



HAL
open science

Learners of English and Conversational Proficiency

Pascale Gouteraux

► **To cite this version:**

Pascale Gouteraux. Learners of English and Conversational Proficiency. Granger, S.; Gilquin, G.; Meunier, F. Twenty Years of Learner Corpus Research: Looking back, moving ahead, Corpora and Language in Use Proceedings of the First Learner Corpus Research Conference (LCR 2011), Presses Universitaires de Louvain, pp.197-210, 2013. hal-01233885

HAL Id: hal-01233885

<https://u-paris.hal.science/hal-01233885v1>

Submitted on 23 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Learners of English and conversational proficiency

Draft version published in Granger, S., Gilquin, G. & Meunier, F., eds. (2013), *Twenty Years of Learner Corpus Research, Looking Back, Moving Ahead*, Corpora & Language in Use series, UCL, Presses Universitaires de Louvain: 197-210.

Pascale Goutéraux,

Sorbonne Paris-Cité Paris-Diderot University, EA 3967 CLILLAC-ARP

Abstract

This study focuses on the inter-relatedness of fluency and complexity as explanatory factors and criteria for the assessment of conversational proficiency within the framework of two current cognitive models. It has been carried out on a cross-sectional corpus of 28 one-to-one conversations between native teaching assistants and French EFL university students from the Diderot-Longdale project.

Key words: learner corpus, psycholinguistic parameters, conversational proficiency

Introduction

Can the conversational proficiency of foreign learners of English be efficiently assessed by using psycholinguistic criteria? The “fifth communicative competence” referred to as “interactive spoken competence” in the Common European Framework of Reference for Languages (or CEFRL) has been part of the French curriculum for the learning of foreign languages in secondary schools since 2004 but it is difficult to find appropriate criteria to assess the conversational skills of a learner. The qualitative aspects of spoken language in the CEFRL (2000: 28) are categorised as ‘range, accuracy, fluency, interaction and coherence’. According to the B2 to C1 descriptors, the learners must be able to “give descriptions”, “express viewpoints” “express (themselves) on a wide range of general topics”. The ‘interaction competence’ refers to the learner’s ability to master turn-taking and functions like initiating and ending a conversation, “get or keep the floor, confirm comprehension”, in keeping with conversational pragmatic research. The key words to describe fluency are “smooth, spontaneous flow”, “clear and coherent discourse”, the main difference being that the B2 level, expected from French secondary school graduates, allows for hesitations while the speaker is searching for patterns and expressions. Yet, such recurrent qualifiers as “sufficient” “controlled”, “suitable” or “appropriate” seem too vague and subjective for assessment. As regards the task-based approach to foreign languages,

the French official directives allow for multiple interpretations of the communicative competence, ranging from the ability to utter a few words to get one's meaning across to sophisticated or/and colloquial effortless mastery of the language (*Nouveaux Programmes Cycle Terminal*, 2010¹). There again we lack specific criteria to assess the oral proficiency of learners of English. Clearly, action-based research on the assessment of oral interactive competence would benefit from a combination of more focused criteria derived from the fields of psycholinguistics, pragmatics and linguistics.

1. Multi componential interactional competence

1.1 Psycholinguistic components

Since the 1990s, studies have been carried out on the relationships between fluency, accuracy and complexity or FAC². These parameters have proved useful for SLA research within the framework of two main models of cognitive development. Skehan's model of limited cognitive capacity for L2 learners (1998) supports the idea of trade-off patterns of cognitive load as a learner's production becomes either more fluent or more accurate or more complex. Robinson's Multiple Resources hypothesis (2001, 2002) claims that learners benefit from cumulative proficiency since progress in one direction, say fluency, scaffolds progress in complexity and accuracy as well. Both models also take into account the impact of external factors (task type, topic familiarity and timed or untimed performance) on increased automaticity. While looking for general trends, researchers have also evidenced that learning remains an individualised process and that variation is typical of combinations of skills displayed by learners. For instance, Robinson *et al* (2002) have explored how learners' cognitive and linguistic aptitudes interact with task conditions in instructed environments to produce various patterns.

To what extent can foreign learners produce fluent, complex and accurate conversational discourse? Do individual differences stem from ingrained aptitudes or do they reflect different steps in the internalization of the developmental process? What is the impact of the task format in instructed environments? These are questions SLA research has been tackling over the past ten years. Learner corpora have proved useful in that regard: Hasselgren (2002), on the 'small words' of oral discourse, De Cock (2007, 2008) and Gilquin (2008) on the *LINDSEI* corpus, Hilton (2008, 2011) on temporal measures of fluency, Goutéaux (2011) on the influence of communicative tasks on secondary school students' oral proficiency, to quote but a few.

1.2 A study of fluency and complexity in advanced learner speech

This paper reports on a preliminary study of the markers of fluency and complexity in advanced learner speech and aims at identifying cumulative or trade-off patterns between the two parameters; it is the first step of a more in-depth analysis of the relationships between all three FAC parameters, from a developmental viewpoint. A quantitative and qualitative approach has been adopted in keeping with Ellis & Barkhuisen's claim (2005) that cumulative measures are best to analyse learner

¹ <http://www.education.gouv.fr/pid285/le-bulletin-officiel.html>

² See special issue of *Applied Psycholinguistics*, April 30, 2009 for an overall review of current research.

speech. The data are a selection of conversations compiled in the Diderot-Longdale database, a part of the international Longdale project (UC Louvain, Belgium). Supported by Paris-Diderot's *Plan Réussite en Licence*, this project also aims at assessing the evolution of a sample of students over 3 years and the researchers plan to use the results to remodel the Bachelor's linguistic curriculum.

Twenty-eight one-to-one conversations between native teaching assistants and non-native volunteer college students were recorded in June and October 2010 (end of the first year or beginning of the second).³ All the students were English majors except for one History major and a German national with English as her second language. The set part of the conversation focused on past experience, dreams, wishes and creativity. The participants had a choice of three topics which favoured individual commitment (Robinson 2002; Dörnyei 2006). The task was unplanned. The 'free' part dealt with their projects and a self-assessment of their first-year academic experience.

- | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> • Topic 1: Suppose you have/had time and money to travel or move to a different country/city, where will/would you go? Why? How will/would you organize your new life? (16) • Topic 2: Can you tell me about an important event, experience or meeting which has made a difference or changed your life in the past six months? (4) • Topic 3: Do you feel creative? Tell me about a work of art you would like to create or participate in: a play, a film, a musical event, a book, a painting, a computer game, <i>etc.</i> How would you go about it? (8) |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Table 1. Set Task: Topic choice by number of students

2. Fluency in a cross-sectional sub-corpus of the Diderot-Longdale database

2.1 Working definitions of fluency

A broad definition of fluency ranges from temporal to lexical measures. Temporal fluency is measured in terms of speech rate (the number of syllables or words per 60s) and number of pauses (filled or unfilled ones or both) and pause location in utterances (at the beginning or the end of propositions or in mid-position). MLU-mean length of units-, or MLT-mean length of turns- in conversational contexts, are used as well for learner corpora research (Hilton 2008, 2011; Gilquin 2008; Iwashita *et al* 2008). For instance fluency and accuracy in the oral performance of dysfluent learners have frequently been linked with hesitations (false starts, repeats, retracing, reformulating, etc.) or pausing in mid-utterance as the speaker is grasping for a lexical item or attempts to solve a form and meaning discrepancy. The same may also feature in native speech, but with fewer hesitations or function word repeats and mostly while the speaker is busy planning the next stretch of discourse. Yet the difference between native and advanced learner's speech is not only a problem of number, frequency and distribution of dysfluent features.

Research also associates fluency in L1 with the frequent and appropriate use of lexical prefabs: formulaic expressions (*cf.* Wray 2002) are stored as lexical chunks in the

³This series was the second wave of conversations; the first one had taken place at the beginning of the first year (2009).

long-term memory: the faster the retrieval process, the higher the speech rate. Oral discourse markers (*Longman Grammar of Spoken & Written English* 1999) such as *well, you know, like, I mean, etc.*, ensuring both fluency and discourse coherence, have been investigated in L2 speech as well (Hasselgren 2002; De Cock 2004, 2007; Gilquin 2008). From a cognitive and interactional viewpoint, *kind of, sort of, like* are said to make on-line controlling of conversation easier for the interlocutors:

Information is less tightly packed in conversation (than in prose or academic discourse), which simplifies both the planning of the speaker and the decoding of the hearer (LGSWE: 117).

From a pragmatic standpoint, vagueness and repetition are characteristic features of informal spoken discourse, with strings of prefabs and cumulative markers like *Yeah, I mean, well, ...okay, so, you know* as conversational starting blocks (LGSWE 1999). The study of politeness in face to face interactions (Goffman 1959) also shows that hedges and vagueness markers (*or something, and so on and stuff like that*), attenuation markers or intensifiers (*a bit of, just, really*) are used by natives to avoid over-assertiveness. Speakers insert *I mean, I don't know, dunno, you know* to introduce expansions, explanations or qualifications and clarify their meaning:

Phraseologies fulfill pragmatic functions, for instance *know, don't know, you know*, contribute interactionally, indicating hesitation and uncertainty, appealing to shared knowledge or understanding, or pre-empting contradiction (Moon 2010).

2.2 Measuring temporal fluency in learners' productions

To assess the levels of fluency reached by non-native college students after their first-year of English studies, two temporal measures of fluency were used, *i.e.* the rate at which words are produced and the number and frequency of filled pauses. Hilton quotes Levelt⁴ (1989) and others who give baseline figures for native speakers' speech rate ranging from 130 to 200 words per minute, with 1/3 of the time spent pausing. In *PAROLE* corpus of LEA (Applied Foreign Language) students, she found that advanced learners produced quite the same number of words per minute as native speakers, from the lower 130 wpm baseline, while dysfluent learners averaged about 50wpm only.

2.2.1 Speech rate

To test this hypothesis again, the beginnings of the recordings were selected: 28 stretches of continuous production ranging from 87 to 290 seconds, ending the interviewer asked a question; and then the number of words per minute was computed (excluding filled pauses but not silent ones). With a mean of 127.81 words per minute and a median of 129.2, none of the Diderot-Longdale students displayed a speech rate equal or inferior to 50 wpm (Hilton's dysfluent learners).

Group 1: 14 less fluent learners	Group 2: 14 more fluent learners
89.68 wpm-129 wpm	132 wpm-175 wpm

Table 2. Fluency measures: speech rate in wpm

⁴ Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: The MIT Press.

The German student stood out with a speech rate of 175 wpm but she usually spoke fast, even in French. At the lower end of the continuum, a Polish-French bilingual student averaged 89.68 wpm. 13 French L1 learners averaged 132 wpm and above.

The relationship between fluency and the cumulated length of stay in an English-speaking country was also investigated since the discussion of fluency often overlaps with the issue of instructed *versus* natural acquisition; the participants' cumulated length of stay in English-speaking countries ⁵ (ESC) was used to study the impact of immersion or direct contact with native speech.

Stays in ESC	Student%	Gr. 1 (89.68-130 wpm)	Gr. 2 (132 wpm-180 wpm)
none	18%	3	2
Less than 1 month	36%	7	3
1-3 months	25%	3	4
4-6 months	14%	1	2
7-12 months	7%	0	3

Table 3. Students' cumulated stays in ESC and speech rate expressed in wpm

54% of the 28 students had either never been to an ESC or had spent less than a month abroad: 10 students in the less fluent Gr.1. The only discrepancy concerned one student who claimed a total of 4-6 months abroad but achieved only a 113 wpm rate and produced a lot of pauses (29) and repetitions.

In the more fluent Gr. 2, 9 had stayed abroad from 1 to 12 months, and the higher speech rate corroborated the impact of immersion on fluency. 4 of the 5 students who had never been abroad or less than one month claimed to watch a lot of TV series and broadcasts, which may have partly made up for the lack of immersion in a natural environment. In most cases, experience abroad and direct contact with native speech had a positive impact on fluency expressed in terms of speech rate.

2.2.2 Number and distribution of filled pauses

The number, frequency and distribution of filled and unfilled pauses are other temporal measures of fluency; in native speech, most pauses naturally occur at sentence or clause boundaries⁶. Hilton (2011) claims that pauses located within clauses, in non-native speech, precede lexical search or lexical errors. She also explains the interaction between fluency and accuracy from a lexical viewpoint. Gilquin (2008) indicates that in non-native speech, hesitations can be due to problems of conceptualization and formulation. While formulation issues can be linked to accuracy (or the lack of it), conceptualization refers to semantic and conceptual complexity, which may be problematic for both natives and non-natives.

An automatic search of the various locations of filled pauses in the Diderot-Longdale corpus did not provide meaningful insights in the misplacement of pauses. But the percentage of filled pauses compared to the total number of words per individual production yielded significant results at both ends of the scale. 7 students within the 130-180 wpm G2 (more fluent speakers) displayed a low 3.48 to 4.38% of filled pauses and 6 Gr.1 students with a lower speech rate (89 to 129 wpm) were in the higher 12.62-18.02% slot. So the figures for 13 students were coherent for both

⁵ Yearly online questionnaires filled out by the students.

⁶ Yet note that 28% of the pauses were located within clauses in Hilton's *Parole* native part of the corpus.

parameters since a high percentage of filled pauses and a low wpm rate are clear indicators of dysfluency while a low percentage of pauses and a high wpm rate are strong indicators of fluency. However 15 students produced mixed results requiring a more qualitative analysis.

As an example, here is an extract of n°38's continuous stretch of speech (468 words, 2mn49s): with a 140.59 wpm speech rate, he belongs to Group 2 and yet if we consider the pause percentage only (12.56%), he would come close to the dysfluent category. In the extract below, the percentage of pauses is even higher (>15%), because of the extra cognitive load he experienced when attempting to retrieve biographical elements pertaining to the creation of his band (Skehan 1998). He paused frequently when trying to remember proper names or types of music or instruments:

I choose to do to (er) speak about music (er) I'm actually doing music since I'm thirteen years old (erm) I started with (er) bass guitar (erm) and then I (er) I (er) sang a lot (er) and I then learned how to play guitar and (er) synthesizers . (er) . so (er) when I was (er) sixteen years old I met (erm) a guy (er) who played piano (er) and who played very very well the pian (?) and piano he was playing since he was like three years old and (erm) and so we started to (er) to play together (erm) and to (er) we decided we decided to (er) to create a band (er) that we called (er) he had a few names . (er) we had this (er) weird name because we were young and so it was called Mad Decisions[...]

Actually numerous pauses, with *so*, *well* and *but*, were located at junction points or when he shifted to a different idea. More typical of learner speech was his insertion of pauses after a ready-made chunk of *verb+ to*, as part of an infinitive clause (*I was asked to*, *we decided to*, *I chose to do to*, *we started to*). Fluency indicators may map onto semantico-syntactic complexity with pausing after a chunk when a syntactic decision has to be made. Yet fluency and accuracy do not overlap systematically as fluent students often make grammatical errors without breaking stride.

I mostly enjoyed the first semester because I I *work hard not really really hard but I I *work every day I I read I was *concentrate *on class (n°14).

3. Complexity in learner speech

Complexity measures apply to more various linguistic sub-fields today than in the 1960s-70s⁷ (Ellis and Barkhuisen 2005):

- lexical complexity encompasses frequency and density (Type-Token Ratio), richness, range, conceptual and semantic complexity,
- interactional complexity is often measured in number and length of turns (among others),
- syntactic measures of oral discourse include propositions, clauses or A-S Units (Foster *et al* 2000).

Yet complexity is said to feature less prominently in conversations than in written text:

Conversation has a strikingly low lexical density *plus* a low degree of grammatical elaboration (LGSWE: 1041)

⁷ T-units (main and subordinate clauses) used to be the only measures of complexity to be relied on.

Indeed this may vary with the degree of formality (interviews vs informal conversations). Learner speech may also skew conversational turns towards more academically stilted forms than ordinary native conversations would.

3.1 Lexical fluency and complexity

De Cock *et al* (1998) tested Kjellemer's idea "that native speakers rely heavily on multiword items (MWIs) or prefabs like *by the way* and *if you see what I mean* that allow for coasting and save cognitive resources for important choice points". One explanation of dysfluency offered by Hilton and others is the poverty of the foreign learner's lexicon and the lack of formulaic phrases. This may be quite true of beginners but the advanced learners of my corpus did not exhibit such a correlation between dysfluency and lexical poverty. Actually they used substitution strategies rather expertly.

As regards lexical proficiency, the frontier between complexity and fluency is not so opaque: it depends on whether connectors (prefabs or not) are considered as fluency indicators associated with hesitations, gap fillers and semantically empty tokens, or as rich lexical markers which fill pragmatic and syntactic functions. De Cock *et al* (1998), De Cock (2004, 2007) and Cobb (2003) have shown that advanced learners can use as many precasts as native speakers but that their speech is characterized by a more limited variety. Indeed non-natives tend to overuse some strings of words and underuse others. Although the Diderot-Longdale corpus exhibits patterns of overuse and underuse of speech markers, these "preferred sequences" do not always tally with the results of De Cock's 2007 study of advanced Belgian learners of English: the patterns are the same but not the strings of words. For instance the study looked into the use of *sort of* (more British), *kind of* (more American) and *like*, which is fairly common in native interactional discourse (Aijmer 2002) and reportedly less frequent in non-native speech. Occurrences of *sort of* (7 tokens/3 students) were as limited as for the Belgian students but *kind of* and *like* appeared quite frequently: *Kind of* (44 tokens) took on a variety of grammatical and pragmatic roles typical of conversational discourse (for instance as an understatement device):

- before nouns (kind of job, work, landscape, music)
- with pro-forms (and that kind of thing, of place)
- as an adjective modifier (kind of weird, mysterious, difficult, sad)
- to mitigate feelings and commitment (I was kind of disappointed, n°41/ I kind of miss speaking Spanish, n°79)

Like, which has become a jack-of-all trades word in spoken American English or in youth speak, featured 327 times. In the total frequency list, *like* was the first lexical word (14th rank) after such function words as pronouns, deictic forms, articles, auxiliary *be* and basic prepositions. 30% of the students used *like* as a preposition associated with pro-forms *things like that*, *stuff like that*, *something like that* (24 tokens). As a stance marker, *like* appeared 84 times (16 students, 40%), which testifies to the internalization of today's most frequent sociolinguistic marker in oral discourse (LGSWE). The structure *pronoun + be + like* frequently introduced direct speech⁸:

We were like oh my God, this is so great! (n°108)

⁸ Natives frequently use *like* (*I was like/he was like*) to report conversations, anecdotes and memories as well as *he went*, *he goes*, but the latter never appear in our corpus.

We were like oh there's no bag left (n°64)

Students also used *like*, stream of consciousness-like, to express thoughts and feelings:

When I came back, I was like oh I loved it but (n°64)

When I came to Charles 5, I was like yes I think I speak good English (n°106)

Like (n°20: 25, n°34: 14, n°64: 6, n°38: 8, *etc.*) was quite overused by those students who had spent time in an ESC. Inserted clauses *I mean*, *dunno*, *you know* also featured: *I mean* was used by 7 students (25%) with a peak for n°27 (15 tokens) and (*I dunno* was even more frequent (61 tokens):

He could but (er) he (er) I dunno he didn't want to so (n°34)

The teacher doesn't (erm) (em) I dunno thrill you (n°68)

[...] poems I'm thinking about (er) I dunno (eh) Tennyson's Lady of Shalot (n°106)

40% students produced a total of 79 *you know* (with overuse for n°24, n°98), which is quite coherent with the conversation format since *you know* often functions as an interactional appeal to the experiences and beliefs of one's interlocutor:

[The *you* approach is a means] to generalize the speaker's experience to build empathy with other participants or fulfill the ritualized intersubjective function of codas in conversational narratives (Scheibman 2007: 132-133).

There again, these results do not quite tally with De Cock (2007 : 223) who points out the underuse of *I mean* and *you know* in the *LINDSEI* corpus. Similarly, research has shown that *in fact* is typically overused by French learners as a result of transfer from French '*en fait*'. The 24 tokens of *in fact* in the corpus were all produced by 4 students with little or no experience of an ESC. Conversely, 12 students displayed proper use of *actually*, either as a stance marker or with an adversative value, with a peak for 3 students with long experience in ESCs (n°20, 9, n°38, 7, n°79, 8).

Thus overuse of linguistic sequences typical of spoken language does occur, as a possible side effect of immersion, or as part of a natural stage of development (the overgeneralization phase of U-Shaped interlanguage models) and as the result of individual choices and preferences. Yet preferred sequences can vary from corpus to corpus, depending on such external factors as the type of English students have been exposed to. The effects of globalization on the evolution and standardization of spoken preferences among young people should also be taken into account.

3.2 Lexical richness

Another question is whether trade-off effects take place with frequent oral speech markers being selected at the expense of lexical richness. To test this hypothesis, both frequency word lists and *Wordsmith* standardized TTR were used on a corpus pruned of filled pauses, repeats and false starts. The standardized TTR ranged from 0.51 to 0.66 (higher intermediate to advanced learners), with a mean of 58.2, a median of 58.5, and 10 students over 0.6. There was no significant relationship between the length of stay in an English-speaking country and lexical richness. This required looking for other causes: the interviewers' personal style and way of questioning may have influenced the degree of lexical complexity of the students' productions. On the other hand, the choice of a familiar or motivating topic made it easier to draw upon lexical and conceptual resources that were already anchored in their long-term memory and

limited lexical risk taking. Lexical input in instructed environment (a prominent part of the linguistic curriculum in L1) may also have played a part.

The dispersion of results makes it impossible to establish a strong connection between fluency (in terms of speech rate) and lexical complexity (in terms of richness and sophistication). Actually, these advanced students entered varied fluency/lexical complexity combinations: very fluent learners like n°20 (TTR 0.53) and n°79 (TTR 0.58) who had spent a lot of time in ESC (7 to 12 months) did not display the highest rates of lexical richness but repeatedly used a lot of oral discourse markers. Another student (n°66) ranked high in terms of TTR (0.61) but low in wpm rate (99.26); she did not produce so many filled pauses (7.52%) but her speech quality (fluency and accuracy) decreased with time as haphazard pausing and phonological and grammatical errors increased. On the other hand, the student with the highest TTR (n°106) had only spent 1-3 months abroad but ranked in the second slot for both fluency measures (153.8 wpm) and limited pausing (7.3%). Such heterogeneous configurations support the hypothesis of multi-competence portraits of L2 learners (Cook 2002; Robinson 2002; Dörnyei 2006) although most proficient students might ultimately score high on all measures of lexical complexity and fluency.

3.3 Syntactic and grammatical complexity in the Diderot-Longdale Corpus

Considering the high level of automaticity of native speech and the fact that native speakers seldom take time to retrace, recap and compute syntactic choices, we may expect learner speech to display an even lower level of complexity. However, complex syntactic structures are as frequent as in native speech:

- 50% produced *would like* + infinitive (25% *I'd like*), 80% used *want* + infinitive (*wanna*, 6)
- numerous causatives (*because*, 306, variants *cos* or *coz*, 6)
- temporal subordinates (*when*, 111): +preterit to report on memories/+present to project into the future
- multiple *if*-clauses (23conditional clauses, 11concessive *even if*, 27indirect speech).

Non-restrictive relatives (30% out of 40 relative clauses) functioned well in topic + comment order to provide additional information:

We had the result, which was that it was a cancer (n°30)

I could finish with New York, which I already visited (n°45)

Linguistics and phonetics, which are totally new for me (n°79)

Oh yes, history, which is also one of my favorite subjects (n°108)

Yet utterances could be inaccurate when the referent and the relative were separated:

a book by Barjavel *who was er which was er which dealt with er (n°45)

I was with this teacher, Mr. M, *which is brilliant (n°38)

Pre-predicate and extra-posed clauses were also less frequent than in native speech:

One reason why I would like to do that is because (n°79)

That's what I want to do later (n°106)

That's why I wanted to go sort of to discover (n°48)

The students found it difficult to manage elliptic and anaphoric structures, which presuppose sharing implicit representations with their interlocutors: the lack of common culture and representations and the absence of social bonding certainly prevented them from selecting such structures. The implicit reference to linguistic agents or objects located in the co-text or context made it tricky to retrieve and activate those conversational English structures which may not have been stored as such. So both cognitive and syntactic factors were at play to curb the production of these complex forms. Very few anaphoric forms appeared in the corpus: *I think so/don't think so/ hope so* (11tokens) */I, they want to* (2). Most elliptic answers were erroneous:

*no yeah I wanted (n°54) *I don't want *it (to stay in the camper van) (n°64).

Conclusion

The study confirms the need for cumulative temporal measures of fluency and for a qualitative re-interpretation of the results that might seem incoherent from a quantitative viewpoint. Speech rate is strongly influenced by automaticity, which is highest among those who spent time in ESC. Frequent pausing is clearly linked with dysfluency but the increasing pressure of task complexity and memorizing difficulties are to be taken into account for both advanced learners and native speakers. Although temporal fluency and accuracy are connected, it is not clear whether grammatical and syntactic accuracy are concerned to the same extent as lexical accuracy: pausing may signal that the speaker is actively searching for the appropriate structure or grammatical form instead of blithely plodding on, totally unaware of producing erroneous speech. This analysis also confirms that advanced learners make pragmatic motivated linguistic choices in conversation and have recourse to appropriate speech markers, with possible overuse of idiomatic forms. The lexico-semantic specificity of patterns may be partly dependent on individual conditions of acquisition and exposure to a variety of Englishes as well as a byproduct of transfer from the L1, hence a possible variation from corpus to corpus.

The links between fluency and lexical richness are not so clear-cut: the corpus shows patterns of coherence for the most proficient learners, possible discrepancy for some other learners and low scores on lexical complexity measures for the less fluent ones, which yields a rather composite picture (or pictures) of learners' proficiency. Weak points in speech complexity (anaphoric forms, ellipses, extra-positions, pre-predicates) have also been noticed: they are not a priority for secondary school EFL teaching, where more academic forms are emphasized. The lack of authenticity of communication in instructed environments also prevents the activation of common cultural and linguistic representations in the foreign language, a pre-requisite for the use of elliptic, anaphoric and cleft-structures, grounded in natural pragmatic contexts. Nevertheless, compensating pedagogical strategies can be devised such as consciousness-raising tasks to increase the students' awareness of the specific features of oral discourse, for instance:

- extracting tokens from spoken corpora and regrouping them in linguistic, psycholinguistic and pragmatic categories,
- using the native non-native conversations to point out and explain phenomena of overuse, underuse and misuse,
- drawing (self) portraits, correlating linguistic and psycholinguistic patterns,

- studying how natives informally deal with complex structures.

For instance, in the 28 conversations, the native assistants used rising intonation, dislocation and juxtaposition as frequently as more academic canonical questions:

And you can't do this in an English speaking country or? (11A)

So the next Longdale in in November or whenever you'll be able to come back and sort of talk about how you've written your play↑(14A)

So what did the other so the four of you were the organizers of the festival I imagine there was a lot of preparation like... (20A)

This study points out directions to further investigate the links between all three psycholinguistic parameters at work in native non-native conversation. A longitudinal analysis of the database under completion will give more information on the developmental patterns of use and the fluctuations of psycholinguistic trends at different intervals. Whether learners tend towards increased proficiency in all three parameters in the long run or keep relying on trade-off strategies remains to be seen.

References

- Aijmer, K. (2002). *English Discourse Particles, Evidence from a Corpus*. Amsterdam. Benjamins.
- Aijmer, K., Altenberg, B. & Johansson, M. (Eds.) (1996). *Languages in Contrast. Text-based Cross-linguistic Studies*. Lund: Lund University Press, 37-51.
- Biber, D. *et al* (1999). *The Longman Grammar of Spoken and Written English*. London, Longman.
- Cobb, T. (2003). Analyzing Late Interlanguage with Learner Corpora: Québec Replications of Three European Studies. *Canadian Modern Language Review*, vol.59, n°3, 393-423.
- Cook, V. (2002), *Portrait of the L2 User*. Clevedon: Multilingual Matters.
- De Cock, S., Granger, S, Leech, G., & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. In Granger, V. (ed.), *Learner English on Computer* London: Longman, 67-79.
- De Cock, S. (2004). Preferred sequences of words in NS and NSS speech sequences, *Belgian Journal of English Language and Literatures New Series*, 2, 225-246.
- De Cock, S. (2007). Routinized building blocks in native speaker and learner speech. In Campoy, M-C. & Luzón, M-J. (eds.), *Spoken Corpora in Applied Linguistics*. Bern: Peter Lang, 217-233.
- Dörnyei, Z. (2006). *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. New Jersey: Lawrence Erlbaum.
- Ellis, R. & Barkhuisen, G. (2005). *Analysing Learner Language*. Oxford: OUP.
- Foster, P., Tonkyn, A.& Wigglesworth, G. (2000). Measuring spoken language: a unit for all reasons, *Applied Linguistics*, n 21/3, 354-375.

Gilquin, G. (2008). Hesitation marks among EFL learners: pragmatic deficiency or difference? *Pragmatics and Corpus Linguistics*. Berlin: Mouton de Gruyter.

Gilquin G., De Cock, S. & Granger, S. (2010) *LINDSEI, Louvain International Database of Spoken Interlanguage*. Belgium: i6doc.com, Presses Universitaires de Louvain-la-Neuve.

Goffman, E. (1959) *The Presentation of Self in Everyday Life*. New York: Anchor Books, Doubleday.

Goutéraux, P. (2011). Complexité des discours d'apprenants et des tâches en classe d'anglais. In Chini, D. et Goutéraux, P. (eds), *Intégration de l'altérité dans l'apprentissage des langues: formes didactiques et procédures psycholinguistiques*. Rives n°3, Paris : L'Harmattan, 38-57.

Hasselgren, A. (2002). Learner corpora and language testing: small words as markers of learner fluency. In Granger, S., Hung, J. & Petch-Tyson, S. (eds.). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: Benjamins, 143-173.

Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency, *Language Learning Journal*, Vol. 36, n°2, 153-166.

Hilton, H. (2011). Psycholinguistique de la production orale, aisance et disfluenza en L2. In Chini, D. et Goutéraux, P. (eds). *Intégration de l'altérité dans l'apprentissage des langues: formes didactiques et procédures psycholinguistiques*. Rives n°3. Paris : L'Harmattan, 75-94.

Iwashita, Noriko *et al* (2008). Assessed levels of second language speaking proficiency, *Applied Linguistics*, 29, issue 1, 24-49.

Kjellmer, G. (1991). A mint of phrases. In Aijmer, K., & Altenberg, B. (eds). *English Corpus Linguistics*. London: Longman, 111-127.

Moon, R. (2010). What can a corpus tell us about lexis? In O'Keeffe, A. & McCarthy, M. (eds). *The Routledge Handbook of Corpus Linguistics*, chap. 15, London: Routledge, 186-201.

Robinson, P. (ed) (2001). *Cognition and Second Language Instruction*, Cambridge: CUP.

Robinson, P. (2002). *Individual Differences and Instructed Language Learning*. Amsterdam: Benjamins.

Scheibman, J. (2007). Subjective and intersubjective uses of generalizations in English conversation. In Englebretson, R. *Stancetaking in Discourse, Subjectivity, Evaluation, Interaction*, Amsterdam: Benjamins, 111-138.

Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: OUP.

Wray, A. (2001). *Formulaic Language and the Lexicon*. Cambridge, UK: CUP.

Websites

Common European Framework of Reference for Languages:
http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf.

Diderot-Longdale: <http://www.clillac-arp.univ-paris-diderot.fr/projets/longdale>

Longdale Louvain Project: <http://www.uclouvain.be/en-cecl-longdale.html>

PAROLE corpus <http://talkbank.org/SLABank/PAROLE/>