Entrevista

Paula Cortezi Schefer Cardoso
paula.cortezi@hotmail.com

Ana Cristina Ostermann
aco@unisinos.br

John Swales
jmswales@umich.edu

A fala institucional e a linguística de corpus: Michigan Corpus of Academic Spoken English (MICASE)

Institutional Talk and corpus linguistics: Michigan Corpus of Academic Spoken English (MICASE)

Entrevistado

John Swales é professor emérito do departamento de Linguística da Universidade de Michigan. Analista do discurso, conhecido internacionalmente especialmente por seu trabalho com gêneros e inglês para fins acadêmicos, John Swales foi diretor do *English Language Institute* da Universidade de Michigan de 1985 a 2001 e é atualmente co-diretor do projeto "Michigan Corpus of Academic Spoken English" (MICASE). É autor de inúmeros artigos e livros que se tornaram referências para a pesquisa e o ensino de gêneros, entre os quais destacamos: *Research Genres: Explorations and Applications* (2004), *Genre Analysis:* English in academic and research settings (1990) e *Academic Writing for Graduate Students:* Essential Tasks and Skills (publicado originalmente em 1994 e atualmente em sua 3ª. edição revisada, 2014).

Entrevistadoras

Paula Cortezi Schefer Cardoso é mestranda em Linguística Aplicada na Universidade do Vale do Rio dos Sinos (UNISINOS) e bolsista da Fundação de Amparo à Pesquisa do Rio Grande do Sul (FAPERGS). Especialista em gestão estratégica e inovação pelo Unilasalle e Licenciada em Letras pela Universidade do Vale do Rio dos Sinos (UNISINOS), Paula Cardoso é membra do grupo de pesquisa Fala-em-interação em Contextos Institucionais

e Não-Institucionais, coordenado pela Professora Ana Cristina Ostermann.

Ana Cristina Ostermann é Professora Titular no Programa de Pós-Graduação em Linguística Aplicada da Universidade do Vale do Rio dos Sinos (UNISINOS) e Bolsista Produtividade do CNPq. Ao longo de seu doutoramento na Universidade de Michigan, foi privilegiada com a aprendizagem proporcionada pelas trocas e convívio com o grupo de pesquisa coordenado por John Swales, que foi também seu professor em diferentes disciplinas e orientador de pesquisa de 1995 a 1998. Coordenadora do grupo de pesquisa Fala-em-interação em Contextos Institucionais e Não-Institucionais.

Esta entrevista aqui apresentada aconteceu por e-mail e foi especialmente pensada a partir da palestra ministrada pelo professor John Swales na Universidade do Vale do Rio dos Sinos (UNISINOS) em 2014.

Paula Cortezi Schefer Cardoso (PCSC) and Ana Cristina Ostermann (ACO): Since the middle 1960s you have been an active member of the English for Academic Purposes (EAP) movement. Could you tell us the important aspects of introducing EAP to linguistic studies?

John Swales (JS): I took a masters-level course in ELT and linguistics in the mid sixties. Ever since, I have been interested in both the grammar as well as the discourse of EAP texts. After my course in Leeds, where

---

I was taught syntax by Michael Gregory (Halliday's Category and Scale Grammar), I went to lead the small English section at the Faculty of Engineering at the University of Libya. This first real EAP experience led to my first textbook in 1971 (*Writing Scientific English*), which contains a fair amount of specialized grammar, aspects of which are still not fully resolved. Such as: Is the choice of "the underlined words" or "the words underlined" a free choice, or are they functionally different? Another question was names in science. We say "Bright's disease", Boyle's law", but "a Rorsarch test", "a Bunsen burner". Is this matter of discoveries versus inventions, respectively? Or something more? A recent one is the direction of code glosses (parenthetical clarifications): when do they go from the lesser known to the better known, and when the converse? A difference between textbooks and articles?

PCSC and ACO: Despite being aware of the extension and depth of your research on academic writing, in this interview we would like to focus mostly on your work on *spoken* interaction – the topic of your lecture at Unisinos in 2014. The Michigan Corpus of Academic Spoken English (MICASE) was a successful project of the English Language Institute (ELI) at the University of Michigan that initiated in 1997. This was a pioneer research project as by that time most of the research corpora on academic language focused on writing. Could you tell us what the challenges in producing a corpus based on speech – in particular, academic speech – were?

JS: Well, we had a very good team, with Dr. Rita Simpson as project manager, and some excellent undergraduate linguistics majors as assistants. However, building a speech corpus is time-consuming and expensive. We reckoned it took about 30 hours of work to get to a really good quality transcript of a 60-minute speech event, and we needed to have quality transcripts if we were going to make both the sound-files and transcripts freely available. In fact, one of my jobs was to go to the speakers and make sure we had got the proper names correct. I am not sure of the total cost but something in the region of $300,000. We couldn't have done it without using some of the examination fee income from the ELI's testing division. Another difficulty was getting permissions from the speakers, some of whom allowed us to transcribe their words, but not to make their recorded words available. So, this led to some serious editing challenges. Others involved training the transcribers. Overall, it was a complex, but rewarding experience over about five years.

PCSC and ACO: When building the MICASE corpus, you recorded different speech events, such as lectures, lab sessions, service encounters, and seminars. What are the benefits of having a large number of speech events in different environments for teaching academic speech?

JS: As we were getting started with MICASE, we went to a corpus linguistics conference in Boston, where Michael McCarthy said to Rita Simpson and me: "Don't just collect the easy stuff, like lectures". So, we tried to capture speech events that involved students from as broad a range as possible, not only in range of genres, but also in range of disciplines, as well in status from first year lectures to PhD dissertation defenses. Some were difficult to get, such as the defenses and research group meetings, but we did manage to get four of each. I used both sets to write most of the chapters in *Research Genres* on defenses and research group meetings. A German scholar used the office hour collection to study teacher-student interactions. A Swiss scholar looked at comprehension checks in small lectures. In the *MICASE Handbook* by Rita Simpson and Sheryl Leicher, they offer a very nice way of measuring the interactivity of academic speech events, and this has implications for cross-cultural education.

PCSC and ACO: What is the role of MICASE in designing teaching materials about different speech events in academic settings? Could you tell us about the developments?

JS: Unfortunately, much of the supporting material for MICASE is not available at the moment because the server on which it was sited crashed and couldn't be recovered. I have been trying for the past year to speed up progress on reconstructing the website. However, I have to tell you that nobody really listens to retired professors! After the corpus was completed, I and a bunch of undergraduates developed quite a series of "Kibbitzers", some dealing with grammar points, and some dealing with surprises, and most offering something for the language classroom. For example, what do you think is the third most common adverb in MICASE after "only" and "really"?[1] *The MICASE Handbook* also has some pedagogic explorations and excursions that are well worth looking at—and perhaps adapting. I have appended some material on how one might explore pragmatic elements of academic speech in Appendix A.

PCSC and ACO: What larger contributions for language teaching (not limited to *academic* language teaching) a corpus based on academic writing and speech have?

JS: In the Math study group, the students in a great episode criticize their instructor for the amount of homework she sets. It becomes clear that they are doing this not to complain, but in order to bond together, a feature of workplace conversations that Eggins and Slade found to be common in Australia.

---

[1] The third most common adverb is "basically".

PCSC and ACO: In Brazil, most of the English textbooks do not use naturalistic interactions to teach speaking. What is your opinion about the contributions and challenges of using naturalistic data?

JS: Usually, fully authentic dialogue is hard to use in an ESL classroom, because (a) it is extracted from its context, (b) it will likely contain references to local issues that will be obscure, and (c) it will probably contain more disfluencies, hesitations and repetitions than needed. So, I would advocate edited versions of "real" spoken data.

PCSC and ACO: In your article "*English as a Tyrannosaurus Rex*" (1997), you discussed the notion of "audience-design" as a way of adjusting the level of specificity and detail depending on the context. The notion of audience design has been largely discussed in Conversation Analysis as well, as a principle with which we, competent speakers and members, operate all the time in our daily interactions. What is the relevance of such notion to teaching academic speech?

JS: My sense of audience design does not, in fact, come from CA's 'recipient design', but from the 'audience analysis' that was fashionable in Technical Communication when I came to the U.S. in the mid 1980s. Two of the leaders here were Mathes and Stevenson, who were Technical Communication professors in the Faculty of Engineering here at the University of Michigan. A recent, relevant development has been the 3MT movement (3 minutes to talk about your doctoral research). My close colleague, Christine Feak, is the co-organizer of the competition here, and she believes that learning to target a general, educated audience by finding a way to summarize complex projects in three minutes in a clear and convincing way is very valuable training experience for the contemporary world.

APPENDIX A. How to do MICASE research: Starting with a functional category - John M. Swales

Starting with a lexical item (such as we have seen with CONCERN) is relatively straightforward, at least initially, because one can be fairly sure of capturing all the tokens in the MICASE database. Starting with a functional category, in contrast, means searching the grammatical and pragmatic literature as well as racking one's brains in order to come up with a list of possible realizations.

Making a start

The category I have chosen to exemplify possible procedures is that of **making suggestions.** Now, of course, what 'counts as' a suggestion presents itself an immediate problem. As a way in, as it were, I have opted for a broad category that includes suggestions themselves, giving advice, and, a bit more dubiously, making recommenda-

tions. So the first task is to assemble a list of potential candidates for these roles. Here are mine (using a stable constructed example in each case):

> (1) I suggest you drop the course
> (2) My suggestion is/would be to drop the course
> (3) I'd/I would drop the course
> (4) I advise dropping the course
> (5) My advice would be to drop the course
> (6) If I were you, I would drop the course
> (7) Why don't you drop the course?
> (8) Why not drop the course?
> (9) You might wanna/want to drop the course
> (10) You could drop the course
> (11) I recommend dropping the course
> (12) My recommendation is to drop the course
> (13) How about dropping the course?
> (14) What about dropping the course?

So we have 14 possible realizations, some using nouns, some lexical verbs, some modals, and some fixed expressions such as "how about..?". (Getting this list together probably took me about two hours.)

Concordancing and quantification

The next stage involves scrutinizing the concordance lines in order to establish which and how many fall within what we might call the *suggest complex.* This is easier with a program such as Wordsmith Tools that has a *delete* function, but it can be done via manual coding on the website. This is going to be a time-consuming task; in fact, so time-consuming that I have decided not to bother with #3 ("I'd/I would") because there are more than 1150 examples. (This decision may come back to haunt me.) The procedure I have outlined is also not going to be an exact science. Sometimes, we don't have enough context to make a firm judgment; at others, there is room for reasonable doubt. So, we do the best we can. The results on Table 1 contain (a) the target structure; (b) the total number of tokens; (c) the number of tokens that fall within the "suggest complex"; and (d) the percentage of (c) in terms of (b).

This took me at least eight hours, including an hour for the write up. So let's hope the effort was worth it! To test this out, we will first see what might be concluded from the table and from the examples I jotted down as I went along.

Commentary on the table

(a) There seem to be three common ways of making suggestions in MICASE: *you could*, *why don't you/we*, *you might wanna/want to*. However, the 24% percentage rate for *you could* show that it is highly multifunctional. Most of the time, it operates to discuss possibilities or general options:

Table 1. Results of the concordance lines.

Target item	Total	Suggest-complex	Percentage
you could	796	193	24%
what about	209	12	6%
why don't you/we	141	120	85%
how about	106	39	37%
you might wanna	104	104	100%
suggest	96	46	48%
suggestion/s	65	4	6%
why not	57	15	26%
recommend	40	24	60%
recommendation/s	23	2	9%
advice	17	3	18%
advise	5	2	40%
if I were you	2	2	100%

(1) you could allow harvesting beyond a certain size if that's somehow correlated with age class
(2) you could actually call it a conditional
(3) you could ask that question, I guess

In contrast, the other two formulae are much more closely associated with the *suggest complex*—almost entirely so with *you might wanna*, largely so with *why don't you/we*, where the exceptions are typically rhetorical questions, as in:

(4) why don't we get the stability at just one?

(b) There is a considerable difference in the percentages of "suggest" uses for the two prefabs *how about…?* and *what about…?*. The great majority of the *what about…?* consist—as might be expected—of instructor-initiated questions. Suggestions are rarer, and largely fall into one of these three syntactic frames:

(5) what about the two thirty slot on Tuesday? (NP)
(6) yeah but what about getting a double degree? (V-ing)
(7) yeah what about if if we change the order of presentations (if-clause)

In fact, a similar set of syntactic possibilities seems to be available for the more frequent use of *how about..?* as a way of making a suggestion:

(8) how about a show of hands…
(9) how about figuring out initial Q, again.
(10) okay. how about if we start with Erin, and just go around the room…

(c) A next observation might invoke negative evidence. For example, we can notice that the nouns *suggestion*, *recommendation* and *advice* are very rarely used to make suggestions (4, 2 and 3 tokens respectively). Here is the complete list for *advice*:

(11) my advice to you would be to, uh, try and read the stuff before lecture..
(12) one of the most important pieces of advice I could give you…is to try to..
(13) pick your fights hm that's my best advice.

Although there are more examples of the verb *suggest* to make suggestions (46 in all), these are in fact under 20% of the total occurrences of the lemma. Overall, the data is pretty clear; these nouns and verbs are not used to *make suggestions*, but to either *report* them, as in:

(14) also we suggested to her, that she look to local businesses..
(15) a constant friction suggests something closer to a solid than to a liquid
(16) Pat was suggesting a much more scientific mhm writing format

Pragmatic differences

(a) why don't you/you might wanna

With the frequency data nailed down, we can now turn to possible pragmatic differences. Look at these examples of *why don't you*:

(17) why don't you grab a chair and join us?
(18) why don't you look it up, find out?
(19) why don't you zip through this introduction?
(20) why don't we take a five minute break?

As these examples show, this form of suggestion-making tends to be used when what is being suggested is simple, straightforward and not particularly onerous.

Now consider the alternate *you might wanna* formula:

(21) you might wanna talk about this difference…
(22) and since you missed some stuff you might wanna do all the homework
(23) you might wanna check with your T-A about this
(24) uh I think you might want to probe a little deeper into this

As we can see, these appear to be somewhat more "imposing" and time-demanding. In consequence, a more polite form is adopted. The *might wanna* is already hedged, so a further hedge, such as in 24), is not that common. But here is an exceptional example (my emphases):

(25) you might *just* wanna *like* again *sort of* draw out some conclusions from this *a little bit.*

(b) why not/how about

In a 2002 paper Svenja Adolphs examines these two prefabs and concludes: We find, for example, that

'why not' is mostly used to address some wider issue in the context of a discussion, with the aim of complaining or lamenting. 'How about', on the other hand, is used in suggestions directed towards an identified problem related to other participants in the conversation (2002, p. 58).

Her data is restricted to just a few examples and is based on British talk. So, how does this interesting distinction pan out in MICASE? Certainly, the 15 relevant examples of 'why not' take place in largely lecture contexts and are not related to individual problems of the interlocutors, as in:

> (26) so if you can have the cliffs along the beach and they are, undergoing,this slow gradual change, why not organisms as well?…so somebody had come up with this, this idea, and you've probably all heard of Lamarck...

On the other hand, as many as seven of the 15 are strawman arguments setting up a later denial of the 'why not' proposition, as in:

> (27) …one could ask why not just a dump a lot of this stuff in your engines, …um uh they will work, but one has to be careful because they work, ultimately, by corroding the metal…

In contrast, the 39 'how about' suggestions tend to occur in more interactive contexts, and when they do occur in lectures they tend, like examples 8 and 10, to be making suggestions about classroom procedures. In the more discursive contexts, they rarely deal—as we might expect—with personal issues and problems, but with suggestions about word choices, about calculations, or about other rather small-scale issues:

> (28) how about improving or how about um, enhancing, protecting…?
>
> (29) how about figuring out initial Q again?
>
> (30) how about if you couple it with glucose?

Distribution

A final issue that we can explore is the distribution of suggestions. If we just focus on the two formulae in the original list that are most predictive of suggestions (*why don't you/you might wanna*), then there are eleven speech events where they occur more than 0.3 times per thousand words (using Wordsmith's plot feature). These consist of:

- 3 office hours
- 2 advising sessions
- 2 small lectures
- 1 lab
- 1 meeting
- 1 seminar
- 1 service encounter

As this list shows, most of these fall towards the highly interactive end of the MICASE continuum.

Conclusions and implications

We have now reached the preliminary end of the story—a story that has taken some 16-18 hours to construct. It shows that suggestions form a fairly broad category and that they have the overall function of trying to change a situation, often one raised in the immediately preceding discourse, and hopefully for the better. In the MICASE data, there are a limited number of formulae for doing this, and these rarely employ the lexical items like *suggest* that we might have expected. We have also seen that *don't you* and *you might wanna* tend to be used in different contexts of imposition, and that *why not* and *why don't you* also function in very different ways. All these are points that can usefully be brought to the attention of those who are attempting to acculturate to the spoken discourse in a US university.

References

ADOLPHS, S. 2002. Genre and spoken discourse: Probabilities and predictions. *Nottingham Linguistic Circular,* **17:**47-60.

SWALES, J.M.; FEAK, C.B. 2004. *Academic Writing for Graduate Students: Essential Tasks and Skills*. Ann Arbor, The University of Michigan Press, 432 p.

SWALES, J.M. 2004. *Research Genres: Explorations and Applications.* New York, Cambridge University Press, 328 p. http://dx.doi.org/10.1017/CBO9781139524827

SWALES, J.M. 1997. English as a Tyrannosaurus Rex. *World Englishes*, **16**(3):373-382. http://dx.doi.org/10.1111/1467-971X.00071

SWALES, J.M. 1990. *Genre Analysis: English in academic and research settings*. Cambridge, Cambridge University Press, 274 p.

SWALES, J.M. 1971. *Writing Scientific English*. Sunbury, Thomas Nelson, 167 p.

Paula Cortezi Schefer Cardoso
Universidade do Vale do Rio dos Sinos
Av. Unisinos, 950, 93022-000, São Leopoldo, RS, Brasil

Ana Cristina Ostermann
Universidade do Vale do Rio dos Sinos
Av. Unisinos, 950, 93022-000, São Leopoldo, RS, Brasil

John Swales
Universidade de Michigan
Rua Tappan, 611, 48109-1220, Ann Arbor, Michigan, Estados Unisinos