



Spoken language grammar and discourse-pragmatics

Regina Weinert

University of Sheffield/Ikerbasque

Spoken language grammar and discourse-pragmatics

Regina Weinert

University of Sheffield/Ikerbasque

Research Background

- ✧ Spoken language is primary in unimpaired human beings.
- ✧ Spontaneous, unplanned spoken and formal written language overlap, but also differ substantially in the nature of their structures.
- ✧ Written language in general and spoken language in general cannot be clearly differentiated; genre and socio-psychological factors, including level of education are important.
- ✧ Spontaneous spoken language has specific grammatical structures as well as interactional features. These require to be analysed in relation to the conditions of spoken language production. The typical features found in spoken language are therefore not considered deviations from “the (written) norm”, due to performance limitations.
- ✧ The nature of spoken language has implications for linguistic theory, language typology, analysis of individual languages, language acquisition and applied linguistics.
- ✧ Spoken language research is data-driven (by necessity) and most approaches adopt a “usage-based” theoretical model of language. They show how linguistic structures can be shaped by communicative and cognitive factors and by experience. Data-based, usage-based or corpus-based study is also a practical way of addressing teaching and learning issues (spoken and written).

Conditions of spontaneous spoken language production

- ✧ Spoken language is linear (we cannot delete or change what we have already said), fast (compared with writing), and subject to working memory constraints.
- ✧ Spoken language is supported by prosody, non-verbal signals, context and interaction.

Features of spontaneous conversation

Broad overview based on English, German, Russian, French and other languages. The exact details of the structures and their functions are often complex and have filled many book and article publications.

- ca 80% of all clauses are main clauses
- units below clause size are common
- most “subordinate” clauses follow the main clause
- syntactic structures are often only loosely integrated
- single “subordinate” clauses are used for pragmatic purposes
- center embedding is infrequent
- 60% or more of noun phrases are pronouns
- most other NPs consist of noun or 1 determiner + noun
- formulaic language is common: *it's a big business; I get the impression; how could that be; I know it's tough for X*
- deictics and discourse markers are very frequent
- fillers and hearer signals are common

Native spoken language grammar

Text 1: Academic consultation

Speaker B is a lecturer and Speaker A is a doctoral student. We can assume that both speakers have a high level of education. The doctoral project is about people who are trying to give up smoking and who are worried that they may eat too much as a consequence.

Underlining indicates overlapping speech
- indicates a change in structure

- A1: I didn't think keeping a diary of everything they ate was really a feasible
B1: they're not going to do it
A2: no exactly
B2: are they
A3: no
B3: but you - they might have something - you might have a question like ehm how much did you eat today ehm less than usual about the same as usual
A4: yeah something like that
B4: yeah
A5: something really simple which they can just pop it in
B5: yeah I really like this I think that cigarette diaries are really good I can see the results here you've got the three groups and you could have
A6: mhm
B6: eh total cigarette consumption in one week
A7: yeah
B7: and week two and week three
A8: yeah that's what I thought I thought it would be quite nice

(see Appendix for analysis)

Relative clauses

- ✧ The relative pronoun, form and grammatical function
- ✧ The head noun (the noun referred to by the relative pronoun)

Relative Pronoun

In spoken English **that** and **zero** are the most common relativisers:

- (1) I liked the book that you gave me
- (2) I like the book you gave me

Who and **which** also occur:

- (3) that is the actor who refused to perform in front of a noisy audience
- (4) I like novels which teach you something

Whose and whom or **preposition + whom** are extremely rare. If a pronoun is used, alternatives are preferred:

- (5) I liked the man who I met yesterday
- (6) I liked the man who I talked to yesterday

It is not uncommon in spoken English to have **resumptive** pronouns:

- (7) they give you a thing that you don't want it

Head noun

Head nouns are very rarely subjects (in both written and especially in spoken language). This avoids center embedding:

- (8) The woman [who phoned yesterday] is no longer interested in the house
- (9) I spoke to the woman [who phoned yesterday]

One way of avoiding center embedding is to insert a pronoun (but this is not always the only or main reason for doing so):

- (10) The woman [who phoned yesterday] she is no longer interested in the house

Adverbial clauses

because/cos

Two orders are possible:

- (1) the students stopped talking because the strict teacher came into the room
- (2) because the strict teacher came into the room the students stopped talking

The above examples can be written and spoken, but in spoken language the order in a) is highly preferred.

The order of clauses also affects their internal structure, which means that the clauses are not exact mirror images, and not all structures are reversible (* means “unacceptable” or “ungrammatical”):

- (3) the students stopped talking because into the room came the strict teacher
- (4) *because into the room came the strict teacher the students stopped talking

Because-clauses which follow the main clause are less integrated, often prosodically independent.

In (1-4) there is a *semantic* relationship. But the relationship can also be *pragmatic*:

- (5) there must be a fire ↯ cos I can see smoke rising

This is sometimes referred to as “epistemic” *because*. In other cases the reason is given for uttering a speech-act such as a question or requests:

- (6) are you going to the post office? cos I have some letters to send
- (7) could you answer the phone please cos my hands are wet

In these pragmatic cases the *because/cos*-clause cannot be placed first.

Because vs cos – semantic vs discourse-pragmatic connection?

A: I gather you're open to persuasion on the eh this course work assignment

B: mhm I am yes

A: because I couldn't get to the lecture last week and eh I had a look at the handouts
(*laughs*)

B: (laughs) and it frightened you did it

A: yeah

If-clauses

It is common for *if-* or *when-*clauses to come before the main clause. **The temporal or logical order is maintained:**

- (1) when you come home we'll go to the cinema
[event 1] followed by [event 2]
- (2) if you behave you will get a nice present
[condition] followed by [consequence]

*if-*clauses also appear as **independent requests:**

- (3) if you'd like to come through (uttered at the dentist's, for example)
- (4) if you just sign here

In some cases, there may be a conditional structure with an *if-*clause and a main clause, but the *if-*clause is also an **instruction**, for example, when we give people directions:

- (5)
Speaker A: if you turn round now
Speaker B: yeah [speaker turns round]
Speaker A: you'll see your surprise present
Speaker B: oh

Cleft constructions

Focus on noun phrase:

Main clause: I want to talk about climate change

WH-cleft: [What I want to talk about] is [climate change]

Reverse WH-cleft: [Climate change] is [what I want to talk about]

Focus on non-finite verb construction:

Main clause: I want to play the accordion

WH-cleft: [What I want] is [(to) play the accordion]

Reverse WH-cleft: [(To) play the accordion] is [what I want]

WH-clefts

WH-clefts tend to be used to introduce topics or points and can open a conversation. They can also signal that the speaker is moving on, providing a reference point. In spoken language *be* can be present, but the structures are often very loose:

- (1) what I wanted to tell you you really do use dried chickpeas
- (2) what I say the danger from whooping cough is far greater than the danger from vaccination
- (3) what you do you twist it like this

Reverse-WH-clefts

This cleft is often used to conclude a part of a conversation and to sum something up.

Written: [another series of new complicated procedures] is certainly not [what we need]

The spoken clefts almost all start with the deictic (demonstrative pronoun) *that*:

- Spoken:
- (1) that's what I meant
 - (2) that's where you stop
 - (3) that's when I realised it was too late

The subject is usually *I*, and the most common verbs are *say, tell, mean, think, wonder, want*.

Non-native spoken language

- ✧ Non-native spoken language has tended to be examined from different perspectives such as accuracy and complexity.
- ✧ Non-native language is only just beginning to be examined explicitly in the light of research on native spoken language (LINDSEI).
- ✧ Studies on pragmatics and discourse markers predominate; syntax has received less attention.

The Ikerbasque Project

Title: The spoken English of advanced foreign language learners with L1 Basque/Spanish

Principal Investigator: Regina Weinert (regina_weinert@ehu.es)

Research Assistant: María Basterrechea

UPV/EHU Associate: María Pilar Garcia Mayo

Aims

The aim of the project is to examine non-native spoken English in relation to typical native speaker *usage* (not in terms of “correctness”, although sometimes the line is difficult to draw). The main areas to be investigated are subordinate/dependent clauses and focusing constructions. The use of pronouns, deictics and discourse markers is also of interest. The work provides quantitative overviews, but it is largely qualitative.

The IkerSPEAK Corpus

22 conversations with an additional short picture task (ca 66 000 words)

22 direction-giving dialogues (ca 17 000 words)

Participants: mostly third and fourth year students, some postgraduate students

Methodology

Corpus data

- ✧ Researching spoken language is very labour-intensive. It requires volunteers, recordings, transcriptions and qualitative analysis. Persuading speakers to be recorded is not easy.
- ✧ Transcribing data takes a long time. One 15-minute conversation can take between 6 and 10 hours to transcribe by an experienced transcriber. This gives you the words and fillers (eh, ehm etc.), overlaps and pauses. If more than two speakers are involved, the work takes even longer. Transcriptions then need to be checked.
- ✧ One 15 minute conversation may be about 3000 words long. For grammatical studies at least 25 000 - 50 000 words are needed for detailed analysis of a genre and 250 000 - 1 000 000 words for general overviews and for lexis.
- ✧ 1 million word corpus of spoken language transcriptions:

$$\begin{array}{l} \text{♀} \quad \text{♂} \quad \quad \quad = 3 \text{ years} \\ \text{♂} \quad \text{♀} \quad \text{♂} \quad \text{♀} \quad \text{♂} \quad \text{♀} = 1 \text{ year} \end{array}$$

Transcription and coding

- ✧ Segmenting spoken language into units is difficult.
- ✧ There is consensus among most researchers that spoken language cannot be divided up into sentences (they are a written language unit). Clauses can be identified (although the beginning and end cannot always very easily be determined) and smaller phrases and units are common.
- ✧ There is much debate about data coding and there are many approaches, but this is part of the analysis and depends on the research purpose.

Applications of Spoken Language Research

- ✧ Deciding which native speaker structures and interactional features should in principle be taught

(standard usage, genre and context specific usage, sociolinguistic issues)

- ✧ Deciding which native speaker features learners underuse or overuse, which aspects are not difficult for learners and which native features maybe considered problematic from a sociolinguistic point of view

(assessing learner language and their communicative intentions)

- ✧ Deciding at which proficiency levels features should be introduced and which ones

- ✧ Deciding how to present spoken language

(authentic transcriptions, adapted transcriptions, scripted conversation)

Appendix

Comments on Text 1

In A1 the speaker probably leaves out “idea” at the end, but Speaker B understands.

In B3 the speaker changes her mind twice (indicated by “- “). There are 10 main clauses; there are 3 complement clauses (A1, B5, A8); all of the main clauses to which they relate are short and involve the verb “think”; two have no complementiser, one is introduced by “that”. There is one contact relative clause - no relative pronoun is used (A1). In A5 there is a relative clause introduced by the relative pronoun “which”; the clause contains a resumptive pronoun (“which they can just pop it in” see the section on relative clauses on p. 5). There is one cleft in A8 (see the section on clefts on p. 8).

There are a few units which are not clauses (B3, A4, A5). The proportion of pronouns is high, despite the topic. In A4 “something like that” looks formulaic. The conversation is interactional, shown by the use of “yeah” and the tag in B2. We would need to listen to this extract to determine some boundaries, for example B5 “I can see the results here you’ve got the three groups” - does “here” go with the preceding or the following clause? In terms of content “here” probably goes with the following clause since there are no results yet of the PhD project. B5 underlines B’s attitude with “really”. There are not many hesitation markers or fillers (eh, ehm).

What can I say?

- (1) there's nine of us in the house
- (2) if I yawn a lot it's cos I didn't get in till quarter to six this morning
- (3) that's the thing about ivy is that it's evergreen
- (4) what you do you first go straight on then at the park you turn left
- (5) it was very interesting because there was a lot of teachers
- (6) I think that when you go to a foreign country you have to get used to the new culture
- (7) there are other people whom I wish I hadn't known
- (8) yes now it makes sense what I have been taught
- (9) which is good for the children
- (10) if you stop at this point

Selected References

- Anderson, A.H., M. Bader, E.G. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson, and R. Weinert (1991). The map task dialogues: a corpus of spoken English. *Language and Speech*, 34.4, 351-366.
- Barlow, M. and Kemmer, S. (eds), (2000). *Usage-Based Models of Language*. Stanford CA: Centre for the Study of Language and Information.
- Biber, D. (1988) *Variation across Speech and Writing*. Cambridge [etc.]: Cambridge University Press.
- Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan (1999). *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Carter, R. and McCarthy, M. (1995) Grammar and the Spoken Language. *Applied Linguistics* 16/2, 141-158.
- Carter, R. and McCarthy, M. (1997). *Exploring Spoken English*. Cambridge University Press.
- Chafe, W. (ed) (1980). *The Pear Stories. Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. Norwood, New Jersey: Ablex Publishing Corporation (*Advances in Discourse Processes*. III.)
- Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (ed), *Coherence and Grounding in Discourse*. Amsterdam/Philadelphia: John Benjamins. 21-51.
- Cheshire, Jenny (2005). Syntactic variation and spoken language. In L. Cornips and K. Corrigan, *Syntax and Variation. Reconciling the Biological and the Social*. Amsterdam/Philadelphia: John Benjamin. 81-106.
- Corrigan, R., Moravcsik, E.A., Ouali, H. and Wheatley, K.M. (eds.) (2009). *Formulaic Language. Volume 1. Distribution and historical change. Volume 2. Acquisition, loss, psychological reality, and functional explanations*. (Typological Studies in Language 83.) Amsterdam & Philadelphia: John Benjamins
- De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literatures (BELL), New Series 2*, 225-246.
- De Cock, S. (2007). Routinized building blocks in native speaker and learner speech: Clausal sequences in the spotlight. In M.C. Campoy & M. J. Luzón (eds) *Spoken Corpora in Applied Linguistics*. Bern: Peter Lang. 217-233.
- Ford, C. (1993). *Grammar in interaction. Adverbial clauses in American English conversation*. Cambridge University Press.
- Gilquin, Gaëtanelle (2008). Hesitation markers among EFL learners: pragmatic deficiency or difference? In Jesús Romero-Trillo (ed.) *Pragmatics and Corpus Linguistics: A Mutualistic Entente*. Berlin, Heidelberg & New York: Mouton de Gruyter.
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London / New York: Longman . (*Studies in Language and Linguistics*. S.) 60-85.
- McCarthy, M. J. (1998). *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.
- McEnery, T. & Wilson, A. (2001). *Corpus Linguistics. An Introduction*. 2nd edition. Edinburgh: Edinburgh University Press 2001. (*Edinburgh Textbooks in Empirical Linguistics*.)
- Miller, J. and Weinert, R. (1995). The function of LIKE in spoken language. *Journal of Pragmatics*, 23, 365-393.

- Miller, J. and Weinert, R. (1998/2009). *Spontaneous spoken language. Syntax and discourse*. Oxford: Oxford University Press.
- Miller, J. and Weinert, R. (2002). Lengua hablada, teoría lingüística y adquisición del lenguaje. In E. Ferreira (ed.) *(In)dependencia entre el análisis de la oralidad y la escritura*. In the Series: Lenguaje, Escritura, Alfabetización. Barcelona/Buenos Aires: Gedisa. 77-110.
- O’Keeffe, A., McCarthy, M. and Carter, R. (2007). *From Corpus to Classroom*. Cambridge: Cambridge University Press.
- Ramírez, M.D. & J. Romero (2005). The pragmatic function of intonation in L2 discourse: English tag questions used by Spanish speakers. *Intercultural Pragmatics* 2, 151-168.
- Romero, J. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics* 34, 769-784.
- Thompson, S.A. and Mulac, A. (1991). The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of Pragmatics*, 15, 237-251.
- Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics* 16.2, 180-205.
- Weinert, R. (1995). Focusing Constructions in Spoken Language. Clefts, Y-movement, Thematization and Deixis in English and German. *Linguistische Berichte* 159, 341-369.
- Weinert, R. (1998). Discourse organisation in the spoken language of L2 learners of German. *Linguistische Berichte*, 176, 459-488.
- Weinert, R. (2004). Relative Clauses in spoken English and German – Their structure and function. *Linguistische Berichte*, 197, 3-51.
- Weinert, R. (ed.) (2007). *Spoken language pragmatics. An analysis of form-function relations*. London/New York: Continuum.
- Weinert, R. (2010). Formulaicity and usage-based language: linguistic, psycholinguistic and acquisitional manifestations. In D. Wood (ed). *Perspectives on formulaic language in acquisition and communication*. London/New York: Continuum. 1-20.
- Weinert, R. and Miller, J. (1996). Cleft constructions in spoken language. *Journal of Pragmatics*, 25, 173-206.

Corpora of spoken language

Basque

corpus collected by Jon Aske:

<http://www.elda.org/catalogue/en/speech/S0123.html>

Spanish

COREC = Corpus de Referencia de la Lengua Española Contemporánea: Corpus Oral Peninsular, director F. Marcos Marín. Available at: www.llf.uam.es/~fmarcos/informes/corpus/corpusix.html (género conversacional)

CREA Subcorpus Oral del Corpus de Referencia del Español Actual at:

<http://www.rae.es/rae.html>

Romance Languages

Cresti, E. and Moneglia, M. (2005). C-ORAL-ROM: integrated reference corpora for spoken Romance languages. Amsterdam/Philadelphia: John Benjamins.

English

British National Corpus (BNC) available at: <http://www.natcorp.ox.ac.uk/>

CANCODE available at:

<http://www.cambridge.org/es/elt/catalogue/subject/item2701617/Cambridge-International-Corpus/>

The British Academic Spoken English Corpus available at:

http://www.reading.ac.uk/AcaDepts/ll/base_corpus/

German

Institut für deutsche Sprache, Mannheim at:

<http://www.ids-mannheim.de/>

Archiv für gesprochenes Deutsch available at:

<http://agd.ids-mannheim.de/html/dgd.shtml>