CHAPTER NINETEEN

DISCOURSE MARKERS IN THE ENGLISH OF FLEMISH UNIVERSITY STUDENTS[1]

LIEVEN BUYSSE,
HOGESCHOOL-UNIVERSITEIT BRUSSEL

1. Introduction

Discourse markers have received a great deal of attention in the study of pragmatics for over twenty years (cf. Schiffrin's 1987 seminal work). Researchers have focussed on a variety of markers and subjects, ranging from descriptive-functional approaches and taxonomic discussions over diachronic studies to investigations of gender differences in the use of discourse markers. All of these studies, however, have concentrated mainly on the language produced by native speakers. A largely peripheral role has so far been reserved for how non-native speakers of English cope with discourse markers.

Why should discourse markers be studied in non-native speaker discourse? Kasper and Blum-Kulka (1993) point out that "learners' distribution patterns of [pragmatic] strategies have been shown to vary from those of native speakers" (1993: 7), which becomes evident in the quality and in the range of linguistic forms, with non-native speakers having a more restricted and less complex inventory of pragmatic items. On the other hand, non-native speakers—and especially those at an intermediate stage of language acquisition—are sometimes found to "waffle" (1993: 9), i.e. they "engage in more speech activity than native speakers" (1993: 9). The explanation for this would be that language learners who are sufficiently proficient in the target language to say whatever they want to say may still feel the need to be more explicit about

[1] Thanks are due to Anne-Marie Simon-Vandenbergen for her most valuable comments to an earlier version of this contribution.

their communicative goals and about how and why they want to achieve them than native speakers.

Aijmer (2002: 3) states that incorrect use or underuse of discourse markers by non-native speakers may lead to misunderstandings, which are characterised by Hansen (1998) as "less significant but certainly far less easy to resolve than the incorrect use of a content word" (1998: 199). Terraschke (2007) believes that, although non-native-like use of pragmatic devices such as general extenders (e.g. *and so on, or something*) is unlikely to impair communication in cross-cultural interactions, speakers may well be perceived differently by their interlocutors depending on whether they are able to employ these tools appropriately. By the same token, Hellermann and Vergun (2007) are convinced that non-target-like use of discourse markers can "mark a speaker as disfluent in the target language in subtle ways" (2007: 161), which enhances that speaker's risk of "being marked as separate from the target speech community, which may then inhibit their chances for continued meaningful interaction in that speech community" (2007: 161).

The opposite is equally true: "the ability to express oneself fluently and confidently in a second language entails the use of those discourse markers that native speakers produce so effortlessly", as Sankoff *et al.* contend (1997: 214). In the same vein Hlavac (2006) notes that native-like proficiency "by definition" (2006: 1872) entails appropriate use of discourse markers, and Terraschke (2007) sees such pragmatic devices as a means for non-native speakers "to create an informal and friendly conversational atmosphere […], to relate better to their native interlocutors and ultimately to fit in better with their native peers" (2007: 158).

After a brief overview of some relevant previous literature in the area of non-native discourse marker research, I shall concentrate on six English discourse markers in a corpus of interviews with Belgian native speakers of Dutch and British native speakers of English. The main aim is to provide a broad perspective on the extent to which some of the most common English discourse markers are used by foreign language learners. The approach will, therefore, be mostly limited to a quantitative analysis, which leads to some interesting findings for which I shall provide a tentative explanation.

2. Discourse markers in non-native speech: Previous research

Fuller (2003a) compares the discourse marker use of native and non-native speakers of English in three speech contexts: an interview, a casual

conversation and an elicited narrative. Her non-native subjects have German, French or Spanish as their mother tongue, and could all be regarded as highly experienced in the language as they are all graduate students or assistant professors who have lived in the US for at least two years. The non-native speakers used all of the discourse markers analysed (*well, y'know, like, oh* and *I mean*) and they use them correctly. However, the frequency of discourse markers among non-native speakers is lower overall and they show less variation in the use of these discourse markers across speech contexts than the native speakers. The non-native group of subjects also turn out to be less homogeneous in their use of discourse markers. Only for *oh* and *well* does Fuller conclude that "non-native speakers do effectively mimic the native speakers" (Fuller 2003a: 206).

In a study of dyadic classroom interaction and of interviews Hellermann and Vergun (2007) investigate how beginning adult second language learners of English use *like, you know* and *well*. Unsurprisingly they find that discourse marker use increases together with a learner's proficiency level. Teachers do not really seem to model these discourse markers, although learners are observed to take over other discourse markers, viz. those which can be regarded as belonging to teacher talk (e.g. *alright, now, so* and *ok*). Hellermann and Vergun have also established that the discourse of learners who have had more class-external exposure to discourse markers shows a higher incidence of these pragmatic devices. They advocate a more extensive attempt by teachers to make learners aware of markers and their functions.

Romero Trillo (2002) looks into the differences between first and second or foreign language acquisition of certain discourse markers. His data has been selected from conversations of both native and non-native (Spanish) children and adults. The discourse markers focused on are the "involvement markers" (which try to enhance the interlocutor's positive face) *you know, you see, I mean* and *well* and the "operative markers" (which try to make the conversation fluent) *look* and *listen*. Native speakers are found to increase their use of the former as they grow older, whereas non-native speakers do not seem to master them appropriately and do not surpass the level of native children. Non-native speakers are, however, able to attain a native-like level of use for the "operative markers". This difference, Romero Trillo suggests, can be due to the latter's link to "the mechanics of the interaction" (2002: 782) as opposed to emphasis of "involvement markers" on the relationship between speaker and hearer. The latter are claimed to be largely unavailable to non-native speakers in an educational context.

Finally, Müller (2005) analyses 70 English dyad conversations of the Giessen-Long Beach "Chaplin" Corpus between university students in which they retell and discuss a silent movie they have been shown. About half of these are native speakers of English and another half are native speakers of German. Four discourse markers are investigated: *well, so, you know* and *like*. The German speakers turn out to employ all four markers but not all to the same extent: frequencies tend to differ as does the usage of individual functions with some being used by English native speakers and not by German native speakers and vice versa. The latter greatly underuse *so*, as Müller records only half the number of tokens as for the native speakers. Results for *you know* and *like* are even more revealing with Germans using them only a fraction of the time of native speakers of English. *Well*, on the other hand, is used twice as often by the German subjects than by the native speakers. What is most striking is not so much the massive underuse of *you know* and *like* but rather the contrasting pattern for *well* and *so* among native speakers of German. Müller (2004, 2005) offers a few tentative explanations. Most importantly, it seems that *well* can substitute for *so* in an avoidance strategy. Because these markers show a partial functional correspondence and can both be translated into German *also*, speakers may use them interchangeably, at least for some functions. However, that does not explain the preference for *well* over *so*. An analysis of German EFL textbooks reveals that *well* is unrealistically overrepresented in course materials, as opposed to *so* (*you know* and *like* are even only scarcely present). Attitude also seems to play a role: some of the German participants in the study were asked later about their feelings regarding *well* and *so*, to which they replied that the former sounds more English or that they believe native speakers use it more frequently.

To sum up, previous research has shown that non-native speakers of English are able to use the same discourse markers as native speakers and to do so correctly. However, they tend to employ them less frequently, and much depends on the individual's level of proficiency and exposure to the language, as well as on the nature of the discourse marker.

3. The corpus

For the current study I have investigated a corpus of informal interviews with 40 Belgian native speakers of Dutch who are in 2nd or 3rd bachelor (age 19-26) at university. Half of the interviewees study English linguistics as a major subject at university (henceforth EL) and the other half are students of commercial sciences at a university college (henceforth CS). Hence, we can distinguish between two types of acquisition: that of

English *linguistics* students (those who study the language and its system, supposedly on a daily basis) and that of English *language* students (those who study the language for instrumental reasons or simply because it is a minor compulsory part of their curriculum). The former are often described as advanced learners of English on the basis of external criteria, viz. the teaching level and the number of hours they have English classes in an average week, both of which are relatively high for students majoring in English at university (see e.g. De Cock 2007, Gilquin 2008, Mukherjee 2009). If these criteria are applied to students of commercial sciences, these learners would have to be considered upper-intermediate learners of English: following their English classes in secondary school their curriculum in tertiary education contains one hour of English classes per week for three years. In the Dutch-speaking part of Belgium pupils at secondary school get 2 to 3 hours of English per week for five years. In class, both in secondary and tertiary education, learners are most often exposed to British English. In the media, however, exposure to American English is inescapable in the many American programmes that are broadcast on television; all of these are subtitled and not dubbed.

The interviews were conducted by members of the teaching staff according to the format of the Louvain International Database of Spoken English Interlanguage (LINDSEI)[2]. They each last about fifteen minutes (with an average word count of 1,587) and follow the same pattern: the interviewee is asked to talk about a travel experience, hobbies, a book, etc., which leads to a conversation with the interviewer; every interview ends with a short picture-based story-telling activity. All interviewees volunteered to participate. Over two thirds of the EL corpus (14 out of 20) are made up of female speech, whereas the CS corpus is almost balanced for gender (11 female interviewees out of 20). These gender ratios reflect those in the student population they represent.

The LINDSEI project also has a native speaker test corpus (LOCNEC, Louvain Corpus of Native English Conversation)[3], from which I selected 20 interviews with the same gender ratio as CS. In total my native speaker corpus (henceforth NS) comprises of 52,036 words of interviewee speech (2,602 on average). All native speakers were British students majoring in English language and/or linguistics at the University of Lancaster.

---

[2] For further details see:
<http://cecl.fltr.ucl.ac.be/CeclProjects/Lindsei/lindsei.htm>.

[3] I am much indebted to the Centre for English Corpus Linguistics at the *Université Catholique de Louvain-la-Neuve* for granting me access to this corpus.

4. Results

4.1. A selection of discourse markers

For this study I have selected some of the most commonly studied (and probably also the most commonly encountered) discourse markers[4]. For each item all non-discourse marker uses were filtered out, which involved some qualitative choices based on relevant literature for each marker. In this section some of their functions are illustrated with examples from the non-native speaker corpus.

4.1.1. *So*

As example (1) demonstrates, *so* is a highly versatile discourse marker, which can take on a number of different guises in one and the same relatively short stretch of speech.[5]

(1) <A> did you see a match <\A>
<B> no we just . passed by by bus **so** <sniffs> I wanted to see the[i:] other one of Celtic but we didn't pass that one erm . he didn't like them huh . **so** he asked me do do you like Cel= eh Glasgow Rangers I said well . I just guessed because in . in Glasgow you're either a Celtic fan or a Rangers fan **so** I said okay I'll go for no <laughs> he said okay I I don't like them either **so** <blows> . good guess erm .. <\B>
(EL20)

It has a basic inferential meaning (cf. among others Schiffrin 1987; Blakemore 1988; Fraser 2006) but is equally able to perform a varied set of functions. *So* has preoccupied many researchers, most of whom have focused either on *so*'s capacity as the prototypical inferential or resultative discourse marker (e.g. Blakemore 1988, 2004; Fraser 1990, 1996, 2006) or on one specific use of *so* (e.g. Johnson 2002 on *so*-prefaced questions; Raymond 2004 on "stand-alone *so*"; Passot 2007 on *so*'s role in a specific structure with *because*). Müller (2005: 68ff.), on the other hand, attempts to provide a comprehensive picture of *so* as it appears in her corpus, much

[4] I do not wish to engage in a taxonomic debate at present. The term *discourse marker* was selected merely because it is probably the most widespread of a host of competing terms (e.g. *pragmatic markers, discourse particles*). The articles in Fischer (2006) provide a nice overview of the most interesting perspectives on the taxonomy of the field.

[5] In all examples <A> is the interviewer and <B> is the interviewee. The transcription conventions are those used for the LINDSEI project.

of which seems to echo Schiffrin's (1987) early findings. She lists nine functions, the most important of which are: marking a result or consequence, taking the discourse back to the main topic after a digression, marking a discourse boundary, and marking a potential turn-transition.

Like Müller (2005) I have excluded some tokens of *so* because they do not perform a discourse marker function, viz. *so* as an adverb of degree or manner, when it expresses purpose, in fixed expressions (e.g. *and so on*) and as a pro-form (e.g. *I think so*). It should be noted, though, that Tagliamonte (2008) does regard some instances of *so* as a degree adverb as discourse marker uses. What she calls "GenX*so*" (2008: 390) – as in *She's so not ready for marriage* – has not been included in the discourse marker category for the present study, because it does not have any relationship with the inferential and discourse-organisational instances mentioned above.

If we calculate the percentage of tokens of *so* that perform a discourse marker function, we can establish its "Index of Pragmatic Use"[6] or "IPU" (Romero Trillo 2002: 776), which clarifies the marker's position vis-à-vis its non-discourse marker uses. The IPU for *so* amounts to 79 for the non-native speakers and the native speakers alike. Thus, in both corpora about four fifths of all tokens of *so* perform a discourse marker function.

4.1.2. *Well*

Excerpt (2) exemplifies two of *well*'s main uses (cf. Müller 2005: 107 ff.): marking an answer to a question in a qualifying way and self-correction or rephrasing.

(2) <A> uhu uhu . and was it realistic <\A>
<B> **well** I thought it was very realistic . I: couldn't imagine that it was eh . could be like that because the film . **well** it took about three hours so it was a very long film . but I had a= a large bag of popcorn <begin laughter> but <end laughter> I didn't eat so much of it because . the film it just . grabs you that you don't realise that the time is going so fast <\B>
(CS01)

Smith and Jucker (2000) claim that *well* acts as a downgrader in the negotiation of common ground in conversation. Aijmer and Simon-Vandenbergen (2003) propose the core meaning of "positive appraisal" for

[6] Romero Trillo's (2002) "Index of Pragmatic Use" is similar, if not identical, to Stenström's (1990) "D-ratio" (1990: 161).

well, with a core function of "signalling awareness of heterogeneity, and more specifically counterexpectation" (2003: 1130).

Following Müller (2005: 107) instances of *well* as the adverb form of *good* and those in phrases such as *as well* were excluded from this study. The IPU for *well* among the non-native speakers is 88 and for the native speakers 70.

4.1.3. *You know*

Jucker and Smith (1998) suggest that *you know* "invites the addressee to complete the argument by drawing the appropriate inferences" (1998: 196), a view later endorsed by Fox Tree and Schrock (2002). In example (3), for instance, *you know* conveys the message: "I don't need to tell you this because as a student I am obviously on a tight budget."

(3) <A> have you studied abroad . anywhere <\A>
<B> no . no I'd like to maybe the journalism class would be in England so .. but I don't know yet because now I ha= it turns out that for my . em thesis . or I don't know what it's called in English . I have to buy a computer and a D V D and . **you know** my budget is like . ooh <laughs> <\B>
(EL07)

Additionally, it can also be used to mark a false start or initiate a repair sequence, introduce a quotation, etc. (cf. Fox Tree and Schrock 2002; Müller 2005). Again all propositional uses of *you know* were discarded. *You know*'s IPU was calculated at 84 for the non-native speakers and at 92 for the native speakers.

4.1.4. *Like*

Discourse marker *like* is considered to mark "new information or focus" (Dailey-O'Cain 2000: 61) with the primary function of indicating "approximation or looseness of meaning" (Fuller 2003b: 367). A distinction is in order between this "focuser" use of *like* and "quotative" *like*, which usually occurs in the structure *be+like* (see e.g. Macaulay 2001; Tagliamonte and D'Arcy 2004; Fox Tree 2006). The first token of *like* in example (4) is an instance of its focusing function, whereas the second illustrates the quotative use.

(4) <B> erm .. I went . to: New York with my parents .. erm it was the first time I went to America . and . it was . in fact a very . erm . strange

> experience in the way that I . imagined Americans were . **like** very er . enthusiast er people and . when I arrived there they were indeed very enthusiast and very . erm . erm overreacting erm for example in the hotel we stayed . er in the morning the san= sandwiches er . there weren't enough sandwiches for all the people and they were **like** I'm gonna sue you . [for a sandwich er <\B>
> (CS02)

Fuller (2003a,b), Dailey-O'Cain (2000) and Müller (2005) deny discourse marker status to quotative *like* because it cannot be left out without rendering the utterance ungrammatical or unintelligible, which is one of the features most often ascribed to discourse markers. It should be noted that there is no consensus in this respect. Romaine and Lange (1991), for instance, distinguish between *like* as a "discourse marker" and as a "discourse marker with quotative function" (1991: 244), and Jucker and Smith (1998: 183) and Andersen (1998:156), include quotative *like* as one of many discourse marker functions of the word. Because of this lack of agreement I will ignore the quotative uses of *like* in the current study[7], as well as its prepositional, conjunctional and verbal uses.

There is quite a discrepancy in the IPU for *like*: only 12 percent of all tokens of *like* show a discourse marker function in non-native speech, whereas that number amounts to 46 for the native speakers.

4.1.5. *Kind of, kinda, sort of*

Kind of, kinda and *sort of* are taken together here. Aijmer (2002: 207-208) points out that they "seem to have the same meaning" and that the only notable difference between them is the extent to which they are used: in British English she observes *sort of* as being much more frequent than *kind of*.[8] These discourse markers can have two types of function: either they act as adjusters, indicating that the word they introduce is not sufficiently adequate to convey the intended message, or they have an affective role in that they downtone or hedge the message (Aijmer 2002: 191ff).

In example (5) both readings might be in play. The speaker has been trying to justify a trip to poor regions in Kenya. He might feel that stating

---

[7] For the sake of completeness: quotative *like* occurs with an overall average NNS frequency of 0.06 tokens per 100 words (CS 0.05 and EL 0.06) and an NS frequency of 0.03.

[8] The native speaker data in the present corpus seem to confirm this tendency for British English: *sort of* is used in more than four out of five instances.

that he "helped" the locals merely by travelling to their region would be an unjustified representation of reality, and hence wishes to downtone that statement. *Kind of* indicates that "helped" does not adequately describe the situation.

(5) <B> actually we **kind of** helped the people it's maybe . erm . how should I say . helps you . erm .. calm your conscience . in a way .. <\B> (EL08)

If *kind of, kinda* and *sort of* mean "a type of", they are left out as non-discourse marker uses. Their IPU amounts to 51 in the non-native speaker corpus and 83 in the native speaker corpus.

4.1.6. *I mean*

Fox Tree and Schrock (2002) describe *I mean* as a means to indicate "upcoming adjustments" (2002: 741). For example, *I mean* is appropriate when speakers want to be more careful or precise than in their prior utterance, or it may introduce a justification (as in example (6)).

(6) <A> well some people think that . you lose some social contact within your family . if you have a dish washer .. [because otherwise <\A>
<B> [yes . that's true yeah you but you lose you lose social contacts by television as well <\B>
<A> uhu [so <\A>
<B> [so <X> <\B>
<A> throw out television <\A>
<B> yeah I wouldn't do it but . **I mean** if <\B>
<A> uhu <\A>
<B> you come with that argument <\B>
(CS05)

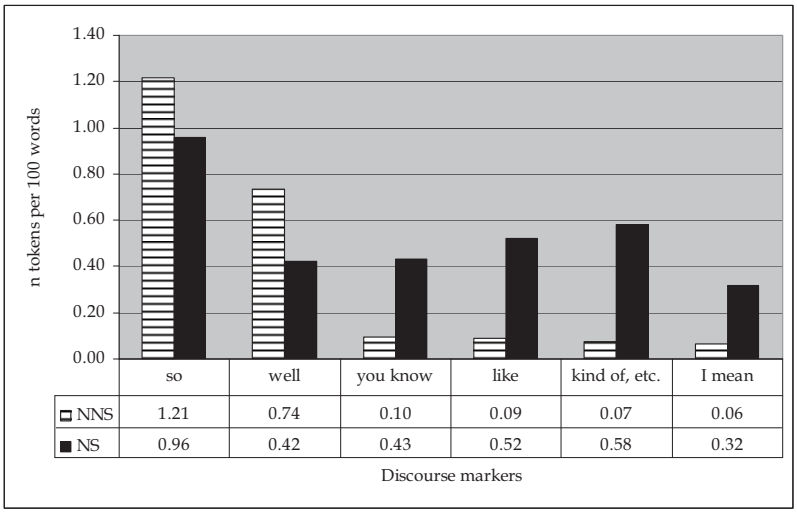
I mean is often considered together with *you know* (e.g. Schiffrin 1987; Erman 1987; Fox Tree and Schrock 2002). Given its focus on the hearer's understanding of the speaker's speech *you know* is expected to be more frequent in dialogues than *I mean* (Fox Tree and Schrock 2002: 744).

The IPU for *I mean* is almost identical for native and non-native speakers: 96 and 98, respectively.

4.2 Discourse marker frequencies

Graph 1 illustrates the frequency of the six discourse markers in the native and the non-native speaker corpora. In the non-native speaker

corpus as a whole there is a striking discrepancy between the extensive presence of *so* and *well*, with a respective incidence of 1.21 and 0.74 tokens per 100 words, and the marginal use of the other markers, none of which reaches an average frequency higher than 0.10 tokens per 100 words. This is in sharp contrast to the native speaker corpus, where *so* equally ranks highest, albeit with an average of 0.96, but where all other markers under consideration are used frequently as well. A log-likelihood[9] test reveals a strong statistical difference (at $p < 0.0001$) between the native and the non-native speaker frequencies for all discourse markers.



	so	well	you know	like	kind of, etc.	I mean
NNS	1.21	0.74	0.10	0.09	0.07	0.06
NS	0.96	0.42	0.43	0.52	0.58	0.32

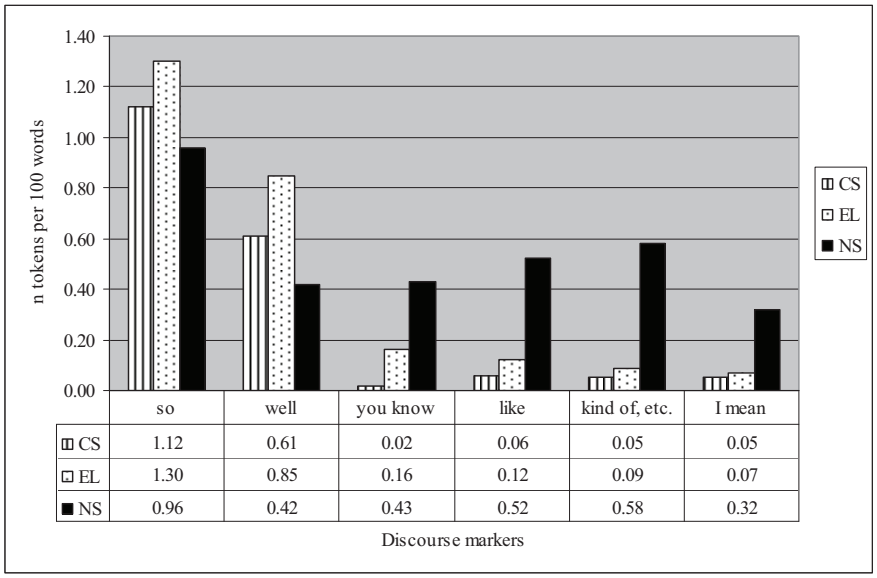
Graph 1: Native versus non-native discourse marker use

So is by far and across the board the most popular discourse marker in this study, in both the native and the non-native corpora. It is followed by

[9] The log-likelihood ratio indicates the degree of certainty of an observed difference representing a true difference or being due to chance. A 100 per cent guarantee is unattainable but the level of significance (represented by the letter *p*) should never exceed 0.05 ($p<0.05$) in corpus linguistics, which means that we can be more than 95 per cent confident that the difference is real (McEnery *et al.* 2006: 55). Log-likelihood is especially useful if the expected frequency is lower than 5, in which case other tests, such as chi-square, are unreliable (Rayson and Garside 2000).

well in the non-native speaker corpus and by *kind of/kinda/sort of* in the native speaker corpus.

Graph 2 splits up the numbers for the non-native speaker corpus into those for the language students (CS) and those for the linguistics students (EL). For *so* and *like* the difference is statistically significant at $p < 0.05$, for *well* at $p < 0.01$ and for *you know* at $p < 0.001$; according to the log-likelihood test the differences for *kind of/kinda/sort of*[10] and for *I mean* are not statistically significant.



	so	well	you know	like	kind of, etc.	I mean
CS	1.12	0.61	0.02	0.06	0.05	0.05
EL	1.30	0.85	0.16	0.12	0.09	0.07
NS	0.96	0.42	0.43	0.52	0.58	0.32

Graph 2: Different types of learners compared: English language (CS), English linguistics (EL) and L1 (NS) students

Overall it should be clear that both CS and EL interviewees make rather extensive use of *so* and to a lesser extent of *well*, but whereas in EL the other discourse markers still occasionally surface, these are virtually absent in CS. There is definitely a huge difference in incidence of *you know* between CS and EL, it being almost inexistent in the former whereas the latter interviewees make moderate use of the marker. Interestingly, that

[10] For the sake of completeness, these are the individual numbers: *kind of/kinda* NNS 0.05 (CS 0.03, EL 0.06) and NS 0.08; *sort of* NNS 0.03 (CS 0.02, EL 0.03) and NS 0.50.

difference does not apply to *I mean,* which is even more frequent in CS than *you know*.

It should be noted, however, that these numbers are small for the non-native speakers in general. This makes the results particularly vulnerable to distortion by individual interviewees' performances. For instance, in both CS and EL two interviewees together account for more than half of all tokens of *like* in their respective subcorpora. Similarly, in both of these subcorpora one student produces over half of all *I mean* tokens. This phenomenon cannot be detected for discourse markers with a higher frequency in the non-native speaker subcorpora (viz. *so* and *well*) nor for the native speaker corpus.

Compared to the native speaker interviewees, EL subjects more heavily overuse *so* and *well* than the CS participants, while their numbers for the other discourse markers more closely resemble native speaker frequencies, although they still far from approximate them.

5. Towards an explanation

5.1. Two types of discourse markers

The most striking observation in the quantitative analysis above is that compared with the native speaker corpus the non-native speakers significantly overuse *so* and *well* whereas they significantly underuse most other discourse markers. How can this be explained? The answer to that question probably lies in the different nature of these items. Discourse markers are often regarded as functioning on one of two levels. The first of these is the structural or textual level (Aijmer 2002: 40), on which markers have a clear role to play in the organisation of the discourse. For instance, they mark a transition (between topics, conversations, etc.), introduce or close a digression, act as a self-correction device, indicate agreement, etc. As Lenk (1998) puts it: they "signal for the hearer how the speaker intends the present contribution to be related to preceding and/or following parts of the discourse" (1998: 52). Another level on which discourse markers can function is the interpersonal or phatic (Aijmer 2002: 48; cf. Bazzanella's notion of "phatic connectives", 1990). They express "attitudes, feelings and evaluations" (Aijmer 2002: 39) of speakers towards hearers or refer to epistemological concepts, e.g. they are hedges or politeness devices.

All markers perform a variety of functions, and one discourse marker may contribute to the structural domain in one instance and to the interpersonal in another. Hence, the relationship between these markers is

best represented by a cline (rather than a clear-cut polarisation) from more to less textually used discourse markers. On one side *so* and *well* would then be among the most typical textual markers, and *like* and *sort of* would be on the other, with *you know* and *I mean* situated somewhere in the middle. A further difference is of a more formal nature: *so* and *well* occur clause-initially and indicate a connection between the upcoming information and previous discourse. In contrast, *you know* and *I mean* are positionally much more versatile and their function is primarily involved in establishing or enhancing interpersonal rapport. *Like* and *kind of/kinda/sort of*, finally, usually take one phrase in their scope, which they "hedge".

5.2. Stigmatisation

Interpersonal discourse markers – which is a convenient short-hand for "interpersonally used discourse markers" – occasionally suffer from stigmatisation. Although few would agree with the final part of their statement today, O'Donnell and Todd (1991) refer to *you see, you know* and *I mean* as "phrases which occur with varying frequency in informal speech, or with unskilful speakers" (1991: 69). In their volume on "bad language" Andersson and Trudgill (1990) deem it necessary to devote a separate chapter to defending "small words" like *sort of* and *y'know* because they "belong to the stigmatized elements of English grammar" (1990: 94). They denounce the argument that these expressions "are unnecessary, that they are used simply as fillers – words that jump out of the mouth while figuring out what clever things to say next" (1990: 95). Schiffrin observes a similar tendency for *y'know* and *I mean*:

> both are markers which are socially evaluated and negatively sanctioned. […] [U]se of both *y'know* and *I mean* could run counter to standard beliefs about the appropriate division of labor in conversation: use of *y'know* can be interpreted as overdependence on the hearer, and use of *I mean* can be interpreted as overinvolvement with the self. And it could be for these reasons that these markers are stigmatized. (1987: 310-311).

Underhill (1988) describes the discourse marker uses of *like* as "intrusive" (1988: 234), and reports that it is considered "entirely ungrammatical in standard English and makes sentences seem disjointed to many listeners" (1988: 234). More recently, Andersen (1998) reports that *like* is "commonly accused of being redundant and without meaning" (1998: 150), whereas Fox Tree (2007) claims that "*[u]m/uh, you know,* and *like* are often treated as a trio to avoid at all costs" (2007: 298) and

Miskovic-Lukovic (2009) points out that *kind of* and *sort of* are "typical of informal style, and nonstandard language varieties" (2009: 602) and that "more often than not, these forms have borne the stigma of bad usage through prescriptivism and bias" (2009: 602).

5.3. Setting

The interview setting in which the conversations were conducted can be characterised as that of a formal conversation (or an informal interview). Östman (1981) describes *you know* as "a very salient stylistic marker" (1981: 20), which aims at "attaining a Camaraderie relation between the speaker and the addressee" (1981: 19). Hence, Jucker and Smith (1998) find that *like* and *you know* are used more often among friends than among strangers. Macaulay (2002) confirms these findings for *you know*. On the other hand, Fuller (2003a) concludes that the native speakers in her study use *you know* and *like* more often in interviews than in conversations, because interviewees frequently have to explain their actions and opinions which is bound to lead to mitigation performed by interpersonal discourse markers. Moreover, she believes that since the interviewer is a relative stranger the interviewee has to make more intense an effort to establish common ground, "and thus there is increased need to qualify utterances and check the information status of the hearer" (2003a: 205). Probably the truth lies somewhere in the middle. As Östman (1981: 19-20) reports, the need for markers like *you know* diminishes as the degree of rapport increases. For instance, in his study of dinner table conversations he finds that a conversation with relatives tends to contain fewer instances of *you know* than one with friends, but the latter would have more than one with strangers.

As the current study does not include any speech contexts other than the interview, we cannot compare discourse marker use in different settings. However, the native speaker frequencies would suggest that discourse markers like *you know, like* and *kind of/kinda/sort of* are used rather extensively by native speakers in an interview context. The opposite holds for the non-native speakers. The explanation for this might be that the foreign language learners in the study were aware of the informality typically associated with these interpersonal discourse markers, yet not – or at least not sufficiently – of their capacity to enhance common ground in a conversation between relative strangers.

In a conversation with native speakers or with English teachers language learners will try to speak the target language "properly". For many that still involves avoiding all signs of informality, and as Aijmer

(2002) notes about *sort of*, discourse markers are "symptomatic of informal speech" (2002: 190). If foreign language learners were made aware, however, of the beneficial role such pragmatic devices can play, the results might look differently. Romero Trillo (2002) believes that "the development of pragmatic competence demands a (pseudo-)natural foreign language context that is often impossible to produce in formal education" (2002: 770). Fung and Carter (2007) acknowledge that the linguistic input foreign language learners get is often unnatural and that "the traditional grammar-centred pedagogic focus […] has been geared towards the literal or propositional (semantic) meanings of words rather than their pragmatic use in spoken language" (2007: 433). Yet, Bardovi-Harlig and Griffin (2005) are still convinced that pragmatic awareness and competence are developed gradually and specific classroom activity is needed to stimulate the process. Watts (2000) contends, however, that the aim of foreign language education is still strongly biased towards written language and practising writing skills, especially at more advanced stages of the acquisition process such as in higher education. By the same token, Barron (2000) claims that "the interpersonal function of language […] is often sacrificed in advanced language classes, in particular in favour of the more prestigious referential function" (2000: 16). In other words, the emphasis is on getting a message across rather than on establishing rapport with an interlocutor. This educational setting makes it difficult for foreign language learners to familiarise themselves with pragmatic devices such as interpersonal discourse markers.[11]

5.4. L1 influence?

One could argue that these non-native speakers of English should already be familiar with similar devices in their mother tongue. Indeed, (Belgian) Dutch has equivalents which even closely resemble their English counterparts in that they can be considered almost literal translations: *weet je (wel)* for *you know*, *(ge)lijk* for *like* (*zoiets hebben van* for quotative *be like*) and *ik bedoel* for *I mean*. For want of relevant research on the matter

[11] Interestingly, Nikula (2002) notes that scarce use of what she calls "modifiers" – linguistic items like discourse markers – "might well be an important pragmatic principle of classroom interaction rather than an interactional problem" (2002: 463). That would signify that there is little need to express interpersonal relations in a classroom setting and hence students are hardly ever exposed to markers of such relations, be it in their mother tongue or in a foreign language. The language classroom would then become utterly unsuited as a locus to develop pragmatic competence.

informal observations of students with the same profile as the participants of the study show that they use these markers with a high frequency in everyday conversation. However, that does not mean that these students are aware of that. Andersson and Trudgill (1990) point out that discourse markers "may pass unnoticed for a long period, but once they are discovered (or, rather, brought to our awareness) we hear them all the time" (1990: 94), which may lead to stigmatisation. Just like in English, these interpersonal discourse markers are often viewed negatively in Dutch and are on occasion publicly scoffed at. Thus, if native speakers of Dutch are aware of how these expressions are used in their mother tongue, they must also be familiar with how the public 'ear' tends to assess them.

In an interview with a relative stranger in a foreign language, in which the interviewer can be regarded as ranking higher than the interviewee (after all, the former is a member of the teaching staff whereas the latter is a student), language learners may prefer to remain on safe ground and hence push the interpersonal aspects of the conversation into a secondary role.

5.5. The opposite picture

Contrary to interpersonal discourse markers, textual discourse markers like *so* and *well* are vital to establishing coherence in the conversation, and are therefore viewed as useful aids to attaining the main goal of the interaction, viz. to get a message across in a coherent fashion. Language learners do not feel reluctant about using them, because these linguistic items are rather uncontroversial.

Moreover, Lorenz (1999) and Anping (2002) have found that the textual discourse marker *so* is highly overused by language learners in written texts, suggesting that they certainly do not associate it with informality. As Hellermann and Vergun (2007) demonstrate, language learners, even in the early stages of the acquisition process, seem to experience few problems mimicking textual discourse markers, because these are typical of teacher talk. Hence, students are constantly exposed to a limited set of textual discourse markers, which offers them a small inventory of such devices, which they then use time and again. In other words, they want to convey their message in a coherent way but (1) their paradigm of lexical choices is still limited, as a consequence of which they always resort to the same strategies; and/or (2) they want to make absolutely certain that their interlocutor interprets the discourse appropriately, to which end they supply an abundance of clues.

Furthermore, the Dutch equivalents of *so* (*dus*) and *well* (*wel, nou (ja), nu (ja)*) are also extremely frequent in spoken language without their being stigmatised. Because of this positive (or at least neutral) attitude towards *well* language learners may resort more to this discourse marker to perform functions which it shares with interpersonal discourse markers. For instance, both *well* and *I mean* can be used as a self-correction device.

5.6. Types of learners

How can the differences between the two types of foreign language learners in the present corpus be explained? It does not come as much of a surprise that the English linguistics students' behaviour conforms better to that of the native speakers for the interpersonal discourse markers. After all, compared to the English language students, they are exposed more intensively to the target language, are made more sensitive to the workings of the language system, and are more familiar with a wider variety of registers. It should be stressed, however, that the numbers for these interpersonal discourse markers are too small to make any conclusive statements.

What is more striking, however, is that the linguistics students divert from the native speaker use of textual discourse markers: both types of learners considerably overuse these markers but the linguistics students do so to a significantly larger extent than the language students. Further in-depth research which qualitatively scrutinises all instances of these discourse markers in the corpus as a whole will have to shed light on the reasons behind this phenomenon.

6. Conclusion

Traditionally discourse markers have been studied in native speaker environments. Only recently has some research surfaced on how non-native speakers deal with such pragmatic devices. In this contribution I have tried to provide a quantitative overview of how *so*, *well*, *you know*, *like*, *kind of/kinda/sort of* and *I mean* occur in English interviews with Dutch-speaking undergraduates majoring either in English linguistics or in commercial sciences. After a qualitative analysis setting off the discourse marker functions of six linguistic items from their non-discourse marker uses, a quantitative analysis has revealed some differences between the frequency in which these non-native speakers of English employ these markers and native speaker practice. The discourse markers with a predominantly interpersonal use (*you know, like, kind of/kinda/sort of* and

I mean) were hardly used by the foreign language learners, in contrast to their native speaker peers. Interpersonal discourse markers tend to be stigmatised in their target language community and are associated with informality. Therefore, the interviewees may have been reluctant to use them in an interview setting due to a lack of pragmatic and register awareness. Most foreign language learners do not get the opportunity to acquire the language in an educational environment that leaves much room for pragmatic language components: the emphasis tends to be on writing skills, and students are taught how to steer clear of informal elements in an effort to keep their language as "neutral" and inoffensive as possible. The textual discourse markers *so* and *well*, on the other hand, are massively overused by the non-native speakers in this study. These allow them to convey their message coherently, and are constantly exemplified in the foreign language classroom as a part of teacher talk. The learners' mother tongue may also play a role: they know from their L1 that interpersonal discourse markers are sometimes to be avoided, whereas the use of textual discourse markers is on the whole unproblematic. Across the board, English linguistics students use discourse markers more often than students of commercial sciences. Much work remains to be done. A thorough investigation of all of the (sub)functions of *so* and *well* falls beyond the scope of the present paper, but would undoubtedly clarify in greater detail how and why the non-native speakers overuse these markers to such an extent. A longitudinal study, in its turn, might reveal whether the virtual absence of interpersonal discourse markers in the interviews with non-native speakers who are in a relatively late stage of their acquisition process, is a case of Romero Trillo's (2002) concept of "pragmatic fossilization" (2002: 770). In other words, have these language learners reached a point where they are unable to get a firm grasp of certain pragmatic elements of the target language? Before this question can be answered more information as to the nature of the different stages of the foreign language learning process is required. One hypothesis would be that the acquisition of native-like pragmatic competence ends where the language learner feels that the difference is no longer salient or that they have reached their final goal, which need not be native speaker performance.

References

Aijmer, Karin. *English Discourse Particles. Evidence from a Corpus.* Amsterdam/Philadelphia: John Benjamins, 2002.

Aijmer, Karin and Anne-Marie Simon-Vandenbergen. "The discourse particle *well* and its equivalents in Swedish and Dutch." *Linguistics* 41.6 (2003): 1123-1161.

Andersen, Gisle. "The pragmatic marker *like* from a relevance-theoretic perspective." In *Discourse Markers. Descriptions and Theory*, edited by Andreas H. Jucker and Yael Ziv, 147-170. Amsterdam/Philadelphia: John Benjamins, 1998.

Andersson, Lars-Gunnar and Peter Trudgill. *Bad Language*. London: Penguin Books, 1990.

Anping, He. On the discourse marker *so*. In *New Frontiers of Corpus Research. Papers from the Twenty First International Conference on English Language Research on Computerized Corpora Sydney 2000*, edited by Pam Peters, Peter Collins and Adam Smith, 41-52. Amsterdam/New York: Rodopi, 2002.

Bardovi-Harlig, Kathleen and Robert Griffin. "L2 pragmatic awareness: Evidence from the ESL classroom." *System* 33 (2005): 401-415.

Barron, Anne. "Acquiring 'different strokes'. A longitudinal study of the development of L2 pragmatic competence." *German as a Foreign Language* 2 (2000): 1-29.

Bazzanella, Carla. "Phatic connectives as interactional cues in contemporary spoken Italian." *Journal of Pragmatics* 14.4 (1990): 629-647.

Blakemore, Diane. "'So' as a constraint on relevance." In *Mental Representations. The Interface between Language and Reality*, edited by Ruth M. Kempson, 183-195. Cambridge: Cambridge University Press, 1988.

—. "Discourse markers." In *The Handbook of Pragmatics*, edited by Laurence R. Horn and Gregory Ward, 221-240. Oxford: Blackwell Publishing, 2004.

Dailey-O'Cain, Jennifer. "The sociolinguistic distribution of and attitudes toward focuser *like* and quotative *like*." *Journal of Sociolinguistics* 4.1 (2000): 60-80.

De Cock, Sylvie. "Routinized building blocks in native speaker and learner speech: clausal sequences in the spotlight." In *Spoken Corpora in Applied Linguistics*, edited by Mari Carmen Campoy and María José Luzón, 217-233. Bern: Peter Lang, 2007.

Erman, Britt. *Pragmatic expressions in English. A study of* you know, you see *and* I mean *in face-to-face conversation*. Stockholm: Almqvist & Wiksell, 1987.

Fischer, Kerstin (ed.). *Approaches to discourse particles*. Amsterdam: Elsevier, 2006.

Fox Tree, Jean E. "Placing *like* in telling stories." *Discourse Studies* 8.6 (2006): 723-743.

—. "Folk notions of *um* and *uh*, *you know*, and *like*." *Text & Talk* 27.3 (2007): 297-314.

Fox Tree, Jean E. and Joseph C. Schrock. "Basic Meanings of *you know* and *I mean*." *Journal of Pragmatics* 34.6 (2002): 727-747.

Fraser, Bruce. "An approach to discourse markers." *Journal of Pragmatics* 14.3 (1990): 383-395.

—. "Pragmatic markers." *Pragmatics* 6.2 (1996): 167-190.

—. "Towards a theory of discourse markers." In *Approaches to discourse particles*, edited by Kerstin Fischer, 189-204. Amsterdam: Elsevier, 2006.

Fuller, Janet M. "Discourse marker use across speech contexts: A comparison of native and non-native speaker performance." *Multilingua* 22 (2003a): 185-208.

—. "Use of the discourse marker *like* in interviews." *Journal of Sociolinguistics* 7.3 (2003b): 365-377.

Fung, Loretta and Ronald Carter. "Discourse markers and spoken English: Native and learner use in pedagogic settings." *Applied Linguistics* 28.3 (2007): 410-439.

Gilquin, Gaëtanelle. "Hesitation markers among EFL learners: Pragmatic deficiency or difference?" In *Pragmatics and Corpus Linguistics: A Mutualistic Entente*, edited by Jesús Romero-Trillo, 119-149. Berlin: Walter de Gruyter, 2008.

Hansen, Maj-Britt Mosegaard. *The Function of Discourse Particles. A Study with Special Reference to Spoken Standard French*. Amsterdam/Philadelphia: John Benjamins, 1998.

Hellermann, John and Andrea Vergun. "Language which is not taught: The discourse marker use of beginning adult learners of English." *Journal of Pragmatics* 39.1 (2007): 157-179.

Hlavac, Jim. "Bilingual discourse markers: Evidence from Croatian-English code-switching." *Journal of Pragmatics* 38.11 (2006): 1870-1900.

Johnson, Alison. "*So…?*: Pragmatic implications of *so*-prefaced questions in formal police interviews." In *Language in the Legal Process*, edited by Janet Cotterill, 91-110. Hampshire: Palgrave MacMillan, 2002.

Jucker, Andreas H. and Sara W. Smith. "*And people just you know like 'wow'*: Discourse markers as negotiating strategies." In *Discourse Markers. Descriptions and Theory*, edited by Andreas H. Jucker and Yael Ziv, 171-201. Amsterdam/Philadelphia: John Benjamins, 1998.

Kasper, Gabriele and Shoshana Blum-Kulka. "Interlanguage Pragmatics: Introduction." In *Interlanguage Pragmatics*, edited by Gabriele Kasper and Shoshana Blum-Kulka, 3-17. Oxford: Oxford University Press, 1993.

Lenk, Uta. *Marking Discourse Coherence. Functions of Discourse Markers in Spoken English*. Tübingen: Gunter Narr Verlag, 1998.

Lorenz, Gunter. "Learning to cohere: Causal links in native vs. non-native argumentative writing." In *Coherence in Spoken and Written Discourse: How to Create It and How to Describe It*, edited by Wolfram Bublitz, Uta Lenk and Eija Ventola, 55-75. Amsterdam/Philadelphia: John Benjamins, 1999.

Macaulay, Ronald. "You know, it depends." *Journal of Pragmatics* 34.6 (2002): 749-767.

McEnery, Tony, Richard Xiao and Yukio Tono. *Corpus-Based Language Studies*. London/New York: Routledge, 2006.

Miskovic-Lukovic, Mirjana. "*Is there a chance that I might kinda sort of take you out to dinner?*: The role of the pragmatic particles *kind of* and *sort of* in utterance interpretation." *Journal of Pragmatics* 41.3 (2009): 602-625.

Mukherjee, Joybrato. The grammar of conversation in advanced spoken learner English: Learner corpus data and language-pedagogical implications. In *Corpora and Language Teaching*, edited by Karin Aijmer, 203-230. Amsterdam/Philadelphia: John Benjamins, 2009.

Müller, Simone. "*Well you know that type of person*: Functions of *well* in the speech of American and German students." *Journal of Pragmatics* 36.6 (2004): 1157-1182.

—. *Discourse Markers in Native and Non-native English Discourse*. Amsterdam/Philadelphia: John Benjamins, 2005.

Nikula, Tarja. "Teacher talk reflecting pragmatic awareness: A look at EFL and content-based classroom settings." *Pragmatics* 12.4 (2002): 447-467.

O'Donnell, W. R. and Loreto Todd. *Variety in Contemporary English*. Second edition. London: Harper Collins Academic, 1991.

Östman, Jan-Ola. You Know*: A Discourse Functional Approach*. Amsterdam: John Benjamins, 1981.

Passot, Frédérique. "*A because B so A'*. Circularity and discourse progression in conversational English." In *Connectives as Discourse*

Landmarks, edited by Agnès Celle and Ruth Huart, 117-134. Amsterdam/Philadelphia: John Benjamins, 2007.

Raymond, Geoffrey. "Prompting action: The stand-alone "so" in ordinary conversation." *Research on Language and Social Interaction* 37.2 (2004): 185-218.

Rayson, Paul and Roger Garside. "Comparing corpora using frequency profiling." In *Proceedings of the Workshop on Comparing Corpora. Held in Conjunction with the 38th Annual Meeting of the Association for Computational Linguistics*. *7 October 2000, Hong Kong University of Science and Technology, Hong Kong*, edited by Adam Kilgarriff and Tony Berber Sardinha, 1-6. Hong Kong: ACL, 2000.

Romaine, Suzanne and Deborah Lange. "The use of *like* as a marker of reported speech and thought: A case of grammaticalization in progress." *American Speech* 66.3 (1991): 227-279.

Romero Trillo, Jesús. "The pragmatic fossilization of discourse markers in non-native speakers of English." *Journal of Pragmatics* 34.6 (2002): 769-784.

Sankoff, Gillian, Pierette Thibault, Naomi Nagy, Hélène Blondeau, Marie-Odille Fonollosa and Lucie Gagnon. "Variation in the use of discourse markers in a language contact situation." *Language Variation and Change* 9.2 (1997): 191-217.

Schiffrin, Deborah. *Discourse Markers*. Cambridge: Cambridge University Press, 1987.

Smith, Sara W. and Andreas H. Jucker. "*Actually* and other markers of an apparent discrepancy between propositional attitudes of conversational partners." In *Pragmatic Markers and Propositional Attitude*, edited by Gisle Andersen and Thorstein Fretheim, 207-235. Amsterdam/Philadelphia: John Benjamins, 2000.

Stenström, Anna-Britta. "Lexical items peculiar to spoken discourse." In *The London-Lund Corpus of Spoken English. Description and Research*, edited by Jan Svartvik, 137-175. Lund: Lund University Press, 1990.

Tagliamonte, Sali A. "So different and pretty cool! Recycling intensifiers in Toronto, Canada." *English Language and Linguistics* 12.2 (2008): 361-394.

Tagliamonte, Sali and Alex D'Arcy. "*He's like, she's like*: The quotative system in Canadian youth." *Journal of Sociolinguistics* 8(4) (2004): 493-514.

Terraschke, Agnes. "Use of general extenders by German non-native speakers of English." *IRAL* 45.2 (2007): 141-160.

Underhill, Robert. "*Like* is, like, focus." *American Speech* 63.3 (1988): 234-246.
Watts, Sheila. "Teaching talk: Should students learn 'real German'?" *German as a Foreign Language* 1 (2000): 64-82.