

## **Corpora and corpus linguistics revisited: an interview with Karin Aijmer**

by **Jolanta Šinkūnienė**

Department of English Philology

Vilnius University, Lithuania

[jolanta.sinkuniene@flf.vu.lt](mailto:jolanta.sinkuniene@flf.vu.lt)

**Karin Aijmer** is Professor Emerita in the Department of Languages and Literatures at the University of Gothenburg. She received her PhD in English Linguistics from Stockholm University (1972). She has served on the Scientific Committee of ICAME (International Computer Archive of Modern and Medieval English). She was a member of the Cambridge Grammar reference panel and a member of the Challenge panel at the ESRC Centre for Corpus Approaches to the Social Sciences (CASS) at the University of Lancaster 2012–2017. From 2004–2013 she served as president of the Swedish Society for the Study of English (SWESSE). She has been elected member of the Royal Society of Arts and Sciences in Gothenburg (*Kungliga Vetenskaps- och Vitterhetssamhället i Göteborg*) since 1998. She is the editor of the *Nordic Journal of English Studies*. Her research is mainly concerned with pragmatics and discourse, in particular with epistemic modality and evidentiality, discourse markers, conversational routines and other fixed phrases. She uses corpus-based methods involving both monolingual and multilingual corpora of English and Swedish for data.

**Jolanta Šinkūnienė (JŠ):** We are meeting at the 38<sup>th</sup> ICAME conference. ICAME is an international organization of linguists and information scientists working with English machine-readable texts. Does this annual conference with a long history have a particular significance to you as a dedicated corpus linguist?

**Karin Aijmer (KA):** My first ICAME conference was in 1985 although this wasn't the first ICAME conference (the first ICAME conference was held in 1979). I did not attend the first ones which were small gatherings with only a few dedicated corpus linguists. But when the conference was organized in Röstånga in Lund by Jan Svartvik, where I was working at that time, I was invited to attend. I remember from this conference that both Nelson Francis and Henry Kučera were there, the compilers of the Brown Corpus and pioneers within corpus linguistics.

**JŠ: How far advanced were corpora at that time?**

**KA:** It should be remembered that the only corpora then were the Brown Corpus and the LOB (Lancaster-Oslo/Bergen Corpus). The London-Lund Corpus was also in preparation and was being digitalized in Lund. The newness of corpora was also reflected in the early ICAME conferences which were kept small without any parallel sessions. This situation has changed and now there are many people wanting to attend ICAME conferences and abstracts are refereed. But it is still a small and friendly conference. I have many fond memories from ICAME conferences. For me a memorable occasion was when we organized the ICAME conference in Gothenburg (the 23<sup>rd</sup> ICAME conference in 2002), because so many famous corpus linguists came together, for example Jan Svartvik, Geoffrey Leech, John Sinclair, Michael Halliday, Michael Hoey, Mike Stubbs, Jack Dubois. So I think that was the occasion I would remember best.

**JŠ: Having participated in your first ICAME conference, did you think that this was the type of conference you would like to attend in the future?**

**KA:** Absolutely! It was a wonderful conference: interesting participants, stimulating ideas, so the conference really strengthened my interest in corpus linguistics.

**JŠ: Did your career start at Lund University?**

**KA:** No, I started out as a research assistant at Stockholm University before moving to Gothenburg. For several years I was then commuting between Gothenburg and Lund where I was working in close co-operation with Bengt Altenberg compiling our English-Swedish contrastive corpus. I have also worked in Oslo, which is close to Gothenburg, for a couple of years. Stig Johansson who was one of the founders of ICAME was a professor there, so working in his department also meant that I learnt more about corpora.

**JŠ: You have been working in corpus linguistics for a number of years exploring a variety of topics. What developments or evolution can you notice in this field?**

**KA:** Well, the field has changed its focus from written corpora to include many different types of corpora. Now we have access to spoken corpora, multimodal corpora, learner corpora, parallel corpora and a variety of specialized corpora to give only a few examples. All these corpora have resulted in a broadening of research. In addition to the more traditional lexicographical studies using written corpora we have seen an explosion of corpus studies in pragmatics, discourse and text which can be linked to the existence of spoken corpora.

**JŠ: Have you worked with learner corpora?**

**KA:** Yes, the Swedish research team (with participants from Lund and Gothenburg) was a part of the ICLE (International Corpus of Learner English) project initiated in Louvain-la-Neuve by professor Sylviane Granger. Our team was responsible for compiling the Swedish component of the ICLE corpus and as a result many of my doctoral students have written dissertations in connection with the corpus. Another project was to compile a spoken correspondence of the ICLE corpus (the spoken component of the Louvain International Database of Spoken English Interlanguage (LINDSEI) Corpus. My own research on learner corpora has dealt with advanced Swedish learners and comparing how they use their modality and discourse markers in comparison with native speakers.

**JŠ: Did you start your corpus-based research from written corpora?**

**KA:** I have mainly worked on spoken corpora. Before becoming professor in Gothenburg I was lucky to work at the English department in Lund together with colleagues who were also doing research on the basis of the London-Lund Corpus. The topics I have been interested in have also focused on spoken data. I have, for example, studied modality and discourse markers in spoken language.

**JŠ: Did you start working on modality in English and then switched to studies on modality in Swedish?**

**KA:** Yes, but modality in Swedish has come in mainly in a contrastive perspective.

**JŠ: You are one of the first prominent linguists to introduce this contrastive perspective based on parallel corpora. It might seem an obvious question, but how do parallel corpora benefit contrastive studies?**

**KA:** I learned, as we went along with our contrastive project, how useful parallel corpora can be if one is interested in similarities and differences between languages. When we started the project on parallel corpora, contrastive linguistics had existed for a long time in applied linguistics, but it had not been successful in predicting learner errors, so we wanted to make a fresh start building parallel corpora in order to compare English and Swedish. In my own work I have been particularly interested in how we can use translations to study discourse markers or modality across languages. One can, for example, get a picture of the multifunctionality of discourse markers such as *well* by studying how they are translated into another language. These results can, for example, be of interest to language teachers.

It is obvious that the study of parallel corpora has grown in a variety of ways since they started to be used. There have been new parallel corpora with many different language pairs and corpus-based research in the areas of syntax, semantics and pragmatics. The first parallel corpora contained translations between two languages only, now we also have multilingual corpora which provide translations from one language into many languages. We are also beginning to see more comparable corpora so that we can now make comparisons between original texts in different languages. This makes it easier, for example, to study the influence of text type on modality in different languages. The area of parallel corpora has had a fantastic growth and development in the last decades. Last year was the 25<sup>th</sup> anniversary of the English-Swedish Parallel corpus and we celebrated this occasion with a symposium together with our Norwegian corpus colleagues.

**JŠ: You have published extensively on a variety of topics – which of your projects/publications do you rate, on the one hand, as the most significant in the field and, on the other hand, as the most interesting to you personally?**

**KA:** For the whole of my career I have been working on discourse markers both in English and contrastively and this is an area which is close to my heart. I have written about discourse markers generally in the book *English discourse particles*. More recently I have returned to this topic in a book on understanding discourse markers where I was more interested in how discourse markers are used in different text types and how they are influenced by the speaker.

Another early and lasting interest is speech acts. One of my first books, *Conversational routines in English*, was on speech acts in the London-Lund Corpus and it involved such areas as creativity, conventionality, politeness – these are areas in which I am still interested, although now I have also become interested in impoliteness. In the research I have been doing lately I have studied how young people use apologies and other speech acts in impolite ways – sarcastic, ironic, humorous.

**JŠ: Will there be any publication on that?**

**KA:** There will be a publication dealing with apologies and how it is used by teenagers. Teenagers use words like *sorry* more than adults, but it is not in order to be polite; they use apologies for teasing and for humour as a way of showing solidarity with their peers. The study was inspired by an earlier article on *please* where I noticed that teenagers used *please* in combinations such as ‘Will you fuck off please’. For both articles I have used a corpus of teenager spoken language (The Bergen Corpus of London Teenage Language, COLT).

**JŠ: Apart from discourse markers, is this the research area that interests you most?**

**KA:** Yes, conversational routines, discourse markers, modality – these are topics that I'm interested in. I have also been doing research on intensifiers in spoken discourse.

**JŠ: I have noticed that in this conference you will be giving a talk on expletives. Would you regard them as intensifiers?**

**KA:** Yes, expletives would be part of a larger project on intensifiers which focuses on new intensifiers especially in the language spoken by adolescents. I use the forthcoming BNC Corpus of spoken English (Spoken BNC2014) released this year in order to study short term changes in this area. I'm, for example, interested in why young people choose to say *fucking mental* instead of *very stupid*. Intensifiers change very rapidly so it is important to have up-to-date spoken corpora. It is also interesting to see how certain combinations with intensifiers are restricted to a particular age group, and that some intensifiers (including many expletives) are used more by women than by men.

**JŠ: So you are moving into the sociolinguistic field, focusing on gender and age group studies, aren't you?**

**KA:** I think there is a general movement to include sociolinguistic factors such as age group and gender in corpus studies since this information is available in many new spoken corpora. You also have many new fields such as sociopragmatics, corpus pragmatics and discourse pragmatics where the use of language by real speakers is in focus.

**JŠ: I was just about to ask if these studies could be covered within the field of corpus pragmatics. How would you define corpus pragmatics?**

**KA:** Corpus pragmatics can be regarded as a combination of the use of corpora and corpus-linguistic methods, on the one hand, and pragmatics, on the other. Corpus-linguistic methods have to do with the use of concordances, keywords, statistical programs and various schemes for pragmatic annotation. Pragmatics is the use of language with characteristic topics such as discourse markers, speech acts, deixis, pragmatic annotation, evaluation. Corpus pragmatics is also interested in principles such as the Gricean maxims, the Relevance Principle in Relevance theory and with politeness principles, so it covers a lot of ground and is quickly becoming a discipline of its own. And let me add, it is not easy to distinguish between corpus pragmatics and discourse pragmatics since both are interested in communication and discourse. What distinguishes corpus pragmatics methodologically is that you have to start with linguistic forms which can be searched for in corpora and move on to function.

**JŠ:** The volume you edited with Christoph Rühlemann in 2015 was called *Corpus Pragmatics*. Is corpus pragmatics an emerging field?

**KA:** Jesús Romero-Trillo edited a book *Pragmatics and Corpus Linguistics*, so the title was in the air, so to speak. The combination between corpora and pragmatics has been around for quite a long time. Corpus pragmatics is still a newcomer but it is covering many different topics in pragmatics which can be studied on the basis of corpora and seems to have a bright future.

**JŠ:** One of the topics would obviously be discourse markers or pragmatic markers. Despite a substantial body of literature on these markers, they still enjoy a growing interest. In your opinion, what accounts for such popularity of these linguistic elements?

**KA:** There is very little agreement about what discourse markers are, what they should be called and how their multifunctionality should be dealt with. We therefore have an ongoing debate about very basic things in the area. Also there has been a widening of the field to include discourse markers in many more languages and to study their variation and change synchronically and diachronically. There has also been more interest in non-prototypical discourse markers such as interjections and vocative. So, if earlier on, you could count studies of discourse markers on the fingers of one hand, the field of study is now wide-ranging, and delimiting it is very problematic.

**JŠ:** Perhaps it is a nice example of how studies on semantically and pragmatically complex items can benefit from the advancement in technology, as well as the use of a variety of corpora and cross-linguistic research.

**KA:** Yes, and there has also been an extensive discussion about the position of discourse markers in terms of whether they are placed in the left periphery or in the right periphery in the utterance. And this, in its turn, is related to larger topics such as grammaticalization, and also to typological differences between languages with regard to where we place discourse markers, the relationship between position and function. So these are different aspects on discourse markers which now make up quite an extensive literature.

**JŠ:** You have mentioned *grammaticalization*. Did you coin the term *pragmaticalization*?

**KA:** I feel I should point out that it is not really my invention. When I was in Lund, I was working on *I think*, which is becoming a sort of discourse marker and I called this process grammaticalization. My colleague Bengt Altenberg insisted that *I think* has

nothing to do with grammaticalization and that it was an example of pragmaticalization. And I said, ok, let's call it pragmaticalization. And afterwards I seem to have become responsible for the term which seems to have caught on.

**JŠ: Since you have promoted the term, I think this is the reason why linguists associate it with you.**

**KA:** Yes, exactly. It has been a useful term in addition to grammaticalization to explain how lexical items achieve pragmatic functions.

**JŠ: If we go back to corpora, in your view, what is the future of corpus linguistics? Which developments in terms of a variety of corpora and corpus-based research can we expect in the years to come?**

**KA:** The future of corpus linguistics is definitely bright and there are many exciting new developments methodologically and theoretically. It has, for example, become customary to use a combination of different methods and to combine corpus linguistic methods with experimental and with ethnographic methods. Another observation is that the development of multimodal corpora provides a new avenue of research. If we take discourse markers as an example, they are used together with different gestures, so there is a lot of research needed to explore how this is done. I also think we can expect to see more co-operation between corpus linguistics and the new discipline of digital humanities in the future.

**JŠ: If someone asked you what your future ideal corpus is, what would you say?**

**KA:** Probably a multimodal corpus since it would make it possible to study spoken language in new dimensions. Another area where corpus linguistics and new corpora are important is Corpus-assisted discourse studies (CADS) which have a critical discourse perspective and study things like climate change, migration problems, hate mail.

**JŠ: What are you currently working on?**

**KA:** I am currently using the new spoken British National Corpus (Spoken BNC2014) to study the variation and change of intensifiers in English. The access to the corpus (while it was still in preparation) has made it possible to study changes in this area which have taken place in the last twenty years or so. I have also continued my research on discourse markers and modal stance markers both contrastively and in spoken corpora representing Englishes spoken in different parts of the world.

**JŠ: So you stay faithful to stance elements.**

**KA:** Yes. I am staying faithful to this area which has been with me since I began my linguistic studies in the Chomskian tradition.

**JŠ: This is an interview for the journal *Kalbotyra* which is published in Lithuania. What future perspectives do you see for Lithuanian linguists in the field of corpus linguistics?**

**KA:** I am well familiar with the research by Lithuanian linguists in the areas of modality and evidentiality on the basis of many different types of corpora. The contrastive corpus work by Lithuanian linguists has had a major impact in contrastive and typological research. I am impressed both with the work of compiling new corpora and the amount of research coming out of this and being published in books and articles. I always look forward to meeting Lithuanian linguists at corpus linguistics conferences and hearing about their research.

**JŠ: Thank you very much for the interview!**

The interview was held on May 26, 2017, at the Charles University in Prague, during the 38<sup>th</sup> ICAME conference.

## **References**

- Aijmer, Karin. 2015. "Will you fuck off please". The use of please by London teenagers. *Pragmática Sociocultural / Sociocultural Pragmatics* 3 (2), 127–149.
- Aijmer, Karin. 2013. *Understanding pragmatic markers: a variational pragmatic approach*. Edinburgh: Edinburgh University Press.
- Aijmer, Karin. 2009. Please: a politeness formula viewed in a translation perspective. *Brno studies in English* 35 (2), 63–77.
- Aijmer, Karin. 2002. *English discourse particles. Evidence from a corpus*. Amsterdam: John Benjamins.
- Aijmer, Karin. 1996. *Conversational routines in English: convention and creativity*. London, New York: Addison Wesley Longman.
- Aijmer, Karin & Christoph Rühlemann. 2015. *Corpus pragmatics*. Cambridge: CUP.
- Romero Trillo, Jesús (Ed.). 2008. *Pragmatics and corpus linguistics: a mutualistic entente*. Berlin, New York: Mouton de Gruyter.