Book review


_The artist is like well I see you l- later you know or something an’ -- so they leave_. This is how Simone Müller’s book, entitled *Discourse Markers in Native and Non-native English Discourse*, opens up. This sentence, extracted from the Giessen-Long Beach Chaplin Corpus, on which the study is based, illustrates the four discourse markers which the author focuses on, namely _like_, _well_, _you know_ and _so_. Her aim is to compare the use of these discourse markers by native American speakers and German learners of English, using comparable spoken corpus data. More precisely, she seeks to answer five questions (p. 24):

a. How does the frequency of the selected discourse markers in German non-native speakers’ (NNS) discourse compare to the frequency in native discourse?
b. Which functions do the selected discourse markers have in the native discourse?
c. Which of these functions also occur in the non-native discourse?
d. How do the frequencies of the functions of discourse markers compare in native and non-native discourse?
e. Which factors related to learning and using English in school and in informal contexts influence the NNS’s use of the discourse markers?

Müller’s study thus combines the frameworks of discourse marker research and interlanguage pragmatics. But it also clearly falls within the scope of corpus linguistics, since her basis to answer the above questions is a spoken corpus of native and non-native narratives and discussions conducted in English, which was collected in large part by the author herself.

The book comprises six chapters. The first one (pp. 1-59) is a long introductory chapter which provides both a literature review of discourse markers and second language acquisition, and an overview of the approach adopted in the book. After emphasising the wide variety of labels used to refer to discourse markers, Müller goes through the different properties proposed in the literature to delimit the phenomenon, showing how most of these cannot be seen as defining criteria. In her own list of criteria, multifunctionality holds an important place. Following other linguists, Müller claims that discourse markers serve a number of (sub-)functions. One of her goals, and one in which she is very successful, is to highlight the various functions of each of the four selected discourse markers as they emerge from the corpus data, and show which of these specific functions display a different distribution across native and non-native speech. In this first chapter, the author also deals with a number of issues related to second language acquisition: how the field developed, how knowledge of a foreign language goes beyond grammar, how interlanguage pragmatics has mainly focused on speech act performance. What is lacking, according to her, is more extensive research into the use of discourse markers in learner language. Her study, precisely, aims to fill in this gap, thus bringing discourse marker research and interlanguage research closer to each other. Her main tool to do so is the Giessen-Long Beach Chaplin Corpus (GLBCC), which she describes in detail. The GLBCC contains transcribed recordings of students retelling and discussing a silent Chaplin film, _The Immigrant_. Most of the recordings involved pairs of students, one of whom saw the whole film and the other of whom watched only half of it. For her own purposes, Müller concentrates on the recordings of 34 American...
native speakers of English and 77 German speakers of English as a foreign language, for a total of approximately 6.5 hours of recording and over 150,000 words. The corpus data are supplemented by questionnaires filled out by the participants and including questions about age, relationship to the partner, first language and, for non-native speakers of English, the context in which they learned English and use it. As Müller explains (and explores in the following chapters), these questionnaires and the information they provide can serve as a basis for analysing the potential influence of a number of (linguistic and non-linguistic) factors on the use of discourse markers. The chapter ends with a presentation of the distribution of the speakers in the corpus according to these factors.

Chapters 2 to 5 each deal with one of the discourse markers selected for study and are all organised according to the same structure. First, the functions of the discourse marker acknowledged in the literature are reviewed. Next, the author describes her own, corpus-derived classification of functions, which she illustrates by means of commented examples from the GLBCC and with reference to the literature when relevant. The third part consists in a quantitative analysis, which compares the frequency of the discourse marker and its various functions in native and non-native speech, and investigates the influence of certain linguistic and non-linguistic factors on the distribution of the discourse marker and its functions. The non-linguistic factors are: gender, age, relationship between the partners (friend, acquaintance or stranger) and role (students A, who retold the first part of the film without a hearer and then participated in the discussion of the film with B; students B, who retold the second part of the film to A and discussed it with him/her; students C, who did not have a partner and had to retell the whole film with no hearer and give their opinion about it). The linguistic factors, which concern only the German speakers, have to do with the acquisition and usage of English in formal and informal environments, as well as the time spent in an English-speaking country (if any), all of which inform an additional factor which Müller refers to as “native speaker contact”. Both the descriptive part and the quantitative analysis of each chapter are followed by a short summary.

Chapter 2 (pp. 61-100) deals with so, which, of the four discourse markers investigated, turns out to be the most frequent one among both native and non-native speakers. Next to non-discourse marker functions (adverb of degree or manner, expression of purpose, fixed expression, direct translation from German and substitute), Müller recognises functions working at the textual level and functions working at the interactional level. The textual functions, which focus on lexical expressions and propositional content, include marker of result/consequence, main idea unit marker, marker summarising/rewording/giving an example, sequential so and boundary marker, whereas the interactional functions, which focus on the relationship between speaker and hearer, include speech act marker (question/request or opinion), marker of an implied result and marker of a transition relevance place. The quantitative analysis reveals that the German learners significantly underuse so as a discourse marker, but that this is mainly due to the underuse of three (textual) functions, viz. marking result/consequence, summarising/rewording/giving an example and sequencing. Some linguistic and non-linguistic factors are also shown to have an influence on the distribution of so and its different functions. Thus, the female American participants use the textual functions more often than their male counterparts, while the female German participants use the interactional functions more often than the German males. Native speaker contact seems to have an impact too, resulting in the higher frequency of the discourse marker (particularly at the textual level) among those learners who have had extensive contact with native speakers.

In chapter 3 (pp. 101-146), it is well that is the focus of attention. Like so, its functions are divided into non-discourse marker functions (adverbial and “in addition”), textual functions (searching for the right phrase, rephrasing/correcting, quoting, moving to the main story, introducing the next scene, concluding) and interactional functions (indirect and direct
answer, response to self-raised expectations, contributing an opinion, continuing an opinion/answer, evaluating a previous statement). As a discourse marker, well is much less frequent than so (in fact, in the native group, it is the least frequent of the four discourse markers) and it does not exhibit any statistical difference in frequency between native and non-native speakers. However, some of its functions do differ significantly between the two groups: searching for the right phrase, concluding, giving indirect answers and continuing an opinion/answer are all overused by the German learners, with the last two mentioned functions not occurring in native speech at all. The analysis of the potential influence of the linguistic and non-linguistic factors on learners’ use of well shows few unified patterns, but it appears that informal interaction with native speakers helps reduce the non-native-like frequency of some functions of the discourse marker.

Chapter 4 (pp. 147-196) discusses the expression you know. Excluding cases where it forms part of the syntax of the clause (e.g. do you know…), you know has functions both at the textual and interactional levels. The former include marking lexical or content search, marking false start and repair, marking approximation, introducing an explanation and quoting, whereas the latter include appeal to the hearer to “imagine the scene” or “see the implication”, reference to shared knowledge, appeal for understanding and acknowledgement that the speaker is right. Of the four discourse markers under investigation, you know is the least frequent one among the German students. It is significantly more frequent among the American students, although its relative frequency in native speech is relatively low compared to that of so and like. Most of the functions of you know are underused by the non-native speakers, and one of them, the quotative function, is never used by the learners at all. The non-linguistic factor that seems to influence the frequency of the expression most is the role of the participants. In particular, the corpus does not contain a single occurrence of you know in the speech of the German participants in role C (that is, with no partner), which suggests that Germans (unlike Americans) do not see the point of using this expression when there is no interaction with a partner. The study of the linguistic factors reveals that exposure to native English results in a lower frequency of you know introducing explanations and a higher frequency of you know asking the hearer to “see the implication”.

The fourth discourse marker, like, is examined in chapter 5 (pp. 197-239). Like so, well and you know, it can be used with non-discourse marker functions (as a verb, as a preposition, as a conjunction and with a quotative function), but unlike the other discourse markers, its discourse marker functions are limited to the textual level. These textual functions are: searching for the appropriate expression, marking an approximate number or quantity, introducing an example, introducing an explanation and marking lexical focus. In terms of overall frequency, like comes second in native speech and third in non-native speech. It is significantly underused by the German learners, and this is true of all its discourse marker functions (and, incidentally, of its quotative function). As for the impact of non-linguistic variables, it is noticeable that like is the only discourse marker that seems to be influenced by the relationship between the partners, with American friends using it more often than strangers to give an example, and German friends using it more often than strangers to search for the appropriate expression, as a lexical focuser and as a discourse marker in general. In terms of linguistic factors, it transpires from the analysis that interaction in American English in informal environments encourages the German students’ use of like, both in its individual functions and generally.

Chapter 6 (pp. 241-252) concludes the book by summarising the main findings and opening up new avenues for research. Interestingly, the chapter also offers some possible explanations for German learners’ behaviour, briefly discussing the possible influence of teaching materials, teacher talk and L1 interference (in the form of avoidance or transfer of frequency).
In what follows, I would like to deal with each of the three main components of this book, namely the literature review, the description of the discourse marker functions and the quantitative analysis, and highlight what I see as their main strengths and weaknesses.

One of the strong points of Müller’s book is its thorough review of the literature on discourse markers in general and on discourse markers in native English in particular. Both in the introductory chapter and at the beginning of each chapter devoted to a specific discourse marker, Müller compares the different points of view expressed in the literature, discusses their pros and cons, and (most of the time) takes a stance on these issues. What is perhaps less well represented is the literature on the use of discourse markers by foreign learners. The author mentions a few studies such as Nikula (1996) or Romero Trillo (2002), but she suggests that this area is largely unexplored, interlanguage pragmatics having focused on speech act performance instead. While this may be so, this seems to be largely due to the types of data that are necessary to investigate these phenomena. Speech acts may be studied by means of elicited data (sometimes in conjunction with corpus data, cf. Schauer & Adolphs 2006). Discourse markers, on the other hand, are best examined in more spontaneous data, such as those found in corpora. Since Müller’s focus is on discourse markers in speech, it is not too surprising that she finds relatively little relevant literature, given the scarcity of spoken learner corpora (the GLBCC and the Louvain International Database of Spoken English Interlanguage [LINDSEI]¹ are among the few existing corpora of speech produced by non-native speakers). But if we consider the written medium, we discover many studies that have dealt with the use of discourse markers (or, more precisely, connectors) by learners of English (e.g. Crewe 1990, Field & Yip 1992, Milton & Tsang 1993, Granger & Tyson 1996, Altenberg & Tapper 1998). Thus so, which according to Müller “has not received much attention in the literature” (p. 61), has been considered in studies of written learner English. It has been pointed out, among others, that so tends to be overused by learners (Granger & Rayson 1998), that this overuse gives learners’ written production an oral tone (Gilquin & Paquot, in preparation) and that learners, unlike native speakers, show a predilection for the sentence-initial position (Granger & Tyson 1996, Lorenz 1999). It is perfectly acceptable that Müller should restrict herself to discourse markers which primarily occur in oral rather than written discourse, but for the sake of completeness, this section of the literature should probably have been given more attention, especially in the case of so, which is common in both written and spoken interlanguage. Another area that is perhaps not developed enough in the literature review is learner corpus research (LCR). The book does not say anything about the (short) history of LCR, nor does it contain references to some of its latest developments. It also fails to use technical terms which have by now gained widespread use among LCR specialists, such as “underuse”, “overuse” or “contrastive interlanguage analysis” (“overuse” is only mentioned once in the literature review and once in the conclusion, and “contrastive interlanguage analysis” merely occurs in a footnote in the conclusion). This uneven literature review creates an imbalance between native speech and pragmatics research on the one hand, and non-native speech and interlanguage research on the other hand, with a focus on the former to the detriment of the latter.

¹ See http://cecl.fltr.ucl.ac.be/Cecl-Projects/Lindsei/lindsei.htm (last accessed on 17 June 2007). The ISLE corpus, which presents itself as a corpus of Italian and German spoken learner English (see Atwell et al. 2003), is actually made up of selected samples of English text and dialogue read aloud by learners of English, and is therefore not a corpus in the sense of “a body of naturally occurring language” (McEnery et al. 2006: 4).
Another noteworthy feature of the book is its detailed and well illustrated description of the various functions performed by the four discourse markers investigated. The fact that the classification of functions was established on the basis of corpus data makes it more reliable than earlier studies of discourse markers based on “invented and anecdotal evidence” (p. 9) – although, of course, it limits the findings to those functions that do occur in the data. It also highlights a situation which may go unnoticed in an intuition-based approach, namely that in some contexts the function of a discourse marker may be unclear and unclassifiable. Thus, the list of functions is followed, for each of the four discourse markers, by a section dealing with “unclassified instances”. In addition, Müller gives indications as to how to identify functions and distinguish them from other, similar functions, so that her classification could quite easily be applied to other corpora, for example to compare her results with data for learners from other mother tongue backgrounds or for other varieties of English. It is slightly to be regretted, however, that, as is the case in the literature review, the author seems to focus mostly on native speech and does not do full justice to the non-native data. It is symptomatic, for instance, that only a quarter of the examples of the discourse marker functions of so, well, you know and like come from the learner subcorpus (although, admittedly, this may be related to the overall underuse of discourse markers by learners). Most of the functions are shared by native and non-native speakers (some exceptions are the use of non-discourse marker so in direct translations of a German expression, conclusive well and well to continue an opinion/answer) and little is said about possible differences between the two groups of speakers, apart from the last couple of lines of each section, which compare the frequency of the function in the two subcorpora. What the reader may miss, in other words, is a qualitative comparison of the use of discourse markers in native and non-native speech. The absence of a significant difference in frequency does not per se imply that the learner uses the discourse marker (or one of its functions) correctly. There may be inappropriate choices of the discourse marker, differences in the preferred position of the discourse marker, differences in its typical collocates, etc. However, because Müller scarcely deals with such aspects, we have to rely solely on quantitative criteria to judge German learners’ performance, which may result in a partial view at best, and an erroneous one at worst.

Turning to the quantitative analysis itself, and in particular the analysis of the different (linguistic and non-linguistic) factors, one must say that it is conducted very methodically – too methodically, sometimes, as the author makes a point of going through each and every factor, even when the analysis proves it to be irrelevant. It is illustrated with many graphs, which makes it easier for the reader to visualise the results. There are, however, a number of problems with Müller’s quantitative approach. It has to be pointed out, first, that although it is in these sections that the reader will expect to find the “nitty-gritty” of the comparison between native and non-native speech (since, as noted earlier, the descriptive sections contain little in the way of actual comparison), the quantitative discussion, on average, takes up hardly more than a fifth of each of the chapters devoted to a given discourse marker (graphs included). This may be somewhat disappointing for readers who wish to learn more about the differences and similarities between learners and native speakers, as they are entitled to in view of the title of the book. Another problem with the quantitative analysis is that it includes the data of the non-discourse marker functions (homonyms such as well as an adverb, like as a preposition or so as a substitute) and of the unclassified instances. While it is a laudable goal to want to take all the data into account, thus adopting of corpus-driven approach, and while this leads to some interesting findings (most notably, the fact that learners tend to use the non-discourse marker functions proportionately more than native speakers), these data do not, in my opinion, belong in the more detailed quantitative analysis. Given that the unclassified category contains miscellaneous cases, it is not very informative that, say, the unclassified uses of well are more frequent among the male German subjects than among the female
German subjects (Figure 3.2a, p. 140). As for the non-discourse marker functions, by being included in the analysis of the factors, they are attributed more importance than is perhaps appropriate in a book on “discourse markers in native and non-native English discourse”. The choice of the factors to be considered is also slightly problematic. On the one hand, some of the factors included in the analysis should have been disregarded or treated differently. As Müller herself admits (p. 55), a factor such as “relationship between speakers” (friend, acquaintance or stranger) may not be very reliable, since people do not necessarily define the term “friend” identically. Also, certain values are grouped together in a way that may appear (linguistically) arbitrary. For the above mentioned factor, “relationship between speakers”, acquaintances and strangers are grouped together “for the sake of comparability, because among both Americans and Germans there were about twice as many friends as strangers and acquaintances taken together” (p. 44). Likewise, learners who spent less than four weeks in an English-speaking country were grouped with those who had never been abroad, because the author “suspected that four weeks might not be enough time yet to pick up discourse marker use” (p. 50, my emphasis). Such arbitrary decisions may not be expected (or even desirable) in an approach that claims to be corpus-driven. On the other hand, some factors are sadly lacking in the quantitative analysis. Thus, it would have been useful to systematically distinguish between the narrative section and the discussion section of the corpus. While Müller studies the influence of the role (A, B or C) on the results, this only partially overlaps with the narrative/discussion distinction, because of the mixed role of the A participants, who retold the first part of the film without a hearer and participated in the discussion of the film with B. More crucially, learners’ proficiency level is not one of the factors examined in the study (we just know that the learners “had mastered English well enough to take classes taught in English”, p. 32). Again, there may be some overlap with another factor, that of age, but this overlap can only be partial, and in view of the paramount importance of proficiency level for the knowledge of discourse markers (as demonstrated for example by Lorenz 1999 for causal markers), it is unfortunate that this factor is totally absent from the analysis. Another feature that could have been added to the quantitative analysis is the study of the interrelation between the different factors. Role is examined in combination with gender and several linguistic factors are grouped together to create the factor of “native speaker contact”, but it would have been interesting to consider all the factors simultaneously, using a statistical technique such as cluster analysis or correspondence analysis. Not only would it have highlighted possible interactions between some of the factors, but it would also have pinpointed the most relevant factors, those that account best for the observed results. Using such a technique, it would perhaps have emerged, for example, that the (counter-intuitive) result according to which the frequency of the discourse marker *like* increases with learners’ occasional usage of English as a primary means of communication but either remains the same or decreases with regular usage of English (p. 236) is best explained by bringing a different factor into the picture.

Overall, *Discourse Markers in Native and Non-native English Discourse* is a commendable book, tackling a very complex topic in a systematic way, presenting a detailed and corpus-driven classification of the functions of four English discourse markers and trying to disentangle the factors influencing their use by German learners. It answers the five questions it sets out to answer, and while, regrettably, none of these really enable the author to identify the actual *reasons* for some of the differences observed between learners and native speakers, the conclusion section contains some extremely interesting results that go beyond the merely descriptive approach of the quantitative analysis and seek to explain the learner data by appealing to teaching-induced and transfer-related factors. One may have wished, in a book aiming to investigate “how non-native speakers of English master the use of discourse
markers” (p. 2), that learner language be given more weight, both in the literature review and in the description of the discourse marker functions, and that factors whose importance is widely acknowledged in second language acquisition such as proficiency level be taken into account in the analysis, but this should not detract from the worth of the book. More generally, Müller’s study is, to date, one of the few large-scale studies into spoken interlanguage, and the GLBCC, one of the few corpora of learner speech. It is to be hoped that this book will inspire many researchers both to collect spoken interlanguage data (and, ideally, make them publicly available) and to carry out detailed quantitative and qualitative analyses on their basis.

GAËTANELLE GILQUIN

FNRS, Centre for English Corpus Linguistics, Université catholique de Louvain

References


