

## INTERVIEW WITH PROFESSOR RANDI REPPEN

Razgovor vodila Višnja Pavičić Takač\*

Filozofski fakultet u Osijeku

U Hrvatskoj je u svibnju 2010. na poziv Hrvatskoga društva za primijenjenu lingvistiku gostovala Randi Reppen, sveučilišna profesorica primijenjene lingvistike sa sveučilišta Northern Arizona, SAD, gdje predaje na magistarskom i doktorskom studiju. Glavno područje njezina znanstvenoga zanimanja je korpusna lingvistika, a posebno se zanima za pitanje primjene rezultata korpusnih istraživanja u poučavanju stranih jezika i izradi nastavnih materijala. Nedavno je objelodanjena i njezina knjiga *Using Corpora in the Language Classroom* u izdanju *Cambridge University Press*-a.

Prof. Reppen je sudjelovala u radu XXIV. međunarodnoga znanstvenoga skupa s temom *Proučavanja diskursa i dijaloga između teorije, metoda i primjene* na kojemu je održala plenarno predavanje naslovljeno *Methods for using corpora to explore language use*. Njezin smo boravak iskoristili i kako bismo joj postavili nekoliko pitanja o korpusnoj lingvistici i poučavanju stranih jezika.

*What is 'corpus linguistics'? Is it a discipline in its own right or is it a methodology? How does it compare to other approaches/disciplines in linguistics?*

For me, corpus linguistics is an approach to investigating language use. It allows researchers to empirically answer questions about linguistic characteristics of language use in various settings. It's another tool we can use to help us understand language and how people use

---

\* Višnja Pavičić Takač, Filozofski fakultet u Osijeku, vpavicic@ffos.hr.

language. It certainly won't give us all the answers to all the questions about language, but like other methods of exploring language and its use, it does help fill in pieces of the puzzle.

*In one of his interviews (cf. Aarts, 2000<sup>1</sup>), Chomsky maintains that corpus does not exist, that is that corpus is a mere photograph of the language which does not provide answers to non-trivial questions. He actually compares corpus studies to butterfly collecting. What is your view on that? What is a 'corpus'? What are its most important features/qualities? Can the Internet be considered as a sort of a corpus?*

There are actually three questions in here: 1) Responding to Chomsky's views on corpora; 2) What is a corpus and the important features and qualities? and 3) Can the internet be considered a corpus?

First, I'll say photographs and butterfly collecting have yielded important and non-trivial information, so I will not take the comparison as negative and continue the analogy. Much like looking at historic photos to see how landscapes and people have changed, there is a rich tradition of corpus studies of historical language. Scholars have learned a lot about language and how it has changed over time. Some of these studies have also yielded important information about authorship attribution, so I would certainly not think of them as trivial. Now to the butterflies, by collecting butterflies we can learn how about they adapt to different environments, how they live and what butterflies live where. This information has been useful to scientists. As language scientists, I think it is important for us to know about the object of our study and collecting and observing it is one way to do that. That essentially is what corpus linguistic methods do. A question is posed by a researcher, teacher, or language student, then data are collected and analyzed. Just like with any method of inquiry, the results are dependent on the quality of the sample, the methods used and the interpretation. Given the many books and journal articles that have been published over the years that have used corpus methods and the information that has been gained from those publications, I would have to say that there is plenty of

---

<sup>1</sup> AArts, B. (2000) Corpus Linguistics, Chomsky, and Fuzzy Tree Fragments. In Christian Mair i Marianne Hundt (ur.). Corpus Linguistics and Linguistic Theory. Amsterdam: Rodopi, 5-13.

evidence for the importance of the information that has been learned from corpus methods.

Now to the second part of the question, a corpus is a collection of texts that represent spoken or written language. As I mentioned above the collection will be shaped by the question(s) being asked. Sometimes corpora can be very small if they are addressing a particular question about a particular type of language use. Say for example analyzing the works of a particular author, or looking at the in-class writing of your students in a particular course. Other times corpora need to be very large as in the case of capturing all the meanings of a particular word, or to make generalizations about language use. But even in large corpora design is still important and will impact the questions that can be explored. Now, the texts that make up a corpus are in electronic form and various software programs are used to get information from the corpus, such as a frequency list of the words or collocation patterns (groups of words that typically go together). In the case of looking at language use it is important to remember that this is a very iterative process. Good corpus research is not simply a matter of dumping a random bunch of texts and then abstractly looking at lists and collocations without going back and seeing those in context. There are many journal articles and books that can give you solid examples of the techniques and processes involved in good corpus research. Looking at those can also give you a sense of the range of questions that can be addressed.

Now for the final part of the question, can the internet be considered a corpus. To me the important part of any research is the question. What do we want to learn? Do we want to find out how blogs differ from e-mail? Then yes, the internet is the source for that. Do we want to compare how different companies or groups use language on their webpages? Again, the internet is the place to look. Do we want to just blindly collect from the internet and then pose a question? No, to me that is not sound. So yes, the internet can be a useful source of texts if they are the texts that are appropriate to answer the questions we are asking.

*Many of our readers are practicing teachers and would like to know how corpora can effectively be used in the language classroom. What are the benefits?*

I think there are several ways that teachers can use corpus information in their classrooms, especially at the intermediate and advance levels of language instruction and certainly in English for specific purposes courses. Let me highlight two main ways. First, knowing how language is used in certain contexts is valuable and can help inform what teachers present. Just as we use our textbooks to help us teach, we can also use information from corpora to help inform our teaching and as a resource for additional materials that we can bring into our classes. Using what we have learned about language from corpora studies is one way to help shape our teaching; using the language from our classrooms is another. We can create corpora from the readings we are asking students to do and we can also create corpora from our students' language. The texts that students produce can be a useful source for identifying areas where we need to reinforce or change certain features. Students can also be involved in comparing their texts with others, for example looking at vocabulary differences, or focusing on a particular feature like transitions. Using class reading texts we can look for a number of issues from vocabulary to lexico-grammar. Using 'key words in context or KWIC activities can be a great way to practice vocabulary, or to tease apart seemingly synonymous words. These are just a few ideas and suggestions.

*Does the use of corpora in the language classroom imply that the grammar we teach has irrevocably gone from "prescriptive" to "descriptive"?*

Absolutely not. Bringing real world examples of language use is just that. It provides a platform for discussing why certain forms are used in some contexts and not in others. Also it raises the awareness of language that is produced in real time, like natural conversation, sometimes has aspects that won't conform to strict prescriptive notions, but that carefully produced language like research articles will have language that conforms more to prescriptive norms. Looking at language change can also help us to realize that sometimes prescriptive notions change...

I do think there is a lot to be gained by looking at natural language and using that to inform our teaching, after all, one of our goals as teachers is for our students to be using language outside our classroom, and we do want them to sound natural. Bringing a new resource into our

classrooms does not mean abandoning everything we did before, it should be added value and help us be more effective, or else we should not bring it into the classroom. From my experiences, bringing corpus activities into classes have been motivating and provided learners with tools to become more autonomous.

*What kind of training or skills do teacher need to use corpora efficiently? Do they need to be familiar with terms such as word-class tagging, syntactic parsing, frequency, concordances, etc.? How can these be 'translated', i.e. operationalised in the language classroom?*

Well, this not a simple question to answer since perhaps different teachers will want to use corpora to a greater or lesser extent, but I'll try to give an answer. Of the terms you mention most are pretty easy to explain – word class tagging and syntactic parsing are just ways for computer programs to identify word classes and grammar – nouns, adjectives, relative clauses, passive voice, etc. I would imagine and hope that most teachers know these word classes and grammatical forms since they are the tools of our trade So, if teachers are already talking about these and using these labels, then there is nothing new here. It is not essential to understand the process by which the words are tagged, just as most of us use word processing programs and do not understand the inner workings of those programs, yet we can use them effectively. Frequency is a pretty transparent term so I think that doesn't pose a problem. Concordance is a term that might be new but it is easy to explain. As you know, it is a tool that allows a user to enter a word and get the occurrences of the word and its context, the words that surround the target word. I've used concordances with beginning level language students so I'm sure that teachers would have no trouble explaining and using this in their classes. Using the tools to do an activity is the way to do this. Having step-by-step screen shots can also be handy. The companion website for my Cambridge book *Using Corpora in the Language Classroom* has a PowerPoint with an activity and screen shots for that activity that teachers can download. This book also goes into much more detail about many of the questions posed here and provides lots of examples of classroom use across a range of different levels and language goals.

*What are, in your experience, common inhibitions that teachers may have when it comes to using corpora? How can these be overcome?*

I think the biggest inhibition teachers have is – how can I add one more thing to my already over-busy schedule and too short class-time. So my advice is to go slowly, see where bringing in a corpus activity allows for more purposeful practice or where it provides more material to demonstrate the language point you are presenting. Starting with vocabulary is always a good way and as teachers we know how important vocabulary is for language learners. Also you do not need to have computers in your classroom. You can prepare the material and bring it in. This is particularly useful if teachers do have any inhibitions since they are not ‘turning students loose’ on the computer. The teacher is still providing the material that students interact with and the teacher can control the vocabulary load and whatever else they present.

*On a more personal note, what do you find fascinating about corpus studies?*

I think language is fun and corpora provide a way to play with language and learn more about the complex puzzle of language that we as teachers are trying to bring to our students. Corpus studies also provide me with information that I think can be useful for language teachers and students. I like the challenge of figuring out how to make that information relevant and accessible to the teachers I train and the language students that we teach.

*Finally, what does the future hold for corpus linguistics?*

I think as technology advances we will see interesting and exciting new tools to look at language. I’m hopeful that in the future we can do more with multi-media corpora. I also expect that the future will bring more access and sharing of corpora. This is already becoming evident with projects such as those at the English Language Institute at University of Michigan, the corpora that Mark Davies has made available on the web and Vienna Oxford Corpus of English as a lingua franca. It’s an exciting time.

*Thank you very much!*