

Hesitation markers among EFL learners: pragmatic deficiency or difference?

Gaëtanelle Gilquin

F.N.R.S – Université catholique de Louvain

I guess what I'm trying to say is, I don't think you can measure life in terms of years. I think longevity doesn't necessarily have anything to do with happiness. I mean, happiness comes from facing challenges and going out on a limb and taking risks. If you're not willing to take a risk for something you really care about, you might as well be dead.

Diane Frolov and Andrew Schneider,
Northern Exposure, Northern Lights, 1993

1. Introduction

A couple of years ago, during a course on “Language and Thought” taught by Professor Herbert Clark at Stanford University, the audience was asked to transcribe a short sentence extracted from a recorded conversation. The sentence was simple enough: “I think that the situation that you describe has gone completely out of hands”. At least, this is how several people (myself included) transcribed the sentence. In reality, the recorded sentence included vocalisations and other (parts of) words which had escaped our notice. What happened was that we concentrated on the semantic content of the sentence, but did not pay attention to what we considered mere interference, namely the markers of hesitation present in the sentence.

This account will probably sound familiar to anyone who has had experience with verbatim transcriptions of spontaneous speech. Since vocalisations, false starts, repetitions and other “smallwords”¹ such as *well* or *I mean* “do not contribute essentially to the message itself” (Hasselgren 2002: 150), they tend to be disregarded. Not only are they disregarded by hearers, in fact, but also, until quite recently, by specialists of language (for an exception to this fact, see Maclay & Osgood’s study, dating from 1959). Talking about smallwords, Hasselgren (2002: 168) refers to an “essential but hitherto largely neglected body of language”. Spoken corpora, by giving access to detailed transcriptions of authentic speech, have made it possible to study hesitation phenomena with a precision and reliability that were

¹ The term “smallword” is borrowed from Hasselgren (2002).

practically unattainable before. Taking advantage of the availability of spoken corpora, and more precisely of a newcomer to the field, the spoken learner corpus, which contains samples of non-native speech (NNS), this paper sets out to investigate the function of hesitation among EFL (English as a Foreign Language) learners with French as a mother tongue, using as a baseline the way the function is performed in native speech (NS). The paper is structured as follows. First the function of hesitation in speech is briefly introduced. Next, the two corpora on which the study is based (LOCNEC and LINDSEI-FR) and the three categories of hesitation phenomena investigated (pauses, smallwords and other devices) are described. The following two sections present the main results of the corpus-based analysis and discuss these results in the light of Foreign Language Teaching (FLT) and English as a Lingua Franca (ELF). Section 7 concludes the paper.

2. The pragmatic function of hesitation in speech

Spoken language, in its purest form, is unrehearsed and spontaneous (see e.g. Burns & Joyce 1997). Interactants build speech as they go along, in a process of on-line planning. As a result, there are times in a conversation when the speaker is inevitably hesitant and does not know what to say next (or how to express it). Such hesitation may manifest itself in various ways. Wiese (1984), for example, mentions filled pauses (e.g. *uh*, *mhm*), repetitions, corrections and drawls. But the literature also recognises a number of smallwords which, among other functions, allow the speaker to “buy time” (e.g. *well*, *I mean* or vague words such as *stuff* or *things like that*). Here, I will adopt a very broad definition of hesitation, which covers (silent and filled) pauses, drawls, truncated words, repetitions, as well as a representative selection of smallwords of hesitation which commonly appear in the literature.²

The function of hesitation is crucial as a conversational strategy. Since speech is dialogic in nature (Burns & Joyce 1997: 13), it is important that a speaker should indicate that s/he needs a moment's reflection, but is still “in control” of his/her turn. Hesitation markers, by signalling a small delay, ensure that the speaker can keep his/her turn in the conversation and is not interrupted by the other participants. As Larcombe (1995) nicely puts it, “[i]n order to keep the floor while we hesitate, we place dummy words in the empty spaces between our

² Corrections (or “repairs”) will only be dealt with as a special case of word truncation, when the speaker starts a word and then goes on to produce another word (e.g. *to a few p= bars*). Other types of corrections (e.g. when the speaker starts with a particular structure and then produces a different one) are not discussed.

words, much as we might drape our coats on a seat at the cinema to prevent others from taking it". Even silent pauses have been demonstrated to play a part in the structure of the message and to contribute to its internal cohesion (Romero Trillo 1994). For foreign (or second) language learners, hesitation is even more crucial. In their search for a formulation which is acceptable in the foreign language, they are likely to experience many planning problems and, therefore, need techniques that enable them to gain time while they are trying to solve these problems.³

Paradoxically, while hesitation is an inherent characteristic of spontaneous conversations and fulfils an important pragmatic function, it goes largely unnoticed in speech (at least native speech).⁴ Talking about the transcription of recorded spoken data, McCarthy (1998: 13) notes that even the best audio typists "often simply 'miss' relevant details" (he gives the example of repetitions) and that they can be "deaf to the presence of discourse markers and other 'little' words which become important the moment one starts to analyse the work they do in the creation of interaction" (see Lindsay & O'Connell 1995 for a clear demonstration of this). Watts (1989), considering the opposite perspective, that of the speaker him-/herself, shows that one may be unaware of the extent to which one uses such little words.

This lack of salience explains why corpora are an ideal resource for the investigation of hesitation markers. Provided they have been transcribed carefully and with some detail, spoken corpora give access to a record of authentic language, including various types of disfluency phenomena, which can be queried automatically and analysed quantitatively and qualitatively. An analysis based on intuition, by contrast, would very probably underestimate the role of hesitation in speech. In the next section, I describe the two corpora that were used in this study.

3. The corpora

In order to investigate the use of pragmatic markers of hesitation by EFL learners, I used a corpus of interviews with advanced French-speaking learners of English, and a comparable corpus of native speech which served as a baseline for the study of the NNS data.

³ De Cock's (2003, 2004) study of recurrent sequences of (two or more) words in native and learner speech has highlighted the particularly large proportion of sequences containing hesitation markers among learners.

⁴ See Chambers (1997) or Temple (2000) on the inequality between native and non-native speakers in this respect, and the fact that disfluency phenomena are more easily stigmatised in non-native speech.

The NNS corpus, LINDSEI-FR, is the French component of LINDSEI (*Louvain International Database of Spoken English Interlanguage*). The LINDSEI project, launched in 1995 at the University of Louvain (UCL) and resulting from the collaboration between a number of universities internationally, aims at gathering oral data produced by advanced learners of English. To date, 11 mother tongue backgrounds are represented: Bulgarian, Chinese, Dutch, French, German, Greek, Italian, Japanese, Polish, Spanish and Swedish. All the components follow the same format so as to make the data comparable. They consist in transcriptions of interviews between a non-native advanced⁵ learner of English and a (native or non-native) interviewer, recorded with the consent of the participants. Each interview lasts about 15 minutes and takes place in three stages. First, the subject is asked to talk for a few minutes about a topic which s/he has chosen⁶ and has had some time to think about before the interview starts (without taking any notes). The conversation then continues informally, with the interviewer asking questions related to the topic chosen by the subject and questions about more general topics (hobbies, life at university, travels abroad, etc). Finally, the subject is presented with four pictures making up a story and is required to recount what happens. Subjects also have to fill in a learner profile questionnaire with information such as gender, number of years of English or knowledge of other foreign languages. The interviews are transcribed orthographically but include indications of pauses (short, medium or long), fillers (*eh*, *er*, *erm*), truncated words, syllable lengthening, overlapping speech, etc.⁷ Each component of LINDSEI contains the transcription of 50 interviews, for a total of over 100,000 words per component. The corpus has been marked up in such a way that the sentences uttered by the learner (B-turns) can be queried independently of those uttered by the interviewer (A-turns), and this is how the corpus will be used here. At this stage, the corpus has not been annotated, but there are plans to have it POS-tagged.

In parallel with LINDSEI, a comparable corpus of native English, LOCNEC (*Louvain Corpus of Native English Conversation*), has been compiled (see De Cock 2003).⁸ Like the

⁵ The proficiency level is established on the basis of an external criterion, namely the number of years of English at university (subjects were in their third or fourth year).

⁶ The topic is chosen among the following three: (i) an experience you have had which has taught you an important lesson; (ii) a country you have visited which has impressed you; (iii) a film/play you have seen which you thought was particularly good/bad.

⁷ The transcription, however, has its limitations when it comes to the (automatic) retrieval of certain hesitation phenomena. While simple programs can be written to extract exact repetitions (e.g. *he knows he knows*), cases of retracing with modification (e.g. *he know he knows...*), are not amenable to such treatment. See Osborne (2006) for an oral learner corpus, the *PAROLE* corpus, annotated using the *CHILDES* system (MacWhinney 2000) and encoded with several hesitation phenomena, including retracings.

⁸ At the moment, access to LINDSEI and LOCNEC is restricted to the project collaborators. However, we expect to release a first version of LINDSEI on CD-ROM in 2008.

LINDSEI components, it contains 50 interviews made up of three parts (set topic, free discussion, picture description), which were transcribed using the same conventions. The interviewees were all native speakers of British English and most of them were undergraduate students in linguistics or English language. In this study, LOCNEC was used as a control native corpus. By comparing the LINDSEI-FR data with data produced by native speakers, it was possible to identify cases where French-speaking learners' use of hesitation markers diverges from that of native speakers, either quantitatively or qualitatively.

Table 1 shows the number of interviews found in each corpus, LOCNEC and LINDSEI-FR, as well as the total number of words for interviewer's and interviewee's turns (A and B turns) and for interviewee's turns only (B turns only). The results presented in this paper are limited to the B turns.

	Nr. of interviews	Nr. of words (A and B turns)	Nr. of words (B turns only)
LOCNEC	50	170,533	125,226
LINDSEI-FR	50	149,127	94,406

Table 1. Number of interviews and words in LOCNEC and LINDSEI-FR

4. Three categories of hesitation markers and their extraction

For the purposes of the analysis, the function of hesitation was divided into three main categories, namely (silent and filled) pauses, smallwords and a miscellaneous category. All markers, except repetitions, were extracted automatically using the programme *WordSmith Tools* (Scott 1999). When necessary, the hits were manually disambiguated.

Silent pauses, which are defined as gaps in the utterance, are probably the most basic way of dealing with problems of formulation. Not knowing what to say, the speaker just remains silent. As pointed out by Fillmore (1979), silent pauses are multifunctional, since they both have a rhetorical function and serve as a marker of disfluency. Because it is almost impossible to identify with any certainty cases where the pause merely has a rhetorical function, however, all silent pauses were taken into account in the analysis. Following the LINDSEI transcriptions, a distinction was made between short pauses (under one second), indicated in LINDSEI by means of one dot, medium pauses (one to three seconds), indicated

by two dots, and long pauses (over three seconds), indicated by three dots. An example is provided in (1).

- (1) I was there for ages and . again I thought right I can't do this so I went out of Breda again <LOCNEC 041>

Alternatively, pauses can be filled by vocalisations, sounds such as *er* or *erm*, as shown in (2).⁹ Discussing *uh* and *um* (the American spelling variants of *er* and *erm*, respectively), Clark & Fox Tree (2002: 75) note that they are “characteristically associated with planning problems”, being used by speakers to announce a delay in speaking. In the LINDSEI transcriptions, the following “fillers” are recognised: *eh* (for a brief sound), *em*, *er*, *erm* and *mm*. Each of them was extracted automatically from the corpora.

- (2) yeah I went to see **erm** another one Jane Eyre and that was really bad it was a a real amateur production <LOCNEC 010>

A number of smallwords can also be used to signal hesitation, as exemplified in (3) with *kind of*. Other examples include *well*, defined by Fuller (2003: 187) as “a delay device when the speaker is not sure how to respond”, *I mean*, which can be used “when pausing to think about what you are going to say next” (*Longman Dictionary of Contemporary English*, Summers 1995: 886), and vague words such as *stuff* and *or something*, which can be used to fill knowledge gaps or lexical gaps (Drave 2002: 26; see also Channell 1994). Most of these words are multifunctional (see Schiffrin 1987: 61 or Aijmer 2002: 19ff). *You know*, for instance, may be used “when you need to keep someone’s attention, but cannot think of what to say next” (Summers 1995: 781), but it may also function as “a speaker appeal for hearer cooperation in a discourse task” (Schiffrin 1987: 63). Often, several pragmatic functions are performed simultaneously by one and the same word and it may prove extremely difficult to disentangle them, despite the help of the surrounding context. Therefore, no attempt was made to disambiguate such cases in the corpus data. By contrast, some smallwords have, next to their pragmatic function(s), a non-pragmatic meaning which can clearly be identified. *Kind of*, for example, illustrated in (3) in its pragmatic function, may also be used with a non-pragmatic function, as a synonym of “type of”, cf. (4). Similarly, *well* has both pragmatic and

⁹ White (1997) counts as filled pauses cases of syllable lengthening (drawls). Here, however, drawls will be considered as a category of their own and dealt with as one the miscellaneous devices (see below).

non-pragmatic functions. Consider the sentence in (5). While the first occurrence signals hesitation, the second one does not, functioning as an adverb together with *as*. Whenever a smallword was used with a non-pragmatic meaning, hence ruling out the function of hesitation, it was discarded manually from the hits returned by the automatic extraction.

- (3) as I say during the year I **kind of** changed my mind about what I wanted to do
<LOCNEC 030>
- (4) it produces just the most fantastic **kind of** sweets <LOCNEC 054>
- (5) yeah well I do enjoy it so I just thought **well** I might as **well** do something I enjoy <LOCNEC 055>

In the miscellaneous category, finally, I included drawls (i.e. syllable lengthening), truncated words and repetitions, which are all signals that the speaker is hesitating (see for example Fox Tree & Clark 1997 on drawls, Temple 2000 on truncations and Wiese 1984 on repetitions). Drawls and truncated words were extracted from the corpora by searching for the symbols used in the transcriptions, namely a colon for drawls and an equals sign for truncated words. Repetitions were identified automatically by means of a program written specifically for this purpose.¹⁰ Examples of drawl, truncated word and repetition are given in (6) to (8).

- (6) they were brilliant so we went skiing and we went **to:** various parts of France
<LOCNEC 008>
- (7) they've had to rethink and they're coming back in their **f=** forties even in their fifties <LOCNEC 011>
- (8) but the last time I got organised and went to Spar first and bought **some some** supper to eat during the film <LOCNEC 035>

Once the different hesitation markers were extracted from the corpus and, when necessary, disambiguated, their frequency in the learner corpus was determined and compared with that in the native corpus. It was thus possible to identify phenomena of overuse and underuse, i.e. cases where learners use significantly more or significantly less of a particular item than native speakers. Statistical significance was tested by means of the chi-square test and probability values of less than 0.05 were considered significant (but most of the results

¹⁰ I thank Marie-Catherine de Marneffe for her help in this matter.

presented are actually significant at the 0.001 level). This quantitative approach was supplemented by a more qualitative analysis, which highlighted differences in use between native and learner English.

5. The hesitation function in native and learner speech

This section presents the main results of the analysis for the three categories of hesitation markers outlined above, namely (silent and filled) pauses, smallwords and a miscellaneous category.

5.1. Pauses

Silent pauses are very frequent, as appears from Table 2, especially short pauses (under one second), which represent the most common device for signalling hesitation, both among native speakers and learners. Medium pauses (one to three seconds) are slightly less frequent and long pauses (over three seconds) are comparatively rare. Learners, however, make heavier use of silent pauses than native speakers, in a way which is statistically significant. This is true of the three lengths of pauses. The words after which pauses typically occur are relatively similar in NS and NNS. In the top four list we find filled pauses (*er* and *erm*), *yes* (or *yeah* in NS) and *and*. This is illustrated in (9) to (11) for native speech.

- (9) he m= he was **erm** .. a lecturer at I think it's Oxford it might have been Cambridge one of one of the[i:] Oxbridge universities <LOCNEC 021>
- (10) yes **yeah** . you you don't really think about anything else you're just sort of sitting there thinking oh gosh <laughs> <LOCNEC 020>
- (11) I think she acted very well **and** . again you really felt for her and erm .. you could kind of imagine yourself in her shoes <LOCNEC 035>

	LOCNEC	LINDSEI-FR	X ²	+/-
Short pause (.)	2087.43	4391.67	958.04	+
Medium pause (..)	1636.24	3153.40	554.10	+
Long pause (...)	54.30	295.53	198.57	+

Total	3777.97	7840.60	1702.24	+
--------------	----------------	----------------	----------------	----------

Table 2. Relative frequency of silent pauses in native and learner speech (per 100,000 words)¹¹

Like silent pauses, the category of filled pauses is, generally speaking, more characteristic of learner speech than of native speech, as shown in Table 3. The individual fillers follow the same tendency, except for *erm*, which is underused by learners. It is interesting to note that, of all the items listed in Table 3, learners prefer *eh* and *er*, which come closer to the normal filled pause in French, *eu*.¹² The fillers involving “m”, by contrast, are less frequent in learner speech.

	LOCNEC	LINDSEI-FR	X²	+/-
<i>eh</i>	81.45	1326.19	1361.18	+
<i>em</i>	42.32	381.33	329.60	+
<i>er</i>	754.64	5260.26	4172.45	+
<i>erm</i>	1266.51	498.91	341.84	–
<i>mm</i>	216.41	414.17	70.05	+
Total	2361.33	7880.86	3636.22	+

Table 3. Relative frequency of filled pauses in native and learner speech (per 100,000 words)

5.2. Smallwords

While French-speaking learners tend to overuse silent and filled pauses to express hesitation, they do not exploit the full range of smallwords that may be used to perform this function, as is clear from Tables 4 and 5, which give the relative frequency of a number of markers of hesitation regularly discussed in the literature. Table 4 lists the smallwords that are shared by native speakers and learners. For some of them the difference in frequency is not statistically significant between the two groups, but in the majority of cases (12 out of 18), there is a

¹¹ In this table and the following, a plus sign signals a case of significant overuse, while a minus sign signals a case of significant underuse. Sometimes, the values were too small to perform a chi-square test; this is indicated by means of “n.a.” (“non-applicable”).

¹² In fact, going back to the original sound files, it turns out that the pronunciation of these fillers is often closer to French *eu*h than to English *eh* or *er*. This is in line with Clark & Fox Tree (2002: 93), who claim that “[s]peakers of English as a second language often import the fillers from their first language”. They add that this is “one reason they continue to be heard as non-native speakers”.

significant underuse among the learners.¹³ Particularly striking is the underuse of *like*, illustrated by (12), which is extremely common in native speech (527.05 occurrences per 100,000 words) but is hardly ever found in learner speech (6.36 occurrences per 100,000 words).

- (12) I don't wanna swim any more people come out with **like** bruises all over their legs where they've hit rocks at the bottom <LOCNEC 025>

	LOCNEC	LINDSEI-FR	X ²	+/-
<i>all right</i>	12.78	3.18	5.73	–
<i>all that</i>	3.99	2.12	0.59	
<i>anyway</i>	45.52	33.90	1.79	
<i>I mean</i>	352.16	152.53	80.75	–
<i>in a way</i>	15.97	3.18	8.41	–
<i>just</i>	746.65	274.35	222.07	–
<i>kind of</i>	86.24	67.79	2.34	
<i>like</i>	527.05	6.36	482.73	–
<i>or so</i>	3.19	7.41	1.91	
<i>or something</i>	43.92	13.77	15.81	–
<i>right</i>	99.02	5.30	80.55	–
<i>something like</i>	27.95	40.25	2.45	
<i>sort of</i>	456.77	34.96	348.65	–
<i>stuff</i>	67.88	5.30	51.47	–
<i>thing</i>	142.14	49.78	44.86	–
<i>things like</i>	55.90	23.30	13.66	–
<i>well</i>	415.25	1076.20	338.60	+
<i>you know</i>	479.13	190.67	126.57	–
Total	3581.52	1990.34	484.37	–

Table 4. Relative frequency of smallwords of hesitation shared by native speakers and learners (per 100,000 words)

¹³ Influence of the context in which the interview took place may not be totally excluded. In LINDSEI-FR the interviewer was one of the participants' teachers, whereas in some of the LOCNEC interviews the interviewer was a fellow student. This difference may have resulted in a more relaxed atmosphere in the latter case, which in turn may have led to a heavier use of some smallwords (see Stubbe & Holmes 1995: 66). However, the number of interviews by a fellow student in LOCNEC is too small to be solely responsible for the observed differences.

A notable exception to learners' tendency to underuse smallwords of hesitation is *well*, which, as Table 4 shows, is significantly overused by learners (sentences (13) and (14) are just a couple of illustrations taken from LINDSEI-FR). To paraphrase Hasselgren (1994), one could say that *well* is a “pragmatic teddy bear” for learners, who cling to it because it is familiar,¹⁴ safe and widely useable. This over-reliance on *well* probably explains why learners do not feel the need to use other smallwords of hesitation, which results in an overall underuse of this category of hesitation markers.

- (13) but <laughs> er **well** when I've seen the number of my room . **well** I I noticed that it wasn't the case <LINDSEI-FR 002>
- (14) er . with the school .. the first time I was **well** about fifteen or sixteen . secondary school .. er . the second time . **well** in fact .. er I've been studying in a: teachers training college before I was here <LINDSEI-FR 048>

Table 5 lists smallwords which occur in LOCNEC but are never found in LINDSEI-FR (at least not with the function of hesitation). Most of these belong to what Stubbe & Holmes (1995) call “set marking tags”, e.g. *and things, or anything*:

- (15) and so we spend the days doing things like climbing mountains and going swimming in lakes **and things** and the nights at the pub drinking you know <LOCNEC 023>
- (16) it wasn't like a test **or anything** it was just like to see what they got right . how much they understand of it <LOCNEC 019>

French-speaking learners' failure to use common expressions including the word *thing* (*and things, all that kind of thing*) and their underuse of the word *thing* and the expression *things like* (see Table 4) are quite surprising in view of the fact that, in writing, they (like learners from most mother tongue backgrounds) overuse the word *thing*, which, because of its vague character, tends to be avoided by native writers.¹⁵ In fact, it appears that French-speaking

¹⁴ As underlined by Mukherjee & Rohrbach (2006: 216), *well* is one of the only discourse markers not to be underrepresented in ELT (English Language Teaching) textbooks and materials.

¹⁵ The relative frequency of the word *thing* in the French component of the *International Corpus of Learner English* (Granger et al. 2002), a corpus of essays written by advanced learners of English, amounts to 37.92 per

learners use the word *thing* almost as often in writing as in speech (37.92 vs. 49.78 per 100,000 words, $X^2 = 1.95$, non-significant), which is in stark contrast to the situation among native speakers, who use it much more often in speech than in writing (113.69 vs. 11.01 per 100,000 words, $X^2 = 12509.11$, $p < 0.001$).¹⁶ This suggests that learners, as a rule, fail to distinguish between genres, using a single type of English, undifferentiated between speech and writing. Put more simply, learners tend to write as they speak and speak as they write (see also Granger & Rayson 1998), which implies, in the latter case, a lack of hesitation markers typical of native spoken interactions.

	LOCNEC
<i>and things</i>	39.13
<i>and stuff</i>	24.76
<i>stuff like that</i>	18.37
<i>or anything</i>	16.77
<i>a bit of a</i>	14.37
<i>or anything/anybody like</i>	3.19
<i>all that kind of thing</i>	2.40
<i>what not</i>	2.40
<i>all that kind of stuff</i>	0.80
<i>all that sort of stuff</i>	0.80
<i>all that stuff</i>	0.80

Table 5. Relative frequency of smallwords of hesitation restricted to native speech (per 100,000 words)

Not only do learners fail to exploit the whole gamut of smallwords of hesitation, but they also use them in a way which may be quite different from their use in native speech. Consider Table 6, which compares the types of phrases modified by *sort of* in native and non-native speech. In both cases, *sort of* most frequently modifies a noun phrase, as in (17) and (18). However, in LINDSEI-FR, *sort of* is regularly used when the learner does not know a word, as illustrated in (19) and (20), which is not the case in native speech.

100,000 words, against 11.01 in the academic component of the *British National Corpus, World Edition* (Burnard 2000), a difference which is statistically significant at the 0.001 level.

¹⁶ The results for native English come from a comparison of the spoken component and the academic component of the *British National Corpus, World Edition* (Burnard 2000).

- (17) the teacher one of the[i:] English teachers out there she was amazing she was like a **sort of** second mom she just adopted me <LOCNEC 008>
- (18) erm .. and then there in fact we: .. we rent a: a **sort of** apartment <LINDSEI-FR 025>
- (19) yes yes .. yeah because of the: **sort of** eh vapeur I don't know how you say it <LINDSEI-FR 040>
- (20) it's a: a round instrument with er . and you er bat= beat against it (...) it's a **sort of** drum but er (...) it's big= it's bigger and it and the sound is <LINDSEI-FR 003>

	LOCNEC	LINDSEI-FR	X ²	+/-
NP	41.08%	72.73%	12.76	+
VP	30.94%	12.12%	5.27	–
AdjP	5.59%	3.03%	0.40	
AdvP	2.45%	0.00%	n.a.	
PrepP	0.35%	0.00%	n.a.	
Other	19.58%	12.12%	1.12	
Total	100%	100%		

Table 6. Types of phrases modified by *sort of* in native and learner speech

In addition, premodification of a noun phrase is much more common among learners, who show a strong preference for this type of phrase to the detriment of others. Native speakers, on the other hand, use *sort of* with a wider range of functions (see also De Cock 2004), including the modification of a verb phrase, e.g. (21), an adjectival phrase (22), an adverbial phrase (23) and a prepositional phrase (24), or the use as an adverb, cf. (25). In the same way as learners fail to exploit the whole repertoire of smallwords of hesitation, it thus turns out that they also fail to exploit the different functions of the smallwords they do use.

- (21) and basically .. the . er the young lion gr= grows up .. he **sort of** runs away sort of thing <LOCNEC 002>
- (22) I do know they're looking for somebody so that's that's **sort of** quite positive I know that nobody else has applied yet <LOCNEC 015>

- (23) we had to start at eight o'clock in the morning and we finished at **sort of** one
<LOCNEC 013>
- (24) so yeah and er .. the rhinos were **sort of** on the track in front of us <LOCNEC
050>
- (25) it's quite a personal thing I think isn't it **sort of** <LOCNEC 041>

Another illustration of the different use of a smallword by native speakers and learners concerns the position of *well* in utterances. Table 7 shows that, while native speakers favour the utterance-initial position, e.g. (26), learners use *well* “all over the place”, with a clear preference for the utterance-medial position, as in (27).¹⁷ This confirms the “teddy bear” nature of *well*, which, like Hasselgren's (2002) lexical teddy bears, tends to be overgeneralised (from a mainly initial position to a variety of positions).

- (26) **well** yes .. you don't have to sort of worry about getting a bus home in the
middle of the night <LOCNEC 016>
- (27) and then . eh we've some family in Portugal so I go there very often . er to
Spain to France **well** Europe in general I like sun so <laughs> I go where the
sun is <LINDSEI-FR 002>

	LOCNEC	LINDSEI-FR	X ²	+/-
Utterance-initial	58.85%	31.89%	103.31	–
Utterance-medial	36.73%	63.68%	100.77	+
Utterance-final	4.42%	4.43%	0.00	
Total	100%	100%		

Table 7. Position of *well* in native and learner speech

Finally, it should be pointed out that LINDSEI-FR contains three French smallwords of hesitation, namely *enfin* (‘well’), *allez* (‘you know’) and *hein* (‘right’), illustrated in (28) to (30). The frequency of *enfin* is particularly striking, being higher than the frequency of most of the English smallwords listed in Table 4.

¹⁷ The results are highly significant, except for the utterance-final position of *well*.

- (28) I I don't meet man= meet many: . many students who are in **enfin** with me during the: the lectures and er .. **enfin** it's not <LINDSEI-FR 020>
- (29) it was . it was really a pleasant experience because it was . they were . **allez** em quiet and .. <LINDSEI-FR 022>
- (30) there are two coasts in Italy **hein** there are the[i:] . the[i:] east coast so on the si= on the side of er the[i:] Ad= .. Adriatic <LINDSEI-FR 017>

	LINDSEI-FR
<i>enfin</i>	70.97
<i>allez</i>	7.41
<i>hein</i>	2.12

Table 8. Relative frequency of French smallwords of hesitation in LINDSEI-FR (per 100,000 words)

The presence of this type of interference in the oral production of even advanced learners is a common phenomenon. While for French-speaking learners the most conspicuous example is *enfin*, it is not rare to hear Dutch-speaking learners sprinkle their foreign talk with smallwords like *allez* or *nou*, or German-speaking learners, with *ach* or *also*. Such smallwords seem to escape learners' conscious attention (see also De Cock 2004: 235), which makes them prime examples of what Færch & Kasper (1986: 59) call "automatic transfer", which "involves the activation of highly automatized sub-routines from a secondary area of declarative knowledge, in situations in which attention is concentrated on something else". That automatic transfer occurs in relation to the hesitation function should come as no surprise, since moments of hesitation are typically moments when one's attention is concentrated on the content of one's speech (i.e. what to say next) rather than on its form, which gives free rein to the transfer of markers which are, to use Blackwell's (2000: 5) words, "at the bottom of the metalinguistic awareness hierarchy".

5.3. Miscellaneous

This section examines a number of other devices that may be used to express hesitation, namely drawls (syllable lengthening), truncated words and repetitions. A cursory look at Table 9 reveals that all these three devices are overused by learners.

	LOCNEC	LINDSEI-FR	X ²	+/-
Drawls	344.18	1958.56	1365.54	+
Truncated words	304.25	1113.28	543.90	+
Repetitions	1801.54	4152.28	1088.38	+
Total	2449.97	7224.12	2853.54	+

Table 9. Relative frequency of miscellaneous hesitation markers in native and learner speech (per 100,000 words)

Let us start with drawls and note, first, that the figures in Table 9 include the lengthening of the article *a* in [ei], but do not include the lengthening of *the* in front of a vowel, since lengthening in such cases is in principle obligatory and, hence, does not (or at least not necessarily) indicate hesitation. Tables 10 and 11 show the words which are most frequently lengthened in LOCNEC and LINDSEI-FR, respectively. We notice that in learner English, the word that is most often lengthened is *to*, both as an infinitive marker, cf. (31), and as a preposition, cf. (32). In both corpora, function words are the most likely candidates for syllable lengthening, especially articles (*a* and *the*) and personal pronouns, e.g. (33) and (34).

- (31) no because you know when you you dance rock it's always the man who has **to:** to do everything <LINDSEI-FR 030>
- (32) well I'm thinking about doing erm a a another year but I'm not sure where in Brussels or I want to go **to:** England but I I have to try and convince my boyfriend to do so <LINDSEI-FR 005>
- (33) before I went I thought that I might not cos I thought **the[i:]** .. **the[i:]** sort of . the problems with . disease and <LOCNEC 012>
- (34) **she:** gets up she sees the picture and she doesn't look too impressed actually with what he's done <LOCNEC 008>

	LOCNEC		LINDSEI-FR
<i>the</i>	151.73	<i>to</i>	542.34
<i>to</i>	67.08	<i>the</i>	451.24
<i>a</i>	43.92	<i>a</i>	296.59
<i>she</i>	12.78	<i>we</i>	83.68
<i>so</i>	9.58	<i>she</i>	80.50

<i>we</i>	7.19
<i>you</i>	5.59
<i>they</i>	5.59
<i>yeah</i>	3.19
<i>my</i>	2.40

Table 10. Most frequent syllable lengthenings in native speech (relative frequencies per 100,000 words)

<i>very</i>	59.32
<i>they</i>	57.20
<i>I</i>	52.96
<i>you</i>	38.13
<i>he</i>	34.96

Table 11. Most frequent syllable lengthenings in learner speech (relative frequencies per 100,000 words)

It is interesting to focus on the two articles, *a* and *the*, and examine the ways in which they are lengthened. Both of them may be lengthened in two ways, namely by prolonging the final vowel, [ə], or by using a different sound, [i:] for *the* and [ei] for *a*. These four possibilities are exemplified for native English in (35) to (38).

- (35) when **the:** the fat kid falls in the chocolate river they do this song about how it's not wise to be greedy <LOCNEC 054>
- (36) it's all set up through **the[i:]** erm the research <LOCNEC 031>
- (37) she was **a:** Spanish teacher no French teacher but she spoke Spanish <LOCNEC 036>
- (38) Macbeth himself was actually **a[ei]** erm a gangster <LOCNEC 047>

Table 12 shows that, while the difference in frequency between native and learner speech is not statistically significant in the case of *the[i:]* (followed by a consonant) and *a[ei]*, it is significant with *the:* and *a:*, which are both overused by learners. In fact, *the:* and *a:* are learners' preferred drawls on the articles (to be compared with native speakers, who favour *a:* and *the[i:]*).¹⁸ It is probably not a coincidence that these two forms correspond to the normal mode of lengthening for the French articles, namely the lengthening of the final vowel ([ə] for

¹⁸ In the case of *a*, however, the sound of the lengthening is similar to that of the filled pause *er*, so that the presence of *a:* may sometimes just be the result of an interpretation on the part of the transcriber. While in (37) above there is little doubt that the speaker indeed had the article in mind (as otherwise the sentence would be ungrammatical), in (i) the sound [ɜ:] could also have been transcribed as *er*.

(i) yes yeah it's very big erm . you'd need **a:** a good five days to see all of it really <LOCNEC 006>

the definite article *le* and [œ̃] for the indefinite article *un*).¹⁹ Learners seem to shy away from the special draws (those that use a different sound) and show a predilection for the mode of lengthening they are used to in their mother tongue.

	LOCNEC	LINDSEI-FR	X ²	+/-
<i>the</i> [i:] (+C)	134.16	111.22	2.28	
<i>the</i> :	17.57	353.79	376.01	+
<i>a</i> [ei]	8.78	3.18	2.65	
<i>a</i> :	43.92	296.59	225.64	+

Table 12. Relative frequency of prolonged *the* and *a* in native and learner speech (per 100,000 words)

Truncation, like syllable lengthening, is overused by French-speaking learners. The difference between native and non-native speakers, however, is not only quantitative but also qualitative. In native speech, most of the truncations involve just one letter, as in (39). This represents 85% of all truncations, against 65% in non-native speech. Learners, by contrast, regularly pronounce whole morphemes before actually truncating the word, as exemplified in (40) to (42). This tendency seems to point to a high degree of decomposition of words in learners' mental lexicon, which echoes the claim made by Kjellmer (1991: 124) about word sequences that learners' "building material is individual bricks rather than prefabricated sections".

- (39) and just it was a bit boring but just working in the shop .. and s= eh selling people newspapers and magazines <LOCNEC 013>
- (40) it was a few years ago and er I had a **boy**= a boyfriend but eh my mother is extremely severe strict with eh with me <LINDSEI-FR 003>
- (41) then I decided yeah .. you you've done more more Dutch than German in in **second**= in secondary school so now you m= you're doing you're making some some German but em <LINDSEI-FR 007>
- (42) and the parents were er . just er behind them and er .. we I was **say**= saying the last words of eh the weekend er say goodbye <LINDSEI-FR 042>

¹⁹ These are the masculine forms of the articles. For the feminine form of the definite article, *la*, the final vowel [a] would be lengthened. For the feminine form of the indefinite article, *une*, things are slightly more complex, as either the initial vowel [y] or the final vowel [ə] could be lengthened. In some cases, even the consonant [n] may be lengthened.

Native and non-native speakers also display a difference in terms of the types of truncations they have recourse to. In Table 13, a distinction is made between three types of truncations, namely stutter, when the complete word comes immediately after the truncation as in (43), delay, when the complete word comes later in the utterance as in (44), and repair, when the complete word does not occur in the utterance at all as in (45).

- (43) and it just has roads it doesn't have any other form of transport there's no tube no tram . **n=** no buses <LOCNEC 050>
- (44) yeah .. so erm .. yeah we arrived in Italy . and erm .. we **st=** basically we sort of stayed one or two nights in each town <LOCNEC 014>
- (45) and eh . one of my **teach=** one of my tutors who was there .. erm called erm . <name> she: . was familiar . with Lancaster University and the Linguistics Department <LOCNEC 002>

	LOCNEC	LINDSEI-FR	X ²	+/-
Stutter	48.56%	46.24%	0.60	
Delay	31.23%	28.34%	1.20	
Repair	20.21%	25.57%	4.28	+
Total	100%	100%		

Table 13. Types of truncations

The three categories are ordered identically in NS and NNS, with stutters coming first, delays second and repairs last. Stutters and delays occur with a frequency which is not statistically different between native speakers and learners. What differs is the proportion of repairs among the two groups: learners use significantly more of them than native speakers. LINDSEI-FR abounds with examples such as (46) to (48), where the learner wants to utter a particular word (presumably *each*, *tricky* and *said*, respectively, in the examples), but goes on to use another one, which s/he considers more appropriate or perhaps safer (cf. *tricky*, which is presumably less familiar to the learner than the adjective *complicated*, which has a cognate equivalent in French). Paradoxically, this “self-correction” is not always for the better. In (47) the speaker fails to select the correct form of the word, using the verbal form instead of the adjectival one (*complicated*), and in (48) s/he starts with what looks like the correct verb, *say*, but then decides against it and uses the inappropriate verb *tell*.

- (46) they used to cook cos I I used to love Sunday lunch <laughs> you know cook Sunday lunch **ea**= every . every Sunday so I .. with Yorkshire pudding and everything <LINDSEI-FR 005>
- (47) and er . one day a woman arrives at er . her husband's house and her name is Shag and in fact er it is her husband's mistress .. it's very very **tri**= enfin complicate <LINDSEI-FR 038>
- (48) she was not stressed er .. and she **s**= she told yes er I I I cannot **c**= not cry here so er it's er .. **i**= it never enfin .. er it helps . nothing <LINDSEI-FR 042>

The last hesitation device to be examined is the repetition of words or sequences of words. Table 14 gives the results for one-, two- and three-word repetitions in native and learner speech.²⁰ In the three cases, repetitions are used much more often by learners than by native speakers. Not unexpectedly, the longer the sequence of words, the less likely it is to be repeated (see also Biber et al. 1999: 1055).

	LOCNEC	LINDSEI-FR	X ²	+/-
One-word repetitions	1395.08	3540.03	1094.15	+
Two-word repetitions	348.97	527.51	40.48	+
Three-word repetitions	57.50	84.74	5.78	+
Total	1801.54	4152.28	1088.38	+

Table 14. Relative frequency of repetitions in native and learner speech (per 100,000 words)

Table 15 gives an overview of the most frequent (sequences of) words that are repeated in the two corpora. With a few exceptions the rank orders are different, but many repetitions are common to the two groups. By and large, they also correspond to the types of repetitions mentioned in Biber et al. (1999: 1055ff) as particularly frequent, with the presence of nominative personal pronouns, the definite article *the* or personal pronouns + verb contractions. One difference with Biber et al.'s results, however, is that prepositions figure quite prominently in the repetitions. This is especially true of learner English, which has seven prepositions in the forty most frequent one-word repetitions (*to, in, for, of, on, with, at*),

²⁰ Contractions are counted as one word, following Biber et al.'s (1999: 1061) suggestion that, since they are likely to be processed as single words, they should be treated as such for the purposes of studying disfluency phenomena.

but this is also the case in native English, with its four prepositions in the top forty (*to, in, of, for*). Some typical repetitions are shown in (49) to (52) for LINDSEI-FR.

- (49) **I I** liked it very much of course but I'm you know after five years you see it's eh a bit like eh routine <laughs> <LINDSEI-FR 029>
- (50) and also **the the** telephone er booth . it was very astonishing you could er you could telephone **to: to:** anyone er .. er without paying anything <laughs> it was gratis er I don't know if you can say that in English <LINDSEI-FR 015>
- (51) we also went **in a in a** pub and they were singing eh lovely Irish songs <LINDSEI-FR 006>
- (52) maybe **I don't know I don't know** but it's very impressive because you walk along this wall and you see all these names .. it's ve= really very impressive <LINDSEI-FR 040>

LOCNEC		LINDSEI-FR	
<u>One-word repetitions</u>			
I I	149.33	I I	530.69
yeah yeah	129.37	to to	433.24
it's it's	91.83	the the	342.14
no no	77.46	a a	233.04
and and	73.47	yes yes	170.54
a a	71.07	no no	146.18
it it	63.09	they they	123.93
the the	62.29	we we	110.16
you you	44.72	you you	106.98
to to	36.73	in in	102.75
<u>Two-word repetitions</u>			
it was it was	34.34	in a in a	19.07
I don't I don't	11.18	no no no no	19.07
I think I think	11.18	in the in the	18.01
I was I was	7.99	it was it was	13.77

that was that was	7.99	on the on the	10.59
<u>Three-word repetitions</u>			
I don't know I don't know	2.40	I don't know I don't know	11.65
you have to you have to	2.40	and so on and so on	5.30

Table 15. Most frequent repetitions in native and learner speech (relative frequencies per 100,000 words)

Combined with the results from the preceding two sections, all this gives a picture of French-speaking learners neglecting smallwords of hesitation and, instead, favouring alternative devices such as (silent or filled) pauses, drawls, truncations and repetitions. We have also seen that, when compared with native English, the learners' use of hesitation markers reveals a number of idiosyncrasies. It is time, now, to assess the impact of such "differences" on learners' discourse and to determine whether they deserve a place in the FLT (Foreign Language Teaching) curriculum.

6. Assessing learners' use of the hesitation function

Non-native speakers are more likely to hesitate than native speakers. This is because, next to the question of what to say next ("conceptualisation"), speakers have to work out how to say it ("formulation"),²¹ and given that the language in which learners express themselves is not their mother tongue but a – usually imperfectly acquired – foreign or second language, this second stage normally involves more difficulties for them than for native speakers. The above analysis, however, has shown that not all categories of hesitation markers are overused by learners, as one may have expected. Learners overuse pauses and other such non-lexical devices, but smallwords, on the other hand, tend to be significantly underrepresented in learner speech, *well* being a notable exception. One may wonder whether these differences between native speakers' and learners' use of the hesitation function are just that – differences – or whether they should best be viewed as pragmatic deficiencies, which should somehow be remedied. This is the issue that is addressed in this section.

²¹ See Kempen (1977) on the distinction between conceptualisation and formulation.

Pragmatic differences have been given considerable attention in the literature on English as a Lingua Franca (ELF), i.e. English as a means of communication between speakers with different mother tongues (see e.g. Seidlhofer 2005). According to the advocates of ELF, only those features which cause misunderstanding should be eradicated. Features which differ from native English but allow mutual intelligibility, on the other hand, are tolerated (or even promoted). In this context many cross-cultural encounters are claimed to be successful, and according to Aston (1993: 245), “interlanguage pragmatics should operate with a difference hypothesis rather than a deficit hypothesis”. Hesitation phenomena such as those investigated here do not normally lead to misunderstanding or communicative breakdown. They are at best “‘ripples’ on the pragmatic surface” (Seidlhofer 2001: 147). As such, they should not qualify for the label of “deficiencies”, but should instead be considered as mere differences, which are “non-fatal” (Jordan & Fuller 1975) to the conversation. In what follows, however, I would like to argue that markers of hesitation may have a role to play in the success (or otherwise) of interactions, and that it is precisely those markers that are overused by learners which may be detrimental to the conversation, whereas the markers they underuse help make the pragmatic “ripples” smoother.

Let us consider silent pauses. Not only do they fulfil the function of hesitation, but they may also indicate that the speaker has finished his/her turn and that the floor has become empty. Silences, therefore, may be misinterpreted, and the learner who overuses them runs the risk of losing his/her turn, while s/he was just trying to gain some time. This is especially true of long pauses (three seconds or more), where speaking may be “declared to have stopped rather than merely paused” (Griffiths 1991: 346). Pauses of one second or less are comparatively well tolerated, one second being, according to Jefferson (1989), the “standard maximum silence” in interactions. It should also be noted that the position of the silent pause (not examined here) may be relevant, as Lennon (1990: 393) points out, with pauses occurring at major syntactic boundaries being more easily accepted (and, I would add, less likely to be misinterpreted) than pauses occurring within syntactic units. Yet, whatever their length or position, silent pauses have a feature, shared by other non-lexical markers of hesitation such as fillers or drawls, which makes them undesirable in interactions, especially when they are overrepresented: they are, in Möhle’s (1984: 36) words, “communicatively disturbing”. More precisely, these markers, often referred to as “temporal variables” (Grosjean 1980), have been shown to contribute to the impression of non-fluency among EFL speakers (see Lennon 1990). In comparison, the (native-like) use of smallwords of hesitation enables the speaker to

hold the floor and stall for time, but in addition, gives an impression of fluency, as convincingly demonstrated by Hasselgren (2002).

The key issue here seems to be fluency, that is, “the ability to contribute to what a listener, proficient in the language, would normally perceive as coherent speech, which can be understood without undue strain, and is carried out at a comfortable pace, not being disjointed, or disrupted by excessive hesitation” (Hasselgren 2002: 148). Although fluency would not be considered as one of the “core” features of ELF, since it is not crucial to intelligibility (it just helps to be “understood without undue strain”), it is nonetheless an important aspect of oral language. As Lennon (1990: 391-392) explains, “fluency reflects the speaker’s ability to focus the listener’s attention on his or her message by presenting a finished product rather than inviting the listener to focus on the working of the production mechanisms”. In other words, it makes it possible for the listener to concentrate on what should be central to an utterance, namely its content. Fluency is crucial in the acquisition of a foreign/second language, as witnessed for example by the *Common European Framework of Reference for Languages* (Council of Europe 2001: 129), which lists spoken fluency as one of the two generic qualitative factors determining the functional success of the learner (the other one being propositional precision) and requires that learners at the C2 level be able to “express [themselves] at length with a natural, effortless, unhesitating flow”. To the question of whether learners’ use of the hesitation function is “deficient” or merely “different”, I would therefore argue for the former. If learners are to achieve native-like proficiency – and despite the claims made by ELF, this is still a goal pursued by many of them (Mukherjee 2005) and one that is pedagogically sound (Kuo 2006) – they have to learn how to deal with hesitation (which is part and parcel of any unplanned spoken interaction) in a way which does not impair fluency.

It therefore seems important to incorporate the function of hesitation into the (advanced) FLT curriculum, at least in the form of awareness-raising activities. Until recently, students were invariably presented with “aseptic” spoken texts (both in reading and listening comprehension tasks), from which all hesitation markers had been removed.²² The textbook extract in Table 16 is a case in point. Lately, mainly under the impetus of corpus linguistics and the ensuing wave of “authenticity”, hesitation markers have started to creep into textbooks, as illustrated by the extract in Table 17. Because of the lack of salience of such markers, however, simple exposure is not enough to raise students’ consciousness. Hesitation

²² See Römer (2004) on the general lack of authenticity of dialogues in textbooks.

markers need to somehow “emerge” and be brought to students’ attention by means of appropriate activities. This could involve addressing issues such as the non-universality of fillers,²³ the variety of hesitation markers, the multifunctionality of smallwords or the role of hesitation as a politeness strategy (in the extract of Table 16, for example, Sue’s answer could have been made more polite simply by adding a hesitation marker before “I can’t actually”).²⁴ Learners should be taught to rely less on pauses and other non-lexical devices, overused and “communicatively disturbing”, and to have recourse, instead, to smallwords, since these are less disruptive and “oil (...) the wheels of verbal interaction” (Stubbe & Holmes 1995: 63).²⁵

JOAN:	Do you think you could lend me £10, Sue?
SUE:	I can’t actually. Sorry but I’ve only got £20 for the weekend.
JOAN:	OK. Thanks anyway.
SUE:	Sorry about that.
JOAN:	It’s OK. Don’t worry about it.

Table 16. Textbook extract 1 (Mills 1990: 176)

J:	What? You ring people up and say, give me so much money and I’ll jump out of an aeroplane?
S:	Yeah, yeah, because I think people are more willing to do that because the, you know because it’s something that kind of is right at the edge of the sort of comfort zone, so people are going, ‘Oh I wouldn’t do that, I’ll, I’ll give...’

Table 17. Textbook extract 2 (Gomm & Hird 2001: 132)

7. Conclusion

Hesitation phenomena are inherent in spontaneous speech, both native and non-native. As noted by Lennon (1990: 392-393), it is therefore not the presence vs. absence of such features that distinguishes between NNS and NS performance, but their frequency and distribution. The corpus analysis carried out on the basis of LINDSEI-FR and LOCNEC has revealed that, while advanced French-speaking learners of English overuse pauses and other non-lexical devices, they tend to underuse smallwords such as *like*, *I mean* or *you know*. This is quite unfortunate since non-lexical hesitation markers are precisely those that give an impression of non-fluency, whereas smallwords “keep our speech flowing” (Hasselgren 2002: 150). For this

²³ Cf. Clark & Fox Tree (2002: 92), who list the most common fillers in several languages.

²⁴ The politeness function of hesitation markers is briefly discussed in Stubbe & Holmes (1995). See also Brown & Levinson (1987) on the politeness function of vague language.

²⁵ See also Romero Trillo (2002) on the need to teach such discourse markers.

reason, learners' use of the hesitation function has been described as "deficient", rather than just "different", and it has been suggested that the function deserves a place in the (advanced) FLT curriculum. The idea is not to eliminate hesitations, which are inseparable from spontaneous speech, but to equip learners with techniques of hesitation that are less disruptive to the interaction.

Fox Tree & Clark (1997: 166-167) note that "[s]pontaneous speech is replete with signals about the actual process of production". They add that "[a]ny model of production will be incomplete until it accounts for these signals, including how they are planned and produced on the fly". This paper has gone some way towards accounting for such aspects in non-native speech. Many more avenues need to be explored, however. To give but two examples, one could examine whether any influence of the mother tongue is noticeable in learners' use of the hesitation markers, or whether interviewers react differently to native and non-native speakers' ways of hesitating. Along these paths too, as was the case for this study, corpora should turn out to be valuable resources, able to shed light on areas which had hitherto largely remained in the dark. And this, we can assert without much hesitation.

References

- Aijmer, Karin. 2002. *English Discourse Particles. Evidence from a Corpus*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Aston, Guy. 1993. 'Notes on the interlanguage of comity'. In Gabriele Kasper and Shoshana Blum-Kulka (eds) *Interlanguage Pragmatics* (pp. 224-250). New York: Oxford University Press.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow, Essex: Pearson Education Limited.
- Blackwell, Susan. 2000. 'Looking up *look*: discourse markers in the Bank of English'. In John M. Kirk (ed.) *Corpora Galore. Analyses and Techniques in Describing English* (pp. 3-16). Amsterdam & Atlanta, GA: Rodopi.
- Brown, Penelope and Stephen C. Levinson. 1987. *Politeness Universals in Language Usage*. Cambridge: Cambridge University Press.

- Burnard, Lou. 2000. *Reference Guide for the British National Corpus (World Edition)*. Available at <http://www.natcorp.ox.ac.uk/docs/userManual/> (last accessed on 27 February 2007).
- Burns, Anne and Helen Joyce. 1997. *Focus on Speaking*. Sydney: National Centre for English Language Teaching and Research.
- Chambers, Francine. 1997. 'What do we mean by fluency?'. *System* 25(4): 535-544.
- Channell, Joanna. 1994. *Vague Language*. Oxford: Oxford University Press.
- Clark, Herbert H. and Jean E. Fox Tree. 2002. 'Using *uh* and *um* in spontaneous speaking'. *Cognition* 84: 73-111.
- Council of Europe. 2001. *Common European Framework of Reference for Languages*. Cambridge: Cambridge University Press.
- De Cock, Sylvie. 2003. *Recurrent sequences of words in native speaker and advanced learner spoken and written English*. Unpublished PhD Thesis. Louvain-la-Neuve: Centre for English Corpus Linguistics, Université catholique de Louvain.
- De Cock, Sylvie. 2004. 'Preferred sequences of words in NS and NNS speech'. *Bell* 2: 225-246.
- Drave, Neil. 2002. 'Vaguely speaking: a corpus approach to vague language in intercultural conversations'. In P. Peters, P. Collins and A. Smith (eds) *New Frontiers of Corpus Research. Papers from the Twenty First International Conference on English Language Research on Computerized Corpora* (pp. 25-40). Amsterdam & New York: Rodopi.
- Færch, Claus and Gabriele Kasper. 1986. 'Cognitive dimensions of language transfer'. In Eric Kellerman and Michael Sharwood Smith (eds) *Crosslinguistic Influence on Second Language Acquisition* (pp. 49-65). New York: Pengamon Press.
- Fillmore, Charles J. 1979. 'On fluency'. In Charles J. Fillmore, Daniel Kempler and William S.-Y. Wang (eds) *Individual Differences in Language Ability and Language Behavior* (pp. 85-101). New York: Academic Press.
- Fox Tree, Jean E. and Herbert H. Clark. 1997. 'Pronouncing "the" as "thee" to signal problems in speaking'. *Cognition* 62: 151-167.
- Fuller, Janet M. 2003. 'Discourse marker use across speech contexts: a comparison of native and non-native speaker performance'. *Multilingua* 22: 185-208.
- Gomm, Helena and Jon Hird. 2001. *Inside Out. Advanced. Teacher's Book*. Oxford: Macmillan Education.

- Granger, Sylviane, Estelle Dagneaux and Fanny Meunier. 2002. *The International Corpus of Learner English. Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Granger, Sylviane and Paul Rayson. 1998. 'Automatic profiling of learner texts'. In Sylviane Granger (ed.) *Learner English on Computer* (pp. 119-131). New York: Addison-Wesley Longman.
- Griffiths, Roger. 1991. 'Pausological research in an L2 context. A rationale, and review of selected studies'. *Applied Linguistics* 12(4): 345-364.
- Grosjean, François. 1980. 'Temporal variables within and between languages'. In Hans W. Dechert and Manfred Raupach (eds) *Towards a Cross-Linguistic Assessment of Speech Production* (pp. 39-53). Frankfurt am Main: Peter Lang.
- Hasselgren, Angela. 1994. 'Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with vocabulary'. *Applied Linguistics* 4(2): 237-260.
- Hasselgren, Angela. 2002. 'Learner corpora and language testing. Smallwords as markers of learner fluency'. In Sylviane Granger, Joseph Hung and Stephanie Petch-Tyson (eds) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 143-173). Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Jefferson, Gail. 1989. 'Preliminary notes on a possible metric which provides for a "standard maximum" silence of approximately one second in conversation'. In Derek Roger and Peter Bull (eds) *Conversation. An Interdisciplinary Perspective* (pp. 166-196). Clevedon: Multilingual Matters.
- Jordan, Brigitte and Nancy Fuller. 1975. 'On the non-fatal nature of trouble: sense-making and trouble-managing in lingua franca talk'. *Semiotica* 13(1): 11-31.
- Kempen, Gerard. 1977. 'Conceptualizing and formulating in sentence production'. In Sheldon Rosenberg (ed.) *Sentence Production: Developments in Research and Theory* (pp. 259-274). Hillsdale, NJ: Lawrence Erlbaum.
- Kjellmer, Göran. 1991. 'A mint of phrases'. In Karin Aijmer and Bengt Altenberg (eds) *English Corpus Linguistics* (pp. 111-127). London: Longman.
- Kuo, I-Chun (Vicky). 2006. 'Addressing the issue of teaching English as a lingua franca'. *ELT Journal* 60(3), 213-221.
- Larcombe, Michael. 1995. 'Why do people say "um" and "er" when hesitating in their speech?'. *New Scientist* 06 May 1995. Retrieved from <http://www.newscientist.com/backpage.ns?id=lw125> (last accessed on 27 February 2007).

- Lennon, Paul. 1990. 'Investigating fluency in EFL: a quantitative approach'. *Language Learning* 40(3): 387-417.
- Lindsay, Jean and Daniel C. O'Connell. 1995. 'How do transcribers deal with audio recordings of spoken discourse?'. *Journal of Psycholinguistic Research* 24(2) : 101-115.
- McCarthy, Michael. 1998. *Spoken Languages and Applied Linguistics*. Cambridge: Cambridge University Press.
- Maclay, Howard and Charles E. Osgood. 1959. 'Hesitation phenomena in spontaneous English speech'. *Word* 15: 19-44.
- MacWhinney, Brian. 2000. *The CHILDES Project: Tools for Analyzing Talk. Third Edition*. Mahwah, NJ: Lawrence Erlbaum.
- Mills, Martin. 1990. *Nexus. English for Advanced Learners*. Oxford: Heinemann International.
- Möhle, Dorothea. 1984. 'A comparison of the second language speech production of different native speakers'. In Hans W. Dechert, Dorothea Möhle and Manfred Raupach (eds) *Second Language Productions* (pp. 26-49). Tübingen: Gunter Narr Verlag.
- Mukherjee, Joybrato. 2005. 'The native speaker is alive and kicking: linguistic and language-pedagogical perspectives'. *Anglistik* 16(2): 7-23.
- Mukherjee, Joybrato and Jan-Marc Rohrbach. 2006. 'Rethinking applied corpus linguistics from a language-pedagogical perspective: new departures in learner corpus research'. In Bernhard Kettemann and Georg Marko (eds) *Planing, Gluing and Painting Corpora: Inside the Applied Corpus Linguist's Workshop* (pp. 205-232). Frankfurt am Main: Peter Lang.
- Osborne, John. 2006. 'Umiming and erring: analysing disfluencies in an oral learner corpus'. In *Proceedings of the 7th Conference on Teaching and Language Corpora (TaLC), Université Paris 7 – Denis Diderot, 1-4 July 2006* (pp. 108-110).
- Römer, Ute. 2004. 'Comparing real and ideal language learner input: the use of an EFL textbook corpus in corpus linguistics and language teaching'. In Guy Aston, Silvia Bernardini and Dominic Stewart (eds) *Corpora and Language Learners* (pp. 151-168). Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Romero Trillo, Jesús. 1994. 'Ahm, ehm, you call it theme? ... A thematic approach to spoken English'. *Journal of Pragmatics* 22: 495-509.
- Romero Trillo, Jesús. 2002. 'The pragmatic fossilization of discourse markers in non-native speakers of English'. *Journal of Pragmatics* 34: 769-784.
- Schiffrin, Deborah. 1987. *Discourse Markers*. Cambridge: Cambridge University Press.

- Scott, Mike. 1999. *WordSmith Tools 3*. Oxford: Oxford University Press.
- Seidlhofer, Barbara. 2001. 'Closing a conceptual gap: the case for a description of English as a lingua franca'. *International Journal of Applied Linguistics* 11(2): 133-158.
- Seidlhofer, Barbara. 2005. 'English as a lingua franca'. *ELT Journal* 59(4): 339-341.
- Stubbe, Maria and Janet Holmes. 1995. 'You know, eh and other "exasperating expressions": an analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English'. *Language & Communication* 15(1): 63-88.
- Summers, Della (ed.). 1995. *Longman Dictionary of Contemporary English. Third Edition*. Harlow: Longman Group Ltd.
- Temple, Liz. 2000. 'Second language learner speech production'. *Studia Linguistica* 54(2): 288-297.
- Watts, Richard J. 1989. 'Taking the pitcher to the "well": native speakers' perception of their use of discourse markers in conversation'. *Journal of Pragmatics* 13: 202-237.
- White, Ron. 1997. 'Back channelling, repair, pausing, and private speech'. *Applied Linguistics* 18(3): 314-344.
- Wiese, Richard. 1984. 'Language production in foreign and native languages: same or different?'. In Hans W. Dechert, Dorothea Möhle and Manfred Raupach (eds) *Second Language Productions* (pp. 11-25). Tübingen: Gunter Narr Verlag.