and classification, three categories can be presented:

- Fresh herbs (Persil, oignon, ail, basilic, échalote, ciboulette, poireau, menthe, estragon, fenouil, feuilles d'aneth)
- Meat (Porc, bœuf, jambon, veau, agneau, poulet, steak)
- Others (concombre, chocolat, cantal, saumon)

In French culinary culture, garlic, parsley, onion, and shallot are classified as aromatics, hence their placement in the first category. Types of meat form the second category. Beyond meats and aromatics, "HACHÉ" can also

describe concombre (vegetable), chocolat (cocoa bean), saumon (cheese), and salmon (fish), which belong to categories distinct from the first two. This justifies grouping them into a third, more general category. Learners should be informed that items in the third category are less frequently used than those in the first two. After revealing the answers, learners should be given time to complete and correct their own category lists.

4. Conclusion

Through the comparison with FLP and FS, FOS is defined by highly targeted objectives, intensive short-term training (ranging from a few days to weeks), a focus on specific tasks and situations, and the need for customized teaching materials, often developed in collaboration with industry professionals. FOS primarily targets non-French-speaking learners and may include students or professionals seeking to acquire skills for specific contexts. FOS addresses immediate and contextual needs (e.g., an assignment abroad). Therefore, FOS courses require tailored development based on specific needs.

Corpus linguistics can be applied not only to corpus collection and extraction but also to corpus-based teaching. In this paper, targeting A2-level students, the teaching of the lexeme "HACHÉ" in the culinary domain serves as an example to demonstrate how corpora can be utilized in instruction. The tools for corpus collection and extraction introduced in this study, combined with the advancements in large language models (LLMs), suggest that future educators should possess basic programming skills to apply corpora more flexibly in FOS teaching.

References

Antoniadis Georges, Fairon Cédric, Granger Sylvianne, Medori Julia & Zampa Virginie, (2006). *Quelles machines pour enseigner la langue ?*, The 13th Conference on Natural Language Processing (Leuven),

- p.795-805. <https://telearn.archives-ouvertes.fr/hal-00197393/document>.
- Anthony Laurence, (2005). AntConc: design and development of a freeware corpus analysis toolkit for the technical writing classroom, *Professional Communication Conference Proceedings*, p. 729-737. <http://ieeexplore.ieee.org/document/1494244/?tp=&arnumber=1494244&url=http://ieeexplore.ieee.org%2Fstamp%2Fstamp.jsp%3Ftp%3D%26arnumber%3D1494244>.
- Azzopardi Sophie, (2010). Présentation : La linguistique « de » corpus au-delà des champs disciplinaires : questions et enjeux transversaux, *Cahiers de praxématique*, n°54, p. 11-24.
- Baroni Marco & Bernardini Silvia (2004). BootCaT: Bootstrapping corpora and terms from the web, *Proceedings of LREC 2004*, Universidade Nova de Lisboa (Lisbonne).
- Biggs Patricia & Dalwood Mary, (1976). *Les orléanais ont la parole*, Longman (London).
- Boulton Alex, (2009). Testing the limits of data-driven learning: language proficiency and training, *ReCALL*, vol. 21(11), p. 37-51.
- Boulton Alex & Tyne Henri (2014). *Des documents authentiques aux corpus : démarches pour l'apprentissage des langues*, Didier (Paris).
- Condamines Anne, (2005). Linguistique de corpus et terminologie, *Langages*, n°157, *La terminologie : nature et enjeux*, p. 36-47. https://www.persee.fr/doc/lgge_0458-726x_2005_num_39_157_973
- Cuq Jean-Pierre & Gruca Isabelle, (2005). *Cours de didactique du français langue étrangère et seconde*, Grenoble, PUG.
- Frankenberg-Garcia Ana, (2005). A peek into what today's language learners as researchers actually do, *International Journal of Lexicography*, vol.18(3), p. 335-355.
- Gaskell Delian & Cobb Thomas, (2004). Can learners use concordance feedback for writing errors? *System*, vol. 32(3), p. 301-319.
- Gerdes Kim, (2014). Corpus collection and analyse for the linguistic layman: The Gromoteur, actes de 12ème Journées internationales d'Analyse statistique des Données Textuelles (Paris). <https://gerdes.fr/papiers/2014/Gromoteur-jadt2014.pdf>
- Gougenheim Georges, Michea René, Rivenc Paul & Sauvageot Aurélien Aurélien (1956). *L'élaboration du français élémentaire : étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, Didier (Paris).
- Habert Benoît, Nazarenko Adeline & Salem André, (1997). *Les linguistiques de corpus*, Armand Colin (Paris).
- Johns Tim, (1991). Should you be persuaded: two examples of data-driven learning, dans Johns Tim & King Philip (Éd.), *Classroom Concordancing. English Language Research Journal*, n°4, p. 1-16.
- Kennedy Graeme, (1998). *An introduction to Corpus Linguistics*, Longman (London & New York).
- Kirschner Paul, Sweller John & Clark Richard, (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching, *Educational Psychologist*, vol. 41(2), p. 75-86.
- Leech Geoffrey, (1992). Corpus and theories of linguistic performance, in Svartvik Jan (Éd.), *Corpus Linguistics, Proceedings of Nobel Symposium*, Mouton de Gruyter (Berlin/New York), p. 105-122.

- Landure Corinne & Boulton Alex, (2010). Corpus et autocorrection pour l'apprentissage des langues, *ASp*, n°57. <http://asp.revues.org/931>
- Loewen Shawn, Li Shaofeng, FeiFei, Thompson Amy, Nakatsukasa Kimi, Ahn Deongmee & Chen Xiaoqing, (2009). Second language learners' beliefs about grammar instruction and error-correction, *Modern Language. Journal*, vol. 93(1), p. 91-104.
- Mangiante Jean-Marc & Parpette Chantal, (2004). Le français sur objectif spécifique : de l'analyse de besoins à l'élaboration d'un cours, Hachette.
- Mourlhon-Dallies F.,(2006).Penser le français langue professionnelle, le français dans le monde, n°346, FIPF et CLE International, pp.25-28.
- O'sullivan Ide & Chambers Angela, (2006). Learners' writing skills in French: Corpus consultation and learner evaluation, *Journal of Second Language Writing*, vol. 15(1), p. 49-68.
- O'sullivan Ide, (2010). Using corpora to enhance learners' academic writing skills in French, *Revue française de linguistique appliquée*, n°15, p. 21- 35.
- <https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2010-2-page-21.htm>
- Scott Mike & Christopher Tribble (2006). *Textual Patterns: Key Words and Corpus Analysis in Language Education*. John Benjamins (Amsterdam/Philadelphia).
- Sinclair John McHardy, (1991). *Corpus Concordance Collocation*, Oxford University Press.
- Sinclair John McHardy, (1996). Preliminary Recommendations on Corpus Typology, Rapport technique, *Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale*. <http://www.ilc.cnr.it/EAGLES96/corpusustyp/corpusustyp.html>
- Tognini-Bonelli Elena, (2001). *Corpus linguistics at work*, John Benjamins (Amsterdam/Philadelphia).
- Tyne Henri, (2009). Corpus oraux par et pour l'apprenant, *Mélanges CRAPEL*, n°31, p. 91-111. <http://www.atilf.fr/spip.php?article3728>
- Yang Tong, (2016). Cuisitext : un corpus écrit et oral pour l'enseignement, communication présentée au *Langues sur objectifs spécifiques : perspectives croisées entre linguistique et didactique - LOSP*, Novembre 2016, Grenoble (France). <http://losp2016.u-grenoble3.fr/>.
- Yoon Hyunsook & Hirvela Alan, (2004). ESL student attitudes towards corpus use in L2 writing, *Journal of Second Language Writing*, n°13, p. 257-283.