

Teaching and learning French with the help of a French Caribbean corpus

Régis Kawecki
Université de Bretagne-Sud
Laboratoire HCTI-LiCoRN
Lorient, France

Abstract

This paper describes and partly analyses the French Learner Corpus which was built at the Centre for Language Learning at the University of the West Indies, Trinidad and Tobago. It is part of a Ph.D. research project currently being completed at the University of Bretagne-Sud in Lorient, France. Its primary objective is to improve the way French is taught and learned in the Trinidad and Tobago context using the tools developed for Learner Corpora. It is original in the sense that it targets learners with a low or intermediate proficiency in French. Learner Corpora can be of great help in the teaching and learning of foreign languages. They help researchers identify the most recurrent linguistic problems faced by specific populations of students in their endeavour to learn a particular second/foreign language, a set of difficulties generally called their interlanguage or provisional grammar. Identifying such problems is important in adapting the teachers' pedagogy to the characteristic difficulties their students encounter. It makes possible the creation of supplementary material to support textbooks which are often far too generic in their approach. Such corpora with some of their associated analysis tools can also be made available to learners for in-class or independent learning.

Keywords

Corpus Linguistics, Foreign Language, French, Interlanguage, Language Acquisition, Learning, Teaching

RESEARCH BACKGROUND AND AIMS

This project originated in the somewhat partial dissatisfaction with the Communicative Approach currently being used in foreign language teaching in settings such as tertiary language centres and Alliance Française branches. This approach emphasizes the communicative possibilities of the language to the detriment of the more formal aspects such as grammar, pronunciation and vocabulary. In doing so, it tends to sideline the written skills (reading and writing) in favor of the oral ones (listening and speaking). This is reinforced by the tutors' impression that teaching time being scarce, reading and writing should be best practiced at home through homework or independent learning. This in itself is not a problem if tutors give proper feedback to their students on what they have been doing on their own. But it seems that this is not always the case, especially in the case of writing. Tutors - often

part-timers, with sometimes other professional commitments – lack the time and/or the theoretical tools that would allow them to reflect more efficiently on their students' production. It is also worth noting that learners in their desire to master speaking and listening in the foreign language tend to neglect the writing skill as well with the inherent risk of seeing their errors becoming fossilized. It was therefore decided to collect samples of the students' output in order to build a learner corpus with one theoretical and two practical objectives in mind: to understand the Trinidadian students' interlanguage when beginning to start writing French as a foreign language in order to improve - in Trinidad and Tobago - the way French is taught (teacher focused) and also learned (student focused).

FOREIGN LANGUAGE LEARNER CORPORA

A corpus being 'a large amount of written and sometimes spoken material collected to show the state of a language' (Cambridge Advanced Learner's Dictionary), it is no wonder that it quickly found its place inside the classroom. Excerpts from native language corpora are for instance used for teaching English for specific purposes or for translation studies (Partington, 1998). By being exposed to a corpus of anthropology articles, students can be shown the specific ways in which anthropologists disseminate their research and results. A parallel corpus of texts in one language with their corresponding translated version in another tongue appearing simultaneously on the same screen is an invaluable tool to train future translators in the difficult task of transferring meaning from one idiom to another. As Sinclair mentioned, "Corpora are now part of the resources that more and more teachers expect to have access to" (Sinclair, 2004).

But it is only quite recently that publishing companies and academics began collecting data derived from learners of second/foreign languages in order to build learner corpora. The first learner corpora were concerned primarily with the teaching of English as a second (ESL) or foreign (EFL) language. Well-known examples of commercially based English learner corpora include the *Cambridge Learner Corpus* and the *Longman Learners' Corpus*. Sylviane Granger of the University of Louvain la Neuve in Belgium initiated the *International Corpus of Learner English* (ICLE), a non-commercially based corpus made up of essays written by advanced EFL students in a variety of countries (Granger, 1998). This trend has become global and EFL/ESL learner corpora have been developed in a variety of countries worldwide with Asia being probably one of the main contributors. The Hong Kong University of Science and Technology for instance is home to a large Corpus of (Chinese) Learner of English.

Learner Corpora about other languages chronologically came last and remain, on the whole, still relatively small in size. Among a few recent additions, let us mention the *International Corpus of Learner Finnish* collected at the University of Oulu in Finland and the *Falko Learner Corpus of German as FL* developed by the Institut für deutsche Sprache und Linguistik in Germany. As for French, learner corpora remain quite scarce. In the Caribbean region for instance – the geographical area we are interested in – there exists only one: the *University of the West Indies Learner Corpus* (Peters, 2009) built in Jamaica, a small collection of spoken texts.

But one characteristic of the majority of the above-mentioned learner corpora is worth mentioning: they are made of texts written and/or spoken by learners with a fairly high

proficiency in the Foreign Language. There is a real scarcity of Learner Corpora that deal with beginner or intermediate students.

TRINIDAD AND TOBAGO

The corpus described and analyzed in this paper was build at the Centre for Language Learning (CLL) of the Trinidad and Tobago campus of the University of the West Indies (UWI). The Republic of Trinidad and Tobago is made of the two most southern islands in the long Caribbean chain that separates the Caribbean Sea from the Atlantic Ocean. They are very close to Venezuela, a mere 15 km away.

The CLL was established in 1997 as the department at UWI responsible for teaching Spanish and French to its students and staff. In addition to Spanish and French – still the two most popular languages – the CLL now offers eight other foreign languages (FL) to undergraduate and postgraduate students not reading for a FL. These courses – which are also made available to the wider community – are offered at three levels of proficiency representing six semesters of study.

CLL students being directed towards the learning of a foreign language for practical purposes related to operating in a real world, the Centre follows a communicative approach to language learning integrating the four language skills and aimed at being able to interact with speakers of the target language to realize specific communicative objectives.

As for the linguistic situation found in this Caribbean nation, let us turn to Prof. V. Youssef who, on the International Corpus of English Web Site, describes it in the following terms:

Trinidad, the larger island, was French- and French-Creole-speaking until the 19th century. Today the islands share a local variety of Standard English as well as an English-based Creole... Standard English still predominates in formal and official communication, while Creole remains primarily associated with informality, emotion and humour.

LEARNER CORPORA AND FOREIGN LANGUAGE ACQUISITION

Learner corpora – these electronic collections of texts produced by foreign language learners (Granger, 1998) – can be of great help in understanding the process of foreign language acquisition and in improving the way languages are taught. They allow for the study of a particular population's learner language either synchronically or diachronically. A synchronic analysis describes the peculiar ways in which learners speak or write a foreign language at a certain stage of their learning process. A diachronic approach makes possible the study of the learners' output over time.

These analyses of foreign language learner corpora define the language acquisition concept of Interlanguage also referred to as Learner Grammar. Interlanguage points to the type of language produced by non-native speakers in the process of learning a second or foreign language. Learners develop, for the most part unconsciously, a personal set of rules to explain

the way the new idiom they are discovering organizes the world, a set that evolves as the instruction continues.

For the most proficient FL speakers, studies of learner corpora often rely on frequency data that “include the learners’ over use or under use of lexical or grammatical forms” (Barlow, 2005). As for less advanced students, they mainly concentrate on explaining the high frequency of certain erroneous forms or, in the case of longitudinal studies, they detail the different chronological stages through which some specific syntactic features are slowly mastered.

To come up with frequency data requires that the learner corpora be annotated with error tags (Granger, 1998) in a very consistent manner. This mark-up is most often done manually and is therefore very much time-consuming. The computer tools available for annotating native corpora are not really helpful when used with learner corpora and specially designed learner corpora mark-up tools are still very sketchy.

DESIGNING AND COLLECTING THE CORPUS

The corpus is made up of written essays – varying in size from 8 words to approximately 470 words – done by all CLL students learning French from the CEFR A1 to the B1+ levels (the Common European Framework of Reference for languages). These essays were done during the two in-course tests taken by CLL students each semester and were collected during the period September 2007 – July 2009 (5 different semesters). The corpus is made up of 143,661 words in total, corresponding to 914 written essays and 356 different students. Most students (92%) contributed at least 2 essays and about a quarter of them (24%) contributed 4 or more texts. The corpus has been rendered anonymous and was compiled with the express permission of the CLL learners, an authorization that was renewed each semester.

Below is an extract from an essay written by a CLL student at the B1 level. It was transcribed as it was written, including all the lexically or grammatically incorrect forms:

Il y a plusieurs monuments remarquables dans P.O.S. Mon favorit est le château Stollemeyer – un édifice grand et gris à côte de la Savanne. La Savanne est grande et elle s'appelle "Queen's Park Savannah". Il y a beaucoup d'arbres et on peut jouer au sports et se promener la. Quelques fois les agents de police se promene au cheval dans les rues de la capitale. L'embouteillage n'arret pas les cheveaux !

The CLL learner population is very homogeneous. The only variable has to do with proficiency. The majority of CLL students are from Trinidad and Tobago with a few coming from the other English-speaking Caribbean states. Some non-Caribbean learners of French – mainly from Venezuela and Colombia – have been discarded so as not to bias the results.

All students were taught French using the same series of textbooks called *Breakthrough French 1-3*. The production tasks and setting were remarkably stable. All essays were written in class without any outside help or reference materials, such as textbooks or grammars. Students were only allowed 30 minutes to complete their essay, which were for the most part concerned with giving personal information and views on very general topics related to everyday life.

Using text searching software such as XAIRA developed by the Oxford University Computing Services and Simple Concordance Program, the corpus is being scanned for lexical and structural difficulties.

The main Interlanguage features characteristic of this population (from Barlow, 2005) are:

- Influence of the textbooks: *formulaic parroting*,
- Interference of the mother tongue (L1): English,
- Interference of another prevalent FL (L2): Spanish,

These overall features influencing the way students write French are presented and illustrated below.

FORMULAIC PARROTING

This expression coined by the American Council for the Teaching of Foreign Languages (ACTFL) refers to the fact that students starting learning a new language are communicating “*minimally and with difficulty by using a number of isolated words and memorized phrases limited by the particular context in which the language has been learned*”. Formulaic parroting is characteristic of the first stages of foreign language acquisition during which beginners learn by heart a series of utterances with which to cope in some basic initial interactions, something similar to the phrase lists published by survival guides for travellers. This parroting is an integral part of the learning process. As students grow more proficient and feel better at ease in the FL, it will diminish and will gradually be replaced by more original and personal productions.

This tendency to mimic what is found in the instruction book can be traced in the writings of a sizable number of learners of this Caribbean corpus. Even at later stages, some intermediate students still lack the confidence to deviate from the model, something that might be due to the fact that these papers were written for grading purposes.

There are for instance a surprisingly high proportion of “accountants” in the corpus beginners’ output. Eating “flank steak”, a fairly rare Caribbean dish, also seems to be unusually common.

- **comptable (CEFR A1)**
 - Je suis **comptable**. Je n'aime pas ma travail
 - trente ans. Je suis **comptable** depuis dix ans
- **bavette (CEFR A2)**
 - très délicieux, surtout la **bavette**. J'aimais qu'on
 - délicieux. La **bavette** était excellent et la gateau

When describing their own island, intermediate students focus on themes introduced in the manual: they tell about its “lushness” – something not entirely irrelevant in Trinidad - and when referring to its historical buildings, most of them mention a particularly rich cultural heritage.

- **luxuriante (CEFR A2+)**
 - Le **végétation** est très **luxuriante** et quand nous
 - La **végétation** était riche et **luxuriante**. J'adore
- **patrimoine (CEFR B1)**

- est **très riche sur le plan du patrimoine**. Il y a plusieurs edifices
- **très riche sur le plan du patrimoine**. Il y a beaucoup de

INFLUENCE OF ENGLISH

As is usually the case with beginners, learners model their output on their mother tongue. Contrary to the instruction received, they will introduce syntactic features in their writing which are direct translations from English.

As the few concordance lines below will show, students spontaneously try using the English structure “to be + present participle” when French asks for the simple present tense:

- **“Je suis + verb”**
 - **traveller pour Noël. Je suis aller** à Sainte-Lucie
 - **tu travail exactement ?/Je suis venir** à Paris en
 - **en ce moment. Je suis apprend** le français.

Another instance of the influence played by English would be the adverb “*aussi*” meaning “*also*” used at the sentence initial, found repeatedly in the corpus throughout the different levels of proficiency: 142 instances in total.

- **“Aussi” at the sentence initial**
 - Je préfère la danse. **Aussi**, j'aime regarder des
 - Je suis au chômage./**Aussi**, je suis célibataire. J'ai
 - Je ne travaille pas. **Aussi**, je ne suis pas mariée,

Although not recommended in formal English, “*also*” is found commonly in the BNC at the sentence initial. In French, “*aussi*” can be found in this position as well but with the meaning of “*therefore*” or “*and so*”. It expresses a consequence of what has been said in the previous sentence. It is part of the formal registry and requires the inversion of the following subject-verb order:

- **me faisait redouter ce moment de la journée. Aussi** m'arrivait-il de me réfugier
(Sketch Engine French Web Corpus)

INFLUENCE OF ENGLISH REINFORCED BY SPANISH

When writing French, the combined influence of English and Spanish can be found in the lexicon used by students, especially when cognates in these two languages are morphologically similar. The following two examples show that it is not restricted to the beginners only:

- **Expected French words : étudiant(e)/étudier (CEFR A1)**
 - Je ne travail pas. Je suis **estudiante** à UWI. Je
 - films./Je n'aime pas **estudier** ou travailler. J'aime
 - je ne travaille pas. Je suis **estudie** en française à
- **Expected French adverb : spécialement (CEFR B1+)**
 - les politique. Et, J'**especiament** aime le magazin
 - la grande respecte **especialement** celles qui ont
 - un travail, **especiament** si on es une femme

An instance of syntactical confusion is shown below. Learners overuse the spatial preposition “*en*” when explaining when one is, lives, stays or works although they have been exposed to

numerous instances of the correct French preposition “à” to be used instead. This assimilation is again probably reinforced by Spanish which requires the preposition “en” in the same syntactical contexts.

- **Overuse of preposition “en”**
 - Nous sommes **en Londres** pour une semaine
 - la retraite **en Paris** et mon père habite **à Rouen**
 - travaille **en Port of Spain**. Il est mécanicien
 - Ma mère habite **en Port-of-Spain**. Elle est

In the second concordance line above, there are two such uses, the first one being erroneous and the second correct. This should be expected in a beginners population: rules are first accepted at the theoretical level but require time to be internalised before becoming part of its interlanguage. In the intermediate stages of their linguistic development, learners tend to mix both correct and “incorrect” forms in equal numbers. It is only in the most advanced classes that we find a steady decrease of this erroneous usage.

LONGIDUTINAL STUDY

The kind of slow internalization described in the previous paragraph can often be found in the corpus. One more example would be the learners’ use of the French quantifiers such as “*une variété de*”, “*assez de*” (enough), “*un kilo de*” which in French are not followed by definite articles. This is also the case of “*pas de*” expressing the zero quantity. It is especially true of the phrase “*beaucoup de*” (a lot of/many), a students’ favourite in their essays. The following table demonstrates the slow pace at which this particular quantifier correct usage is becoming part of the learners’ provisional grammar.

	« BEAUCOUP DE »				
	SEM.	CORRECT FORMS	INTERMEDIATE FORMS*	ERRONEOUS FORMS	% CORRECT FORMS
YEAR 1	1	8		6	57.14
	2	171	6	77	63.69
YEAR 2	3	46	4	22	69.44
	4	30		11	73.17
YEAR 3	5	63	3	29	69.47
	6	68	2	18	79.55

* *Intermediate forms correspond to the absence of any definite article but without any elision of the “e” as in “beaucoup de eau” (expected correct form: “beaucoup d’eau”)*

BENEFITS TO TUTORS

Identifying and analysing such frequency problems are important in understanding the interlanguage associated with a particular learner population in their endeavour to learn a specific foreign language. Chinese learners of French certainly do not share the same interlanguage as Caribbean learners of French. Even British learners of French do not resemble the Caribbean learners of that same language. One obvious reason lies in the fact that different linguistic repertoires lead to different interferences with the learning process.

Understanding this interlanguage can lead to improved pedagogical strategies more adapted to suit the original difficulties encountered by local learner populations. Tutors will have access to convincing evidence with which to evaluate their textbook which are – for commercial reasons mainly – far too broad in their approach of the foreign language.

This Caribbean French Learner Corpus will give tutors a definite sense of the aspects of French syntax or lexicon that require additional attention for the particular community of learners given to their care. Reinforcement can then be an integral part of the pedagogical approach through the design of specially targeted exercises.

BENEFITS TO THE STUDENTS

Learners should not be restricted to their own output but should have access to numerous students' productions. Some regular in class direct contact with the corpus could be organized in order to show students instances of their syntactical/lexical errors in context.

The corpus with some carefully chosen associated analysis tools could furthermore be made accessible to students, either on line or on CD-ROMs. It would make possible an autonomous self-discovery of their distinctive interlanguage (Granger, 2001) through data-driven learning as described by Johns (1994).

If both options are carefully monitored in order to avoid that “learners rehearse their own errors” (Seidlhofer, 2002), these direct contact with and hands-on approach to the corpus should provide students with a precious feedback on their internal French Learner Grammar and the way it develops over time, cornerstones of the normal foreign language learning process (Hanzeli, 1975). It would certainly motivate learners further by giving them a sense of the improvements achieved by students at the more advanced levels.

REFERENCES

- Barlow M. 2005. “Computer-based analyses of learner language”, in Ellis R. & Barkhuizen G. (eds.), *Analysing Learner Language*, Oxford University press, pp. 335-357.
- Boulton A. 2008. “Esprit de corpus : Promouvoir l’exploitation de corpus en apprentissage des langues”, *Texte et corpus*, n°3, pp. 37-46.
- Ellis R. & Barkhuizen G. (eds.) 2005. *Analysing Learner Language*. Oxford University Press.
- Ghadessy M., Henry A. & Roseberry R.L. 2001. *Small Corpus Studies in ELT, Theory and Practice*, John Benjamins, Amsterdam.
- Granger S. 2001. “Didactique des langues étrangères, linguistique de corpus et traitement automatique des langues”, in Marquillo Larruy M. (ed.), *Questions d’épistémologie en didactique du français (langue maternelle, langue seconde, langue étrangère)*, Poitiers : Cahiers FORELL.
- Granger S., Hung J. & Petch-Tyson S. (eds.) 2002. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, John Benjamins, Amsterdam.
- Granger, S. (ed.) 1998. *Learner English on Computer*, Longman, Harlow.

- Hanzeli, V. E. 1975. "Learner's Language: Implications of Recent Research for Foreign Language Instruction", *The Modern Language Journal*, vol. 59, n° 8, pp. 426-432.
- Johns T. 1994. "From printout to handout: grammar and vocabulary in the context of data-driven learning", in Odlin T. (ed.) *Perspectives on Pedagogical Grammar*, Cambridge University Press, pp. 293-317.
- Marquillo Larruy M. (éd.) 2001. *Questions d'épistémologie en didactique du français (langue maternelle, langue seconde, langue étrangère)*. Poitiers : Cahiers FORELL.
- Milton J. & Chowdhury N. 1994. "Tagging the interlanguage of Chinese learners of English", in Flowerdew L. & Tong K. K. (eds), *Entering Text*, Hong Kong University of Science and Technology.
- O'Malley J.M. & Chamot A.U. 1995. *Learning Strategies in Second language Acquisition*, Cambridge University Press.
- Odlin T. 1989. *Language Transfer*. Cambridge University Press.
- Partington A. 1998. *Patterns and Meanings: Using Corpora for English Language Research and Teaching*, John Benjamins, Amsterdam.
- Peters H. 2009. "Développement d'un corpus oral d'apprenants : Apport à la didactique du français langue étrangère dans les Caraïbes anglophones", *Cuadernos de Lingüística de la Universidad de Puerto Rico*, vol. 2, n° 2, pp. 21-32.
- Rybak S. & Hill, B. 2003. *Breakthrough French 1 (4th ed.)*, Palgrave Macmillan, Basingstoke, Hampshire.
- Rybak S. & Hill, B. 2003. *Breakthrough French 2 (2nd ed.)*, Palgrave Macmillan, Basingstoke, Hampshire.
- Rybak S., Ollerenshaw, J. & Hill, B. 2003. *Breakthrough French 3 (2nd ed.)*, Palgrave Macmillan, Basingstoke, Hampshire.
- Scott M. & Tribble C. 2006. *Textual Patterns: Key words and corpus analysis in language education*, John Benjamins, Amsterdam.
- Seidlhofer B. 2002. "Pedagogy and local learner corpora: working with learning-driven data", in Granger S., Hung J. & Petch-Tyson S. (eds), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, John Benjamins, Amsterdam.
- Sinclair J. 2004. *How to Use Corpora in Language Teaching*, John Benjamins, Amsterdam.