Annotations and tools for an activity based Spoken Language Corpus

Jens Allwood, Leif Grönqvist,
Elisabeth Ahlsén and Magnus Gunnarsson

1. Introduction

The paper contains a description of the Spoken Language Corpus of Swedish at the Department of Linguistics, Göteborg University, and a summary of the various types of analysis and tools that have been developed for work on this corpus. The corpus is an incrementally growing corpus of spoken language which presently consists of 1,3 million words from about 25 different social activities. It is based on the fact that spoken language varies considerably in different social activities with regard to pronunciation, vocabulary, grammar and communicative functions. The goal of the corpus is to include spoken language from as many social activities as possible to get a more complete understanding of the role of language and communication in human social life. The corpus is based on audio (50%) or video/audio (50%) recordings of naturalistically occurring interactions. The recordings have been transcribed according to the transcription standard Modified Standard Orthography MSO, which is a standard for transcription which is more faithful to spoken language than Swedish standard orthography but less detailed than a phonetic or phonematic transcription would be. In MSO, standard orthography is used unless there are several spoken language pronunciation variants of a word. When there are several variants, these are kept apart graphically. Although the goal is to keep transcription simple, MSO includes features of spoken language such as contrastive stress, overlaps and pauses. MSO also includes procedures for anonymizing transcriptions and for introducing comments on part of the transcription. Besides describing our corpus and standard of transcription, we will also describe several tools we have developed for using the corpus. The corpus has been used for various kinds of quantitative and qualitative analysis which will be briefly reported. A book of frequencies of Swedish spoken language has been produced. The book contains word frequencies both for the words in MSO format and in standard orthographic format. It also contains comparisons between word frequencies in spoken and written language. There is statistics on the parts of speech represented in the corpus, based on an automatic probabilistic tagging, yielding a 97% correct classification. The corpus has also been the basis for work using various kinds of manual coding, e.g. communication management (including hesitations, changes, feedback and turntaking), speech acts, obligations, misunderstandings, etc. Finally, we point to ways of using the corpus for other types of qualitative analysis, e.g. for CA-related sequential analysis. The corpus is continuously being digitized using CD:s with Mpeg compression. Each CD contains both transcriptions and recordings.

2. Corpus Description

The Göteborg Spoken Language Corpus consists of spoken language samples from several languages. So far the following languages are included. Table 1. Over and above this we also work with other spoken language corpora collected by other teams.

Table 1. Spoken language corpora at Göteborg University, Department of Linguistics (Pars of the corpora are Multimodal.)

- Göteborg Spoken Language Corpus (Kernel Corpus - adult first language Swedish), 1.3 million words
- Adult language learners of Swedish, 2 million words
- Aphasic speakers
- Child language corpus (Swedish and Scandinavian), 0.75 million words including the adults
- Educational progress, 416 interviews, 2 million words
- Non-Swedish adult spoken language corpus
  * Chinese (70 000 words)
  * Bulgarian (25 000 words)
  * Arabic
  * English (10 000 words), BNC
  * Finnish
  * Italian (3000 words)
  * Norwegian (140 000 words)
  * Spanish
- Wizard-Of-Oz Corpus, Bionic
- Intercultural communication corpus

The kernel corpus of adult first language Swedish speakers is the corpus we will focus on in this article. In Table 2, below, we present basic data on this corpus. The corpus is organized on the basis of social activities rather than for example on the basis of dialects or categorizations of speakers such as social class or gender. However, regroupings of, or selections from, the corpus according to criteria such as these are also possible. The limitations which exist in our ability to create subcorpora are dependent on the fact that we do not always have the relevant information about individual speakers.

Table 2. The Göteborg Spoken Language Corpus

Type of social activity	**No. of recordings**	**Average number of speakers / recording**	**Numbe r of sec-Tions***	**Tokens (including pauses and comments)**	**Audible word tokens uttered in recording**	**Duration****
Auction	2	6.0 *	111	26 776	26 459	3:14:11
Bus driver/ passenger	1	33.0 *	20	1 360	1 345	0:13:33
Consultation	16	3.0 *	239	34 865	34 285	2:44:25
Court	6	5.0 *	79	33 401	33 261	3:58:33
Dinner	5	8.0 *	30	30 738	30 001	2:49:54
Discussion	34	5.8 *	255	240 426	237 583	17:19:24
Factory conversation	5	7.4 *	48	29 024	28 860	2:19:47
Formal meeting	13	9.7 *	186	219 352	215 582	15:45:54
Hotel	9	19.2 *	183	18 950	18 137	6:47:50
Informal conversation	22	4.4 *	152	94 490	93 436	7:48:41
Information Service	32	2.1 *	40	14 700	14 614	0:13:40
Interview	58	2.9 *	1 031	396 758	393 907	30:34:27
Lecture	2	3.5 *	3	14 682	14 667	1:38:00
Market	4	24.2 *	38	12 581	12 175	2:18:37
Religious Service	2	3.5 *	10	10 273	10 234	1:10:45
Retelling of article	7	2.0 *	7	5 331	5 290	0:42:00
Role play	2	2.5 *	7	5 702	5 652	0:39:16
Shop	49	7.4 *	139	36 385	34 976	6:40:46

Task-oriented dialog	26	2.3 *	46	15 475	15 347	2:05:20
Therapy	2	7.0 *	8	13 841	13 529	2:04:07
Trade fair	16	2.1 *	16	14 353	14 116	1:12:46
Travel agency	40	2.7 *	112	40 370	40 129	5:53:57
Total	**353**	**4.9 ***	**2 762**	**1 310 284**	**1 204 029**	**118:15:53**

*A section is a longer phase of an activity with a distinct subordinate purpose. The bus driver/passenger recording, for example, has 20 sections, where each section involves talk with a new passenger.

** The duration has, in most cases, been estimated on the basis of the number of word tokens. The estimate is conservative and probably under-represents actual duration by about 30 hours.

3. Storage

Around 50% of our 1.3 million tokens corpus is stored on audio tapes and the rest on video tapes (Umatic, VHS or BetaCAM). In order to preserve the recordings, tapes are being copied to newer tapes, while simultaneously being digitized. There are several possible formats for storage: (i) Analog video: BetaCAM is probably the best analog video format but VHS is almost the only one used nowadays. (ii) DV (digital video): One mini DV-tape takes 60 minutes or a DVCam 180 minutes. This format requires a fast computer. (iii) Mpeg: We have tried to use a constant data rate of around 200 kb per second and this will give a fair quality and the format may be used on almost any PC/Mac. For a phonetic analysis the sound should not be compressed with mpeg but with some non-destructive method.

4. Description of MSO (Modified Standard Orthography) - the corpus transcription standard

The transcription standard we have used is called MSO (Modified Standard Orthography). It can perhaps most rapidly be explained through exemplification. Consider the example below:

Example 1. Transcription according to the MSO standard with translation.

§1. Small talk

$D: säger du de{t} ä{r} de{t} ä{r} de{t} så besvärlit då

$P: ja ja

$D: m // ha / de{t} kan ju bli så se{r} du

$P: < jaha >

@ <ingressive>

$D: du ta{r} den på morronen

$P: nej inte på MORRONEN kan ja{g} ju tar allti en promenad på förmiddan [1 å0]1 då vill ja{g} inte ha [2 den]2 medicinen å0 sen nä ja{g} kommer hem möjligtvis

$D: [1 {j}a]1

$D: [2 nä]2

$D: oh I see is it it is so troublesome then

$P: yes yes

$D: m // yes / it can be that way you see

$P < yes >

@ <ingressive >

$D: you take it in the morning

$P: no not in the MORNING I always take a walk before lunch [1 and]1 then I don't want [2 that]2 medicine and then when I get home possibly

$D: [1 yes]1

$D: [2 no]2

The example shows the following properties of the transcription standard:

(i) Section boundaries paragraph sign (§). These divide a longer activity up into subactivities. A doctor-patient interview can, for example have the following subactivities. (i) greetings and introduction, (ii) reason for visit, (iii) investigation, (iv) prescribing treatment.

(ii) Words and space between words.
(iii) Dollar sign ($) followed by capital letter, followed by colon (:) to indicate a new speaker and a new utterance.
(iv) Double slash (//) to indicate pauses. Slashes /, // or /// are used to indicate pauses of different length.
(v) Capital letters to indicate contrastive stress.
(vi) Word indexes to indicate which written language word corresponds to the spoken form given in the transcription (å0 corresponds to written language *och*). In the cases where spoken language variants can be viewed as abbreviated forms of written language, we use curly brackets {} to indicate what the standard orthographic form would be, e.g. de {t} = *det*.
(vii) Overlaps are indicated using square brackets ([ ]) with indices which allow disambiguation if several speakers overlap simultaneously.
(viii) Comments can be inserted using angular brackets (< > to mark the scope of the comment and @< > for inserting the actual comment). These comments are about events which are important for the interaction or about such things as voice quality and gestures.

5. Tools which have been developed

The following tools have been developed to aid work related to the corpora.

5.1 TransTool

TransTool is a computer tool for transcribing spoken language in accordance with the transcription standard (Nivre 1999). It will help the user to transcribe correctly and make it much easier to keep track of indices for overlaps and comments (cf. Nivre et al. 1998).

5.2 The Corpus Browser

The Corpus Browser is a tool that makes it possible to search for words, word combinations and phrases (as regular expressions) in the Gothenburg Spoken Language Corpus. The results can be presented as concordances or lists of utterances with as much context as you wish and with direct links to the transcription.

5.3 TRACTOR

Tractor is a coding tool which makes it possible to create new coding schemas and annotate transcriptions. Coded segments can be discontinuous and it is also possible to code relations. A coding schema can be represented as a tree with strings on all nodes and leaves, and a coding value is a path through the tree. That model is similar to the file and folder structure on a computer harddisk. This framework makes it easy to analyze the codings in a Prolog system, but it is not possible to order the codings or code a coding, because a coding only consists of two discontinuous intervals and a coded value.

5.4 Visualization of codings with FrameMaker

This document describes a toolbox that makes it possible to visualize coding schemas and coding values with colors. bold, italics, etc. directly in the transcription as a FrameMaker document. Different parts of the transcription may also be marked (or removed!) to get a legible view of it without all details you are not interested in.

5.5 TraSA

If you have a corpus transcribed according to the Göteborg Transcription Standard, TraSA it is very easy to calculate some 30 statistical measurements for different sections and/or speakers. You will be able to count things like number of tokens, types, utterances, theoretical vocabulary.

5.6 SyncTool

SyncTool is developed (as a prototype for MultiTool) for synchronizing transcriptions with digitized audio/video recordings. It is also meant to be a viewing tool allowing the researcher to view the transcription and play the recording without having to manually locate the specific passage in the recording.

5.7 Work on a synchronizing tool – MULTITOOL

Work has been done on a tool for synchronizing dialog transcriptions with audio and/or video files for the same dialogs (cf. Nivre et al., 1998).

MultiTool is an attempt to build a general tool for linguistic annotation and transcribing dialogues, browsing, searching and counting. The system can handle any number of participants, overlapped speech, hierarchical coding schemes, discontinuous coding intervals, relations and synchronization between codings and the media file. **The internal state:** The fundamental idea is to collect all information in an internal state containing only codings and synchronizations. Even the transcription is made by codings. The internal state can be visualized with a number of different views. **The views:** The Standard View shows one utterance on each line, overlaps and other details that the user wants are marked. The Partiture View has one line for each participant and the codings are viewed in chronological order along the x-axis. This will give a clear view of the dialogue structure and the overlapping sections. The Coding View shows the tree structure of all coded values so far, and their frequencies. Each value can be expanded to the next level in a similar way as Windows Explorer. The Media Player will play audio and video. The user can navigate through the media file to find interesting sections. The Time Scale shows the codings in linear time and the sound waveform which is very useful when aligning coding points and media. **Why views?** One important detail is that the views can be synchronized to show the same sequence when the user scrolls in one of them. The internal state contains all information so it is possible to have many views of the same kind, showing different parts of the dialogue. Changes made in one view will immediately change in the internal state and the other views. **Codings:** A coding consists of two discontinuous intervals (lists of starting and ending coding points), one list of speakers, and a coded value. It should be interpreted as a relation between the two intervals. Transcribed words is a special case where the first interval is continuous and the second an empty list. A synchronization indicates that a specific coding point corresponds to a specific time. **Implementation:** MultiTool is written in JAVA+JMF which makes it platform independent, and the interpreters are rapidly getting more efficient so the performance will probably be good enough on the major platforms very soon. A second prototype is now finished and in use. The architecture makes it easy to expand the system with new type of views. **Download:** Multitool as well as examples and The MultiTool User's Manual may be downloaded from http://www.ling.gu.se/SDS/multitool.

6. Types of quantitative analysis

Using the information provided by the MSO compliant transcriptions, we have defined a set of automatically derivable properties which include the following:

(i) **Volume**: Volume comprises measures of the number of words, words of different lengths, pauses, stresses, overlaps, utterances, turns relative to speaker, activity and subactivity.
(ii) **Ratios:** Various ratios can then be calculated based on the volume measures.

For example:

MLU	=	words/utterances
% pauses	=	100*pauses/(words+pauses)
% stress	=	100*stressed words/words
% overlap	=	100*overlapped words/words
speed=		words/duration

Alternatively, pause, stress and overlap can be given per utterance. All of these measures can then be relativized to speaker, activity or subactivity.

(iii) **Special descriptors**: One example of a special type of descriptor is "vocabulary richness" as measured through type/token, Guiraud, Über, Herdan or "theoretical vocabulary", cf. van Hout & Rietveld (1993). Other descriptors we have constructed are "stereotypicality" which looks at how often words and phrases are repeated in an activity, "verbal dominance" and "verbal equality", "liveliness" and "caution", and overlap in different utterance positions.

(iv) **Lemma:** We also implemented a simple stemming algorithm which enables us to collect regularly inflected forms together with their stem.

(v) **Parts of speech**: Parts of speech are assigned using a probability based statistical (Viterbi - trigram) parts of speech tagger which has been adapted to spoken language. Using this, a parts of speech coding has been done for the whole Göteborg Spoken Language Corpus, roughly 1.3 million transcribed words. The correctness of the coding is about 97% (cf. Nivre & Grönqvist, 1999). Words subdivided according to parts of speech can then be assigned to speaker, activity or subactivity.

(vi) **Collocations**: All speakers, activities and subactivities can be characterized in terms of their collocations, sorted by frequency as complete utterances or by "mutual information".

(vii) **Frequency lists**: Frequency lists can be made for words, lemmas, parts of speech, collocations, and utterance types.

(viii) **Sequences of parts of speech**: Utterances of different length can be characterized as to sequence of parts of speech. This allows a first analysis of grammatical differences between speakers, activities and subactivities.

(ix) **Similarities:** Similarities between activities are captured by looking at the extent to which words and collocations are shared between activities.

7. Types of qualitative analysis

7.1 Overview

In order to increase reliability, qualitative analysis in Göteborg has often resulted in the development of coding schemas. The following provides an overview of the Göteborg coding schemas (cf. Alwood 2001):

1. Social activity and Communicative act related coding
   - 1.1 Social activity
   - 1.2 Communicative acts
   - 1.3 Expressive and Evocative functions
   - 1.4 Obligations
2. Communication management related coding
   - 2.1 Feedback
   - 2.2 Turn and sequence management
   - 2.3 Own Communication Management
3. Grammatical coding

3.1 Parts of speech (automatic, probabilistic)
3.2 Maximal grammatical units
4 Semantic coding.

7.2 Contributions, utterances and turns

Following Grice (1975), Allwood, Nivre and Ahlsén (1990) and Allwood (1995), the basic units of dialog are gestural or vocal *contributions* from the participants. The term *contribution* is used instead of *utterance* in order to cover also gestural and written input to communication. Verbal contributions can consist of single morphemes or be several sentences long. The term *turn* is used to refer to the right to contribute, rather than to the contribution produced during that turn. One may make a contribution without having a turn and one may have the turn without using it for an active contribution, as demonstrated in the example below, in which B's first contribution involves giving positive feedback without having the turn (square brackets indicate overlap) and his second contribution involves being silent and doing nothing while having the turn.

A: look ice cream [would] you like an ice cream
B1: [yeah]
B2: (silence and no action) [51]

Contributions, utterances and turns are not coded since they are obtainable directly from the Göteborg transcription standard – MSO.6 (Modified Standard Orthography, version 6).

7.3 Coding related to Social activity and Communicative acts

7.3.1 Social activity

Each transcription is linked to a database entry and a header containing information on.
(i) The purpose, function and procedures of the activity
(ii) The roles of the activity
(iii) The artefacts, i.e. objects. furniture, instruments and media of the activity
(iv) The social and physical environment
(v) Anonymous categorical data on the participants, such as age, gender, dialect and ethnicity.
In addition, the major subactivities of each activity are given.

7.3.2 Communicative Acts

Each contribution can be coded with respect to one or more communicative acts which can occur sequentially or simultaneously. The communicative acts make up an extendible list, where often used types have been provided with definitions and operationalizations. Some often used types are the following: Request, Statement, Hesitation, Question, Answer, Specification, Confirmation, Ending interaction, Interruption, Affirmation, Conclusion, Offer.

7.3.3 Expressive and evocative functions

In accordance with Allwood (1976, 1978, 2000), each contribution is viewed as having both an *expressive* and an *evocative* function. These functions make explicit some of the features implied by the communicative act coding. The *expressive* function lets the sender express beliefs and other cognitive attitudes and emotions. What is "expressed" is made up of a combination of reactions to the preceding contribution(s) and novel initiatives. The *evocative* function is the reaction the sender intends to call forth in the hearer. Thus, the evocative function of a statement normally is to evoke a belief in the hearer, the evocative function of a question is to evoke an answer, and the evocative function of a request to evoke a desired action.

7.3.4 Obligations

If the dialog and communication is to be cooperatively pursued, whether it be in the service of some activity or not, they impose certain obligations on both sender and receiver. With regard to both expressive and evocative functions, the sender should take the receiver's perceptual, cognitive and behavioral ability into consideration and should not mislead, hurt or unnecessarily restrict the freedom of the receiver. The receiver should reciprocate with an evaluation of whether he/she can hear, understand and carry out the sender's evocative intentions and signal this to the interlocutor. The sender's and receiver's obligations can be summarized as follows (see also Allwood 1994):
Sender: 1. *Sincerity*,. 2. *Motivation*, 3. *Consideration* (cf. Allwood 1976)
Receiver: 1. *Evaluation*,. 2. *Report*, 3. *Action.*

7.4 Communication management related coding

7.4.1 Introduction

The term "communicative management" refers to means whereby speakers can regulate interaction or their own communication. There are 3 coding schemas related to communication management (cf. Nivre, Allwood & Ahlsén 1999): 1) Feedback coding, 2) Turn and sequence management coding, and 3) Own Communication Management (OCM) coding.

7.4.2 Feedback coding schema

A feedback unit can be described as "a maximal continuous stretch of utterance (occurring on its own or as part of a larger utterance), the primary function of which is to give and/or elicit feedback concerning contact, perception, understanding and acceptance of evocative function" (Allwood, 1988). All feedback units are coded with respect to "Structure", "Position/Status" and "Function". Coding structure means coding grammatical category (part of speech, phrase or sentence) and also "structural operations". "Structural operations" is subdivided into "phonological", "morphological" and "contextual" operations, each of which have different values.

7.4.3 Turn and sequence management coding

Turn and sequence management coding encompasses the following phenomena:

(A) Overlap and interruption: Overlap is coded in the transcriptions and can be extracted automatically. Interruption is a code for those overlaps which aim/at or succeed in changing the topic or taking away the floor from another speaker.
(B) Intended recipient: This type of coding has 4 self explanatory values
(i) particular participant
(ii) particular group of participants
(iii) all participants
(iv) no participant (talking to oneself).
(C) Marking of the opening and closing of subactivities and/or the interaction as a whole.

7.4.4 OCM coding schema

OCM means "Own Communication Management" and stands for processes that speakers use to regulate their own contributions to communicative interaction. OCM function coding concerns classifying whether the OCM unit is:

- choice related - helps the speaker to gain time for processes concerning continuing choice of content and types of structural expressions, or:
- change related - helps the speaker to change already produced content, structure or expression.

OCM units are also coded with respect to structure of the OCM related expression. This structure can be divided into "basic OCM features", "basic OCM operations" and "complex OCM operations". Pauses, simple OCM expressions such as hesitation sounds etc and explicit OCM phrases count as basic OCM features. Basic OCM operations are: "lengthening of continuants", "self interruption" and "self repetition". The category "Complex OCM operations" stands for different ways to modify the linguistic structure. The OCM coding schema is described in Allwood, Ahlsén, Nivre & Larsson (1997).

7.5 Grammatical coding

There are also ways of coding grammatical structure. One of these is an automatic coding of parts of speech. Another is a coding of "The Maximal Grammatical Units", a coding schema is described in Allwood (2001). When coding Maximal Grammatical Units, one should primarily try to find as large units as possible, the largest unit being complete sentences. Sentences are subclassified by using the schema "**sentences**". In spoken language, there are many utterances that are not sentences, so secondarily, one should try to find complete phrases, which should be coded in the schema "**phrases**". If it isn't possible to find either complete sentences or complete phrases, single words should be coded by parts of speech in the schema "Parts of speech". Each one of the three mentioned schemes contains different categories.

8. Conclusions and Future Directions

In this paper we have described work done at the Department of Linguistics, Göteborg University to collect, transcribe and store spoken language material. We have also described some of the tools we have developed in order to aid work on analyzing the data both automatically and manually. Finally, we have described some or the results obtained so far. Future work will include incremental expansion of the corpus both to obtain data from new social activities and in order to equalize the size of the material

from different activity types. We will also be making increased efforts to make the corpus more multimodal by making the audio and video recordings on which the transcriptions are based more available. Work on tools for analyzing the corpus will continue. The most immediate goal is to complete MULTITOOL which will hopefully give us a better possibility of working with multimodal data. Similarly, work on qualitative and quantitative analysis will be continued. An ambitious goal is to work toward a grammatical description of spoken language and toward a systematic description (perhaps not a grammar) of multimodal face-to-face communication.

9. References

Allwood, J. (1976). Linguistic Communication as Action and Cooperation. *Gothenburg Monographs in Linguistics 2.* Göteborg University, Department of Linguistics.

Allwood, Jens (1977). A Critical Look at Speech Act Theory. In Dahl (Ed.). Logic, *Pragmatics and Grammar*, Lund, Studentlitteratur, pp. 53-69.

Allwood, Jens (1978). On the Analysis of Communicative Action. In Brenner (Ed.) *The Structure of Action*, Basil Blackwell, Oxford.

Allwood, Jens (1987). A Semantic Analysis of Understanding. In V. Rosén (Ed.) *Papers from the Tenth Scandinavian Conference of Linguistics,* University of Bergen, Dept of Linguistics and Phonetics.

Allwood, Jens (1988). Feedback in Adult Language Acquisition. Final Report II. *Ecology of Adult Language Acquisition* (ESF).

Allwood, Jens (Ed) (1996 and later editions). *Talspråksfrekvenser*, Ny och utvidgad upplaga. Gothenburg Papers in Theoretical Linguistics S21. Göteborg University: Department of Linguistics.

Allwood, Jens, Nivre, Joakim & Ahlsén, Elisabeth (1990). Speech Management: On the Non-Written Life of Speech. *Nordic Journal of Linguistics*, 13, 3-48.

Allwood, Jens (1994). Obligations and Options in Dialogue, *Think*, Vol 3, May, ITK, Tilburg University.

Allwood, Jens & Hagman, Johan (1994). Some Simple Measures of Spoken Interaction. In Gregersen, F. And Allwood, J. (Eds.) Spoken Language, *Proceedings of the XIV Conference of Scandinavian Linguistics.*

Allwood, Jens, Ahlsén, Elisabeth, Nivre, Joakim & Larsson, Staffan (1997). *Own communication management.* Göteborg University, Department of Linguistics.

Allwood, Jens (1999). Some Frequency based Differences between Spoken and Written Swedish. In *Proceedings of the 16th Scandinavian Conference of Linguistics*, Turku University, Department of Linguistics.

Allwood, Jens (2000) "An Activity Based Approach to Pragmatics". In Bunt, H. & Black, B.(eds.) *Abduction, Belief and Context in Dialogue; Studies in Computational Pragmatics.* Amsterdam, John Benjamins

Allwood, Jens (2001) Dialog Coding – Function and Grammar: Göteborg Coding Schemas. *Gothenburg Papers in Theoretical Linguistics* GPTL 85. Dept. Of linguistics, University of Göteborg.

Grice, H. Paul (1975). Logic and conversation. In P. Cole and J. L. Morgan (Eds.) *Syntax and Semantics* Vol. 3: Speech Acts. New York: Seminar Press, 41-58.

Nivre, Joakim (1999). Transcription Standard. Version 6. Göteborg University. Department of Linguistics.

Nivre, Joakim, Tullgren, Kristina, Allwood, Jens, Ahlsén, Elisabeth, Holm, Jenny, Grönqvist, Leif, Lopez-Kästen, Dario & Sofkova, Sylvana. Towards multimodal spoken language corpora: TransTool and SyncTool. *Proceedings of ACL-COLING 1998*, June 1998.

Nivre, Joakim & Grönqvist, Leif (1999) Tagging a corpus of Spoken Swedish. Forthcoming in *International Journal of Corpus Linguistics.*

Nivre, Joakim, Allwood, Jens & Ahlsén, Elisabeth (1999). *Interactive communication management - Coding manual V1.0.* Göteborg University, Department of Linguistics.

Van Hout, Roeland & Rietveld, Toni (1993) *Statistical Techniques for the Study of Language and Language Behaviour.* Berlin & New York: Mouton de Gruyter.