

# PRAGMATIC MARKERS IN SPOKEN INTERLANGUAGE

*Karin Aijmer, Göteborg University, Sweden*

*Aspects of language which always seemed to linguists to be far from the bread-and-butter side of language are now being seen as the backbone of the enterprise. (Stubbs 1986: 23)*

## **Introduction**

Computer-based corpora have facilitated the study of native speakers' use of English in speech and writing. Recently we have witnessed the emergence of several new areas for corpora, for example language acquisition and foreign-language teaching. Traditionally, second-language research has been less concerned with authentic learner data. As Granger points out (2002: 7), the reason is the difficulty of controlling all the factors affecting learner output. The situation is now changing and there is an increasing interest in the description of how learners write and speak English (Hunston 2002). In particular, there are corpora composed of the speech and writing of learners of English which can be used to study how learners actually use language. The most influential work has been done by Sylviane Granger from the Université Catholique de Louvain-la-Neuve in Belgium. Granger has initiated collaboration between researchers in different countries who are collecting data of advanced students' English (The International Corpus of Learner English; see Granger (ed.) 1998).

The present article takes a first step towards using a corpus of advanced Swedish learners' spoken English. Although advanced Swedish learners of English have a good command of English grammar and lexis, we may assume that their style of speaking differs from that of native

speakers. Learners may overuse or underuse certain devices in comparison with native speakers and therefore sound non-native.

To begin with, it is important that conversation is distinguished from writing and from more formal speech (Chafe and Danielewicz 1987). Conversation is generally unplanned. It is produced under cognitive and processing constraints which are reflected in filled and unfilled pauses, repetition, incomplete grammatical structures - features accounting for what Chafe (1982) describes as the fragmented nature of speech as compared with integration in writing. Certain linguistic items are more characteristic of speech than of writing or occur only in speech. Lexical items 'peculiar to spoken language' are, for example, *well, you know, you see, actually, sort of*, etc. (Stenström 1990). They will here be referred to as pragmatic markers (on the choice of terminology see Aijmer et al. Forthcoming).

Pragmatic markers are also relevant to the learners' communicative needs. Communicative stress can be high for learners, especially in conversations with native speakers which is reflected in the use of markers. The question which will be asked here is whether a particular use of markers is characteristic of learners. In order to find out whether this is the case, we need to compare learners and native speakers in order to identify similarities and differences between the two groups. Do learners overuse or underuse pragmatic markers compared to native speakers? Do they use markers for the same purposes as native speakers? I was also curious to find out more about pragmatic markers by studying their use in learner corpora. Do we get a one-sided picture of their functions by looking only at native speakers?

## Material

It is time-consuming to compile a corpus of spoken language. Moreover, it provides the challenge of having to choose a system of

transcription (given in note 2). The corpus is made up of interviews with advanced Swedish learners who were in their third year of studying English at Göteborg University. The learners were interviewed by a native speaker on a topic such as a recent trip or a movie they had seen and were subsequently asked to describe a series of pictures from a comic strip. Each interview lasted for about 15 minutes. The complete material transcribed consists of 50 interviews (c 100,000 words). The corpus will be put into electronic form together with other spoken learner corpora to form a sister corpus of the International Corpus of Learner English (see De Cock et al 1998).

The data in this exploratory study is fairly small - only about 10,000 words. Moreover, I have not been able to make a comparison with a similar group of native-speaker students. Instead, the data has been compared with a similar amount of conversational material from the London-Lund Corpus of Spoken English (Greenbaum and Svartvik 1990).<sup>1</sup> Learners' English is in focus since I believe that we need to find out more about the strategies learners use when speaking in a foreign language and the cognitive stress is particularly taxing.

### **Swedish learners' interlanguage – an illustration**

Non-native spoken discourse is illustrated below (with the system of transcription given in note 2).

<A> is the (male) interviewer. He is a native speaker of English.

<B> is the Swedish learner, a 21-year-old woman. B describes a trip she made to the Dominican Republic with her family:

(1)

Turn 1

<B> in Sweden *you know* everyone paint the= their houses they're painted *you know* in <breath> in a <breath> special <breath> *kind of* red colour *and all that* <breath> and this= *this was like* neon yellow and pink and . eh *and all sorts of* orange and so it was really bright and may be just too em *I don't know* an attempt to: make their miserable lives <begin laughter> a bit <end laughter> brighter *I don't know* <\B>

<A> did you get off the bus or= or you [know <XX> <\A>

Turn 2

[yeah we got off the bus and <swallows> and we walked around a bit and it's a city called Prerto<?> Plata <\B>

<A> mhm <\A>

Turn 3

<B> it's in eh *I think* that's the *sort of sort of* a capital <breath> erm and *it's really* but *it is a bit* eh *I think* because my . my mother's boyfriend's son <breath> he's also been to Cuba <breath> <\B>

<A> aha <\A>

Turn 4

<B> and he said that that it is very similar <breath> with eh *sort of* American influences *and all that* [it's <\B\> (SW027)

Markers (*italics*) such as *you know* or *I think* are pervasive in informal conversation. Formally they are phrases (*you know, and all that*) or single words (*like, well*); they are flexible and can occupy different positions in the utterance. Their contribution to the interpretation of the utterance cannot be described in truth-conditional semantics and they are not part of the proposition. Thus, the propositional content does not change with *you know* (or other markers), but the marker has the function of signalling

that the information is shared in order to involve the addressee in the interpretation of the utterance. Markers are characteristically multifunctional with a variety of pragmatic or discourse functions which depend on the context. Therefore the search for a core meaning of pragmatic markers constraining their multifunctionality is an important issue in research on the semantics/pragmatics interface (see e.g. Aijmer 2002 (and the references there)).

Much of the literature in discourse analysis has described how participants in natural conversation use such expressions to reach an understanding or an interpretation of what the speaker means (e.g. Schiffrin 1987; Stenström 1994; Jucker and Ziv (eds.) 1998). Therefore, we have a good picture of the pragmatic and discourse functions of markers. Pragmatic markers such as *you know, I think, sort of, actually, and that sort of thing* have the function of checking that the participants are on the same wavelength, or of creating a space for planning what to say, making revisions, etc. Informal conversation is largely phatic and the markers in informal conversation mainly perform a phatic function (Bazzanella 1990: 630). However, the question of whether native speakers and learners use markers for the same purposes is open for investigation. This is therefore a question to which I will return in the discussion below.

## Results

The markers used by learners are listed in Table 1 with combinations of markers listed separately:

Table 1

*Pragmatic markers in the spoken learner corpus*

Type of marker	Number of markers
I think	40
sort of	38
well	38
I don't know	28
actually	26
you know	23
like	14
I mean	13
yeah (not as an answer to a question)	13
or something	11
kind of	8
I guess	5
and all that	5
and everything	4
and stuff like that	4
or anything	3
really (final position only)	3
and stuff	2
or something like that	2
∅ something	1
and things like that	1
and all sort of	1
something like that	1
or whatever you want to call it	1
and everywhere	1
or so	1
∅ stuff like that	1
or anything like that	1
and that kind of stuff	1
Total	290

*Patterning of pragmatic markers*

well right

well yeah

well I guess

well actually

actually well

eh yeah

but yeah

yeah so

God yeah

you know like ... and stuff

sort of more or less

really sort of

just sort of

sort of just

sort of ... or whatever

sort of I don't know

sort of more or less

sort of like

sort of something

really sort of I don't know ... really

pretty sort of

very sort of

sort of thing

very sort of thing

kind of ... or something

kind of ... and all that

like ... or something

like maybe ... something like that

I don't know actually

I don't know I don't know ... or something

I don't know I think

like you know ... or anything

a bit you know

you know it was sort of like

I mean ... or anything

sort of ... or something

The corresponding data for native speakers are shown in Table 2 (see the opposite page).

*Patterning of pragmatic markers*

sort of ...or anything

just sort of

sort of ... and things

sort of particularly

too sort of

sort of rather

Table 2

*Pragmatic markers in the LLC (S1.8, S1.12)*

Type of marker	Number of markers
I think	77
you know	68
sort of	41
well	35
really	19
I mean	19
you see	13
and so on	5
or something	4
I suppose	3
actually	2
or anything	2
like	2
and that sort of thing	1
or anything of that sort	1
and suchlike	1
and things	1
Total	294

well actually

well I think

oh you know

you know... and things

I think you know

I think actually

I think really

I think you see

Second-language research has been more interested in less advanced learners' use of pragmatic markers and the results have therefore been different. Hasselgren (2002) investigated 14-15-year-old pupils with a comparable group of native-speaker pupils carrying out similar tasks. The considerable underuse in the non-native group of 'smallwords' (especially among the less mature learners) was correlated with their lack of fluency (cf. also De Cock et al 1998). In my material both native speakers and learners used pragmatic markers. The major difference between learners and native speakers has to do with the frequency of individual markers. *I think, you know, sort of, I mean, well, actually, really* were frequent in both groups. However, only the learners used *I don't know* and *yeah* and only native speakers *you see*.

Hedges like *I think* or *I guess* signal that the speaker is uncertain (termed 'shields' in Prince et al 1992). Other hedges such as *sort of* introduce fuzziness within the proposition (termed 'approximators' in Prince et al 1992). *And everything, and stuff (like that/ or something/ or anything* differ from *sort of* and *kind of* since they are normally placed at the end of the utterance.

One reason for being uncertain or vague is politeness. But there may be several reasons why a hedge is used depending on who the speaker is. In native speaker conversation markers have interpersonal function and are associated with face-saving, politeness and indirectness rather than with imprecision, approximation or uncertainty (Brown and Levinson 1987). Therefore, it is possible that other uses of pragmatic markers than those relating to face and politeness have been neglected. In academic discourse

for instance, markers are used epistemically where less accuracy is appropriate (Mauranen. Forthcoming). In the learner data, markers often co-occur with pauses and are best explained in terms of cognitive and verbal planning problems or as uncertainty devices.

*Well* was for instance often used inside the turn in the learner corpus as a pause-filler or before a reformulation:

- (2) which was: . quite a good experience I would say *well* I changed family . first the first family I got to was really . they were really horrible . so I left after five days <breath> and . *well* then I got to: . eh just a completely different family not from the upper class or anything so . I don't know if that matter but they were really nice to me and there was a single mother <breath> <\B>  
(SW023)

Moreover, *yeah* was used as a pause-filler where *well* would have been expected:

- (3) I don't know it sort of . *yeah* as I said it just became: like ordinary life <\B>  
(SW 023)

*Sort of* and *kind of* were mostly before the word or phrase they modified in both groups. In the learner data it was also frequent (9 examples) without a head, for instance, before a restart:

- (4) <B> mm but then we had . we *sort of* you got the tips after that which was more than the wages . [so <\B> (SW023)

*Like* was poorly represented in the London-Lund Corpus which may be due to the fact that the corpus was compiled almost thirty years ago (cf.

Andersen 2001 on the frequency of *like* in present-day adolescent speech). The following example is from an interview with a learner:

(5) yeah. it did. cos it was just. we used to go to this pub *like* every night  
(SW023)

*I think* and *you know* both have high frequencies, which is to be expected when the conversation is informal:

(6) <B> [<XXX> he's a *I think* he is from <breaths> eh Austria *I [think* <\B>  
(SW024)

*You know* is difficult to distinguish functionally from *I think* and *sort of*. However, *sort of* was more frequent among learners. Both learners and native speakers prefer *I think* to *you know*. The markers can occur in different turn positions and this is one way in which learners and native speakers can differ. For example, *I think* was more frequent in mid or end (parenthetical) position than initially. In the LLC material, only six out of 77 examples were not placed first in the utterance (or after *and*, *but*, *because*). When *I think* is not placed first it always expresses uncertainty (Aijmer 2001). Assuming that learners generally express more uncertainty in conversation, this result is not surprising.

Among formally and functionally similar words or phrases such as *and all that*, *and everything* we find a great deal of variation.

There were also non-native-like tags such as *or whatever you want to call it* in the Swedish data. French learners of English seem to underuse utterance-final tags. In the corpus compiled by De Cock et al (1998) native speakers used almost four times as many vagueness tags as learners, although the French learners also overused some tags (*and so on*).<sup>4</sup> My

sample was probably too small to establish whether there is a similar difference between Swedish learners and native speakers.

*I guess* and *kind of* in the learner corpus were probably due to American influence. In addition, learners use more clustering and collocations. This is not surprising since learners are likely to feel more communicative stress.

#### *Repetition, stranding, clustering and collocation of pragmatic markers*

Markers can be repeated or stranded; they cluster together or collocate with each other. Repetition indicates non-fluency and leaves the hearer time to plan what to say next or to choose a new orientation of the discourse (*sort of sort of*).

Stranding is illustrated in turn 3 (example 1 above) where *it's really* and *it's a bit* are used without a following head phrase. Clustering of markers is illustrated in (7):

- (7) <A> do you think portraits very rarely look *like you know* the people they are supposed to [represent <\A> (SW023)

When markers cluster this is a sign that they have a similar function. Unlike collocations, there is no internal ordering between the words in clusters. In collocations, i.e. co-occurrence of words forming a single marker, we also find combinations of elements with contradictory meanings such as *really sort of*:

- (8) <B> for . I was there from ninety-five till ninety-eight .. *sort of more or less* the whole time but I <breath> always we= went home to went back to Sweden in the summer and ..during the holidays at . Christmas. er .. but yeah . it was *really sort of* .. I don't know if it impressed me really [<laughs> <\B> (SW 023)

Larger patterns with pragmatic markers are illustrated in (9)-(11):

(9) <B> yeah . I like more eh if you say *like* pop arts *and stuff* <\B> (SW025)

(10) <B> er yeah I guess . but not not like the ordinary stuff *like you know*  
Rembrandt *or any- [thing* <\B> (SW025)

(11) when you= going to a room and there's a short video sequence <\B>  
<A> [aha <\A>  
<B> [of *like* one person *or something* and that's I I like that kind of art [you  
know <\B> (SW025)

The possibility for markers to cluster suggests that they have little function in themselves. Both learners and native speakers use clusters of markers to get more time for planning what to say next, to make a new start, or to reformulate what they have just said. This may in fact be the dominant or only function of markers in learner speech while native speakers also use clustering to reinforce the phatic function of the markers.

### *I don't know*

*I don't know* suggests that speakers are not taking full responsibility for what they are saying (see Tsui 1991). Learners make frequent use of *I don't know*, which makes them sound more uncertain than native speakers. The uncertainty may be underlined by repetition and by other markers (*sort of, as I said*):

(12) <A> would you go back to live there <\A>  
<B> no . not to live there no <\B>  
<A> why not <\A>

<B> <breath> *I don't know it sort of . yeah as I said it just became: like ordinary life* <\B>

<A> hm <\A>

<B> it wasn't that exciting any more to: live there <\B>

In (13), *I don't know* is followed by an expression in which the speaker expresses his uncertainty:

- (13) but-t-t . em . I stayed with a family who had Maori relatives *I don't know* <XX> *I think the husband and family was half Maori or something* <swallows> and he spoke very warmly about the culture and <breaths> and I think . think they are they are they . I think they want to they want to preserve it <\B> (SW 024)

*I don't know* in particular is a device helping the speaker to achieve fluency in the conversation. In (14), *I don't know* is placed between words in a phrase filling a pause while the speaker tries to think of the right word:

- (14) and I got a bit *I don't know homesick* . I wouldn't say homesick . but I went back because my sister had a baby <\B> (SW023)

*I don't know* is used as the equivalent of a pause before a new start:

- (15) and you could *I don't know you could* <breath> have a nice garden with lots of fruit (SW024)

In the data from native speaker conversation looked at by Tsui, *I don't know* introduced a turn component and was frequently used to signal disagreement and to avoid commitment in addition to being a marker of

uncertainty. When used by learners, however, *I don't know* functioned only as an uncertainty device or 'filler'.

### Conclusion

Irrespective of the small size of the corpus, there are still some conclusions we can draw from this study. By comparing learners' conversation with native speakers we get a picture of the problems students have in communicating in a foreign language. The fact that the student is unaccustomed to the interviewing situation may also contribute to this uncertainty.

The type of spoken language studied in this project is informal spontaneous speech. As Östman points out (1982: 161), the same social and psychological causes may produce both informal conversation and pragmatic markers. Since learners and native speakers are not in the same psychological situation as conversational partners, we may expect them to use markers for different reasons. Learners use vague and uncertain markers to express uncertainty or hesitation and not for face-saving or to signal politeness. Markers are also used as strategies when the learners have communication problems. For example, markers were typically stranded in the conversation, leaving it to the hearer to complete the message. Clustering of markers was another characteristic feature of learner language with the function of filling a space in conversation. The non-native speaker generally used the same markers as native speakers. An exception is *I don't know*. In my material, learners made frequent use of *I don't know*, which makes them sound more uncertain than native speakers. Thus the phrase occurred before, between, and after constituents as well as in combination with other markers.

## Notes

1. Conversations S.1.8 and S.1.12. Since the conversations in the learner corpus were shorter, six learner conversations have been used (SWO 22, 23, 24, 25, 27,29).

### 2. Transcription conventions

The end of each turn is indicated by either <\A> or <\B>

Empty pauses are marked by dots corresponding to the length of the pause.

Filled pauses are marked e.g. as *eh*, *er*, *erm*, etc.

<X> represents one unclear word (or a syllable)

<XX> represents two unclear words

<XXX> represents three or more unclear words

3. Tag questions and interjections have not been included either in the native speaker or the non-native speaker corpus.

4. De Cock's corpus of French learner language was compiled according to the same principles as the Swedish corpus. The native speaker corpus is more directly comparable to the learner material than the London-Lund Corpus.

## Works cited

- Aijmer, K. 2001. *I think* as a marker of discourse style in argumentative Swedish student writing. In: Aijmer, K. (ed.) *A wealth of English. Studies in honour of Göran Kjellmer*. Gothenburg Studies in English 81. 247-257.
- Aijmer, K. 2002. *English discourse particles. Evidence from a corpus*. Amsterdam & Philadelphia: John Benjamins.
- Aijmer, K. & A. Foolen & A. M. Simon-Vandenberg. Forthcoming. Discourse particles in the perspective of heteroglossia. In: Fischer, K. (ed.). *Approaches to discourse particles*. Amsterdam & Philadelphia: John Benjamins.
- Andersen, G. 2001. *Pragmatic markers and sociolinguistic variation*. Amsterdam & Philadelphia: John Benjamins.
- Bazzanella, C. 1990. Phatic connectives as interactional cues in contemporary spoken Italian. *Journal of Pragmatics* 4. 629-47.
- Brown, P. & S. C. Levinson. 1987. *Politeness. Some universals in language usage*. Cambridge: Cambridge University Press.
- Chafe, W. 1982. Integration and involvement in speaking, writing, and oral literature. In: Tannen, D. (ed.). 1982. *Spoken and written language: Exploring orality and literacy*. Norwood: Ablex. 35-53.
- Chafe, W. & J. Danielewicz. 1987. Properties of spoken and written language. In: Horowitz, R. & S. J. Samuels. *Comprehending oral and written language*. New York: Harcourt Brace Jovanovich. 83-113.

- De Cock, S. & S. Granger & G. Leech & T. McEnery. 1998. An automated approach to the phrasicon of EFL learners. In: Granger (ed.). 67-79.
- Granger, S. (ed.) 1988. *Learner English on computer*. London & New York: Longman.
- Granger, S. 2002. A bird's-eye view of learner corpus research. In: Granger, S. & J. Hung & S. Petch-Tyson (eds.). 3-33.
- Granger, S. & J. Hung & S. Petch-Tyson (eds.). 2002. *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam & Philadelphia: John Benjamins.
- Greenbaum, S. & J. Svartvik. 1990. The London-Lund Corpus of Spoken English. In: Svartvik (ed.). 11-45.
- Hasselgren, A. 2002. Learner corpora and language testing: smallwords as markers of learner fluency. In: Granger, S. & J. Hung & S. Petch-Tyson (eds.). 143-73.
- Hunston, S. 2002. *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Jucker, A. H. and Y. Ziv (eds.). 1998. *Discourse markers. Descriptions and theory*. Amsterdam & Philadelphia: John Benjamins.
- Mauranen, A. Forthcoming. *They're a little bit different... Observations on hedges in academic talk*. In: Aijmer, K. & A.-B. Stenström (eds.). *Discourse patterns in spoken and written English*. Amsterdam & Philadelphia: John Benjamins.
- Östman, J.-O. 1982. The symbiotic relationship between pragmatic particles and impromptu speech. In: Enkvist, N. E. (ed). *Impromptu speech: A symposium*. Åbo: The Research Institute of the Åbo Akademi Foundation. 147-77.
- Prince, E. F. & J. Frader & C. Bosk. 1982. On hedging in physician-physician discourse. In: di Pietro, R. J. (ed.) *Linguistics and the professions*. Norwood: Ablex.
- Schiffrin, D. 1987. *Discourse markers*. Cambridge: Cambridge University Press.
- Stenström, A.-B. 1990. Lexical items peculiar to spoken discourse. In: Svartvik (ed.). 1990. 137-175.
- Stenström, A.-B. 1994. *An introduction to spoken interaction*. London & New York: Longman.
- Stubbs, M. 1986. 'A matter of prolonged fieldwork': Notes towards a modal grammar of English'. *Applied Linguistics* 17. 1-25.
- Svartvik, J. (ed.). 1990. *The London-Lund Corpus of Spoken English: Description and research*. Lund: Lund University Press.
- Tsui, A.B.M. 1991. The pragmatic functions of *I don't know*. *Text* 11. 607-622.