

Hesitation disfluencies in a speech technology perspective

M. Horne, J. Frid and M. Roll

Language technology in Sweden 13/5 2004





The role of function words in spontaneous speech processing

SPEECH TECHNOLOGY PROGRAM PROJECT

Funding: 2003-2004

PROJECT LEADER

MERLE HORNE

JOHAN FRID

GÖSTA BRUCE

BIRGITTA LASTOW

MIKAEL ROLL

ADINA SVENSSON

www.ling.lu.se/projects/ProSeg2.html

DEPT. OF LINGUISTICS, LUND U.



Function words:

- Important information in the segmentation of speech



Hesitation disfluencies

(1) jag spelar **i EH PAUSE** Bohusläns landslag

'I play **in EH PAUSE** Bohusländs regional team'

(2) men ändå **att PAUSE** det kan hända så mycket

'but still **that PAUSE** so much can happen'



Psycholinguistic hypotheses to be tested using spontaneous speech

- **Commit- and- Restore hypothesis:**
Stranded function word reflect
“syntactic commitments” (Clark & Wasow (1998))

- **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))**



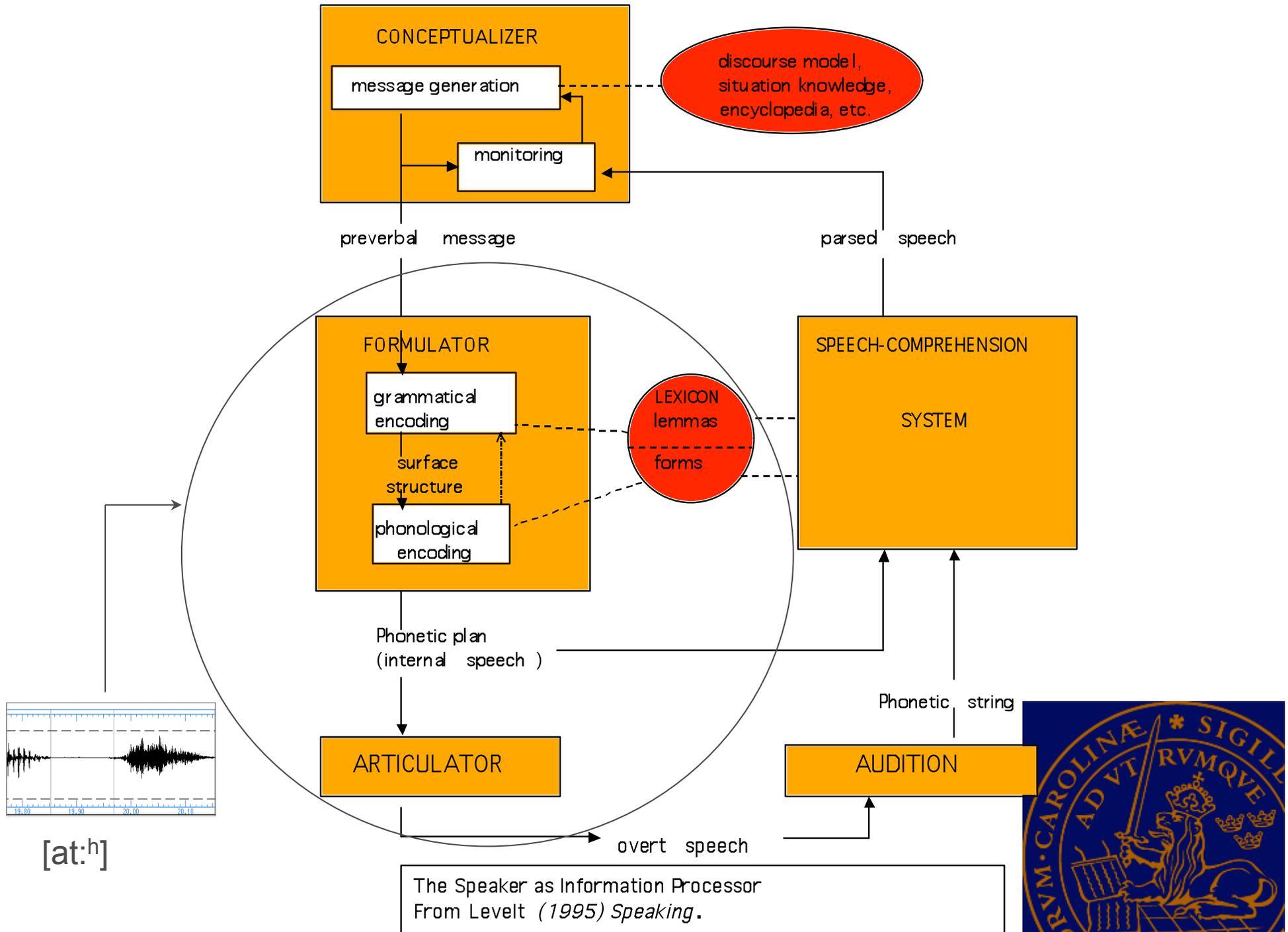
Stranded function words are phonetically prominent (marked)

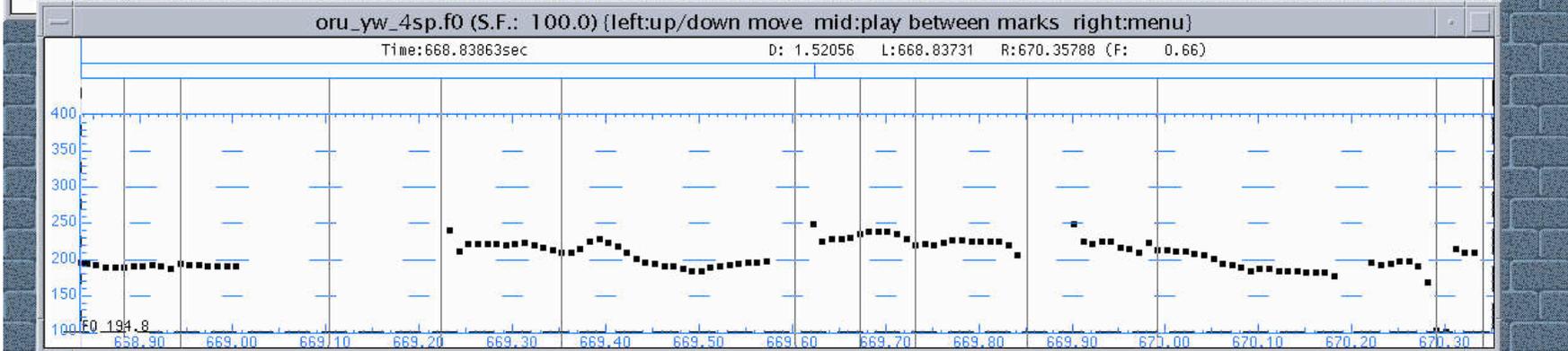
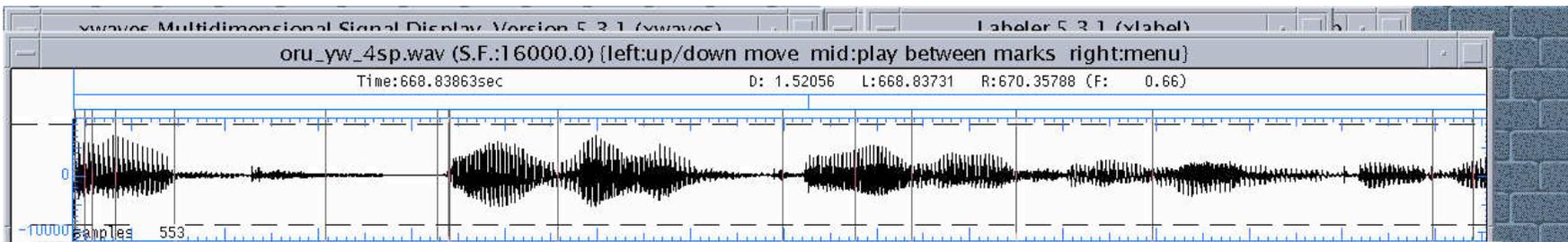
'Hesitation disfluencies in Swedish: prosodic and segmental correlates'.

2003. *Proceedings ICPHS* (Barcelona)

M. Horne, G. Bruce, J. Frid, B. Lastow & A. Svensson







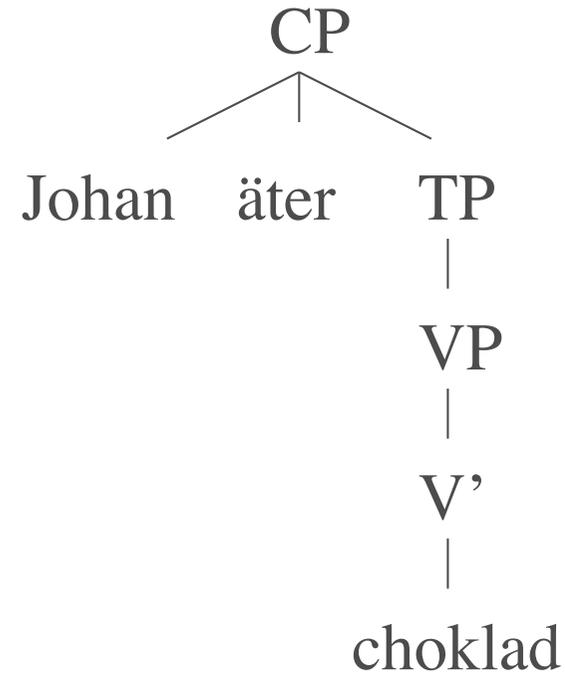
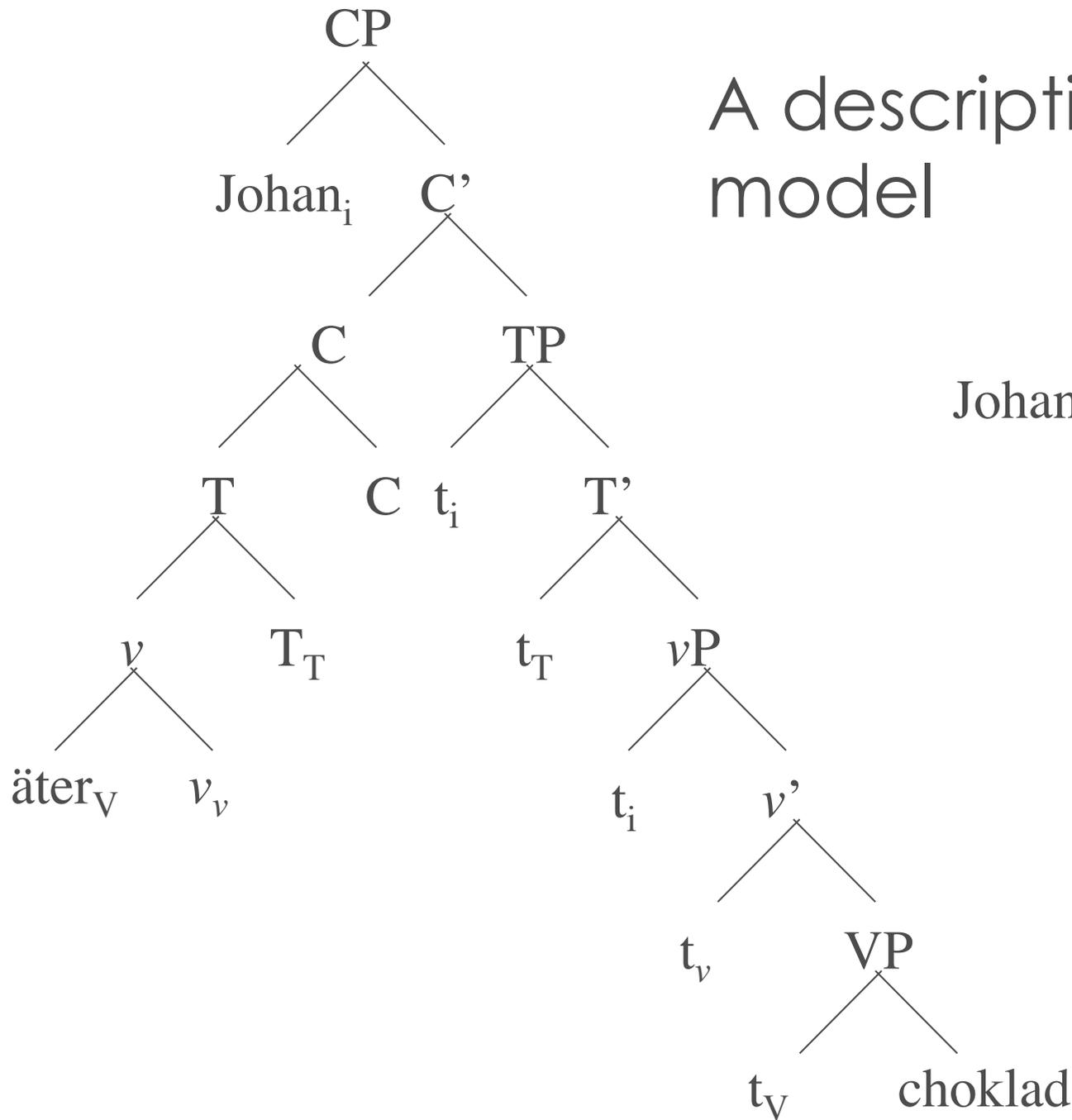
oru_yw_4sp

s	v	o	ATT	s
sASP	eASP			
	ATT	PAUS		där vet jag ju vad som händer
<p2>	ATT	PAUS	där	vet jag ju vad som händer </p2>
</t2><t2>				

System tray area containing icons for network, date (Aug 25), printer, and a taskbar with buttons labeled One, Two, Three, Four, and EXIT.



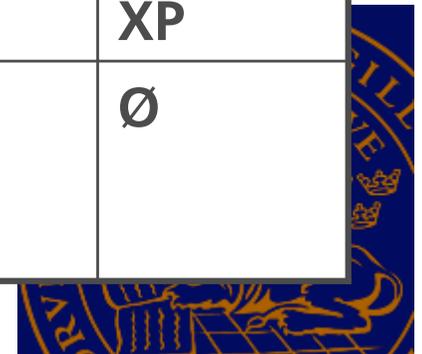
A descriptive syntactic model



Syntactic categories

Category	Head	Phrase
Adjective	A	AP
Adverb	ADV	ADVP
Auxiliary verb	AUX	
Complementiser	C	CP
Conjunction	CONJ	CONJP
Determiner	D	DP
Discourse Marker	DM	
Expletive pronoun	EXPL	
Infinitive Marker	IFM	
Noun	N	NP
Negator	NEG	NEGP
Preposition	P	PP

Category	Head	Phrase
Pronoun	PRON	
Particle	PTL	
Quantifier	Q	QP
Trace	t	t
Tense		TP
Intermediate verb phrase		V'
Copula	VCOP	
Lexical verb	VLEX	
Wh-pronoun	WH	WH
Undefined	X	XP
Phonologically empty or absent	∅	∅

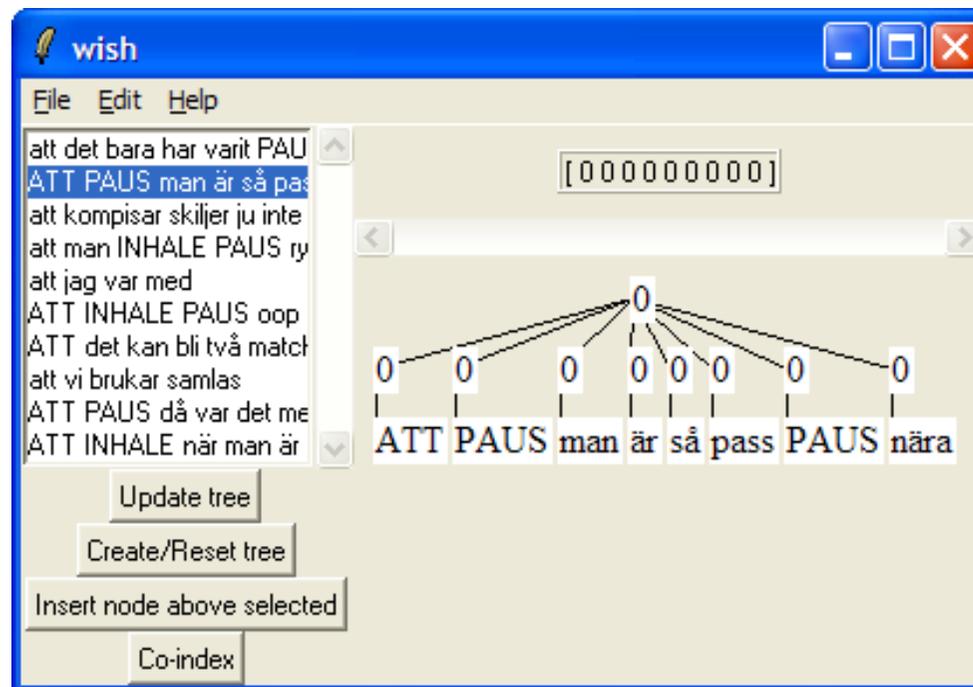


Domain for syntactic analysis:

A clause following att, from the first element of its first constituent to the last element of its last argument linearly. Adjunct phrases and clauses to the right of the matrix clause are ignored.

Grammmal

- Each unit is assigned an unanalysed structure...

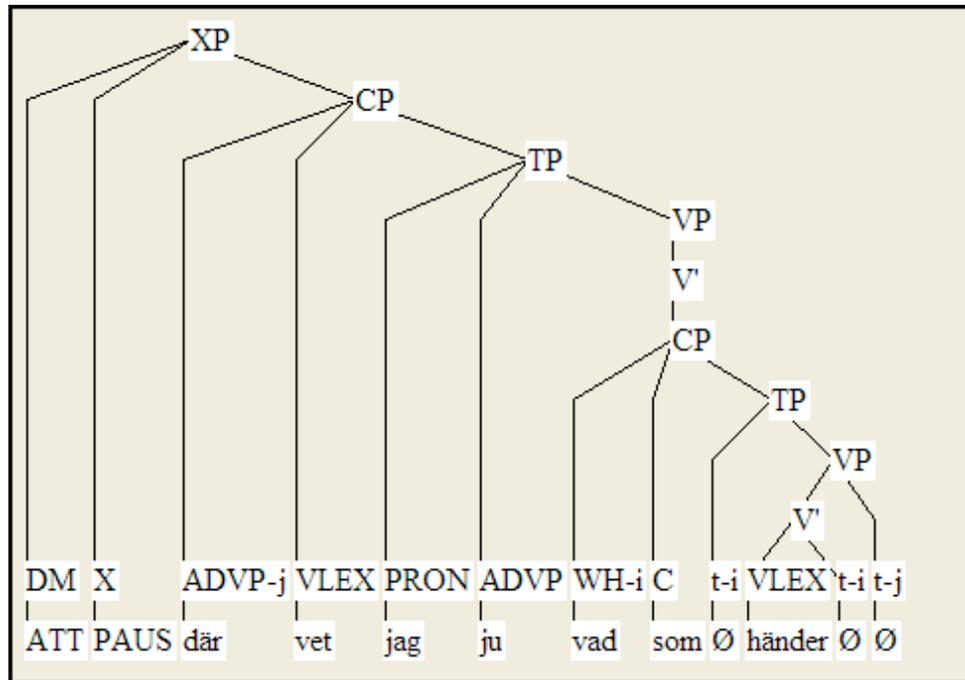


Complexity measurements

- **Depth: sum of the depth of all nodes**
- **Levels: number of levels (max depth)**
- **Nodes: number of nodes**
- **Phrases: number of phrases (all non-terminal CP TP VP DP QP NP AP CONJP XP PP ADVP NEGP)**
- **Words: number of words**
- **Nodes with X or t do not count**



Complexity: example



depth	levels	nodes	phrases	words
68	10	17	7	8



Results

	depth	levels	nodes	phrases	words
"att" (n=64)					
Mean	44,20	6,89	13,41	5,13	7,16
Stdev	26,05	1,56	4,23	1,70	2,48
Median	37	7	13	5	7
"ATT" (n=44)					
Mean	74,23	8,36	16,84	6,73	8,59
stdev	59,47	3,02	7,10	2,97	3,64
median	49	8	14,5	6	8
t-value	3,075	2,919	2,837	3,175	2,257

($p \leq 0.05$ for all)



Another hypothesis to be tested:

Timing restrictions on working memory

Baddeley (1997): part of working memory where speech coding takes place (phonological loop) has a time limit of around 2 seconds.

- Should be reflected in segmentation of spontaneous speech.**



Support:

- often a pause after 2-2.5 sec speech**
- often a F0-reset after 2-2.5 sec speech**
- often final lengthening after 2-2.5 sec speech**
- often a constituent-boundary (often clause boundary) after 2-2.5 sec speech**



oru_yw_4sp.d (S.F.:16000.0) {left:up/down move mid:play between marks right:menu}

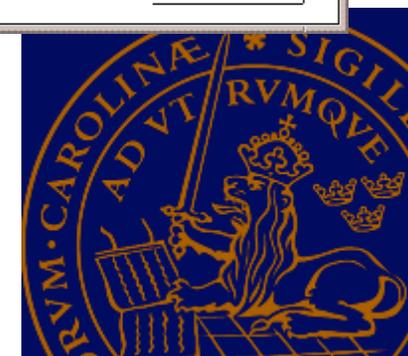
Time: 34.94250sec D: 2.35000 L: 34.65250 R: 37.00250 (F: 0.43)

oru_yw_4sp.f0 (S.F.: 100.0) {left:up/down move mid:play between marks right:menu}

Time(f): 12.00250sec D: 2.35000 L: 34.65250 R: 37.00250 (F: 0.43)

oru_yw_4sp

ORI									
#cp									
xwa									
#Är									
ser									
ser									
ser									
1									
Wrc									
	ATT								
	eASP								
	ATT PAU\$		man är så pass			PAU\$		NÄRA	
	<n?>	PAU\$	man	är	så	pass	PAU\$	nära	
	ATT							</p2>	
	</t2></t2>							</t2></t2>	
								Elaboration	



Conclusions

- The phonetic form of the function word ATT in hesitation contexts is different from that in fluent contexts: useful in speech recognition
- Syntactic complexity after hesitations can be measured and thus predicted after hesitations: useful in models of speech recognition and understanding
- Evidence for 'Phonological loop': useful in segmentation of speech

