

# A Large Reference Corpus for Spoken French: ESLO 1 and 2 and Its Variations

---

Noëlle Serpollet,<sup>1</sup> Gabriel Bergounioux,<sup>2</sup>  
Annie Chesneau<sup>3</sup> and Richard Walter<sup>4</sup>

## Rationale for the colloquium

The use of corpus data has drastically changed our approach to language. However, few oral corpora are at the linguist's disposal. Moreover, the methods and techniques used to exploit written data are not always very suitable to spoken data. And even if now new technologies are available to record spontaneous speech, digitize the sound and transcribe the files, we still have to note that a large corpus of spoken French which would be available to the scientific community as a whole does not exist.

Our project aims at filling this gap and our objective is to create, analyse and distribute the ESLO corpus (see below).

The Enquête Sociolinguistique à Orléans (Socio-Linguistic Survey in Orléans: ESLO 1) was undertaken from 1968 to 1971 by British scholars. Their aim was to record spontaneous interactions to teach French as a foreign language at secondary school level. The data gathered constitute an important oral corpus of about 300 hours of speech (4,500,000 words), with interviews and other recordings.

A new survey, ESLO 2, has been undertaken by the CORAL (Centre Orléanais de Recherche en Anthropologie et Linguistique) in order to constitute, forty years on, a corpus which may be comparable in terms of data gathering and archiving. The objective was set to 400 hours or so of sound documents, that is about 6,000,000 words. Put together ESLO 1 and 2 will form a collection of 700 hours of recording and more than 10 million words, which is today considered as a reference value for the processing and investigations planned.

This colloquium will bring together four papers describing and presenting different aspects of the VARILING project, an ANR (National Research Agency)-funded initiative which aims at processing the linguistic variations within two French oral corpora.

- The first paper (G. Bergounioux) will present the corpus under investigation, and the corpus being created.

---

<sup>1</sup> CORAL, Variling Project, Centre Orléanais de Recherche en Anthropologie et en Linguistique, University of Orléans

*e-mail:* noelle.serpollet@univ-orleans.fr

*e-mail:* richard.walter@u-paris10.fr

<sup>2</sup> CORAL, Variling Project, Centre Orléanais de Recherche en Anthropologie et en Linguistique, University of Orléans

*e-mail:* gabriel.bergounioux@univ-orleans.fr

<sup>3</sup> CORAL, Variling Project, Centre Orléanais de Recherche en Anthropologie et en Linguistique, University of Orléans

*e-mail:* Vaslin.chesneau@wanado.fr

<sup>4</sup> CORAL, Variling Project, Centre Orléanais de Recherche en Anthropologie et en Linguistique, University of Orléans

*e-mail:* richard.walter@u-paris10.fr

- The second one (A. Chesneau) will report on a comparative study of the same speakers, 40 years on, and analyse the diachronical variations observed.
- The third paper (N. Serpollet) will aim at answering the following questions: How do the variations in the pronominal form of a question influence the choice of the pronoun used in the answer ? And could we predict the gender of an interviewee from his/her answers ?
- And the fourth paper (R. Walter) will tackle the technical aspect of how to create an audio corpus, then structure and process the data in order to make them available to the scientific community.

# Variling Project: ESLO2

## From a Reference Corpus to a Prototypical Corpus

---

Gabriel Bergounioux<sup>1</sup>

### 0. Introduction

The Variling project meets a demand which is recurrent in linguistics: to increase an expertise on languages – in this instance French – which integrates the variations and makes it possible to analyse them, may it be the recognition of data, their tagging or the various processings they can be submitted to. Technologies have been mainly developed using written corpora (e.g. Frantext<sup>2</sup>) and to be able to devise new knowledge, applicable to oral data, the scientific community cared about creating collections of spoken material gathered within a database.

### 1. Processes of language change

#### 1.1 Competition of variants

Recent studies show that Present-day French (or « français contemporain », contemporary French, FC) exhibits increasing linguistic differentiations (Gadet, 2003). However, not all such changes are nowadays phenomena. Some features characteristic of contemporary speech have been increasing in frequency for centuries. For example, the fading of the first part of the negative marks, noticed since the XVIth century.

Example 1 : Negative marks in CF

- |                 |                          |                     |
|-----------------|--------------------------|---------------------|
| - 'ne...pas'    | > 'pas'                  | (not)               |
| - 'ne...jamais' | > 'jamais'               | (never)             |
| - '-ne...plus'  | > 'plus' /ply/ vs /plys/ | (no more) vs (more) |

For thus, there is a lot of micro-level studies of individual features involved in social variations, but not so many macro-level studies that consider the co-variation of a large number of linguistic features. According to the variationnist approach, the aim of Variling is to concentrate on the use of language as the locus of language change, assuming that one variant progressively prevails over other variants. So, there is a competition of variants over time, as A. Chesneau's contribution will highlight it.

While some indications exist that some original changes occur (e.g. the liaison without linking (Encrevé, 1988)), most changes are assumed to be changes in frequencies, a given construction becoming significantly more frequent, or indeed being used much less frequently than before (e.g. the declining use of the future tense in FC).

Example 2 : Future tense in CF

(a) Future > Present

VA 436 je **mettra**i de l'huile dans ma poêle

CS       oui

VA 436 et puis quand mon huile **est** chaude je **mets** une euh mon om(elette mes œufs

---

<sup>1</sup> Department of Linguistics / CORAL, Université d'Orléans

*e-mail:* gabriel.bergounioux@univ-orleans.fr

<sup>2</sup> <http://www.atilf.fr/atilf/produits/frantext.htm>

(b) Future > Progressive future

**RC** alors cette question peut-être **va vous faire rire** monsieur mais je la **pose** quand même comment est-ce qu'on fait une omelette chez vous ?

**GJ 131** euh on fait une

**RC** ah c'est pourquoi je vous pose la question [rire]

**GJ 131** # [rire] elle allait en faire une justement c'est ce qu'**on va manger** ce soir #

Changes in frequency in language can only be investigated reliably on the basis of sufficiently large amounts of linguistic material of a controlled character. In order to be able to trace language change through social classes and through time, the corpus has to contain interviews from both men and women from all social classes (which implies to answer the question : how social classes can be defined today ?).

In many cases, what can be observed are changes that, until the advent of corpus-linguistics and reliable historical corpora (analyzed by specific tools (Habert, 2006)), could not systematically be investigated.

## 1.2 Vertical and horizontal processes

There are two processes of language change: vertical transmission, or inheritance, and horizontal transmission, or diffusion, reflecting multiple social and language histories manifested by different linguistic subsystems as equally valid traces of the biography, as speaker and as hearer, of CF speakers. Linguistic change is at least in part determined by the kinds of linguistic change that are traceable to the sociolinguistic method. However, it is controversial to acknowledge how language contact (substrate effects, code-switching etc.) spread over a whole community.

Vertical and horizontal modes of transition form the extremes of an alternative. However, different linguistic sub-systems behave differently. Regular sound change is normally associated with vertical transmission.

Example 3 : A phonetical change in CF  
milliard : /miljaR/ > /mijaR/

But there are various linguistic subsystems which allow reconstruction of different kinds of sociolinguistic processes (e.g. where particular processes can be shown to favour vertical or horizontal transmission types). The comparison with two interviews (1969 / 2006) of the same speaker demonstrate some effects of the latter one. On a larger scale, ESLO1 and ESLO2 give the same insight about the horizontal diachronic variations.

## 2. Two reference corpora: ESLO1 and ESLO2

### 2.1 ESLO1

There are few recordings of every-day CF all the twentieth century long (Blanche-Benveniste and Jeanjean, 1987). The most important testimony on spoken CF before 1980 is the Socio-Linguistic Survey in Orleans (Enquête Socio-Linguistique à Orléans, ESLO1), a collection of one hundred and fifty-seven interviews with several references (sociological characterization, see below), completed by recordings within a professional or private context, which adds in total to 300 hours of speech and to a corpus estimated to 4,500,000 words, including forty-three “get in touch” again, thirty-six recordings in professional context, fifty one phone calls,

twenty-nine debates or conferences, *etc.* (Bergounioux, 1992). This survey has been realized by a British professors team.

Example 4 : Sociological references (ESLO1)

- interviewee (name, first name)
- interviewer (initials)
- date and place of birth
- education level
- job
- INSEE and ESLO codes
- family status
- politic opinion
- notes about the interview
- contents
- date and place of interview
- length
- acoustic quality
- transcription (full, extracts)
- other recordings with the same interviewee

There are transcriptions carried out by linguists from Amsterdam, Leuven and now Orléans universities. See below for an example of the principles of Orléans transcriptions (no punctuation, no capital letters except on proper nouns...).

Example 5 : Extract (omelette interview 010)

<b>Omelette : 010</b>	<b>Témoïn (interviewee) : JI 306</b>	<b>Durée (Length) : 2 min</b>
-----------------------	--	---------------------------------------

**MB:** je vais vous poser une dernière question euh avant de demander de () quelques petits renseignements euh très très brefs euh comment est-ce qu'on fait une omelette d'après vous comment est-ce qu'on fait une omelette ?

**JI 306:** # oh mais i(l) y a différentes façons de faire une omelette par exemple j'en ai fait une dimanche soir #

**MB:** # dites-moi oui oui # racontez-moi alors

**JI 306:** # alors j'ai fait une omelette aux rognons aux rognons et aux champignons #

**MB:** # ah ah oui #

**JI 306:** # #

**MB:** # #

**JI 306:** # alors euh j'ai don(c) acheté deux rognons de porc et je les ai coupés p() je les ai fendus par la moitié j'ai retiré la partie grasse qui est au milieu du rognon p(u)is je les ai coupés en petits dés que j'ai fait revenir dans la poêle avec du beurre salés poivrés quand i(l)s ont été cuits je les ai mis à part dans un plat au chaud et j'ai j() avant j'avais épluché mes champignons lavés comme i(l) faut je les avais coupés j'ai fait cuire mes champignons sauter mes mes champignons c'est-à-dire que je les fait je fais évacuer euh l'eau du champignon p(u)is alors cette eau de champignon qui sent le champignon #

**MB:** # hm hm hm hm #

**JI 306:** # #

**MB:** # #

**JI 306:** # je la mets dans mes oeufs cassés vous voyez je casse mes oeufs et je je mets cette eau de champignon dans mes oeufs alors j'ai fait cuire mes champignons avec de l'ail du persil et puis salé poivré alors mes oeufs que j'avais cassés euh enfin une quantité soit huit neuf selon le nombre de personnes je les ai battus comme i(l) faut je les ai salés je les ai poivrés et p(u)is alors j'ai remis dans la poêle mes rognons avec mes champignons #

**MB:** # hm hm hm hm #

**JI 306:** # #

**MB:** # #

**JI 306:** # j'ai attendu que ma friture soit assez chaude et là j'ai versé mes oeufs dessus alors tout doucement j'ai fait prendre mon omelette avec mes rognons et mes champignons c'était délicieux oui #

**MB:** # hm hm hm [rire] #  
**JI 306:** recette à emmener  
**MB:** très bien je vous remercie infiniment

(Transcription Linda Hriba / CORAL, Université d'Orléans)

## 2.2 ESLO2

A new survey, ESLO 2, has been undertaken since 2003 by the CORAL (Centre Orléanais de Recherche en Anthropologie et Linguistique, EA 3850, Orléans Centre for Anthropological and Linguistic Research) in order to constitute, forty years on, a corpus which may be comparable in terms of data gathering and archiving. The objective was set to 400 hours or so of sound documents, that is to say about 6,000,000 words. The aim of the project is to further develop the theory of language change, which takes into account the full spectrum of language variation, namely social, register-related, dialectal variation as well as language contact. Moreover, some specific projects deal with languages in contact, spokespersons and personalities (mayor, bishop, political party and trade unions leaders, *etc.*).

Actually, the language contact situation in Orléans provides vital insights into language variation at different levels. That is the reason why a subsidiary program has been planned, LCO (Langues en Contact à Orléans / Languages in Contact in Orleans), devoted to throw light on this point. One of the main debates in contemporary linguistics concerns the issue of language change (Labov, 2001). Theories of change usually focus on one dimension of language variation : social variation (often related to register and style variation), or situations of contact between different languages, or dialect contact. However, only very few studies have taken into account these dimensions together and focused on the interaction of the aforementioned phenomena.

ESLO 2 has for characteristics to be the only programme in France (there is one equivalent in Quebec (Thibault and Vincent, 1990)) which offers to rationalize the sociolinguistic dynamics of French while benefiting from an horizon of retrospection of four decades. Today, there is no real corpus of reference in French which would be open to all kinds of search and queries. Recordings from ESLO 1 had been constituted in the 1960s for that very reason. By suspending dialectal variation and allophones' acquisitions, ESLO 2 chose to establish a questionnaire which will remain accessible to studies which will have in common to be interested in the French language without any restrictive condition, whether it come from the level of analysis or the type of object.

To control the opening of the survey, it seemed necessary to make of this reference corpus a prototypical one; that is a product controlled at each of the following stages in order to preserve the qualities required by different types of research and to enrich it with properties that will ensure its availability and its use : conditions of recording, of digitalization, of indexing but also of preservation, availability on the Net or anonymisation (Baude, 2006). An exhaustive presentation is accessible on the Net<sup>3</sup>.

This study has the ability to integrate diachronic and social variation and their interaction; it is open to a dialectal comparison with the survey undertaken in Lormont and led by the CELITH-MODYCO (Centre de Linguistique Théorique de l'Ecole des Hautes Etudes en Sciences Sociales / Theoretical Linguistics Centre of the Social Sciences High Studies School – Paris), and other studies too.

---

<sup>3</sup> <http://www.univ-orleans.fr/eslo/>

### **3. A prototypical corpus**

#### **3.1 A reference within a domain**

From early on, ESLO 2 has been designed in order to prefigure a reference within a domain which, at the international level, is getting structured and where the adoption of a standardized format for collecting, archiving, processing and analysing data is confronted to the multiplicity of developments and norms. The certification is first built by taking into account the practices in use regarding the manufacturing of oral corpora in linguistics. From the exchanges between the actors of the research, a synthesis of the recommendations will be made by the consensus of the users and the experts, which will determine a conception of the processing of ESLO 1 and of the manufacturing of ESLO 2. This conception will make it possible to exemplify these two databases for future corpora (and their processings).

Actually, the methodology of ESLO project draws on well-established corpus-linguistic procedure. The corpus is compiled within the framework of the survey. The interviews and other records will be converted to text files collected into an electronic corpus. Computerized taggers and parsers will be used. ESLO2 takes an interdisciplinary approach to investigate aspects of language variation. The objective is to gain insights into the nature of language change by using methods from computer science, sociolinguistics and linguistics.

#### **3.2 How to do ?**

The present project attempts to achieve the following :

- give a suitable overview of the state-of-the-art, especially in French linguistics,
- include large data samples to ensure token frequencies which will allow statistical analyses on all levels of the investigation,
- include both language-external (primarily sociolinguistic, e.g. variety/register, age, sex, education, etc.) and language-internal (morphosyntactic, semantic-lexical, discourse-related) factors,
- describe in detail how and why factors are chosen and coded,
- employ statistical procedures to show which factors can possibly trigger linguistic changes and which of the factors have the strongest (explanatory/predictive) effect on the phenomena overall,
- incorporate theories and methodologies from various sub-disciplines of linguistics (primarily morphosyntax, discourse theory, and sociolinguistics), with a special attention to natural language processing solutions.

Transcriber has been chosen for assisting the manual annotation of speech signals as it is more specifically designed for the annotation of broadcast news recordings and for creating corpora used in the development of automatic broadcast news transcription systems<sup>4</sup>. ESLO2 is interested in :

- annotation of named entities,
- synchronization signal / text,
- types and sizes of segmentation,

---

<sup>4</sup> <http://trans.sourceforge.net/en/features.php>

- speech turns and overlapping speech,
- *etc.*

### 3.3 What for ?

If ESLO 2 has a cumulative aim (to increase the quantity of data in order to provide comparisons with other corpora), the survey is also reflexive (to back up the survey, the processing and exploitation with an analysis of the experience in order to contribute to a definition of the norms). This conception concerns :

- a futurology on the exhaustivity of the uses with a calculation of representativeness,
- an inventory of the collection techniques (recording and digitalization formats),
- a policy to train the interviewers and to inform the interviewees in order to integrate within the variation criteria the variation linked with the interviewer. This would lead to a summer school on this theme,
- a collection of data along with the enrichment in meta-data coded for relevant extralinguistic parameters (genre, age, gender and social characteristics...),
- a coding and a cataloguing anticipating the main requests emerging in linguistics, but also in sociology, in anthropology, in history, in information and communication, *etc.*
- an aligned transcription within a perspective of normalization,
- an identification by means of searches using state-of-the-art software which will be used for designing a new software devoted to automatic transcription (in collaboration with the Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI) / Computer Sciences Laboratory for Mechanics and Engineering Sciences, a CNRS laboratory associated with Paris-6 and Paris-11 Universities),
- a tagging, with categorization and lemmatisation (in particular, research on the persistent problems of disfluences within speech and of anaphorical co-reference when in a situation of spontaneous speech, in collaboration with the Laboratoire d'Informatique / Computer Laboratory, Tours University),
- an anonymisation procedure (the identification of the questions that the anonymisation will put forward will make it possible to draw up a vade mecum of the elements to be taken into account at this stage of the work, from a research on the detection of naming entities (and not only named entities),
- storage, with archiving and indexing,
- a procedure to make the data available for public use: the building and the maintenance of the site must provide a free access via the Internet (with a conviviality and an ergonomics of the applications, if possible in a multilingual version),
- shared data: interoperability and protections, linked with the proposals expressed within the framework of the CatCod programme aimed at cataloguing and coding corpora and which extends the work of the EPML-50, « Corpus d'interaction langagière » (Corpus of linguistic interaction). The purpose is to develop consensus on best current practice for the digital archiving of language resources in French. The selected specifications will be proposed to the « Text Encoding Initiative » (TEI) consortium<sup>5</sup>.

---

<sup>5</sup> <http://www.tei-c.org/>



Beyond all this, the implementation of the monitoring (maintenance and safety) and the applications will be outlined. It is not a question of anticipating the analyses, but of allowing them.

#### **4. Analysis of linguistic variation**

Language change is a commonly acknowledged phenomenon. However, little empirical scholarship has been devoted to the simultaneous empirical study of various dimensions of variation. Analysing and explaining language change is highly complex at all the level of language as it raises questions about minimal and maximal units of cross-linguistic variation as well as about the issue of different uses of phonology and grammar (*i.e.* of the same phonology and of the same grammar). Given that information structure is an important component of grammatical variation, captured in formal and informal spoken language, the question is : how to link information and expression, from a sociolinguistic point of view ?

##### **4.1 Recipes for the omelette**

Pilot studies of samples of the material will be carried out in order to determine which linguistic parameters are the most suitable for each linguistic and each sociolinguistic feature examined. To complement the quantitative procedure, close studies of records will be carried out to uncover situational, pragmatic variation that may be difficult to code for within an exclusively quantitative framework. A first insight about the different points of view has been sketched out : ninety-six extracts, from ESLO1, when interviewees answered the question « Comment fait-on une omelette chez vous ? » ('How do you cook an omelette at home ?').

Example 6 Basic actions and developments.

Macroaction	Expansion de la macroaction	Métadonnées	Autre
je casse les oeufs  et dans la poêle très chaude	dans un bol ou dans un saladier  il y en a qui ajoutent du lait	ça se fait	voilà l'omelette pour moi
sel que je bats avec les oeufs	poivre des petits morceaux de beurre que je mets avec les  il y en a qui il y en a qui mettent une petite goutte d'eau il y en a qui mettent un peu d'huile dedans  je mets des tranches de bacon dedans	oui il y en a qui en mettent moi non non ma mère le faisait	c'est moi qui les fais toutes  au régiment on on mettait deux louches d'huile à la place des oeufs et on se faisait des petits entremets avec les oeufs
je tourne tout			madame a horreur de faire cuire les oeufs

## 4.2 Linguistics and sociology

The aim is to investigate language change in its social embedding to show that can be traced practically concerning all the major grammatical changes through the social classes (Bourdieu, 1979). Do these changes proceeding from « above » (in Labov's terms, i.e. changes that were consciously endorsed and prescribed) or « below », i.e. natural changes that first occurred in the vernacular, and found their way into more general acceptance despite school syllabus and grammatical recommendations ?

After the classification of the data, aggregates will be computed in order to ascertain whether variation and change can be attested in the data (Biber, 1985). Then appropriate differences in distributions will be tested for statistical significance to increase the reliability of the conclusions. The next step is to identify the factors that underlie the patterns attested in the data. In addition to the factors that will have been coded for in the data base, factors such as prescriptive statements and societal changes will be emphasized in the discussion.

## 5. Conclusion

Widely, this project reveals a strong ambition. Its aim is to constitute a corpus which will be a prototype, within all the steps of its construction, and which will be able to get to the same qualitative and quantitative level (including by its patrimonial dimension) as the big spoken corpora already built or being built in Europe and in the world. Within three years, the carrying out of this project must contribute, in a decisive way, to the structuring of the community (in a partnership with all the research laboratories), within a process associating the build-up of knowledge and the critics of its constitution, the analysis of tools and the

availability, the dialogue between researchers and the exchange of corpora which would have become interoperable around a theme: the consideration of variations within the language.

## References

- Baude, O. *et al.* (2006) *Corpus oraux. Guide des bonnes pratiques*. Orléans: PUO.
- Bergounioux, G. *et al.* (1992) « Enquêtes, corpus, témoins ». *Langue Française* 93.
- Biber, D. (1985) *Variations across spoken and written language*. Cambridge: CUP.
- Blanche-Benveniste, C and C. Jeanjean (1987) *Le français parlé: transcription et édition*. Paris: Didier-Erudition.
- Bourdieu, P. (1979) *La Distinction*. Paris: Minuit.
- Encrevé, P. (1988) *La liaison avec et sans enchaînement*. Paris: Seuil.
- Gadet, F. (2003) *La variation sociale en français*. Paris: Ophrys.
- Habert, B. (2006) *Instruments et ressources électroniques pour le français*. Paris: Ophrys.
- Labov, W. (2001) *Principles of Linguistic Change : Social Factors*. Oxford: Blackwell.
- Thibault, P. and D. Vincent (1990) *Un corpus de français parlé. Montréal 84: historique, méthodes et perspectives de recherche*. Québec: Département de langues et linguistique de l'Université Laval.

# Going back over the Survey “Language: the looking glass of society”

---

Annie Chesneau<sup>1</sup>

This ESLO project is part of the University of Orléans CORAL laboratory. The centrepiece of this paper is a study carried out with the same participants forty years later.

The participants of the 1968 study will be contacted so that they can take part in the new study. The new survey while meeting ESLO’s standards will also be updated.

Transcriptions of these interviews will lead to a diachronic analysis for socio-linguistic variations. Are there significant language characteristics involving diachronic variations for the same group of people over a forty-year period? Can these be observed in each individual? To what extent do they co-exist with a lexicon and structures which might appear dated for a generation of young people today? Does the variation result in a change in overall language or is it linked specifically to a well defined social group at a certain moment in time?

## **1. Formation of the group**

Following extensive research, forty-four people who participated in ESLO1 were located. Initial contact was made based on a group taken from a series of ten interviews carried out by the English team between 1969 and 1972. The same ten people were interviewed by the current author in 2005 and 2006.

### **1.1 Interviews**

The interviews on two dates over an almost forty-year gap took place under identical criteria.

### **1.2 Questionnaire**

The most recent questionnaire was reduced in size in order to avoid one-word answers. It still contained the same headings; work, hobbies, education, language, culture, society. Some questions were updated or deleted such as the following one “can you tell me what happened in May 1968?”

### **1.3 The interview**

The form of the interview was “semi open-ended”. The questionnaire was used as a guide to obtain the desired answers. Extra questions were sometimes necessary in order to complete an answer. Conversely, an interviewee could cancel out a question by answering it before it was asked.

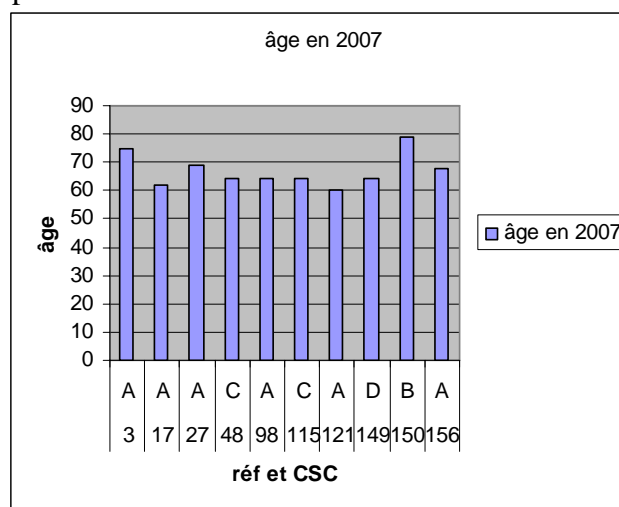
---

<sup>1</sup> Université d’Orléans, CORAL, Variling Project  
e-mail: Vaslin.chesneau@wanadoo.fr

As in the first study after 1968 the interviews took place in the participants' homes which explains the occasional presence of a third speaker e.g. a family member of the interviewee. The total length of recordings for ESLO1 and ESLO2 is 15 hours and 10 ½ hours respectively. The number of actual words in an interview varied between 5,500 and 15,700. These figures include the interviewer's words. ESLO1 interviews lasted longer time wise. However, ESLO2 actually contained more individual words than ESLO1.

## 1.4 Participants

Two women and eight men. They were divided up according to socio-economic class (SEC) as defined by Alix Mullineaux<sup>2</sup>. The two women had experienced social mobility, one going from SEC group B to group A, and the second from SEC Group D to C. No male participant had changed SEC group.



① Age profile of interviewees

The above chart shows a gap of 19 years between the oldest and youngest participant. The horizontal line illustrates each person's socio-economic class (SEC) and number used for the recordings. Six participants belonged to the highest SEC, one to group B, one to group C and the remaining two to group D.

This paper will deal with a cross-section of participants, all in the same age group and most belonging to the highest SEC (using Mullineaux). Variations in other participants' speech will be used for comparative purposes only, due to insufficient numbers. Sociolinguistic differences between male and female will not be included as only one woman belonged to SEC A. Any differences will be mentioned without drawing any general conclusions.

## 1.5 Transcription

All the recordings needed to be transcribed. "Transcriber" software was used to do this. This provided both sound and saving the inscription on an XML print-out. Having carried out the transcription myself I was able to think of the layout for the diachronic variation.

<sup>2</sup> A. Mullineaux, 1982

## 2. Diachronic variation

Henriette Walter<sup>3</sup> states in her article “Le français en mouvement” [French On The Move] that ① “lexicon has experienced phenomenal change and considerable enrichment” ② “(in France) phonology is moving towards a vocalic system favouring contrasts in timbre, abandoning contrasts in length and towards a vocalic system reduced to three nasal units” ③ “grammar is experiencing... the loss of certain verbal forms and an increasingly generalised syntactic freedom”.

Does the same person show such linguistic changes, especially in lexicon, at two distinct moments of his life? Would a group of speakers show the same changes? What impact would a forty-year gap have? Can traces of such changes be observed from interviewing the same ten people twice with a forty-year gap in interviews?

By transcribing the two recordings from 1969-72 and 2005-06 I was able to provide an analysis of both phonological and lexical changes. The analysis of diachronic variation of lexicon will centre on four main lines, a person’s view of society and the vocabulary this involves, language register for the same type of speech over a forty-year period, loan-words and the form they take on and neologisms. Examples will be given for morpho-syntactic and phonological changes. This could develop into wider research on several linguistic fronts.

### 2.1 A test

A lexical variation in diachrony could be gleaned from a short experiment. I got forty applied languages students listened to audio extracts from ESLO1 and ESLO2. These were not dated. The audio extracts were of the following question “what do you think of women at work, are you for or against it?”

The students were asked to classify each extract under 1969 or 2005 explaining their choice. This worked out perfectly as the lexicon contained crucial reference points. The evolution of society leads to changes in vocabulary for discussions on the same topic. These changes go beyond shifts in opinion on the topic as the majority of those interviewed in ESLO1 were in favour of women at work.

## 2.2 How society is represented

### 2.2.1 The child in society

“L’enfant”(child) and “le gosse” (kid) are polysemic. They express two different registers of language. “Enfant” is standard French while “gosse” is colloquial. They can be used generically or to mean “son”, “daughter”, “pupil”. The chart below gives the different uses of these words in ESLO1 and ESLO2. The use of synonyms is also noted here.

Using Gougenheim, the following frequency was observed; “enfant” (child) 305, “gosse” (kid) 110, “gars” (lad) 88, “fils” (son) 69, “fille élève” (pupil - f) 41, “jeune fille” (girl) 26. Proportionally this gives 3 “enfant” for every “gosse” (3:1). ESLO1 has 134 uses of “enfant” and 106 of “gosse”. These uses seem to demonstrate equal frequency. “Gars” was used only once. The word is almost obsolete.

ESLO2 gives 75 “enfant” and 15 “gosse” proportionally 5 “enfant” for every “gosse”. The use of kid peaked after 1968. It is currently less used than in 1958.

---

<sup>3</sup> H. Walter, 2001, p.8

<b>ESLO1</b>	3	17	27	98	121	150	156	Total
enfant générique (child)	4	6	0	3	6	6	7	32
enfant fils (child son)	12	29	2	0	8	7	19	77
fils fille (son daughter)	0	11	0	4	5	0	9	29
enfant élève (child pupil)	1	5	3	5	4	0	9	27
élève (pupil)	1	0	9	3	3	1	0	17
gosse générique (kid)	3	5	0	5	50	0	0	63
gosse fils fille (kid son daughter)	1	7	1	12	0	0	0	21
Gosse élève (kid pupil)	0	11	0	8	3	0	0	22
filles garçon (girl boy)	0	13	0	0	4	0	0	17
Total	22	87	15	37	83	14	44	305

<b>ESLO2</b>	3	17	27	98	121	150	156	Total
enfant générique (child)	1	4	1	3	1	9	0	19
enfant fils (child son)	6	7	6	4	5	5	14	47
fils fille (son daughter)	3	0	3	0	3	24	26	59
enfant élève (child pupil)	2	0	0	2	0	2	3	9
élève (pupil)	0	1	8	3	24	0	55	91
gosse générique (kid)	1	0	0	0	0	0	0	1
gosse fils fille (kid son daughter)	1	0	0	0	1	0	0	2
Gosse élève (kid pupil)	0	0	0	7	5	0	0	12
filles garçon (girl boy)	0	0	0	0	0	0	0	0
Total	14	12	18	19	39	40	98	240

② Incidents of the words “enfant” (child) and “gosse” (kid)

On a synchronic level a significant difference in the use of “gosse” (kid) can be observed. However, the declining use of this word in ESLO2 is generalised. Using “gosse” to mean “son/daughter” or “pupil” is less frequent. This is particularly true of teachers. Is this “distancing” by words a manifestation of the crisis in our schools?

### 2.2.2 Society

Ten years before ESLO, the “Français Fondamental” report noted the following stratification of society: count, countess, baron, baroness, intellectual, middle-class, communist, worker, working-class, small farmer.

The table below lists the words used in ESLO1 and ESLO2 to talk about the structure of society, urban and rural.

Ref	SEC	Class of origin	ESLO1	ESLO2
3	A	A		rurale
17	A (ESLO2)	E		personnes âgées, stressant, des HLM, un self, le médecin
27	A		petits bourgeois classe bourgeoise, prolétaires, milieux ruraux, le patois, milieu bourgeois, Monsieur le proviseur	l'économie souterraine, agriculteur
98	A	D	campagnards, paysan, fils de bourgeois, la classe bourgeoise, ouvriers	une démarche écologique,
121	A	A	paysans, fils de bourgeois, la classe bourgeoise, cadre supérieur, un paysan	agriculteur, le verlan, le paupérisme culturel, milieu favorisé, le monde agricole
150	B	E	le français moyen	DRH, restaurant d'entreprise, milieu rural
156	A	A	français moyen classe moyenne	second cycle

③ Social class groupings

Words used to represent society have also evolved since ESLO1. Words such as “classe moyenne” (middle-class), “favorisée” (advantaged), “agriculteurs” (farmers), “ruraux” (rural) have replaced “classe bourgeoise” (bourgeois), “de ceci” “de cela” (these/those people), “classe prolétaire” (working class), “ouvrier” (worker), “paysan” (peasant). The decline in Marxism is also reflected in changes in the lexicon of class.

Class differences are now defined by economics, finance and culture. A person’s place of birth or residence (town/country) is no longer crucial. As well as the “official” economy there is now a “hidden” economy.

Age is dealt with differently. “Personnes âgées” (Elderly) has replaced “les petits vieux” (old folks) and les “seniors” (senior citizens). The “malheureux” (needy) have become “les pauvres” (the poor).

## 2.3 Language reference points are different

ESLO1 participants distinguish between “good French” and “provincial dialects” (with various regional accents).

People come to learn French in Touraine and the Orléans hinterland because this is where the “best” French is spoken. “In regions like Berry, the South they speak “bad” French” (17). There “are regions that apparently speak better French than other ones” (27).



“Accents, regional pronunciations, lilted words, the old way of speaking a dialect” (150) are being abolished judges one participant.

Social class and the people one frequents influence the way people speak. “People with little education mispronounce certain words”, “a person who’s worked as a maid in bourgeois houses will speak good French. My mother and I have noticed this” (17).

Today the following observations “my country family with its strong Picard accent, Touraine French” are still made. However, it is young people’s French which is considered unusual. “There are too many departures from the norm, too many new words” “hard, meuf”. “It’s a way of talking tribally” (121), “you wonder if it’s really French or just gibberish” (98). The intonations remind one participant of the “French spoken in Morocco” (3). It is this French which influences language and not frequenting the bourgeoisie (who are no longer talked about). Cultural and economic capital defines social classes. Society is no longer structured on divisions like aristocrat, bourgeois, worker, small farmer. The interviewees are aware of this. This diachronic variation in lexicon is directly linked to changes in society. Hence, the different vocabulary used to talk about them irrespective of the interviewees’ social class or political beliefs. Changes in register not linked to speech were also observed. This remained the same for both surveys.

## **2.4 Changes in language register**

The register used by one social class has shifted. Those who belonged to the middle class in 1968 were expected to use standard if not affected French.

Words previously judged to be “more or less vulgar used by certain groups in society” (3) are now used by those who used to criticise them.

In ESLO1 the word “cochonnerie” (filth) appeared once in a sixty-minute interview, coming across as being risqué. The same speaker today uses words and sentences whose register is colloquial e.g. “ils me l’ont sucré” (they taken it off me).

The family man who used no colloquial words in ESLO1 now uses colloquial indeed vulgar terms such as “emmerdent la gueule” (piss me off).

In 1968, only participants 17 and 98 who belonged to the SEC A but were originally from the most modest SEC (D and G) employed colloquialisms. Twelve hours recording for ESLO1 produced a dozen colloquialisms from two participants. Today all the speakers used such expressions. ESLO2 produced five times more of these words. Is this just like the word “gosse” (kid) a fashion or will it last? 1968 heralded in greater tolerance and the table below shows this.

Ref	SEC	Class of origin	ESLO1	ESLO2
3	A	A	des cochonneries, rester dans les jupons de	le gourbis, un gars, faire le singe, des tas un gamin, ils me l'ont sucré, boulotter, un copain
17	A(ESLO2)	E	engueuler (conjugué ou non), éjecter des gens, ils ne sont pas très chauds, 15000 balles, des bourdes, un type, se faire avoir	je me suis défoncée, super, il est axé là-dessus, des trucs
27	A			elle a débarqué, faire la loche
48				des trucs
98	A	D	un tas, c'est dingue, elle est potable, écrire comme des cochons	truc, machin, boulot, une fumisterie, on se fait sonner les cloches, se foutre du monde, crever de faim, à tout casser, ça me gonfle, on est mal barré, la flemme, j'ai été boulé, le bazar dans la classe, on bosse, costaud, une mémoire d'enfer
121	A	A	écrire comme des cochons	un tas, un bouquin, des paperasses, je pousse une gueulante, elle n'a pas digéré
150	B	E		il emmerde les profs, être au placard, j'ai jamais été fichu, sa gueule, un gars
156	A	A	je m'y trimballe, un pot	la fainéantise

#### ④ Social class groupings

Interviewees from SEC C or D did not use a colloquial register. In interviews part of a colloquial word was pronounced but never in its entirety. Post-interview and off the record, participants use colloquialisms such as “la tambouille” (grub), “cinq cent balles” (five hundred quid). Thus an element of self-censorship on the record existed.

## 2.5 Loan words

The bogey question of ESLO1 was “what do you think of franglais?” Various answers were given. Very few speakers were against it.

“We need to get on with it on two levels. From a cultural level it’s no bad thing. Linguists might not like it but it shows that the French and the English can move forward a little bit together” (121).

Others were more measured. “The danger is ending up using clichés or automatically using an English word when a French word already exists” (003).

Others were totally against it. “I can’t stand it” (17).

Other speakers feared for the future. “There needs to be some kind of limit or we’ll end up with... I don’t know... you have to admit that there are more and more English words eh” (98)

“We’ll have to avoid using words that are 50% French and 50% English or distorted words from both languages.

The following chart shows how little “franglais” this generation uses. The percentage of loan words from English as a total for ESLO1 was 0.3% and 0.0% for ESLO2.

	ESLO1	ESLO2	
Ref	Emprunt	Emprunt	Mots français
3	planning, relaxé	winch	courriel, agenda
17	football, water, dancing, speakerine	self sandwich stressant	Animateur
27		speakerine	Journaliste
98	Parking, test, jazz, basket, club, stock, goal, basket-ball	footing foot basket tennis club shoot stop mixer (verbe)	
121	leader hobby	parking camping gadget mail	Logiciel
150	relaxation	budget steak	
156	speaker (2)	Parking, se mixer	Journaliste

⑤ English loan-words

The 1968 “le planning” has been replaced by “writing plans into a diary”. “E-mail” is under pressure from “le courriel”. “Le dancing”, “le speaker” and “la speakerine” have disappeared. Sporting terms have remained as they are universal. They have been adapted appropriately to become new words.

Social class did not influence frequency of loan-words. The French media’s love affair with English loan-words is not reflected here.

## 2.6 Loan words

The following influence the formation of neologisms, loan words especially from English and changes in technology and society. I will first calculate the number of neologisms and then briefly discuss the morphological processes involved in formation of these words.

### 2.6.1 Changes in technology

New terms appear.

Ref	SEC	Class of origin	ESLO1	ESLO2
3	A	A	photos magnétophone	piratage, pirater, courriel, cassette, SMS, TTX, portable, écran, déconnecté, l'ordinateur, l'informatique
17	A(ESLO2)	E		ils captent, ttx,
27	A			informatique, DVD, TNT, numérique, le prompteur, internet un fichier, disquette, courriel, mèle
	48			la logistique, le magnétoscope, le portable, les textos
98	A	D	il est blasé	on serait assez branché, le footing, mixer, remixer
121	A	A		nouvelle technologie, numérique, informaticien, piloter l'ordinateur, le site informatique, software, hardware, courriel, mèle, internet, des disquettes, un fichier
150	B	E		Zapper, disjoncter
156	A	A		des logiciels se mixer exploiter par ordinateur, organigramme

© New words linked to technological changes and terms in fashion

## 2.6.2 Formation of these neologisms

### 2.6.2.1 Adding a morpheme

“Ing” is added to English words. This is not used as a gerund but an action e.g. “le parking”: to park one’s car in a car-park, “le camping”: to camp in a campsite, “le footing”: to run.

“Er” can be added at the end of a word to make a verb, the suffix “age” is also used. This usually conveys the meaning of the word “pirater”: to pirate, “disjoncter”: to cut off, “zapper”: to channel-hop

### 2.6.2.2 Truncation

Football becomes “le foot”. A self-service restaurant becomes “le self”. “Le vélocipède” is a bicycle and then becomes “le vélo” again. “Le professeur” (teacher) is a “prof” (teach), the “télé” (telly) is watched, teachers do “prep” (prep work) and “manip” (lab work) in a labo (lab).

The following table provides the number of truncated words used in ESLO1 and ESLO2 (“télé”, “prof”). Only speaker 27 resists such truncations.

Ref	SEC	Class of origin	ESLO1	ESLO2
3	A	A	0	2
17	A(ESLO2)	E	1	11+ sympa, vélo
27	A		0	0
98	A	D	3	23+ la gym, le foot le basket
121	A	A	4	le bac, le tram,
150	B	E	0	4
156	A	A	0	2+1fac, vélo, manip, prépa,TP, labo

⑦ Incidents of “le prof” (teach) and “la télé” (telly), other truncated words and neologisms

### 2.6.3 Use of acronyms

No acronyms were used in ESLO1. Various acronyms were used by different speakers irrespective of social class in ESLO2. Some examples are CGT/CFDT/PS/C.E./RTT/DRH/SMS/DVD.

Incidents of truncations and acronyms may explain the greater use of words per unit of time in ESLO1 compared to ESLO2.

### 3. Conclusion

A clear diachronic variation is evident. It takes different forms.

As P. Encrevé<sup>4</sup> states, “Contrary to popular wisdom, the language of a subject is not the one spoken but the one listened to”.

The participants were recorded at two moments in their lives. When younger they had listened to their parents’ language. They were immersed in a much standardised pre-1968 French.

The students in ESLO1 (017, 121, 98) tried to reflect the new 1968 freedom in their use of words. This was not the case for all age groups.

At present the language that is heard is dominated by the omnipresent audio-visual media.

All the interviewees have slowly experienced assimilation between the language they hear and the one they speak. To this new change must be added experiences that are embedded in a person’s memory. These include lexicon, syntax, phonology and language “lessons” by parents, teachers and academics. A change in language register is thus a form of emancipation. Employing new words is inevitable due to technological advances including ideas such as globalisation.

Syntax is more resistant to change especially for important structures. Interviewees still say “dont” (of which) and not “que” (which).

Some signs of change are obvious. Participants priding themselves on their use of the double negative is mere illusion. In 1968 there were 50% of double negatives, today they only

<sup>4</sup> P. Encrevé, 1977, p.6

occur sporadically. They are used especially for emphasis “je ne veux pas” (I do not want to”. Past participles in the feminine have become less marked.

Phonological variations such as long/short vowels, open/closed vowels, and liaisons show a great deal of movement.

Will these lead to definitive changes? Other diachronic variation projects will be required to determine this, hence programmes such as ESLO, PFC.

## **Bibliography**

- Ashby, W (1982) ‘The drift of French syntax’. *Lingua* 57.
- Baude, O., C. Blanche Benveniste, M-F. Calas, P. Cordereix, I. de Lamberterie, L. Goury, C. Marchello-Nizia, L. Mondada (2005), Guide des bonnes pratiques pour l’exploitation, la conservation, et la diffusion des corpus oraux.
- Bergounioux, G (1996) ‘Etude Socio Linguistique sur Orléans’. *Revue Française de linguistique appliquée* 1-2, 87-88.
- Blanche-Benveniste, C. (2000) Le français parlé: un regard sur la syntaxe. Histoire de la langue française.
- Delais-Roussarie, E. and J. Durand (2003) Corpus et variation en phonologie du français.
- Deulofeu, J. (2001) ‘L’innovation linguistique en français contemporain’. *Le Français dans le monde, oral, variabilité et apprentissages*, 18-31.
- Encrevé, P.(1977) ‘Linguistique et sociolinguistique’, *Langue française*, 4-15.
- Gadet, F. (1996) ‘Niveaux de langue et registres de la traduction’, *Palimpsestes* 10,17-40.
- Gaudin, F. (2004) ‘La socioterminologie’, *Langages* 157, 81-92.
- Marchello-Nizia, C. (1997), ‘Variation et changement, quelles corrélations?’, *Langue Française* 115, 117-123.
- Mullineaux, A. And M. Blanc (1982), ‘The problem of classifying the population sample in the sociolinguistic survey of Orleans (1969) in terms of socio-economic, social and educational categories’, *Review of Applied Linguistics* 55, 3-37.
- Walter, H. (2001) ‘Le français en mouvement’. *Le Français dans le monde, oral, variabilité et apprentissages*, 8-18.

Gougenheim200/lisezmoi.gougenheim.htm

Tell me how you cook and I will tell you who you are  
How can a question such as “How do you make an omelette?”  
illustrate the morpho-syntactic and sociological variations found in  
the ESLO oral corpus?

---

Noëlle Serpollet<sup>1</sup>

## Abstract

Corpus linguistics has become a methodology totally essential for the study of variation from a pragmatic, morpho-syntactic or even social point of view. In this paper I propose to study these synchronic variations in a sub-part of the ESLO oral corpus composed of ninety-four exchanges ‘question-answer’. I will examine this test sample called “Eslomelette” in order to analyse the answers given to the following question « How do you make an omelette ? ».

I will first observe, at a pragmatic level, the variation in the way this embarrassing question is asked : does the interviewer feel that he/she has to comment on the question or not ? And as a result, what type of pronominal form is used ?

Then, at the morpho-syntactic level, I will describe how this pronominal variation within the question conditions the choice of the pronoun used in the answer.

This will lead me to classify the different answers in two categories (‘type’ and ‘occurrence’-recipes). I will end up by examining the variation at the social level: I will see if it is possible to predict the social origin or the sexual identity of an interviewee from his/her answers.

## 1. Introduction

The development and use of computerised corpus data has drastically changed our approach to the study of language. And for nearly four decades now, corpus linguistics research has grown and expanded around the creation and analysis of written corpora, getting bigger every day. Furthermore, for about twenty years, studies carried out on corpus of spoken data have brought about a new approach to linguistic research.

On all the previous points, I refer the reader to Baude (2006) where the state of the art in corpus linguistics in relation to the constitution, the archiving, the analysis and the diffusion of spoken data is described at length. This guide takes into account not only the large perspectives that the digitalization of spoken language corpora has opened, but also the varied ethical and legal issues that the linguist now has to face.

However, few oral corpora<sup>2</sup> are at the linguist’s disposal. Moreover, the methods and techniques used to exploit written data are not always very suitable to

---

<sup>1</sup> CORAL, Variling Project, Centre Orléanais de Recherche en Anthropologie et en Linguistique, University of Orléans

*e-mail:* noelle.serpollet@univ-orleans.fr

<sup>2</sup> I will use the terms ‘oral corpora’, ‘corpus of spoken data’ or ‘spoken language corpora’ to cover the French term “corpus oraux” defined in Baude (2006: 19) as ordered collections of recordings of oral and multimodal linguistic productions.

spoken data (Habert *et al.*, 1997:13). And even if now, new technologies are available to record spontaneous speech, digitize the sound and transcribe the files, Abouda and Baude (2006) remark that a large corpus of spoken French which would be available to the scientific community as a whole does not exist yet. Contrary to what is available to a researcher when he/she wants to work on the English language (spoken and written) or even on written French, when a linguist wants to analyse spoken French, he/she cannot get hold of a large reference oral corpus which would be open to all kinds of search and queries.

The VARILING project undertaken by the CORAL team (Centre Orléanais de Recherche en Anthropologie et en Linguistique) aims at filling this gap and our objective is to create, analyse and distribute the ESLO<sup>3</sup> corpus.

I will present here an example of the type of research undertaken on the ESLO 1 corpus of spoken French compiled between 1968 and 1971. My aim is to provide a snapshot of one aspect of linguistic analysis applied to oral data. What can be done with a small corpus of spoken language?

This paper will report a study in which I analysed the variations from a pragmatic, morpho-syntactic and social point of view in a sub-part of the corpus called “Eslomelette”. I will aim at answering the following questions: How do the variations in the pronominal form of a question such as « How do you make an omelette ? » influence the choice of the pronoun used in the answer ? And would it be possible to predict the gender of an interviewee from his/her answers ?

## 2. Background

### 2.1 A lack of large corpora of spoken language

As was mentioned in the introduction, it is necessary to remark that as far as spoken French is concerned, even if important work is being carried out, the linguists do not have at their disposal, a large reference corpus in which the variations of the French language would be easily searchable, retrieved and analysed.

But what is a reference corpus? This question is answered in Bergounioux (2007), where the corpus under investigation (ESLO 1), and the corpus being created (ESLO 2) are presented. Moreover, many studies reporting corpus analyses mention the corpora they have used, but these data are seldom available<sup>4</sup> to other researchers, as they are not largely freely distributed. The technical aspect of how to create an audio corpus, then structure and process the data in order to make them available to the scientific community is tackled by Walter (2007) who exemplifies how such a big corpus can be used as a prototype.

This paper is part of the larger project *A large reference corpus for spoken French: ESLO 1 and 2 and its variations*<sup>5</sup> which can be divided into four sub-projects:

- description of the two distinct ESLO corpora forty years apart, which eventually will constitute a reference and prototypical corpus for spoken French,
- observation of the diachronic variations<sup>6</sup> (from ESLO 1 to 2),

---

<sup>3</sup> Enquête Socio-Linguistique à Orléans / Socio-Linguistic Survey in Orléans.

<sup>4</sup> Abouda and Baude (2006: 145) talk about ghost corpora, always mentioned as proof, but never appearing to anyone else but the one talking about them.

<sup>5</sup> See Serpollet (2007).



- analysis of some synchronic variations (within ESLO 1),
- description of how spoken data are gathered and made available to the larger scientific community.

## 2.2 Our project: Variling

Nowadays, the variationist approach is one of the main aspects being taken into account in several projects involving spoken data. Therefore, systematically collected spoken data are of great interest to carry out such observations and to study linguistic practices among the population. And since oral production is much underrepresented in language studies in general, let alone corpus work, the objective of our project is to develop a large reference corpus whose analysis will take into account the variations within spoken French.

The Variling project<sup>7</sup> is an ANR<sup>8</sup> (National Research Agency)-funded initiative meeting a demand which is recurrent in linguistics, that is to increase an expertise on languages which would integrate the variations (in spoken French) and make it possible to carry out their analysis: may it be the recognition of data, their tagging or the various processing they can be submitted to. Technologies have been mainly developed using written corpora and to be able to devise new knowledge which would be applicable to oral data, the scientific community needs to create collections of spoken material gathered within a database and has started to do so with the exploitation of a sound archive from the 1970s and the compilation of a new corpus of spoken French.

The most important testimony on spoken French before 1980 is the Socio-Linguistic Survey in Orléans (Enquête Socio-Linguistique à Orléans, ESLO 1), a collection of around 157 face-to-face interviews completed by recordings within a professional or private context. A new survey, ESLO 2, is being carried out in order to constitute, forty years on, a corpus which may be comparable in terms of data gathering and archiving.

ESLO 2 has for characteristics to be the only programme in France (there is one equivalent programme in Quebec) which offers to rationalize the sociolinguistic dynamics of French while benefiting from a horizon of retrospection of four decades. From early on, ESLO 2 has been designed in order to prefigure a norm within a domain which, at the international level, is getting structured and where the adoption of a standardized format for collecting, archiving, processing and analysing data is confronted to the multiplicity of developments and norms.

Variling will determine a conception of the processing of ESLO 1 and of the manufacturing of ESLO 2 which will enable us to use these two databases as examples for the construction of future corpora and for their processing.

This project reveals a strong ambition as its aim is to constitute a corpus which will be a prototype, within all the steps of its construction and which will be able to get to the same qualitative and quantitative level as the big spoken corpora already built or being built in Europe and throughout the world.

---

<sup>6</sup> Chesneau (2007) reports on a comparative study of the same speakers, 40 years on, and analyses examples of diachronical variations observed from one oral corpus to the other.

<sup>7</sup> Traitement des variations linguistiques dans les corpus / Processing of the linguistic variations within corpora.

<sup>8</sup> Agence Nationale de la Recherche.

### **3. Data**

#### **3.1 ESLO 1**

Lonergan *et al.* (1974:1) state in the catalogue of recordings that the origins of ESLO date back to 1966, at the time of the audio-visual revolution of modern language teaching in Great Britain. The introduction of new techniques, and most of all, the growing importance given to non-literary speech revealed an acute need of authentic samples of spontaneous spoken French.

The Enquête Socio-Linguistique à Orléans was undertaken in the French city from 1968 to 1971 by a group of British scholars, French lecturers and sociolinguists from the University of Essex in Colchester. Their aim was to record spontaneous interactions to teach French as a foreign language in secondary and higher education and to constitute a corpus of recordings of spoken French which would be available to researchers (from linguistics, sociology, language teaching). The data gathered add in total to 300 hours of speech and constitute an important oral corpus estimated to 4,500,000 words, with face-to-face interviews and other types of recordings. A comprehensive presentation of the survey can be found in Blanc *et al.* (1971) and Ross (1974, 1979).

Unfortunately, this survey has not been very much exploited since, and lengthy and thorough analyses have been sparingly carried out... until now.

#### **3.2 ESLO 2**

As was previously explained, the CORAL team now not only aims at analysing these data (digitalization, tagging, exploitation, diffusion) but is also constituting a second survey, ESLO 2, within a variationist diachronic perspective in order to compare spoken French forty years on. This survey will take up the main characteristics found in ESLO 1 but will adapt the sociological and linguistic parameters to the current situation<sup>9</sup>. The objective was set to 400 hours or so of sound documents, that is about 6,000,000 words. Put together ESLO 1 and 2 will form a collection of 700 hours of recording and more than 10 million words, which is today considered as a reference value for the processing and investigations planned.

### **4. Method of analysis**

#### **4.1 Objectives of this study**

I will examine a part of the corpus that I have described above through the analysis of the answers given to the following question « How do you make an omelette ? ». This sub-corpus studied is called “Eslomelette”. It is composed of ninety-four exchanges ‘question-answer’, which object is the description of the recipe for the omelette. Each recording lasts from 30 seconds to 3 minutes and amounts to a total of about 19,000 words.

The object of this paper is the analysis of the pronominal variations within the answer to a ‘specific’ question. I put the word specific in inverted commas because

---

<sup>9</sup> See Chesneau (2007) for a study of the lexical variations across the two corpora.

we will see that the form of the question is not stable but varies, leading to the use of several pronominal forms within the question, which in turns leads to a variation in the personal pronoun used in the response.

I will start by observing, at the pragmatic level, the variations in the way this surprising question is asked : does the interviewer feel that he/she has to comment on the question or not ? And as a result, what type of pronominal form is used ?

Then, at the morpho-syntactic level, I will describe how this pronominal variation within the question conditions the choice of the pronoun used in the answer.

This will lead me to classify the different answers in two categories ('type' and 'occurrence'-recipes) and to see that the two types of recipes generally given by the interviewees are conditioned by specific questions.

I will end up by examining the variations at the sociological level : I will see if it is possible to predict the sexual identity of an interviewee from his/her answers.

## 4.2 Corpus-driven approach

Within the corpus linguistics approach, observation contributes to theory more than theory contributes to observation (Leech, 1992: 111). It is assumed that empirical data such as corpus data enable the researcher to make objective statements and not subjective ones based upon their own internalised cognitive perception of the language. Then one can draw conclusions, make generalisations and rules and proceed from the data to the theory (i.e. the corpus-driven approach which uses corpus data to build up a theory – see on this issue Sinclair, 1991; Hunston & Francis, 2000; Tognini-Bonelli, 2001a and 2001b).

In the case of the sample corpus “Eslomelette”, the researcher was faced with an existing corpus, that had been created almost forty years ago for didactic purposes and which was available for analysis. It was there and ready to use but had not been created, for example, to study the variations within the pronominal forms.

Hence, between the two possible approaches to the analysis of a corpus, one aiming at confirming existing theories and the other letting evidences emerge from the corpus, I chose the latter. The methodology used was thus *corpus-driven* (due to the type of corpus), as it was difficult to have a hypothesis upfront and as the researcher did not know really what to expect from such an original corpus. The only objective was to make an internal comparison of an omelette with another one and to measure the synchronic variations observed. Then after listening to the corpus, I noticed a variation in the types of personal pronouns, not only used in the questions asked but also in the responses given and therefore, I decided to focus on this element.

I will provide a quick overview to highlight the differences between a corpus-driven approach and a corpus-based approach.

Contrary to corpus-based work, *corpus-driven work* attempts to define the categories of description step by step, *in the presence of* specific evidence from the corpus, according to Tognini-Bonelli (2001b). Hence the hypotheses are arrived at inductively on the basis of corpus data. She adds (ibid.) that the “essential methodology of CDL [Corpus-driven Linguistics] is to exercise the researcher’s intuition in the presence of as much relevant data as can be assembled. It is accepted that there is no such thing as a theory-neutral stance, but in CDL the attempt is made to suppress all received theories [...]”.

On the other hand, a *corpus-based approach* is a methodology that uses corpus to expound, test or exemplify existing theories that were produced before large

corpora came into use. It enables the researcher to test his/her hypotheses against corpus data. According to Tognini-Bonelli (2001b) “corpus-based work relates data to existing descriptive categories adding a probabilistic extension to theoretical parameters which are already received, i.e. established without corpus evidence”.

## 5. Results obtained: variations observed

The observations carried out enabled me to draw out a hypothesis that will be developed in 6.1. From what I heard in the recordings of interactions, I was tempted to classify the variety of responses in two types: ‘*type*’-*recipés*, found within an answer using the personal pronoun *on* and describing the different tasks to undertake in order to make an omelette and ‘*occurrence*’-*recipés* in an answer using the pronoun *je* and telling a specific way of making an omelette or recounting a personal experience). I also observed that the two types of recipes given by the interviewees were conditioned by the form of the questions asked. Hence what was left to do was to check what pronominal form used by the interviewer in the question triggered a form in *on* or *je* or other in the answer.

### 5.1 Pragmatic level – Variations in the form of the question

I will first of all describe if the question used was a direct one or was commented on by the interviewer.

This question was asked by both British and French interviewers. Its original wording was « Comment est-ce qu’on fait une omelette ?<sup>10</sup> Pourriez-vous m’expliquer comment on fait ? ».

The first thing that should be noted when one does listen to the recordings is the variety in the forms taken by the question. It can be preceded by a justification helping the interviewee understand why such a surprising question is asked and therefore contains a comment, as in:

- (1) Extract 119 : euh là je vais passer à() à des questions euh / différentes enfin plus précises à des questions sur euh la langue / générale / euh / si euh / oh +[pron=\*] ah oui j’ai d’abord une autre question à vous poser / ça va paraître vous paraître idiot / mais euh [rire] **comment est-ce que vous<sup>11</sup> faites une omelette**
- (2) Extract 60 : je vais vous poser une question qui va vous faire rire mais euh les Anglais sont étonnés [rire] les Anglais sont étonnés par les façons différentes de vivre qu’il y a en France et en Angleterre [bb] et ils trouvent que ça ça joue même pour les choses les plus petites +[pron=pi] alors euh **est-ce que vous pouvez racontez par exemple comment vous vous<sup>12</sup> faites une omelette**
- (3) Extract 124 : bon merci madame maintenant je vais vous poser une question qui peut-être va vous faire rire mais je la pose quand même /**comment est-ce qu’on<sup>13</sup> fait une omelette chez vous<sup>14</sup> madame**

<sup>10</sup> This could be roughly translated by « How do you make an omelette ? – Could you explain to me how to make one? » or « How is an omelette made? – Could you tell me how an omelette is made? ».

<sup>11</sup> *Vous* = you; How do you make an omelette ?

<sup>12</sup> *Vous vous* = the interviewer insists on the pronoun; How do you (yourself) make an omelette ?

<sup>13</sup> *On* = generic indefinite form, meaning one in general, somebody (nobody in particular), people; How is an omelette made? How do people generally make an omelette?

<sup>14</sup> *chez vous* = can mean ‘in your household’ or ‘in France’ / in your country / at home / in the region you live in/come from vs ‘in England’ : in our/their country.

Or it can be asked directly, without any comment, as in:

- (4) Extract 116 : **pourriez-vous m'expliquer comment vous feriez une omelette** / s'il vous plaît
- (5) Extract 117 : **est-ce que vous pourriez m'expliquer comment on fait une omelette**

The second remark is that even when there is no comment made on the question, this one is introduced by a polite request. I found only six cases out of ninety-four in which the question was asked straight away without any attempt to soften its brutality or the surprise it might create. In those cases, the question asked corresponds to the one written in the questionnaire (« Comment est-ce qu'on fait une omelette ? Pourriez-vous m'expliquer comment on fait ? »), whereas the other ones are slightly modified and / or commented on.

- (6) Extract 28 : et **comment est-ce qu'on fait une omelette** pouvez-vous m'expliquer
- (7) Extract 83 : et **comment est-ce qu'on fait une omelette** est-ce que vous pourriez m'expliquer comment on fait

I would like to indicate that I only tackle the problem of the question here in so much as it influences the answer (see 5.2.) and that I mainly work on the choice of the pronominal forms (at the morpho-syntactic level). For a much more detailed analysis of a peculiar question within the interview, see Abouda and Perrot (2006) who concentrate on the variations at the pragmatic level and on the strategies adopted to legitimate such an embarrassing and surprising question. I will rapidly summarize below some of their findings. The main types of comments added to the original question form and the different strategies used to justify the question are the following:

- clarification of the aim of the question with
  - o a cultural explanation: to report on the cultural differences between England and France,
  - o a technical explanation: to check the sound level of the recording,
  - o to relax the situation, change the topic, put the interviewee at ease,
- different types of modalisation are found:
  - o direct characterization of the question by the interviewer: the question is *amusing* ('*amusante*'), *different* ('*d'un genre différent*'),
  - o indirect characterization, the interviewer anticipates the interviewee's reaction: question which will sound *stupid* ('*qui va vous paraître bête*'), which *will make you laugh* ('*qui va vous faire rire*'),
- several marks of politeness, special precautions which add to the comments listed above and which seem to correspond to different forms of distance from the question: *can I ask you...* ('*est-ce que je peux vous poser...*'); *I would like to ask you...* ('*je voudrais vous poser...*'),
- a strategy consisting in placing the question in relation to others to come : *before* talking about the trade union ('*avant de parler du syndicat*').

### 5.1.1 Question commented or not

I will observe if the interviewer feels that he/she needs to comment on the question or not. It will also be interesting to verify if this presence or absence of comment is linked to their nationality.

<b>Question with comments</b>	<b>56</b>	<b>59.6%</b>
<b>Directly asked question</b>	38	40.4%
<b>TOTAL</b>	94	100%

**Table 1:** Types of question

Table 1 shows that the question holds a comment in the majority of the cases (59.6 percent). Therefore, I wondered if these questions were more embarrassing to ask when the interviewer was British.

Type of Q	Nationality				TOTAL
	British		French		
<b>Commented question</b>	40	71.4%	16	28.6%	56 100%
<b>Direct question</b>	32	84.2%	6	15.8%	38 100%
	72		22		94 100%

**Table 2:** Type of question asked and nationality of the interviewer

We are shown in Table 2 that the comments are made in 71.4 percent of the cases by British interviewers. However the questions are also asked directly by a majority of Britons (84.2 percent). This result is not surprising when we know that 76.6 percent of the interviewers were indeed British.

Type of Q	Nationality			
	British		French	
<b>Commented question</b>	40	<b>55.6%</b>	16	<b>72.7%</b>
<b>Direct question</b>	32	44.4%	6	27.3%
<b>TOTAL</b>	72	100%	22	100%

**Table 3:** Nationality of the interviewer and type of question asked

Table 3 shows that, in fact, 40 out of 72 Britons asked a commented question, which adds up to 55.6 percent: this is the majority of the cases, but not a very high one. If we compare with the French, 16 out of 22 (that is 72.7 percent) also asked a commented question. Therefore, it seems that it is the French and not the British interviewers that find the question awkward and embarrassing to ask. They probably feel that they need to justify it more than their British counterparts and thus they present it as a-typical.

### 5.1.2 Influence of the type of question on the pronouns

I will now check and see the type of pronominal form used in each type of question.

Pronominal form within the commented question	Raw numbers	Percent	Total of forms (including the changes)
ON – chez vous	32	57.1%	
Question reformulated using VOUS	2*	3.6%	
VOUS	12*	21.4%	21 37.5%
VOUS VOUS	5*	8.9%	
ON	2	3.6%	
Question reformulated using VOUS	2*	3.6%	
Infinitive	1	1.8%	
<b>TOTAL</b>	<b>56</b>	100%	

**Table 4:** Pronouns used in the commented questions

We are shown in Table 4 that the personal pronouns used within the questions with comments are varied: we can note a majority of *on – chez vous* (see footnotes in section 5.1. for a clarification about the types of pronouns used) with 57.1 percent of the cases, then *vous* in 21.4 percent of the cases (adding up to 37.5 percent of the pronouns used in a commented question when we take into account the questions in which the interviewer started with one type of personal pronoun and ended the question using *vous* due to a lack of response), *on* in only 3.6 percent of the case and one *infinitive* form.

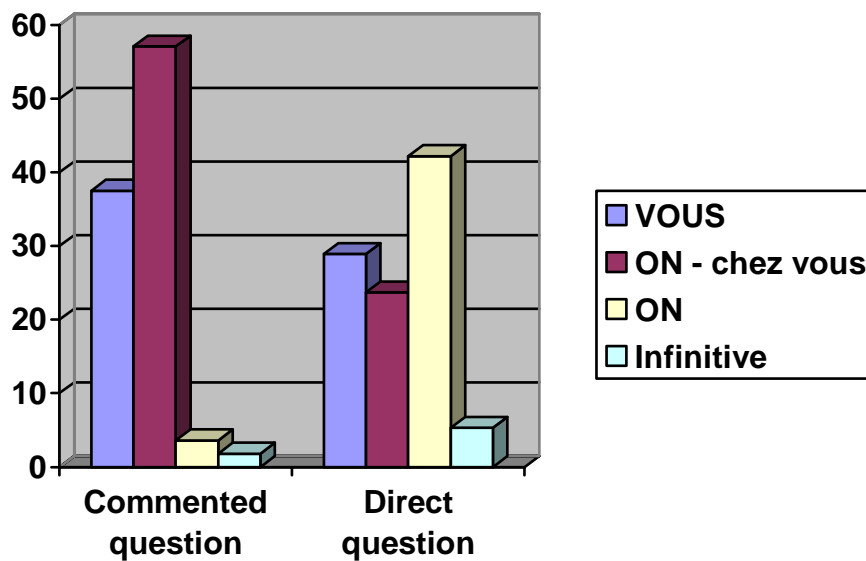
I now need to examine the type of pronominal forms used in the questions bearing no comment.

Pronominal form within the direct question	Raw numbers	Percent	Total of forms (including the changes)
ON	16	42.1%	
VOUS	10*	26.3%	
ON – chez vous	9	23.7%	11 28.9%
Question reformulated using VOUS	1*	2.6%	
Infinitive	2	5.3%	
<b>TOTAL</b>	<b>38</b>	100%	

**Table 5:** Pronouns used in the direct questions

Table 5 shows that regarding the questions asked directly, we find the same pronominal forms, but with a different distribution: the pronoun the most frequently used is *on* in 42.1 percent of the cases, then *vous* (26.3 percent and up to 28.9 percent), *on – chez vous* only appears in 23.7 percent of the questions and we can note also two infinitive forms ('est-ce que vous pourriez m'expliquer comment *faire* une omelette' / Could you tell me how to make an omelette?).

Figure 1 below summarizes the results from the two previous tables.



**Figure 1:** Overview of the distribution of the pronominal forms within two types of questions

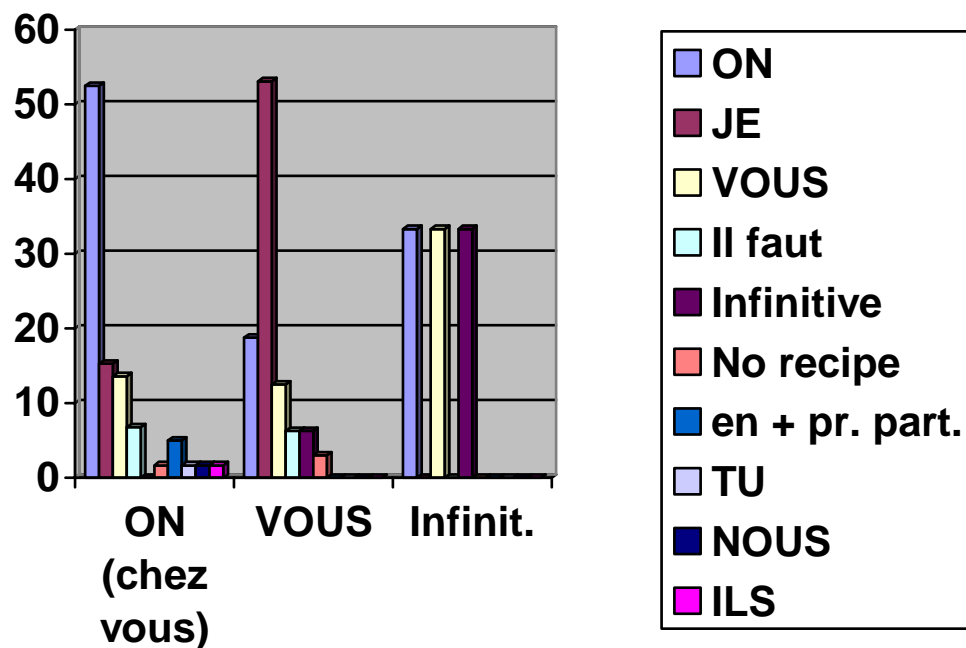
It can be noted that the presence of comments largely favours the form *on – chez vous* in comparison with the use of *vous*, whereas the absence of comment favours the pronoun *on*, then the form *vous* and lastly the form *on – chez vous*.

The form *on – chez vous* presupposes in itself a cultural contrast (*chez vous* vs. *chez nous*) and differs from the classical questions centered on the interviewee’s experience. It then seems logical to find this form along with the different types of comments listed previously. On the contrary, with *on*, a general and more straight-forward question is asked about the way in which one is supposed to make an omelette. With *vous*, the interviewer stays centered on the interviewee, so there is no opposition or break with the rest of the questionnaire (not a formal one anyway, as there is one as far as the contents of the question is concerned).

## 5.2 Morpho-syntactic level – Variations within the answers

What is left to examine is the morpho-syntactic level. I will now describe in what way the personal pronoun used by the interviewer in his/her question influences the choice of the pronominal form used in the interviewee’s response.





**Figure 2:** Three varieties of questions and the pronouns used in their corresponding answers

Figure 2 above details the results obtained:

We need to note first of all that two types of questions have been put together, those using the form *on – chez vous* and those using the pronoun *on* (first cluster of bars in the graph). They amount to 62.8 percent of the questions. The questions with *vous* make 34 percent of the total and the rest contains only three infinitives (3.2 percent for the third cluster).

If we look at the first cluster of bars, we see that a question using *on – chez vous* leads in 52.5 percent of the cases to an answer containing the pronominal form *on* (either general, or more personal, meaning ‘*nous*’ / *we* as a the family, then we find an answer with *je* in 15.3 percent of the cases (these responses relate back to the personal experience of the interviewee, to a specific situation), then with *vous* (13.6 percent) they relate to the interpersonal relationship between the interviewer and the interviewee. We also find answers without any personal pronoun, but with *il faut + infinitive* (you need to / you have to), or an *infinitive* (pour *faire* une omelette... / to make an omelette) or that do not contain any recipe or with *en + present participle* (‘*en cassant les oeufs*’ / (by) breaking the eggs; ‘*en battant des oeufs avec une fourchette*’ / (by) beating the eggs with a fork), and in 1 percent of the cases for each form with the pronouns *tu*, *nous* and *ils*.

The second cluster of bars shows that a question with *vous* is answered by a response with the pronoun *je* in 53.1 percent of the extracts, with *on* in 18.8 percent of the cases and with *vous* in 12.5 percent, etc.

We should note that in 31.9 percent of the cases, *changes* within the pronouns used in the answers are observed: the interviewee starting with *je* and ending his/her recipe with *on*. These changes happen for 76.7 percent in the extracts containing the form *on – chez vous* in the question. This is probably due to the fact that this form is very ambiguous and can either lead to the description of a family recipe (*chez vous*

meaning at home, in your household, in your family) or towards a national recipe (*chez vous* in France vs. *chez nous* in England).

## 6. Analysis of the answers and classification

Throughout the analysis and the observation of the variations within the responses as far as the use of personal pronouns is concerned, it has been possible to notice that a general trend prevailed, i.e. that the recipe to make an omelette described by the different interviewees fell into two categories, according to what pronoun was used in this recipe:

- the answer using the pronominal form *on* seems to contain a ‘type’-recipe listing the several tasks undertaken in order to make an omelette,
- whereas the answers with the pronoun *je* would contain an ‘occurrence’-recipe recounting a personal experience or explaining a particular way of making an omelette: how the interviewee does make the omelette.

### 6.1 Typology of the responses given: two types of recipes

Now that my hypothesis has been reformulated, I will work the other way round: I will classify the responses according to the category they fall into and I will check if the impression I had regarding the use of the pronouns was right and founded. I will thus present a typology of the answers according to the variation in the type of recipe.

I start with the re-examination of the questions containing the forms *on* and *on – chez vous* and I observe the following results in the figure below:

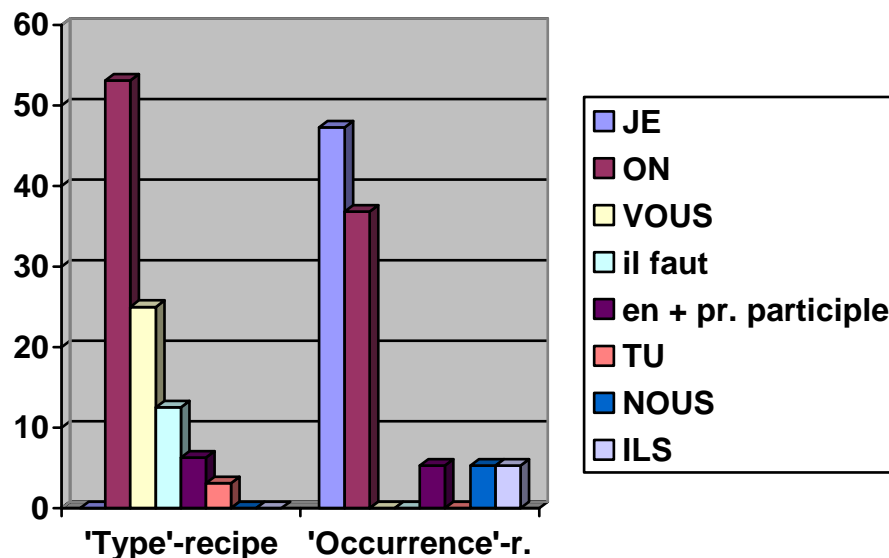


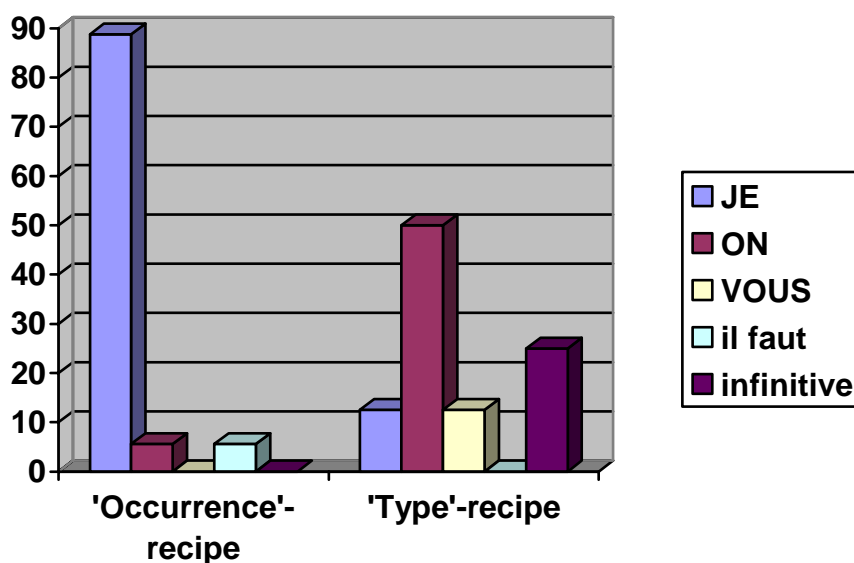
Figure 3: Questions using *on / on – chez vous*: Two types of recipes and the pronouns used

Out of 59 questions asked, 32 responses expressing a ‘type’-recipe, i.e. a general description of what is expected to be done are found, which amounts to 54.2 percent of the cases. It should be noted that in 32.2 percent of the answers, an

'occurrence'-recipe is found, in 11.9 percent we have a variation in the response (starting with a generic recipe and finishing with the expression of a more personal experience) and in one extract we do not have any recipe at all (1.7 percent). In these extracts containing a 'type'-recipe, we can see from Figure 3 that the pronoun *on* is used in the majority of the cases (see the illustration in example 8 below), that is in 53.1 percent of the responses, *vous* in 25 percent, etc.

- (8) Extract 070 : **RC:** alors cette question que je vais vous poser va vous faire rire  
**FS 725:** [rire]  
**RC:** mais je la pose quand même comment est-ce qu'on fait une omelette chez vous  
**FS 725:** [rire] euh j'aime beaucoup faire les omelettes alors euh **on** prend des oeufs **on** les casse **on** met dans un saladier **on** met **on** les bat **on** met du sel du poivre et **on** ajoute un p(e)tit peu d'eau pour qu'elle soit plus légère ensuite **on** met soit du gruyère [pron=pif] soit des pommes de terre soit du jambon selon l'omelette que l'**on** désire ensuite quand elle est très bien battue **on on** verse dans la p(oêle) dans une poêle **on** met **on** la sert assez moelleuse pas trop sèche pas trop #1 baveuse comme on dit [rire] #

I continue with the re-examination of the questions containing the pronoun *vous*.



**Figure 4:** Questions using *vous*: Two types of recipes and the pronouns used

Thirty-two questions use the pronoun *vous* in their formulation and they lead to a response describing an 'occurrence'-recipe, i.e. recalling a recipe specific to the interviewee or a personal experience, in 56.3 percent of the cases. We can also notice that 25 percent of the answers contain a 'type'-recipe, 15.6 percent vary from the type to the occurrence and one extract does not describe any recipe (3.1 percent).

Figure 4 shows that the pronoun *je* is used in 88.8 percent of the 'occurrence'-recipe extracts (illustrated in example 9 below), and that the forms *on* and *il faut* + *infinitive* are found in 5.6 percent of the cases each.

- (9) Extract 55 : CS: [b] oui [rire] bon là on va faire des passer à des des questions qui sont un peu plus / plus précises i(1) y en a qui vont vous paraître un peu bêtes [rire] qui vont vous faire rire / mais enfin c'est / c'est utile au point de vue culture et tout  
 CW 739: / hm hm /  
 CS: est-ce que par exemple vous pouvez nous nous décrire / **comment vous vous y prenez pour faire une omelette** vous devez en faire souvent [rire] #  
 CW 739: / +[rire] / / comment **je** fais une omelette [rire] +[pron=pi] **je** prends une poêle **je** mets du beurre dans ma poêle et [b] **je** casse des oeufs **je** les bats dans séparément naturellement pas dans la poêle dans un saladier à côté p(u)is quand mon beurre est fondu bah [rire] **j(e)** fais mon omelette **j(e)** la fais cuire

## 6.2 Sociological variations

In the last part of this paper I will carry out an observation of the variations at the sociological level: it could be possible to examine different social parameters for each interviewee as we know their age, sex, socio-cultural level (they have been classified according to their socio-professional categories in the catalogue of the recordings, Lonergan *et al.* (1974)) and geographical origin. I will limit myself here to their gender and will check if we can predict the gender of an interviewee from the type of recipe he /she used in his/her answers.

I will thus examine the typology of answers that I have established in order to see how they divide up depending on the sex of the interviewees.

Is there a correlation between the type of recipe described and the gender of the speaker? For example, would a woman be more likely to use an ‘occurrence’-recipe focusing on her personal experience of cooking (the way in which she prepares an omelette) and would a man more often use a ‘type’-recipe describing what is generally required to make an omelette?

This impression is confirmed by the following examples in which the male interviewees provide very short answers. In example 10, the man is not interested in cooking as he is married to a good cook and therefore does not interfere in the kitchen department. In example 11, the answer is very short as he does not know a lot about cooking.

- (10) Extract 122 : **RP 283**: vous savez que je m'intéresse() **j'ai une femme cuisinière** et que **je me suis bien gardé de d'y mett(r)e mon nez je lui laisse le soin de le faire** si j'ai à la faire je sais faire une omelette et je pense que on doit opérer de même j'ai mes enfants qui vont commencer à faire l'omelette alors euh **je ne dois pas me mêler de ces choses-là**
- (11) Extract 150 : **YT 387**: oh là [rire] vous savez en cassant les oeufs ça c'est sûr  
**MD**: [rire]  
**YT 387**: **mais autrement non la cuisine c(e n')est pas mon rayon hein**

However, the same impression is invalidated in examples 12 and 13 in which the interviewees, two men explain at length what different ingredients can be used. Whereas a woman in example 14 is not very loquacious as she does not like omelettes very much.

- (12) Extract 008 : **GJ 131**: # [rire] # ah mais c'est pas compliqué hein pour moi c'est pas compliqué on casse les oeufs on on bat tout ensemble euh on met un peu d'eau j(e) crois on mélange un peu d'eau enfin on assaisonne sel poivre euh en Lorraine on **on découpe des petits des petits morceaux de lard qu'on fait frire avant** [b] et puis on enfin on verse tout ça dans la dans la poêle et puis on tourne jusqu'à temps que ça soit à peu près cuit quoi [rire]

RC: hm hm [pron=pi] on sent déjà l'odeur de votre om() omelette  
 GJ 131: # oui [rire] **seulement ici on ne l'a fait pas au lard parce qu'on ne trouve pas de de charcuterie comme en comme en Lorraine #**  
 RC: # [rire] # non  
 GJ 131: # **on trouve non euh la charcuterie ici c'est pas très fort** et bah on n'a qu'a +[pron=pi] **du lard fumé des des saucisses fumées** vous savez c'est pas du tout fait pareil ici enfin bon moi **je on met des des épices aussi dedans quand même quelquefois des de l'ail de l'oignon** [pron=pif] #

(13) Extract 149 :YR 399: euh une comment qu'on fait une omelette **alors là i(l) y a plusieurs façons de faire les omelettes** euh moi je vais vous parler des om(elettes) des omelettes que je que je connais vous avez l'omelette aux oeufs carrément c'est-à-dire que vous mettez euh du beurre ou de l'huile dans votre poêle quoi vous cassez les oeufs dans une assiette vous les battez bien vous les épicez sel poivre à la rigueur vous pouvez mettre un p(e)tit de d'appétit ou de persil si on veut p(u)is euh vous vous faites cuire ça dans dans la poêle ou alors euh vous avez euh l'omelette aux pommes de terre vous f() vous pelez des pommes de terre vous les faites assez cuire dans la poêle et p(u)is vous mettez euh les oeufs que vous avez battus toujours dans une assiette ou dans un bol ça ça aucune importance que vous mettez avec les pommes de terre ou alors vous avez l'omelette au croûton c'est-à-dire euh les croûtons c'est euh [rire] c'est des morceaux de pain vous voyez des morceaux de pain qui qui restent du pain rassis c'est-à-dire du pain qui n'est pas du de la journée si vous voulez alors c'est p(e)tit morceau de pain vous faites revenir un p(e)tit peu dans dans dans du beurre euh alors euh évidemment i(ls) se ramolissent et p(u)is après i(ls) durcissent et p(u)is vous mettez vos oeufs là-dedans ou alors vous avez l'omelette au jambon c'est-à-dire avec des p(e)tits morceaux de jambon ou alors l'omelette au lard avec des p(e)tits morceaux de lard (il) y a (il) y a pas mal de ch() d'omelette vous avez l'omelette aux pommes **vous avez euh une quantité presque industrielle d'omelette de façon de faire les omelettes**

(14) Extract 101 : MS 530: # une omelette oh bah c'est #  
 BV: # [rire] #  
 MS 530: ça euh +[pron=\*] je bats des oeufs je mets du sel [rire]  
 BV: oui  
 MS 530: p(u)is c'est tout euh une omelette moi ça **j(e) (n')aime pas tellement faire une omelette** [rire]

I have re-used the typology of responses established in the previous part in order to describe how the recipes listed are divided up by genre.

Recipes in responses	Gender			
	Female		Male	
	Raw number	%	Raw number	%
Type	22	51.2	21	48.8
Occurrence	30	81.1	7	19.9
Variation from type to occurrence	7	58.3	5	41.7
No recipe given	0	0	2	100
Total	59	62.8	35	37.2

Table 6: Recipe categories and gender of the interviewee

Table 6 shows that the 'occurrence'-recipes are mostly uttered by women (in 81.1 percent of the cases): out of a total number of 37 'occurrence'-recipes 30 have been uttered by women.

However, one needs to be careful when reading these results, as it should be taken into account that the question « How do you make an omelette ? » or one of its variants was asked in majority to women. In the "Eslomelette" sub-corpus, we count 62.8 percent of women versus only 37.2 percent of men. Table 7 below shows how the responses of female interviewees on the one hand and male interviewees on the

other hand really divide up between ‘type’ and ‘occurrence’-recipes, variation in the recipe and no recipe given at all:

Recipes in responses	Gender			
	Female		Male	
	Raw number	%	Raw number	%
Type	22	37.3	21	<b>60</b>
Occurrence	30	<b>50.8</b>	7	20
Variation from type to occurrence	7	11.9	5	<b>14.3</b>
No recipe given	0	0	2	<b>5.7</b>
<b>Total</b>	59	100	35	100

**Table 7:** Gender of the interviewee and type of response given

The traditional view that women ‘intuitively’ know how to cook and that men only know ‘about’ cooking seems to be holding as we can see in the table above that women tend to answer with an ‘occurrence’-recipe when they explain how to make an omelette (50.8 percent of their answers, contrary to men who use such a recipe in only 20 percent of their responses), whereas a majority of the men interviewed (60 percent) describe how to make an omelette in a ‘type’-recipe.

As regards the extracts where there is some variation within the explanation of the recipe, more women uttered them (7 vs 5), but in proportion of the total number of answers for each gender, men tended to vary more in their description of the recipe. Finally, the only two instances in which we do not have any recipe described are uttered by two men.

The next thing that would be interesting to do – as only 94 interviewees in the sub-corpus “Eslomelette” out of 157 in ESLO1 have been asked the omelette question – would be to check if in ESLO 1 there is the same number of men and women and if yes, to wonder why the cooking question was more likely asked to women (59 on 94). Is this again because of the prejudice mentioned at the beginning of this part, i.e. that women do the day-to-day cooking whereas men do not cook at all or they are well-known chefs?

## 7. Conclusion

I hope that this descriptive and exploratory study of a small corpus of spoken French has shown that however modest in size a corpus may be, the linguist can always make interesting discoveries.

This paper reported the pragmatic variations that could be found in a question that was embarrassing mainly for some of the French interviewees, who had to obtain from the interviewees a description of a recipe, which most of them found surprising.

I also described in this paper the morpho-syntactic variations both in the pronominal form used in the questions and as a result in the responses to the question.

Then I showed that the hypothesis that had arisen from the data was verified and that two types of recipes emerged : a question with the form *on – chez vous* led to a response using the pronoun *on*, this response containing most of the times a ‘type’-recipe, whereas a question with the pronoun *vous* tended to receive an answer with *je*, telling a personal experience in an ‘occurrence’-recipe.

Finally I noticed that the type of recipe used in the interviewee's answer could help predicting his/her gender, and that it was worth noting that the 'occurrence'-recipes were mainly found in women's responses.

## References

- Abouda, L. and O. Baude (2006) Constituer et exploiter un grand corpus oral: Choix et enjeux théoriques. Le cas des ESLO, in F. Rastier and M. Bellabriga (eds) *Corpus en Lettres et Sciences sociales : des documents numériques à l'interprétation*, Actes du colloque international d'Albi, juillet 2006, pp. 143–50. Published by C. Duteil and B. Foulquié. Paris : Texto. Supplément de *Texto!* septembre-décembre 2006 [en ligne], vol. XI, n°3-4. Also available on-line from [http://www.revue-texto.net/Parutions/Livres-E/Albi-2006/Actes\\_ALBI-06.pdf](http://www.revue-texto.net/Parutions/Livres-E/Albi-2006/Actes_ALBI-06.pdf) (accessed: 23 May 2007).
- Abouda, L. and M.-E. Perrot (2006) Une question "embarrassante" en situation d'interview. Modalisation et stratégies de légitimation. Paper presented at the 3<sup>e</sup> Rencontre Fribourgeoise de la Linguistique sur Corpus Appliquée aux Langues Romanes – corpus et pragmatique, Freiburg im Breisgau, Germany, 14-17 September 2006.
- Baude, O. (ed.) (2006) *Corpus oraux. Guide des bonnes pratiques 2006*. Paris: CNRS-Editions and Orléans: Presses Universitaires d'Orléans.
- Bergounioux, G. (2007) From a reference corpus to a prototypical corpus. Paper presented at the Corpus Linguistics 2007 Conference, Birmingham, UK, 27-30 July 2007, in *Proceedings of the Corpus Linguistics 2007 Conference*.
- Blanc, M. and P. Biggs (1971) 'L'enquête sociolinguistique sur le français parlé à Orléans'. *Le Français dans le Monde*, 85, 16–25.
- Chesneau, A. (2007) Going back over the survey – "Language: the looking glass of society". Paper presented at the Corpus Linguistics 2007 Conference, Birmingham, UK, 27-30 July 2007, in *Proceedings of the Corpus Linguistics 2007 Conference*.
- Habert, B., A. Nazarenko and A. Salem (1997) *Les linguistiques de corpus*. Paris: Armand Colin.
- Hunston, S. and G. Francis (2000) *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
- Leech, G. (1992) Corpora and Theories of Linguistic Performance, in J. Svartvik (ed.) *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82*, Stockholm, 4-8 August 1991, pp. 105–22. Berlin: Mouton de Gruyter.
- Lonergan, J., J. Kay and J. Ross (1974) *Etude sociolinguistique sur Orléans, Catalogue des enregistrements*. Colchester : Orléans Archive, University of Essex, Department of Language and Linguistics.
- Ross, J. (1974) Enquête sociolinguistique et description syntaxique (L'étude sociolinguistique sur Orléans, 1967-1973), in A. Verdoodt (ed.) *Applied Sociolinguistics. Proceedings of the Third Congress of AILA in Copenhagen 1972*, Vol. II, pp. 137–54. Heidelberg: Groos.
- Ross, J. (1979) Sampling, elicitation and interpretation: Orléans and elsewhere, in W.C. McCormack and S. Wurm (eds) *Language and Society: anthropological issues*, pp. 231–47. The Hague: Mouton.
- Serpellet, N., G. Bergounioux, A. Chesneau, and R. Walter (2007) A large reference corpus for spoken French: ESLO 1 and 2 and its variations. *Colloquium*

- presented at the Corpus Linguistics 2007 Conference, Birmingham, UK, 27-30 July 2007.
- Sinclair, J. M. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Tognini-Bonelli, E. (2001a) *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Tognini-Bonelli, E. (2001b) 'Towards a Corpus-Driven Approach'. Paper presented at ICAME 2001, Louvain-la-Neuve, Belgium, 16-20 May 2001.
- Walter, R. (2007) The computer construction and the interoperability of the ESLO corpus. Paper presented at the Corpus Linguistics 2007 Conference, Birmingham, UK, 27-30 July 2007, in *Proceedings of the Corpus Linguistics 2007 Conference*.



# The Computer Construction and the Interoperability of the ESLO Corpus

---

Richard Walter<sup>1</sup>

This paper will present the choices which were made within the structure of the ESLO oral corpus project. The first objective of this project is to digitize the sound documents from the original tape recordings and to offer an indexing and a first transcription of these documents in order to put the data in a database where they could be stored but also consulted.

The presentation of this corpus of audio linguistic resources is based on an analysis of the methodologies which the project has to use in order to be able to describe these resources and to exploit them.

Our aim is to illustrate the interest there is to think about the interoperability of the codings and the metadata for the use of an oral corpus. What we then want to do is to simplify the technical creation of audio corpora and thus organise the constitution of linguistic resources, in order to make them usable by enlarged academic and industrial communities.

---

<sup>1</sup> CNRS – MoDyCo, Variling Project, CORAL, Universities of Paris 10 and Orléans  
*e-mail*: richard.walter@u-paris10.fr