LUND UNIVERSITY

Hesitation disfluencies after the clause marker ATT 'that' in Swedish

Horne, Merle; Frid, Johan; Roll, Mikael

2005

Hesitation disfluencies after the clause marker *att* 'that' in Swedish[1]

Merle Horne, Johan Frid and Mikael Roll

1 Introduction

1.1 Function words and hesitation disfluencies

One factor making the processing of spontaneous speech a challenge is the fact that speakers do not always produce complete clauses or complete syntactic constituents of other kinds. The fact that speakers sometimes pause in their speech production, e.g. to access a word from their mental lexicon or to plan a relatively complex utterance has made the study of different kinds of speech 'disfluencies' an important topic for linguists, speech technologists and psycholinguists (e.g. Clark & Wasow 1998, Levelt 1989, Heeman 1997, Eklund 1999, Nordling 1998, Shriberg 1994). Thus, a central issue in research on spoken language is the development of methods for identifying relevant processing units in the stream of speech, i.e. what G. Miller referred to as the 'chunking' problem (Miller 1956). Boundaries corresponding to punctuation marks (periods, commas, etc) do not always have clearly specifiable correlates in spoken language and thus one fundamental problem that has to be solved is: how do different kinds of phonetic, lexical and syntactic form interact in signalling the boundaries of relevant processing units in spoken language?

In the speech technology project our group is involved in, we are investigating function words occurring before hesitation disfluencies. According to Clark & Wasow's (1998) 'Commit and Restore' model of speech production, stranded function words signal that the speaker intends to produce a constituent of the kind signalled by the kind of function word produced, e.g. a clause after a stranded conjunction, a prepositional phrase after a preposition, etc. Thus the recognition of stranded function words (conjunctions, prepositions, pronouns) can be expected to be important for

automatic parsing algorithms. Further, according to Clark & Wasow's 'complexity hypothesis', the probability that a speaker will hesitate in speech production will increase, the more complex the constituent being planned is. Clark & Wasow (1998) equate complexity with *grammatical weight* which is the amount of information on a conceptual level expressed by a constituent. Complexity is further assumed to be measurable in terms of a number of lexico-grammatical parameters, e.g. number of (content) words, number of phrasal nodes, etc.

1.2. Working memory and speech production

According to Levelt's (1989) model of speech production/comprehension, a message is linguistically coded (syntactically and phonologically) in a component called the 'Formulator' which has access to the mental lexicon. The output of the Formulator, a phonetic plan, is then processed by the 'Articulator'. A fundamental question for a model of speech production is then the structure of the phonetic plan and restrictions on the functioning of the Articulator component. The role of working memory in this process is of particular interest. Cognitive psychologists researching on human memory have discussed this kind of process in terms of the notion of 'working memory'. Baddeley (1997) assumes that the phonological processing of speech (phonetic plan and Articulator) is associated with a subsystem in working memory called the "Phonological Loop". Of relevance to speech recognition and to the modelling of speech production in general is the size and nature of the production units (speech chunks) processed by the working memory as well as time restrictions on its functioning. Baddeley claims that the phonological loop stores information for a duration of about two seconds, but that information can be retained for longer periods if refreshed by the articulatory control process through subvocal rehearsal. One can thus assume that the time restrictions on the phonological loop in working memory play a central role in the production of speech and that this is reflected in the chunking of speech fragments. One of our goals is this to investigate the prosodic, syntactic and pragmatic correlates of these 2-2.5 second production units coded in the phonological loop.

2 Prosodic and segmental characteristics of the conjunctions *att* 'that' and *och 'and'* in hesitation disfluencies

Following the reasoning in Clark & Wasow (1998), one could hypothesize that, since Swedish conjunctions such as *att* 'that' and *och* 'and' occur

before major constituents, i.e. clauses, one could perhaps expect that their phonetic form before a hesitation would probably differ from that in fluent speech. In a previous study (Horne et al. 2003) we investigated prosodic and segmental properties of these function words in hesitation disfluency contexts, i.e. before a perceived pause, and compared them to their properties in fluent speech.

Analysis of F0 showed that there was no significant difference in F0-level between *att* and *och* in disfluent and fluent contexts. This finding is interesting since it seems to indicate that an upcoming hesitation after a conjunction has no effect on speakers' fundamental frequency level in the speech fragment being produced up to the point of hesitation. However, measurements of segmental duration comparing fluent and disfluent *att* and *och* showed that: a) duration of *att* and *och* is longer before hesitations (about 130 ms longer), b) vowel duration in *att* and *och* is longer before hesitations (about 20 ms longer), c) duration of the final aspiration phase in *att* and *och* is longer before hesitations (about 20 ms longer for *att*, 10 ms longer for *och*), d) stop occlusion phase in *att* and *och* before hesitations is much longer before hesitations (about 90 ms longer).

Our findings thus indicate that the function words *att* and *och* are characterized by relatively marked segmental characteristics before hesitation disfluencies as compared to their form in fluent contexts.

3 Syntactic and pragmatic complexity in speech fragments following hesitation disfluencies after *ATT* 'that'[2]

Since our previous investigation has provided support for the expectation that hesitation disfluencies at clause boundaries are characterized by nonreduced forms of *att* and *och*, we have continued our investigation in an attempt to characterize the environment following the hesitations. Our hypothesis is that the phonetic form of the function words, together with pause information can provide important cues as to the information status of the speech fragment(s) following the disfluency. In the present study, we limit ourselves to the study of speech fragmants following disfluencies after *ATT* at clause boundaries. Following the 'Complexity Hypothesis', one can expect that the phonetically marked cases of *att* occurring before disfluencies should be followed by 'heavier' or more complex speech

---

[2] In what follows, we will represent disfluent instances of *att* with upper-case letters *(ATT)*.

fragments than weak forms of *att* in fluent contexts. This weight or complexity can be expected to be reflected in the lexico-syntactic structure of the fragment(s) following the hesitation. This lexico-syntactic complexity reflects in its turn pragmatic structure differences in the discourse enviroments after *att*. These differences can be expected to relate, for example, to the topic structure of the discourse.[3]

3.1 Syntactic and pragmatic complexity
3.1.1. Syntactic complexity

Following Levelt (1989), we consider clauses to be the basic unit of linguistic encoding and thus a simple clause can be assumed to be associated with a default level of complexity. We further assume that complex clauses can be encoded in one production unit under certain circumstances. Clause-final phrasal or clausal arguments are probably encoded together with their matrix clauses, since they contribute decisively to the core meaning of the verb phrase, e.g. the concrete vs. abstract meaning associated with *John showed me that block of chocolate / that he is happy*. Clauses in this position increase complexity more than noun phrases, since they are inherently more complex. Thus clause-level recursion in the relevant domain is a significant parameter for determining complexity.

Clause-final phrasal and clausal adjuncts, such as the adverbial clause in *John buys chocolate when he is happy,* on the other hand, do not have a decisive effect on the core meaning of a verb phrase; thus, they can be attached to the matrix clause even after the latter has been produced. This means that they are not likely to appear in the same production unit as their matrix clauses. When an adjunct is promoted to a matrix-clause initial position, however, as in *When John is happy, he buys chocolate*, it must be encoded before producing the rest of the matrix clause, in order to obtain linear order. We therefore assume, with Hawkins (1994), that matrix-clause initial adjuncts increase the complexity of the matrix clause more than they do in final position.

As a result, *the domain for which complexity will be calculated consists of a clause from the first element of its first constituent to the end of its last argument linearly, excluding adjunct phrases and clauses to the right of it.*

3.1.2. *Pragmatic complexity and topic structure*

Syntactic complexity can, in its turn, be assumed to mirror pragmatic complexity, i.e. complexity related to discourse structure. For example, it can be expected that factors related to discourse topic structure such as the given/new status of information and the related speech act type influence discourse production and pro4cessing. In studies on the information status of noun phrases, for instance, disfluencies have been shown to be more likely to appear before noun phrases introducing new referents than before noun phrases that express given referents. Arnold et al. (2003) show that, when confronted with a disfluent realization of a noun phrase (e.g. *thee, uh, candle*) and different possible visible referents, listeners are more likely to look at a new object than at one presented earlier.

The distinction between new and given information is thus closely correlated with speech act type. Different speech acts containing new and given information can therefore be expected to correlate with disfluent and fluent productions of clauses following the conjunction *att*. In the labelling of our data, we have chosen to use the speech acts presented in Nakajima & Allen (1993) in their study on the relation between topic structure and prosodic phrasing. We follow them in their use of five categories: 'new topic', 'topic continuation', 'elaboration', 'clarification' and 'summary'. An 'elaboration' adds new information that is relevant and related to the previous utterance, whereas a 'clarification' clarifies something in a previous utterance that may have been vaguely expressed. We have, however, added another category, namely 'old topic', which is a return to a relatively recent topic which the speaker assumes to be accessible to the listener.

We expect the contextually/pragmatically more independent topic structure categories to be more common in disfluency contexts since they can be assumed to be associated with more cognitive activity/planning. Hence 'new topic' which introduces (a) new referent(s) should be more likely to appear after disfluent *att* than 'topic continuation'; likewise, one would expect that 'elaboration', which adds new information on a certain topic, would occur more often in a disfluent context than 'clarification', which simply restates what is said about a particular topic in the preceding utterance in a clearer way. A 'summary', however, is relatively more complex, in the sense that it summarizes the main point of the preceding utterance as well as others before it.

In addition to these topic-structure based speech acts, we have observed another speech act type, which we have termed 'empathetic quotation', that appears in disfluency contexts. In an empathetic quotation, a speaker expresses what (s)he thinks someone feels regarding a certain situation as if (s)he were reproducing that person's thoughts and emotional expression. Typically, empathetic quotations start with an interjection. An example of an empathetic quotation is presented in the clausal complement in (1).

(1)   If that happened, she would feel that: "Oh God! I can't believe it!"

Gardener et al (1998) show that the cognitive devices handling emotional recognition and other heavily context-dependent functions are collocated in systems to some extent separated from more language-specific resources. The access to representations of emotions supposedly experienced by others can therefore be assumed to add pragmatic complexity to linguistic expressions when compared to empathetically more neutral variants. One would hence expect empathy to be a relevant factor in triggering disfluencies. We regard empathetic quotations therefore as a specific category of the speech act 'quotations'.

In sum, in order to classify the utterances following *att* according to their pragmatic or discourse-related function, we have used the following speech act categories in this study: new topic, topic continuation, elaboration, clarification, summary, old topic, quotation, and empathetic quotation.

3.2 Analysis of syntactic structure

To describe the relevant syntactic units in the speech fragments following *att*, we have used a simplified version of the kind of phrase-structure grammar presented in Chomsky (1995). Following Holmberg & Platzack (1995), subjects and finite verbs are taken to be raised to CP (Complementizer Phrase) when other elements do not prevent this.

The most significant modification we have made to the original model is the removal of the majority of bar-level projections, traces resulting from ordinary subject or verb movement, phonologically empty or affixed heads, and the entire *v*P. We use an intermediate verb phrase V' to differentiate internal arguments from adjuncts. As a result we get the same number of phrasal nodes as in the original model, facilitating complexity calculations. These changes do not imply any theoretical assumptions.

3.2.1 Grammatical Labels

The grammatical categories in Table 1 have been used in the syntactic labelling of the spontaneous speech data. Most of these categories are straightforwardly used as in recent generative studies. We have chosen to differentiate lexical verbs, VLEX, from copulas, VCOP, and auxiliary verbs AUX. The 'discourse marker', DM, is used to label instances of *att* that do not function as subjunctions but rather as discourse markers signalling that an utterance of a clause-like nature is being planned. We occupy X and XP to label unidentified projections and elements at different levels, as well as pauses and editing expressions. Phonologically null and 'missing' elements are marked with ∅, the difference being the category dominating them.

The speech fragments extracted from the corpus for syntactic analysis were delimited with the labels p2 and /p2 in the ESPSWaves speech analysis program. These labels were chosen to mark the beginning and end, respectively, of production units consisting of clauses following *att*, from the first element of their first constituent to the end of their last argument linearly, the domain for which syntactic complexity is calculated.

Table 1. Grammatical categories used in the labelling of speech data.

Category	Head	Phrase	Category	Head	Phrase
Adjective	A	AP	Preposition	P	PP
Adverb	ADV	ADVP	Pronominal	PRON	
Auxiliary verb	AUX		Particle	PTL	
Complementizer	C	CP	Quantifier	Q	QP
			Trace	t	T
Conjunction	CONJ	CONJP	Tense		TP
Determiner	D	DP	Intermediate verb phrase		V'
Discourse marker	DM		Copula	VCOP	VP
Expletive pronoun	EXPL		Lexical verb	VLEX	VP
			Wh-word	WH	WH
Infinitive marker	IFM		Undefined	X	XP
			Phonologically null or missing	∅	∅
Noun	N	NP			
Negator	NEG	NEGP			

3.2.2 Computer-based tool for grammatical analysis

In order to facilitate the investigation of syntactic structure in spontaneous speech, we have developed a computer-based tool, *GRAMMAL*, that enables grammatical analysis and mark-up in the form of tree structures. The tool works by first reading in a file with clause(-like) units (p2 units) from a label file, and then allows the user to produce a syntactic analysis of the words in these sentences. The tool is currently configured to read annotation files in the *xlabel* (Entropic) format, but could possibly be configured to read other file formats. The user starts by selecting such a label file containing transcriptions at the word level as well as clause-based (p2) unit boundaries. A p2 unit is currently defined by inserting the labels <p2> and </p2> around material that is to be considered as a basic clausal production unit (see 3.2.1). These labels provide a means of inserting start and end tags for the units to be analysed.

    An example of an *xlabel* word transcription is shown below. The three columns contain time data, color data (only used when displaying the labels in *xwaves*) and label data.

```
31.765660    76 jag              34.276633    76 </p2>
31.985678    76 tycker           34.487199    76 i och med
32.184990    76 nog              34.806961    76 <p2>
32.292723    76 <p2>             34.810889    76 ATT
32.342964    76 att              34.916691    76 PAUS
32.483011    76 det              35.069410    76 man
32.699607    76 bara             35.181576    76 är
32.760004    76 har              35.336882    76 så
33.247498    76 varit            35.889083    76 pass
33.333194    76 PAUS             36.307548    76 PAUS
33.911275    76 positivt         36.930355    76 nära
34.275948    76 egentligen       36.934060    76 </p2>
```

When read into GRAMMAL, this mark-up will be converted into the following analysis units (because of the <p2> and </p2> tags, enumeration is added for clarification):

```
1. att det bara har varit PAUS positivt egentligen
2. ATT PAUS man är så pass PAUS nära
```

The units to be analysed subsequently appear in the program in a component called *listbox*, from where each p2-unit may be selected using a mouse left click. The current tree-structure analysis of the unit then appears on the right side of the window. Both a graphic rendering of the actual tree and a compact, flat version with square brackets is shown (Figure 2). The user may then modify the structure by renaming, inserting and siblifying the nodes.

As the user modifies the tree structure, both the visual appearance and the flat representation are updated. In the current version of the program the analysis cannot, however, be modified via the flat representation; this is, however, a possible future extension. An example of a partial analysis is shown in Figure 3.

The data that is produced is stored in a file that is updated as soon as a change in the tree structure is made. Following is an example of such a data file:

```
root ;   ; CP ;   ;                node23 ; node19 ; t ; i ;
node1 ; root ; C ;   ;              node24 ; node23 ; Ø ;   ;
node2 ; node1 ; ATT ;   ;           node18 ; node19 ; AP ;   ;
node3 ; root ; X ;   ;              node17 ; node18 ; ADVP ;   ;
node4 ; node3 ; PAUS ;   ;          node9 ; node17 ; 0 ;   ;
node21 ; root ; CP ;   ;            node10 ; node9 ; så ;   ;
node5 ; node21 ; PRON ; i ;         node11 ; node17 ; 0 ;   ;
node6 ; node5 ; man ;   ;           node12 ; node11 ; pass ;   ;
node7 ; node21 ; VCOP ;   ;         node13 ; node18 ; X ;   ;
node8 ; node7 ; är ;   ;            node14 ; node13 ; PAUS ;   ;
node22 ; node21 ; TP ;   ;          node15 ; node18 ; A ;   ;
node20 ; node22 ; VP ;   ;          node16 ; node15 ; nära ;   ;
node19 ; node20 ; V' ;   ;
```

In the future, this data format may be altered so as to conform to the emerging standards for linguistic annotation, e.g. the *Annotation Graphs* described by Bird & Liberman (2001). The program currently runs on Windows and Unix families of platforms and a Mac version is in the pipeline. This multiple-platform compatibility is possible due to the platform-independent nature of the Tcl/Tk programming language.
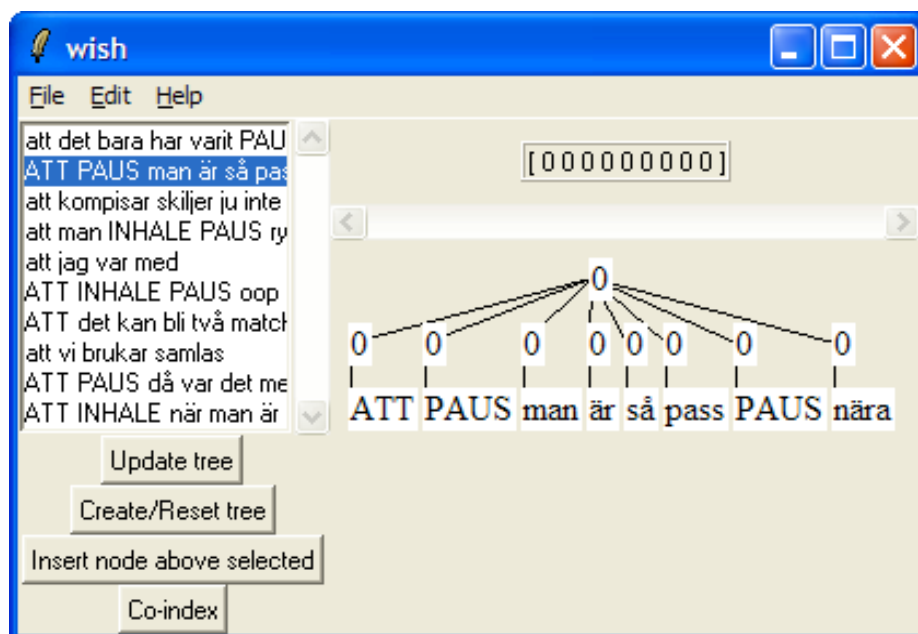
Figure 2. An unanalysed clause-based p2-unit. The listbox with the clauses is shown on the upper left. Below that are a number of control buttons. The analyses appear on the right: at the top, the flat structure and below, the tree structure.
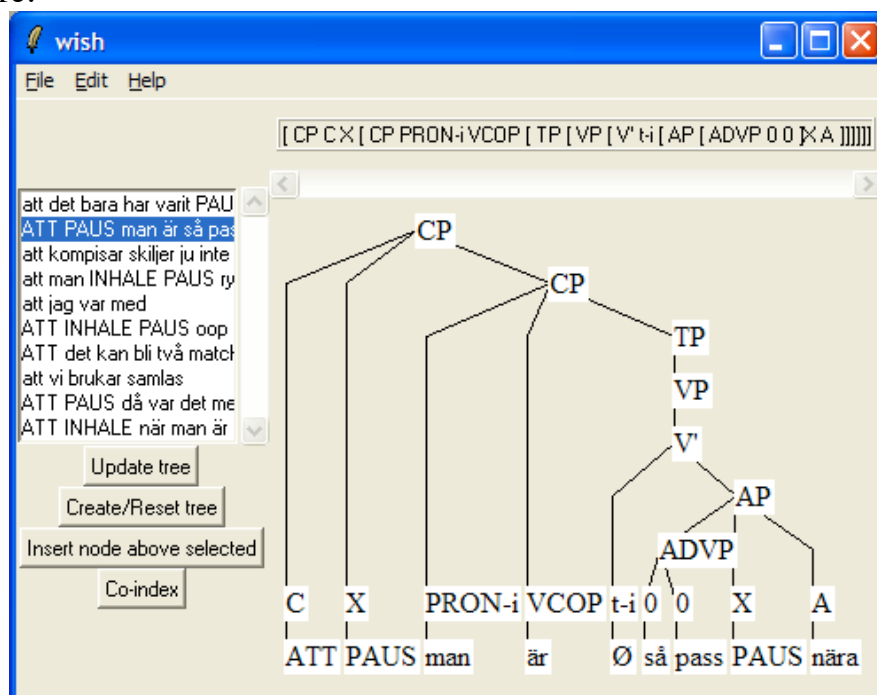

Figure 3. A partially analysed clause.

3.3 Analysis of pragmatic structure

For the analysis of the pragmatic structure of spoken discourse, we have labelled fragments following *att* according to the speech acts discussed above in section 3.1.2.: new topic, topic continuation, elaboration, clarification, summary, old topic, quotation, and empathetic quotation. The first five come from Nakajima & Allen (1993), while last three have been introduced in order to handle our specific data. An old topic is a return to a relatively recent previous topic. A quotation is an attempt to reproduce more or less literally what someone (including oneself) has said or thought, or could say or think. An empathetic quotation is a quotation of someone else's speech or thought, or hypothetic speech or thought, with the intention of also reproducing the person's emotional expression.

3.4 Analysis of timing restrictions on working memory

One idea (cf. section 1.2) that we have begun to investigate in this study is that there are time restrictions on the chunking of speech, e.g. on the linguistic

coding of a message. According to the model of memory assumed by Baddeley (1997), the part of working memory where speech coding takes place has a time limit of around 2 seconds. This is of interest for the analysis of spontaneous speech since one can expect that this timing restriction will be reflected in the segmentation of spontaneous speech. For example, one can expect that phenomena such as prosodic phrasing or grouping will in many cases reflect this constraint on speech processing and thus that pausing phenomena and tonal phenomena such as F0-resets should be observed at the boundaries of phonological loop units.

From the examination of the speech of the speaker chosen for this study on speech production, we hypothesize that the phonological loop has a processing time-span of between 2 and 2,5 seconds. This is based on the following observations related to the prosodic and syntactic form of these timing units: there is often a pause after 2-2.5 seconds of speech, there is often a F0 reset after 2-2.5 seconds of speech, there is often final lengthening after 2-2.5 seconds of speech, there is often a constituent boundary after 2-2.5 seconds of speech. The constituent boundary often corresponds to a clause boundary which is furthermore what one would expect given the fact that a syntactic clause corresponds to a basic proposition, which is considered to be a basic unit
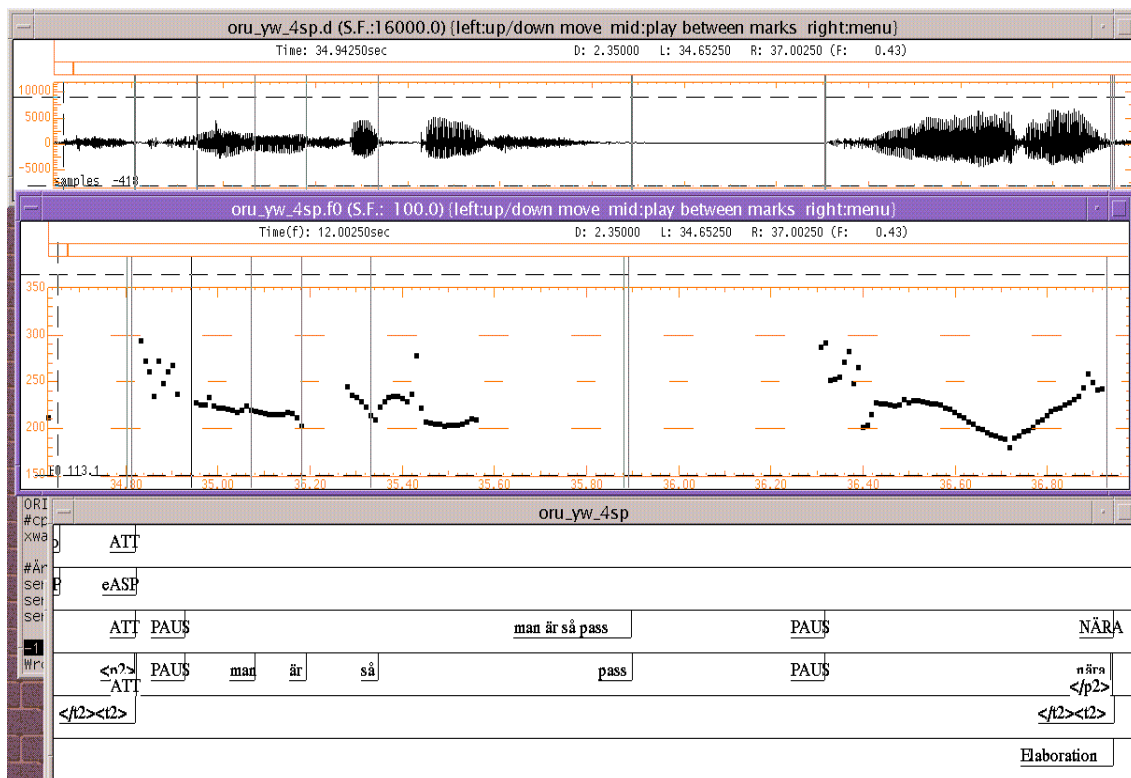
Figure 4. Screendump from ESPSWaves analysis of utterance *att PAUSE man är så pass PAUSE nära* 'that PAUSE one is so PAUSE close'. The six label tiers show, from top to bottom: 1. labels for disfluent *ATT* and fluent *att*. 2. segmentation of *att* into vowel and consonant, including start and end ASPiration phases, 3. transcription without segmentation into words, 4. transcription with segmentation into words and syntactic (/p2) segmentations, input to GRAMMAL, 5. working memory 'phonological loop' timing units (/t2), 6. topic structure labels.

of speech production. The speech data have consequently been tagged with labels <t2> and </t2> to mark the beginning and end, respectively of the assumed 2-2,5 second long timing units in working memory (see Fig. 4).

3.5 Data

The data used for this methodologically-oriented study on our spontaneous data has been limited to one of the 24 speakers from the previous study (Horne et al. 2003). This speaker produced 25 instances of att: 13 instances were followed by hesitation disfluencies (these were labelled 'ATT') and 12 were produced during fluent speech (these were labelled 'att'). The speaker is a female, 22 years old, from Orust in Bohuslän (Götaland).

4. Preliminary results

4.1. Syntactic complexity

Analysis of grammatical structure has shown that clause-level recursion appears to be a relevant measure for determining the degree of syntactic complexity of a clause. Consider the examples in (2)-(6):

(2)   då var det mycket när man var på elitläger och sånt ATT PAUSE [$_{CP}$ då var [$_{TP}$ det mer [$_{ConjP}$ [$_{CP}$ man hade avslappning innan] å [$_{CP}$ tränarn stod å prata] å [$_{CP}$ det skulle va mörkt då i rummet]]]]

'there was a lot when we were at elite camp ATT PAUSE [$_{CP}$ then was [$_{TP}$ there more [$_{ConjP}$ [$_{CP}$ one had relaxing before] and [$_{CP}$ the trainer talked] and [$_{CP}$ it had to be dark in the room]]]]'

(3)   det var egentligen ATT PAUSE [$_{CP}$ [$_{CP}$ när man är runt (PAUSE kan det vara) femton sexton] så får [$_{TP}$ alla klubbar PAUSE i Bohusländal då [$_{VP}$ [$_{DP}$ ∅]]]]

'it was really ATT PAUSE [$_{CP}$ [$_{CP}$ when one is around (PAUSE can it be) fifteen sixteen] then get [$_{TP}$ all clubs PAUSE in Bohusländal [$_{VP}$ [$_{DP}$ ∅]]]]'

(4)   just som inte Autonova då ATT PAUSE[$_{CP}$ där vet [$_{TP}$ jag ju [$_{CP}$ vad som händer]]]

'just like not Autonova ATT PAUSE [$_{CP}$ there know [$_{TP}$ I [$_{CP}$ what happens]]]'

(5) ATT INHALE [$_{CP}$ jag vill [$_{vP}$ kunna [$_{ConjP}$ [$_{VP}$ gå till jobbet] å INHALE [$_{VP}$ veta i princip [$_{CP}$ vad som händer]]]]]] kanske

'ATT INHALATION [$_{CP}$ I want [$_{vP}$ to be able [$_{ConjP}$ [$_{VP}$ to go to work] and INHALATION [$_{VP}$ to know in principle [$_{CP}$ what is happening]]]]] maybe'

(6) men ändå ATT PAUSE [$_{CP}$ det kan ju hända [$_{QP}$ så mycket [$_{CP}$ att du vet ju inte [$_{CP}$ vad du får rycka ut på eller vad som SWALLOW PAUSE händer just då]]]]

'but nevertheless ATT PAUSE [$_{CP}$ it can happen [$_{QP}$ så much [$_{CP}$ that you do not know [$_{CP}$ what you have to do or what SWALLOW PAUSE happens just then]]]]'

The five examples above illustrate clause-level recursion following disfluent *ATT*. In these examples, the clauses following *ATT* all contain a clause in an argument position or inside an argument. The clauses following *ATT* in (2) and (3) appear to be arguments of the preceding clauses, with an expletive *det* 'it/there' occupying the subject position in TP. However, their syntactic and semantic connection to the preceding clauses is weak. This is best seen in that they both have main-clause word order, with an adverbial moved to the left edge of the clause, the adverb *då* in (2) and the adverbial clause *när man är runt (PAUSE kan det vara) femton sexton* 'when one is around (PAUSE can it be) fifteen sixteen' in (3). Even if the clauses are of a kind rarely seen in traditional grammars (they look like predicative constructions but do not contain much of a predicate), subordinate-clause word order would be expected for the internal argument, which is in this case the "logical subject" of the matrix clause.

In (4), the clause following *ATT* (*där vet jag ju vad som händer* 'there know I what happens') modifies the DP *Autonova*. It could be seen as a relative clause inside that DP, but again, main-clause word order, with the adverb *där* 'there' moved to the left edge of the clause, contradicts such an analysis. In the utterance in (5), which directly follows (4) in the recording, *ATT* does not have any syntactic function in relation to the preceding clause, neither as a conjunction nor as a subjunction, but is rather a discourse marker signalling that the speech unit being planned is a clause. Its status as a discourse marker is strengthened by the fact that there is a major shift in topic from the fragment preceding *ATT* to the one following it. In the subsequent fragment (6), *ATT* follows the conjunction *men* 'but' and has to be analysed either as part of a conjunctive expression or as a discourse marker. Here too, the clause following *ATT* has main-clause word order: the

clause-level adverbial *ju* 'you know' and the negator *inte* 'not' precede the subject *du* 'you', instead of following it, as they would have with subordinate-clause word order.

In contrast to the examples above involving fragments after disfluent *att*, the fragments after fluent *att* produced by the same speaker showed only one complex clause, namely the one in (6) (*att du vet ju inte vad du får rycka ut på eller vad SWALLOW PAUSE händer just då…*). However, this complex clause appears in a fragment following a disfluent *ATT*. If disfluent *ATT* functions to signal that a complex speech fragment is being planned, then fluent clause boundariy markers within the complex fragment are quite expected, since the speaker should have had sufficient time to plan the complex clause during the hesitation. The data also show that the whole fragment following disfluent *ATT* in (6) (*det kan hända så mycket att du inte vet vad du får rycka ut på*) appears to have been planned as a unit, since it is articulated during one /t2 (2.5 second) production unit.

4.1.1 Measuring syntactic complexity
Since one of our goals is to be able to obtain a reliable measure of syntactic complexity, we are currently investigating a number of different ways of estimating the syntactic complexity of clauses. Several possibilities exist, e.g. the number of levels in the syntactic tree, the overall number of nodes or the number of phrasal nodes. Simply taking the number of levels would, however, perhaps result in a measure that is too blunt, since embedding would not affect the result. Furthermore, it would also depend on how detailed the analysis is. On the other hand, the overall number of nodes would depend too much on the number of words in the clause and thus is not an optimal measure either. We think, therefore, that the number of phrasal nodes would be the most suitable measure. In order to include embedding, one can weight each phrasal node according to how deep in the tree it is, e.g. a phrase node high up in the tree would weigh more than a phrase node at a lower level.

In Figure 5, we show the syntactic structure of the utterance in (2) illustrating a disfluent *att* followed by a clause which has an internal argument consisting of three conjoined clauses. A possible way of estimating the complexity of this utterance is to count the number of non-terminal phrasal nodes. We exclude the intermediate verb phrase V' since this always lies below a VP. This would amount to a complexity score of 15, since there are 15 phrasal nodes in the utterance. By weighting each phrasal node

according to how deep in the tree it is (i.e. adding a higher figure the deeper in the tree the node is) we get a complexity score of 99 for the same utterance. On this analysis, the top node XP counts as 1, the daughter node CP as 2, the following TP as 3 etc. The deepest node is the CONJP that dominates the nodes VLEX, CONJ and VLEX in the middle of the tree. This counts as 11. It

Figure 5. Syntactic tree structure generated by *GRAMMAL* for the speech fragment after *att* 'that' in utterance (3) showing clause recursion.

is still a matter of investigation which method to use. By using the GRAMMAL tool, however, all these calculations can be made automatically.

4.2. Pragmatic Complexity and Topic Structure

In many cases, complexity in cognitive planning is reflected in the linguistic structure in terms of complex syntactic structures, as in (2). However, the complexity on the level of information structure is not always reflected syntactically. Sometimes it is lexical information that reflects increased complexity in discourse planning.

In (7) below, the clause initiated by disfluent *ATT* is the object of a prepositional expression *i och med*. The clause following ATT does not have a complex syntactic structure, but its role as regards the development of the topic structure of the discourse can, on the other hand be regarded as 'heavy' It bears an explanatory relation to the matrix clause predicate and this explanation contains the most important share of new information in the utterance which is encoded in the predicative AP head, *nära*.

(7)   Jag   tycker nog   att det bara har   varit PAUSE positivt
        I   think   probably   that it   just   have been   PAUSE   positive
        egentligen i och med ATT PAUSE man är så pass PAUSE   nära
        really   in   and with   THAT   PAUSE   one is so PTL   PAUSE   close

Thus utterances that introduce new topics into the discourse can be thought of as being pragmatically complex or 'heavy' and thus can lead to disfluencies in speech due to the extra cognitive activity that can be assumed to be associated with them.

When the fragments following *att* were categorized according to the speech act categories mentioned above, some interesting patterning was observed. We found 3 new topics (e.g. (3), (5) above), 6 elaborations (e.g. 0, (4) and (6) above), 2 summaries (e.g. 0 below), and 2 empathetic quotations (0 and 0 below) after disfluent realizations of *att*. The fragments following fluent *att* were classified as 5 topic continuations, 5 clarifications, 1 quotation, and 1 old topic. In other words, all the speech acts appeared exclusively either with fluent or disfluent *att*. Fluent *att* seems to appear if the speech act following *att* shows a higher degree of coherence with the topic of the preceding utterance, and disfluent *att*, in the opposite context, creating oppositions such as topic continuation : new topic, clarification : elaboration, quotation: empathetic quotation.

9

In (8), *ATT* is part of the colloquial conclusive conjunctive expression *så att* 'so (that)', which introduces a summary of the preceding conjoined sentences.

(8) | Vi | har | match | en | dag | i | veckan | och | så | är |
|---|---|---|---|---|---|---|---|---|---|
| We | have | match | one | day | in | week.the | and | then | are |

vi	ju		ett	B-lag	också	så	ATT	det	kan
we	you.know		a	B-team	also	so	THAT	there	can

ju		bli	två	matcher		också
you.know		be	two	matches		also

Fluent realizations of *så att* that have been observed in other speakers differ from that in (8) as they are not followed by summaries of several clauses, but simple inferences from an immediately preceding clause.

In (9) below, the speaker is describing what she believes that her sister felt in a certain situation, and in (10), she expresses what she believes people think about her own situation of not being able to do some work in the garden. In both cases she employs what we have called 'empathetic quotations'. We use this term to indicate that the speaker recognizes or imagines some other person's feelings regarding a certain situation. The emotional character of the utterances below is strengthened by the opening interjections *happ* and *men gud*.

(9) | det | känner | nog | | hon | mer | ATT | PAUSE | happ |
|---|---|---|---|---|---|---|---|---|
| that | feel | probably | | she | more | THAT | PAUSE | |

		oop						
nu	tog	ju		Sandra	dom			
now	took	you.know		Sandra	them			

(10) | då | tror | dom | ju | | ATT | PAUSE | men | gud | det |
|---|---|---|---|---|---|---|---|---|---|
| then | think | they | you.know | | THAT | PAUSE | but | god | it |

är	jätteskönt	att	slippa
is	wonderful	to	not.have.to

4.3 Prosodic features of disfluencies following att

Although our data is very limited in this study, we have made a number of observations regarding the prosodic features of disfluencies following *att* which can be further tested in follow-up studies. One observation is that all

of the fragments following fluent *att*, except for one (a quotation) were realized within a speech production unit (chunk) which included *att*. This was not the general case for the fragments following disfluent *att*. These fragments were included in a separate speech production unit; they were separated prosodically from *att* by a pause in 12 of the 13 cases. This prosodic independence of speech fragments following disfluent *att* correlates well with their relatively independent status as regards their categorization as regards topic-structure. Another observation regarding the prosodic characteristics of disfluencies following *att* has to do with the occurrence of inhalations duringpauses: inhalations occurred during pauses after disfluent *att* before fragments classified pragmatically as 'new topics' and 'empathetic quotations'. Only one inhalation occurred in a fluent-*att* context, but in that context, it occurred *before* the *att*. Another observation involving pause structure was that the only filled pause that occurred in a disfluency following *att* was before a fragment realizing a 'new topic'. This observation is in line

with the results in Arnold et al. (2003) who correlate disfluencies containing filled pauses with the introduction of discourse-new entities. Swerts et al. (1996) have also observed how filled pauses are more typical in the vicinity of major discourse boundaries (topical units).

5 Conclusion

This study has had as its goal to investigate the relationship between the fluent and disfluent productions of *att* and the structure of speech fragments following them. According to Clark and Wasow's (1998) 'Complexity Hypothesis', we expected more complex speech fragments to occur after disfluent *ATT* than fluent *att*.

In our attempt to develop a methodology for measuring complexity, we have made a detailed investigation of all the productions of *att* in the spontaneous production of one speaker from the Orust dialect of Swedish. We have concentrated on a study of the syntactic structure of the speech fragments following *att* and their relation to the pragmatic structure of the discourse, in particular the fragments' role as regards the topic structure of the discourse. It was further observed that even extralinguistic (emotional) factors contribute to the production of disfluencies. A number of observations regarding the prosodic correlates of disfluencies were also made.

Complexity has thus been been assumed to be related to both structural linguistic categories (syntactic and phonetic form) as well as to pragmatic categories (topic structure and speech act type) and even extralinguistic (emotional states). It is thus obvious that a number of factors interact in triggering a disfluent realization of *att*. Syntactic word order patterns reveal that the pragmatic coherence between two clauses decreases with the use of disfluent *ATT* as compared to fluent *att*. Seven of the thirteen examples of disfluent *ATT* appear with exclusive main-clause word order, while none has obvious subordinate clause word order, even in contexts where it would be highly expected. This indicates that disfluent *ATT* has a function of 'cutting' the discourse. Participating in a conclusive conjunction, disfluent *att* may indicate that the following conclusion does not have a direct pragmatic relation to the immediately preceding fragment, such as being an inference from it, but rather constitutes a summary of a larger part of the preceding discourse. Before clauses containing new information, disfluent *att* signals that the clause is not intimately connected to the preceding fragment as regards information structure. Disfluent *ATT* tends to signal a new topic rather than topic continuation, and an elaboration rather than clarification, where clarification is more strongly bound to the preceding utterance.

Before empathetic quotations – fragments that imply recognition or imagination of other's emotions – disfluent *ATT* may signal a change in the deictic centre as compared to the preceding discourse. Furthermore, at a clause boundary represented by disfluent *ATT*, there seems to be more likely to be a rupture between two speech acts and two production units than at fluent *att*.

As regards prosodic features of disfluencies after *att*, we have observed that disfluent *ATT* is almost always followed by a clear prosodic boundary in the data investigated for this study. In all cases but one, this boundary was marked by a silent pause, in some cases including inhalation. It was also observed that the only filled pause that occurred after a disfluent *ATT* was before a fragment introducing a new topic.

More data is of course needed in order to give more support for the preliminary results presented in the present investigation. We feel, however, that we have now a solid methodological framework to use in the further investigation of disfluencies in spontaneous speech production.

References

Arnold, J., Fagnano, M. & Tanenhaus M. 2003. 'Disfluencies signal thee, um, new information'. *Journal of Psycholinguistic Research* 32, 25-36.

Baddeley, A. 1997. *Human memory: theory and practice*. Hove: Psychology Press.

Bird, S. and Liberman, M. (2001). A formal framework for linguistic annotation. *Speech Communication*, 33, 23-60.

Chomsky, N. 1995. *The minimalist program*. Cambridge, Mass.: MIT Press.

Clark, H. & Wasow, T. 1998. 'Repeating words in spontaneous speech'. *Cognitive Psychology* 37, 201-242.

Eklund, R. 1999. 'A comparative analysis of disfluencies in four Swedish travel dialogue corpora'. *Proceedings of Disfluency in Spontaneous Speech Workshop*, Berkeley, 3-6.

Gardener, H., Brownell, H., Wapner, W. & Michelow, D. 1998. 'Missing the point. The role of the right hemisphere in the processing of complex linguistic materials'. In *Pragmatics –critical concepts*, A. Kasher, ed., vol. 6, 170-192. London: Routledge.

Hawkins, J. 1994. *A performance theory of order and constituency*. Cambridge University Press.

Heeman, P. 1997. *Speech repairs, intonational boundaries, and discourse markers: modelling speakers'utterances in spoken dialogue*. PhD thesis, Univ. of Rochester, N.Y.

Holmberg, A. & Platzack, C. 1995. *The role of inflection in Scandinavian syntax*. Oxford: Oxford University Press.

Horne, M., Frid, J., Lastow, B., Bruce, G. and Svensson, A. 2003. 'Hesitation disfluencies in Swedish: prosodic and segmental correlates'. *Proceedings ICPhS* (Barcelona), 2429-2432.

Levelt, W. 1989. *Speaking: From intention to articulation*. Cambridge, Mass.: MIT Press.

Miller, G. 1956. 'The magical number seven, plus or minus two: Some limits on our capacity for processing information'. *Psychological Review* 63, 81-97.

Nakajima, S & Allen, J. 1993. 'A study on prosody and discourse structure in cooperative dialogues'. *Phonetica* 50, 197-210.

Nordling, J. 1998. 'Reparationer i spontant tal'. *Kandidat* '(B.A.)' term paper, Dept. of Linguistics, Lund University.

Shriberg, E. 1994. *Preliminaries to a theory of speech disfluencies*. Ph.D. thesis, Univ. of Berkeley.

Swerts, M., Wichmann, A. & Beun R-J. 1996. Filled pauses as markers of discourse structure. *Proceedings of the International Congress on Spoken Language Processing (ICSLP)*. Philadelphia, USA, 1033-1036.