Some Simple Automatic Measures of Spoken Interaction

Jens Allwood & Johan Hagman

SSKKII / Dept of Linguistics
University of Göteborg
{jens, hagman}@ling.gu.se

SSKKII Reports #6

Center for Studies of
Language, Cognition, Interaction,
Semantics, Communication, and Information

Abstract

This paper presents a number of simple measures of spoken language and spoken interaction which are based on a standard transcription format and a few simple additions to generally available word processing mechanisms. The measures are divided into (i) volume measures which give the number of words, utterances or other units per recording, speaker or subsection of recording, (ii) density measures which are derived from volume measures to neutralize size and (iii) measures which require minimal coding or a combination of simpler measures. In addition, we discuss ways of visualizing the highly complex data set these measures yield — concerning e.g. dynamics and multidimensional relations.

Contents

1	**Purpose**	**1**
2	**Some simple measures**	**1**
2.1	Transcription	
2.2	Units of spoken language	3
	Utterances	4
	Sequences	
2.3	Simple measures	5
	Volume	
	Density and speed	
2.4	Two examples:	6
	Company meeting and Teenage discussion	
2.5	Comments on the measures	10
3	**Minimal coding**	**10**
4	**More ambitious constructs**	**12**
5	**Dynamics**	**14**
6	**Concluding Words**	**15**
	Appendix I, II, and III *(three pages)*	17

1 Purpose

Spoken language is a fundamental trait of the human species. Its evolution might well be synchronized with that of homo sapiens. For an overview of this discussion, cf. MacNeilage, Studdert-Kennedy and Lindblom (1993). However, spoken language is fundamental also from a social point of view since it is integrated not only with the human brain but also with human society in various not yet totally understood ways.

In spite of its fundamental role, there are still many lacunae in our knowledge of spoken language. These lacunae concern most aspects of linguistic organization, e.g. phonetics, phonology, morphology, syntax, semantics and pragmatics. One of the reasons for this has been the implicit and sometimes explicit focus of linguistics on written language, cf. Linell (1982) and Harris (1980). Another reason has been the relatively recent progress in audio, video and computer technology enabling us to record and analyze spoken language without having to rely on either memory or written language.

In this paper we want to make a small contribution to the study of spoken language by discussing a number of measures of written transcriptions of spoken language which can be obtained very readily by making a few small additions to generally available word processing mechanisms to be described below. We are aware that analyzing transcriptions still involves reliance on a written language version of spoken language. This admittedly is a weakness but we hope that the properties of language we will be discussing will retain their value even after spoken language becomes more accessible for direct measure than today. The measures we suggest can be viewed as alternatives and additions to other quantitative measures such as those discussed, for example, in Biber (1988) or Johansson & Stenström (1991) but have the virtue, we believe, of being slightly more easily available.

2 Some simple measures

2.1 Transcription

Basically, what we want to do is to explore what can be done with units which are directly available in a transcription meeting certain minimal requirements which will be described below, (cf. Hagman & Nivre 1993). We will also discuss what can be achieved if some minimal coding is added to this. The requirements on the transcription are briefly described as follows:

1) Normal ASCII text, including the three Swedish standard letters *å*, *ä*, and *ö*. The orthography is also slightly modified to fit spoken language pronunciation. For example, Swedish spoken *ja* for written *jag* ("I"); *de* for *det* ("it", "that") — which by the way is homographic with *de* ("they"); and *va* for both *var* ("was", "were") and *vad* ("what").

2) Spaces between words (determined through the transcriber's intuition).

3) Indication of speaker by use of Roman capitals and ':', e.g. `A:`, `B:`, etc. For unidentified speakers we always use `X:`.

4) Indication of utterances, where an utterance is operationalized as speech sounds emitted by one speaker bounded by silence or the utterance of another speaker. Silence is primarily important initially or finally in a dialog. Pauses marked by a slash belong to an utterance and are not counted as silence bounding an utterance. In the transcription, utterances are text lines between indications of speakers.

5) Indication of pause: `/`, `//`, `///`, depending on the duration of the pause.

6) Use of capitals in words to indicate stress, e.g. `EMBARRASSED` below.

7) Indication of overlapping speech by matching (and possibly indexed) square brackets: `A:[43...]43 ... B:[43...]43`

8) Comments on the interaction by matching angular brackets `<...>`, where the comment part occurs on a separate line marked by a '`@`' in the margin.

EXAMPLE 1 below shows this type of transcription. The transcription excerpt is taken from the first section of a discussion between students about sex education in school.

EXAMPLE 1 Transcription meeting certain minimal requirements

```
§  Start
M: här får ni nästa ämne
D: <how should sex education be conducted in school> // <ooh>
@  <reads the note about the subject>
@  <deep breath>
M: på engelska nu alltså
B: ja ha
@  <takes the note about the subject and holds it up in front of B>
D: på engelska // USCH va svårt ja kan inte prata engelska
B: <> mja
@  <clears her throat>
C: hm
D: hm /// that is NOT an easy subject <especially not in english>
@  <giggling>
B: nah // it shouldn't be like it was when I was // it was in sixth
   grade or something // and we had a teacher that was so EMBARRASSED
   that she hardly could speak about it <>
@  <giggles>
B: so you just // got more embarassed than you were before
C: [yeah I]
D: [we had] when we were abroad we had a teacher for all <> // the
   school I mean we // all the pupil spent // <eh> all the time in
   the same class-room // I mean from second grade until the gymnasium
   so // and uh // uh she said that / I don't know the teacher // but
   we had some teachers <sometime I don't remember the one> // and she
   told us to write down all the questions on // <like notes> and then
   we put it in a box because everyone is so embarassed
@  <gathering movement with her hands>
@  <outward movement with the hands as if in dispair>
@  <throws out her hands>
@  <shows with her hands>
B: hm hm
.
.
.
```

2.2 Units of spoken language

A precondition for a quantitative treatment is a characterization of the units that are to be counted. At least the following units could be relevant for the spoken part of a dialogue (cf. also Allwood 1992b and Allwood 1994).

1. **Utterances**
   Speakers fundamentally contribute to discourse through utterances which can have various features:

   (i) Phonological: articulatory, perceptual or acoustic features, phonemes, di-/triphones, syllables
   (ii) Lexical: words, phrases
   (iii) Grammatical: parts of speech, construction types such as attribution and predication; phrase types such as NP, VP, S, etc.
   (iv) Functional:
   - Own communication management (OCM)
   - Interactive management (IAM) in:
     - Turn and floor management
     - Feedback management
     - Sequential management
   - Other communicative functions, e.g. statements,questions,requests and exclamations

2. **Sequences**
   A sequence of two or more utterances can form a unit on the basis of at least the following factors (cf. Allwood 1992b):

   (i) turn regulatory mechanisms
   (ii) contextually given functional dependency, e.g. question/answer or statement/positive feedback
   (iii) relevance to a given topic
   (iv) conventionally or functionally motivated subactivity (subsection, phase)

Clearly, the majority of these units and features cannot be quantitatively measured without extensive coding. The phenomena and features that are directly available from a transcription constructed according to the format discussed above are, as already mentioned:

(i) speakers
(ii) utterances (specified as to speaker)
(iii) words
(iv) pauses
(v) overlaps
(vi) stressed words
(vii) comments containing, for example, information about laughter, gestures, and gaze direction

Without too much trouble, it is also often possible to add the following information:

(viii) a division of the transcribed recording into phases or sections dependent on subactivity or topic (cf. Allwood 1992a); such subsections are marked by '§', in the margin, followed by an indication (number and/or name) of the subsection; subsection coding is not a standard part of our transcriptions but can easily be added, as it has been in the transcription above.

(ix) total duration of the transcribed recording

(x) the duration of each subsection (with some trouble).

2.3 Simple measures

On the basis of the units we have discussed above, it is now possible to propose the following measures which can be related to the recording as a whole, to a particular speaker, to a particular section or to a particular speaker in a section.

1. **Volume**

1.	No. of	word tokens	5.	No. of	overlaps
2.	"	word types	6.	"	comments of a specific type (e.g. laughter or gestures)
3.	"	utterances			
4.	"	pauses	7.	"	stressed words
			8.	"	minutes

2. **Density and speed**

By relating volume measures to each other, a number of relative density measures can be defined

	Measure	*Comment*
1.	Type/$\sqrt{\text{token}}$	*Index of Guiraud*, which accounts for the nonlinear growth of the number of tokens with respect to that of types.
2.	Token/utterance	Also known as MLU (Mean Length of Utter.)
3.	Token/minute	Speed
4.	Pause/utterance	Reflection and hesitation
5.	Pause/minute	Reflection and hesitation
6.	Overlap/utterance	Eagerness and/or competition (partial overlap)
7.	Overlap/minute	Eagerness and/or competition (part. overl.)
8.	Utterances without complete overlap	Turns
9.	Utterances with complete overlap	Often "backchannel": utterances which are not turns
10.	*N* most freq. word types	The most frequent vocabulary

Below we demonstrate these measures on two different kinds of recordings; a company meeting and a teenage discussion. TABLE 1 gives some of the volume measures while TABLE 2 shows some of the density and speed measures for the two meetings. In TABLE 3 we then show the volume measures (except duration per speaker) for the company meeting, and in TABLE 4 we exemplify how these measures can also be applied to a single subsection. TABLE 5 finally shows density for the participants of the company meeting. The meeting has been divided into the following 12 subactivities: **I** Öppning (Opening); **II** Dotterföretag (Daughter company); **III** Utgångspunkter (Points of depature); IV Gruppindelning (Group division); **V—VIII** Redovisningar nr. 1—4 (Presentations no. 1—4); **IX** Föreslagna åtgärder (Suggested measures); **X** Konkretiseringskrav (Concrete demands); **XI** Framtida möten (Future meetings); **XII** Åtgärder 2 (Measures 2)

2.4 Two examples: Company meeting and Teenage discussion

TABLE 1 Comparison of volume totals for two recordings

	Company meeting	**Teenage discussion**
Volume		
Word tokens	22.073	5.250
Word types	2.469	1.078
Utterances	1.284	647
Overlaps	164	216
Pauses	177	377
Minutes	139	32

TABLE 2 Comparison of density and speed for two recordings

	Comp. meeting	**Teenage discussion**
Density		
Type/√ token	16.6	14.9
Token/utterance (MLU)	17.2	8.1
Overlaps/utterance	0.1	0.3
Pauses/utterance	0.1	0.6
Speed		
Token/minute	159	164

A comparison of the two recordings shows that the company meeting is longer than the discussion, therefore density measures are to be preferred. If we take a look at these ratios, we see that the type/√token ratio is higher and utterances are roughly twice as long in the meeting as compared to the discussion. The speed of the two recordings differs insignificantly but the teenage discussion has almost three times as many overlaps as the company meeting which could mean that it has a livelier tempo. This hypothesis is perhaps slightly contradicted by the fact that the teenage

discussion has almost four times as many pauses. This, however, could also be an effect of the higher speed and greater number of overlaps in the teenage discussion.

TABLE 3 Company meeting participants (volume and % of totals)

	A	B	C	D	F	G	H	J	K	M	R	X
Wto	5960	2865	2668	2788	1547	1863	1408	200	329	711	1628	106
″ %	*27*	*13*	*12*	*13*	*7*	*8*	*6*	*1*	*1*	*3*	*7*	*0*
Wty	1103	701	603	641	477	536	385	107	161	273	463	77
″ %	45	28	24	26	19	22	16	4	7	11	19	3
Utts	326	162	98	148	79	71	127	33	28	43	126	43
″ %	*25*	*13*	*8*	*12*	*6*	*6*	*10*	*2*	*2*	*3*	*10*	*3*
Ovl	35	21	16	14	10	9	8	13	0	6	11	19
Pau	65	25	15	38	2	10	8	1	0	3	8	2

TABLE 4 Comparison of speakers in a subactivity (co. meeting, sect. III)

	A	B	C	D	F	G	H	J	K	M	R	X	∑
Wto	1339	210	65	199	36	44	3	48	0	30	153	7	2134
″ %	*63*	*10*	*3*	*9*	*2*	*2*	*0*	*2*	*0*	*1*	*7*	*3*	*100*
Wty	359	112	48	118	27	32	3	35	0	28	88	7	414
Utt	58	21	12	21	1	4	1	7	0	3	20	5	146
″ %	*40*	*14*	*8*	*14*	*1*	*3*	*1*	*5*	*0*	*2*	*14*	*3*	*100*
Ovl	16	3	1	2	0	3	0	3	0	1	2	3	34
Pau	21	1	0	1	0	0	0	0	0	0	1	0	24

TABLE 5 Density measures for the company meeting participants

	A	B	C	D	F	G	H	J	K	M	R	X
Type/√tkn	14.3	13.1	11.7	12.1	12.1	12.4	10.3	7.6	8.9	10.2	11.5	7.5
Tkn/utter.	18.3	17.7	27.2	18.8	19.6	26.2	11.1	6.1	11.8	16.5	12.9	2.5
Ovl/utter.	.11	.13	.16	.09	.13	.13	.06	.39	.00	.14	.09	.44
Pau/utter.	.20	.15	.15	.26	.03	.14	.06	.03	.00	.07	.06	.05

If we instead turn our attention to the participants (in the company meeting), TABLE 3 shows that speaker A utters 27% of all words and uses 45% of all word types. A also contributes 25% of all utterances but his utterances are not the longest. He does not overlap more than others but he has the second highest pause rate. If we consider TABLE 4 for subsection III of the meeting, we see that A's word share goes up to 63% and that his share of utterances is 40%. This means at least that A occupies a lot of verbal space (cf. below for a discussion of dominance). We also see that A has the highest type/√token ratio (14.3). This could indicate that his word types vary more than other speakers'. Since the type/token ratio typically increases faster than linearly, type/√token (the index of Guiraud) could be a way of compensating for this, in order to obtain a measure of how "rich"

or “varying” a speakers's language is. In this example, though, this formula yields values which seem quite well correlated to those of tokens (Wto).

TABLE 6 Comparison of most frequent wordtypes in the subactivities of a company meeting

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1	det	det	det	det	vi	det	det	det	det	det	det	det
2	vi	att	är	vi	det	och	och	de	att	vi	är	att
3	att	som	vi	är	med	de	är	och	är	och	vi	är
4	och	och	jag	ja	är	att	vi	att	och	att	då	jag
5	så	är	har	tre	så	vi	att	är	vi	är	så	då
6	jag	så	ja	att	och	jag	har	jag	jag	jag	ja	så
7	har	i	och	så	här	är	de	så	har	som	ska	vi
8	här	man	i	och	som	i	då	som	de	har	jag	inte
9	om	då	att	jag	ska	som	jag	med	så	så	och	ju
10	inte	en	som	då	har	har	så	på	om	kan	möten	för

In TABLE 6, we show the most frequent words, based on a morphologically disambiguated transcription, over the 12 subactivities of the meeting. As we see, there is a remarkable stability in the frequency rates for different words over the 12 subactivities. Most of the highly frequent words in one subactivity are also highly frequent in all other subactivities. Perhaps this points to structurally similar needs in all subactivities.

The frequency list can be filtered in many ways in order to look for the occurrence of specific types of words. One such filtering divides a frequency list into categorematic and syncategorematic words. The distinction categorematic/syncategorematic is of medieval origin and is built on the idea that some words are categorematic, i.e. nouns, verbs, adjectives, and some adverbs. Categorematic words indicate objects, substances, processes, and properties. Other words are *syn*categorematic, indicating relations between categories, i.e. prepositions, conjunctions, articles, pronouns, numerals, and some adverbs.

In order to get a rough indication of the categorematic content of the subsections or a more conceptual domain, we can now let the frequency lists pass through a filter where syncategorematic words are filtered out. TABLE 7 displays the result of such a filtering.

TABLE 7 Comparison of most frequent (categorematic) wordtypes in the subactivities of a company meeting

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1	kommer	kommer	dag	grupp	statliga	kontakt	säga	kontakt	kommer	seminarium	möten	brevutskick
2	tiden	firma	veckan	grup-per	kommer	bra	kontakt	säger	coronado	bestämma	gång	ansvarig
3	stänga	se	står	bra	ka4	mer	bra	säga	brev	säga	fredag	fall
4	straff	pengar	ser	exempel	grupp	stock holm	står	fält	kontakt	kommer	eftermiddag	brev
5	sitta	mer	bra	kundämnen	bra	göteborg	skrivit	pratade	säga	försöka	korta	skicka
6	pratar	företaget	timmar	sätt	rutinerna	behöver	poseidon	företag	seminarium	seminarier	möte	seminarium
7	känner	jobbar	satt	plus	första	samma	kommer	centralt	prata	se	månaden	menar
8	informera	sälja	namn	kommer	trycka	håller	firma	sitter	kullenbergs	utskick	onsdag	komma
9	början	fall	fall	förslag	totala	se	säger	sa	bra	varje	mars	bättre
10	kamran	fairway	stämmer	frågan	sälja	sa	stock holm	bra	sagt	tid	kort	sida

Not surprisingly, the list is full of words which specifically point to activities which are relevant for a consultancy company.

Other measures are also possible to define in a fairly simple way. Three of these are (for reasons of space, we will not discuss their adequacy here):

1. *Communicative function* as expressed by single lexical items. Below, in TABLE 8, we will illustrate this with feedback words. The words are first listed and then their share of the vocabulary of the activity, of a speaker or of a subactivity is calculated
2. *Speaker dominance*, a characteristic which can at least partly be illustrated by combining the following measures: % word token; % non-total overlap; % utterance; % non-total overlap/utterance
3. *Support,* which can be partly captured by a combination of other measures: % total overlap and % FBW or FBWU (see section 4).

2.5 Comments on the measures

As mentioned above, we have four different potential bases for the measures:

1. The recording (activity) as a whole
2. A speaker's verbal output during a whole recording
3. A subsection (subactivity) of the recording
4. A speaker's verbal output in a particular subsection

All measures are probably not of equal interest in relation to all four bases. For example, an individual's share of overlap is probably more interesting than the overlap calculated for the activity as a whole or for a subactivity, if the distribution of overlap over individuals is very uneven. Secondly, while absolute measures sometimes are interesting, mostly we are interested in relative measures. We want to know an individual's share of the word tokens produced, rather than the absolute number of tokens. There is also a lot more to be said about how to avoid various statistical effects. Thirdly, for some measures it would be interesting to have an expected, "normal" value to compare a given measure with. This is so, especially if we want to compare different recordings with each other. We want to say, for example, that the individuals in a given activity speak faster or slower than what is considered "normal" or we want to make statements about the normality of a certain distributional pattern, etc. It might for instance be the case that the chairperson of a meeting normally has a greater share than other participants individually (perhaps 30-50%) of both word tokens and utterances. If he/she has an even greater share (> 50%), maybe we would say that the chairperson is unusually dominant. Conversely, if he/she has a smaller share (< 30%), we could say that the chairperson is unusually "low key".

By comparing with other recordings of a particular activity, the measure "the 100 (or any other suitable number) most frequent word types", as illustrated in TABLES 6 and 7 above, can serve as one way of investigating whether an activity or individual shows a persistence in respect to word types used. For an individual this can also be investigated by looking at his or her variation over subactivities in a given recorded activity. A more fine grained analysis of this type could be done if the semantic field relations between the words were taken account of. In this way we could also get a picture of the persistence of a particular semantic or conceptual domain.

3 Minimal coding

Besides the measures exemplified above, a few more measures can be added with minimal coding effort. The simplest type of coding we have in mind here is based on word type identification. This type of coding can, for

example, be used to select a word type or a set of word types to represent some feature of meaning or communicative function. The following are some possibilities:

Function		Example
1. Negation		*nej, nä, inte, ingen, inga* (*no, not none*)
2. Modality:	necessity:	*nödvändig, tvungen* (*necessary, forced*)
	possibility:	*möjlig, eventuell* (*possible*)
3. Epistemic:	knowledge:	*veta, kunna* (*know, be able to*)
	belief:	*tro, anse, tycka* (*believe, consider, think*)
4. 3 dimensional relations between objects:		*i, in, in i* (*in, into*)
5. Causal relations between states of affairs:		*därför att, ty, för att* (*because, for, since*)
6. Deictic functions: personal:		*ja, jag, du, han, hon, vi, ni, dom* (*I, you, he, she, we, you, they*)
	spatial:	*den, det, den här, denna, detta, den här, det här, den där, det där* (*this, that*)
	temporal:	*nu, då, imorgon, ikväll, idag* (*now, then, tomorrow, tonight, today*)
7. Feedback functions:		*ja, jaa, jo, nä, nej, m, mm* (*yes, no, m*)
8. Own communication management:	morphemes:	*äh, öh, hm* (*eh, er, hm*)
	repetition:	*vi-visste* (*kne-knew*)
9. Any semantic field:		For example the one associated with *natur* (*nature*) containing words like *essens, l andskap, verklighet, kropp*, (*essence, landscape, reality, body*,)
10. Conceptual focus:		All word tokens of a particular field/total categorematic word tokens
11. Certain root fields:		As a special case of a semantic field we could investigate for example all words related to the root *natur* by derivation or inflection like *natur-lig* (*natural*) or *natur-en* (*the nature*).

All of these measures can then be relativized to each other or to the more global measures described above. Here are some examples of some measures that might be applied to feedback words cf. also Allwood (1993):

12. Feedback word tokens/total word tokens (FBW)
13. Feedback word type/total word types (FBW type)
14. Feedback word types / √(feedback word tokens) (FBW type/√tokens)
15. Feedback word tokens/utterance (FBWU)

Use of the feedback measures is illustrated in the following table:

TABLE 8 Feedback in two activities

	Company meeting	**Teenage discussion**
FB Tokens	637	359
FB Types	14	22
FBW (%)	2.9	6.8
FBW type (%)	0.6	2.0
FBW type/√ tokens	0.6	1.2
FBWU	0.5	0.6

We can observe that there is more feedback in the teenage discussion than in the company meeting. 6.8% of all words versus 2.9% are feedback words. We also observe that the teenagers use more different types of feedback and that it occurs in a greater percentage of their utterances. It should be noted that these results are not fully reliable, since the two transcriptions are not based on the same principles and all relevant feedback is not captured. They are intended merely as an illustration of how a (over-) simplified measure of feedback can be constructed.

4 More ambitious constructs

Ideally, we want to characterize spoken interactions using more complex concepts than the ones we have so far discussed. Some of the relevant concepts are the following: power, control, dominance, liveliness, involvement, interest, conflict, attack, support, certainty, confidence, trust, relevance, understanding, empathy, liking. But also some of the possible opposites of these concepts are interesting: subordination, submission, boredom, "low key", passivity, cooperation, help, uncertainty, hesitation, distrust, suspicion, lack of understanding, misunderstanding, "coldness", dislike, etc.

Closer to linguistic communicative concerns, we might want to code the type of :

(i) reference
(ii) predication
(ii) attribution
(iv) concepts
(v) semantic-epistemic operations
(vi) speech acts
(vii) own communication management
(viii) feedback
(ix) turn management
(x) interruptions
(xi) pauses
(xii) sequencing

For both of these types of measures, nothing can, at least not initially, replace a close word by word analysis and coding. We can, however, strengthen intuition-based coding by certain simple measures which often

correlate with the more complex characteristics we are interested in. In doing this, we should also try to experiment in combining several simple measures to throw light on more complex measures. Below are some examples. Again, space does not allow us to discuss or motivate the measures. ";" indicates that the measures are meant to be combined.

(i) *Liveliness:* FBWU; overlap/utterance. The measure could be based on combining rank numbers for both FBWU and overlap. The measure can be used for comparison between recordings, persons or sections. It could also be combined with a relative or standardized normality rating.

(ii) *Dominance:* (= occupation of verbal space) Word share; utterance share. This measure could also make use of a combination of rank numbers and could be used to compare individuals. To use it on the activity or subactivity level would require comparison of distri butions.

(iii) *Own Communication Management* (OCM)
(cf. Allwood, Nivre & Ahlsén 1990)
Pause share; OCM morpheme share;
OCM word constructions.

hesitation: Pauses, OCM morphemes and self repetitions can be divided by the total number of words to give a total word level OCM share (OCW). OCM constructions can be divided by the number of utterances to give an OCMU share. The ranking of the two measures can be added to give a total OCM measure.

(iv) *Support:* Ranks for FBW, FBWU and complete overlap could be added. See TABLES 9A and 9B below.

(v) *"Low key":* Ranks for share of words, share of utterances, share of overlaps are added. The lowest ranking individuals or activities are the most "low key".

(vi) *Uncertainty:* Ranks for OCM and ranks for use of doxastic predicates (believe, think, suggest, etc) could be added

TABLE 9A Support measures from the company meeting

Meas.\Spkr	**M**	**C**	**J**	**F**	**B**	**D**	**R**	**A**	**H**	**G**	**K**
Feedback											
FB tokens	28	80	10	45	79	82	51	160	51	30	6
FBW (%)	3.94	3.00	5.00	2.91	2.76	2.94	3.13	2.68	3.62	1.61	1.82
FBWU	.65	.82	.30	.57	.49	.55	.40	.49	.40	.42	.21
Overlap											
Compl. ovl	3	3	4	2	5	3	2	6	2	2	0
- " - /U (%)	6.98	3.06	12.12	2.53	3.09	2.03	1.59	1.84	1.57	2.82	.00

TABLE 9B Support rankings from the company meeting;
$3 \leq$ support index ≤ 33

Meas.\Spkr	**M**	**C**	**J**	**F**	**B**	**D**	**R**	**A**	**H**	**G**	**K**
Ranks											
FBW	10	7	11	5	4	6	8	3	9	1	2
FBWU	10	11	2	9	6	8	4	7	3	5	1
Complete ovl/Utter	10	8	11	6	9	5	3	4	2	7	1
Support index	**30**	**26**	**24**	**20**	**19**	**19**	**15**	**14**	**14**	**13**	**4**

TABLES 9A and 9B together exemplify a complex measure of support which is derived for each speaker by first calculating the relative feedback share both in terms of tokens and utterances as well as the relative share of overlaps in terms of utterances. On the basis of these measures (TABLE 9A) ranks are assigned to all speakers (TABLE 9B), where the speaker with the highest value gets 11 (which is equal to the number of speakers) and the one with the lowest gets 1. Finally the three ranks are added together and a total support score is calculated. Thus, this will order the speakers within a support index range from 3 to 33 and, as we can see, M turns out to be the most supportive participant of the meeting and A ranks as the eigth most supportive member. Since the score depends on the total number of speakers in the recording it can not be used to compare speakers across recordings.

5. Dynamics

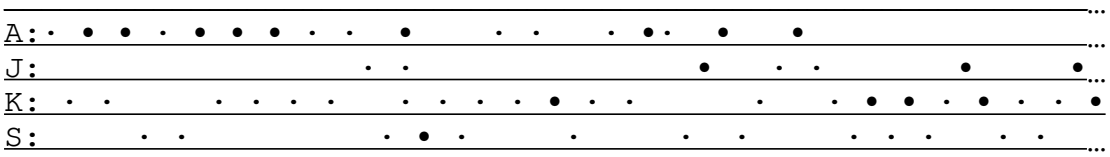
Simple measures can also be used as a point of departure for a study of the dynamics in an interaction. One way to get a picture of the dynamics is simply to arrange the subactivities or phases of an activity along a time scale with some representation of their relative length. The measurements discussed above can then be plotted along the sequence, giving a picture of the subactivity (phase) bound sequential variation in the activity. This will also give a somewhat coarse idea of the temporal dynamics of the activity. In FIGURE 1, APPENDIX I, we exemplify the kind of diagram we have in mind, by showing how verbal dominance, as measured by the number of

uttered words, varies over speakers and subactivities. We regret that the copy presented here does not render the original colour-coded information.

In FIGURE 2, APPENDIX II we can see how a measure like share of word tokens and share of utterances can be aligned to give a picture of the magnitude of the sections in terms of words and utterances.

If we want to study the interaction on a finer time scale, most of the measures are inapplicable since they require accumulation over a period to be interesting. What a study with a finer time scale will yield is a kind of "musical score diagram" of speaker activity. This is interesting if we want to study interaction in a precise way. Cf. FIGURE 3.

FIGURE 3 Speaker activity in the beginning of the teenage discussion with dots whose size represent the utterances' length (· < MLU < •)



Another problem in studying interaction is the question of who interacts with whom. Very often we find that some speakers occupy a more central role than others. This can, for example, be reflected in the fact that more remarks are addressed to a particular person or in the fact that this person's remarks to a greater extent are taken notice of. One way to approach this problem in an automatic way is to keep track of which speaker follows which speaker. To some extent this might reflect interests and alliances among speakers. In TABLE 10, APPENDIX II, we present an example of such a "who follows whom" table and in FIGURE 4, APPENDIX III we show a way to represent the information in the table graphically in a cluster diagram in which the relative spatial positions of the speakers are (automatically) calculated as a function of the first column of each of the eleven column triads in TABLE 10 (the unidentified speaker X was ignored here). In other words, speakers who are close to each other follow each other often. The background shades of gray indicate how the number of produced word tokens varies between speakers (light shades indicate more word tokens).

6. Concluding Words

In this paper we have discussed various ways in which rather simple extensions to commonly available word processing programs can be used to provide simple but still interesting quantitative characteristics of transcribed spoken material. In the future, we would like to extend the work reported by combining the measures in more complex ways to capture richer concepts. We also believe that at the present time insights can be

gained by a combination of a more subjective approach to coding with a use of the more crude measures described here. This is also necessary to get an impression of the strengths and weaknesses of the measures. It would, in fact, be interesting to compare results of complex (but still automatic) measures with results of "interpretative coding" to get a cost/benfit estimation, i.e. to see how crude the automatic measures are and how costly the non-automatic ones are.

Acknowledgements

We would like to thank Elisabeth Ahlsén and Jokim Nivre for valuable comments and discussion.

References

Allwood, J. (1992a) Kommentarer till ett samtal: kommunikativ verksamhetsanalys, in Hellberg, S., Kotsinas, U.-B. och Lindell, I. (eds) *Svenskans Beskrivning* 1992. Lund University Press, Lund.

Allwood, J. (1992b). On dialogue cohesion, *Gothenburg Papers in Theoretical Linguistics*, no. 65, Department of Linguistics, Univer sity of Göteborg: also in L. Heltoft (ed) *Papers from the 13th Scandinavian Conference of Linguistics*, (1994) Roskilde University Centre.

Allwood, J. (1993) Feedback in Second Language Acquisition,. in Perdue, C. (ed.) *Adult Language Acquisition. Cross Linguistic Perspectives*, Vol. II. Cambridge: Cambridge University Press, pp. 196-235. Also in Feedback and language acquisition, *Gothenburg papers in Theoretical Linguistics*. University of Göteborg, Dept of Linguistics, 1992.

Allwood, J. (1994) Nogle Perspektiver på talesprogsforskning in *Nysproglige Studier* no. 18, Dansklærerforeningens forlag, Copenhagen

Allwood, J., Nivre, J. and Ahlsén, E. (1990).Speech Management: On the Non Written Life of Speech.,*Gothenburg Papers in Theoretical Linguistics*, 58. Univ. of Göteborg, Dept of Linguistics. Also in *Nordic Journal of Linguistics*, 13.

Biber, D. (1988) *Variation across Speech and Writing*, Cambridge University Press, Cambridge.

Hagman, J. & Nivre, J. (1993) *Transcription Standards*, Technichal Report, Dept of Linguistics, University of Göteborg.

Harris, R. (1980) *The Language Makers,* Duckworth, London

Johansson, S. & Stenström, A.-B.(1991) *English Computer Corpora*, TiEL, Mouton de Gruyter, Berlin.

Linell, P. (1982) *The Written Language Bias in Linguistics*, SIC 2, Tema Kommunikation, University of Linköping.

MacNeilage, P., Studdert-Kennedy, M. & Lindblom, B. (1993) Hand Signals, in *The Sciences*, Jan/Feb '93, The Academy of Sciences, New York.

FIGURE 2 The twelve sections of the Company Meeting divided proportionally to their share of tokens (upper line) and their share of utterances (lower line).

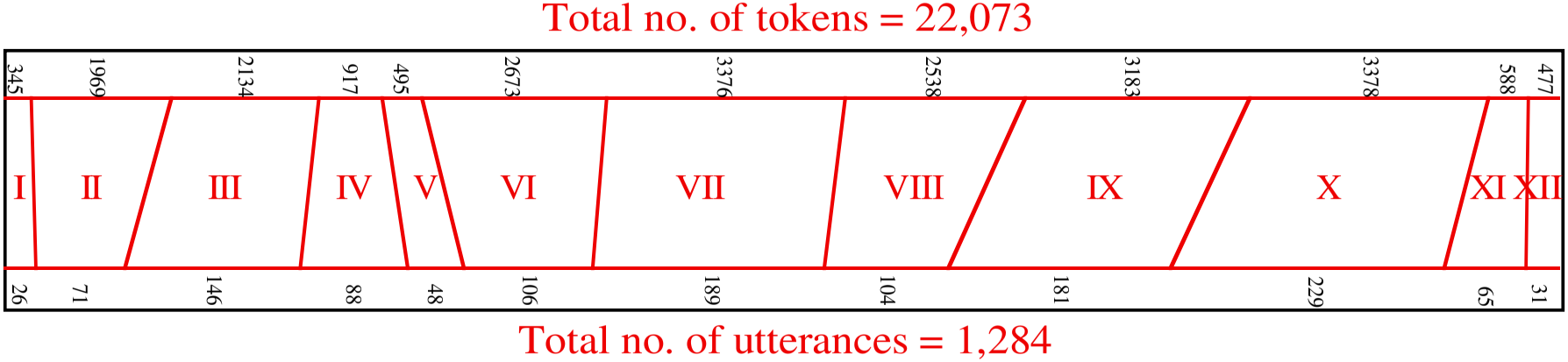


TABLE 10 The speakers' utterance sequence in the Company Meeting

	A			B			C			D			F			G			H			J			K			M			R			X		
A:	6	0	NB!	70	40	176	26	25	106	52	37	139	26	20	131	21	17	121	37	32	115	15	8	178	9	7	126	17	11	155	36	31	115	11	9	120
B:	57	40	143	6	0	NB!	15	12	126	18	18	99	12	10	124	3	8	35	22	16	140	2	4	49	1	3	29	7	5	131	15	15	98	4	4	89
C:	26	25	106	17	12	142	2	0	NB!	10	11	89	11	6	184	11	5	211	7	10	73	3	3	119	1	2	47	4	3	121	3	9	32	3	3	109
D:	52	37	139	17	18	94	10	11	89	2	0	NB!	7	9	77	11	8	139	14	15	95	4	4	104	6	3	184	6	5	120	13	14	91	6	4	143
F:	26	20	131	9	10	93	13	6	218	5	9	55	1	0	NB!	7	4	165	5	8	64	0	2	0	3	2	172	0	3	0	9	8	118	1	2	45
G:	19	17	109	5	8	59	11	5	211	10	8	126	8	4	189	3	0	NB!	5	7	73	3	2	168	0	2	0	1	2	43	3	7	45	3	2	154
H:	40	32	124	12	16	76	8	10	83	18	15	123	5	8	64	4	7	58	1	0	NB!	0	3	0	2	3	71	2	4	46	31	12	251	3	4	83
J:	14	8	166	2	4	49	4	3	158	7	4	182	1	2	49	1	2	56	1	3	30	0	0	0	0	1	0	1	1	88	2	3	62	0	1	0
K:	11	7	154	1	3	29	0	2	0	7	3	215	0	2	0	0	2	0	3	3	107	0	1	0	0	0	0	0	1	0	5	3	182	1	1	124
M:	19	11	173	6	5	112	3	3	91	5	5	100	0	3	0	3	2	129	2	4	46	0	1	0	0	1	0	0	0	0	3	4	71	2	1	162
R:	43	31	137	10	15	65	6	9	64	10	14	70	7	8	91	3	7	45	27	12	218	4	3	123	6	3	218	5	4	118	3	0	NB!	2	4	57
X:	12	9	131	7	4	156	0	3	0	4	4	95	1	2	45	4	2	205	3	4	83	2	1	211	0	1	0	0	1	0	3	4	85	7	0	NB!

(Underlined parts read: A follows himself **6** *actual* times out of **0** *expected* times
D is followed by B **17** *actual* times out of **18** *expected* times ≈ **94**%.)

The (rounded) numbers of expected times were obtained by assuming an even distribution; if S_1 and S_2 are two different speakers, U_{Sn}, U_{exp}, and U_{tot} are the actual, the expected and the total number of utterances respectively, and R is self-succession (marked 'NB!' above) then: $U_{exp}<S_1, S_2>=(U_{S1}-R_{S1})*(U_{S2}-R_{S2})/(U_{tot}-R_{tot})$. **APPENDIX II**

APPENDIX I

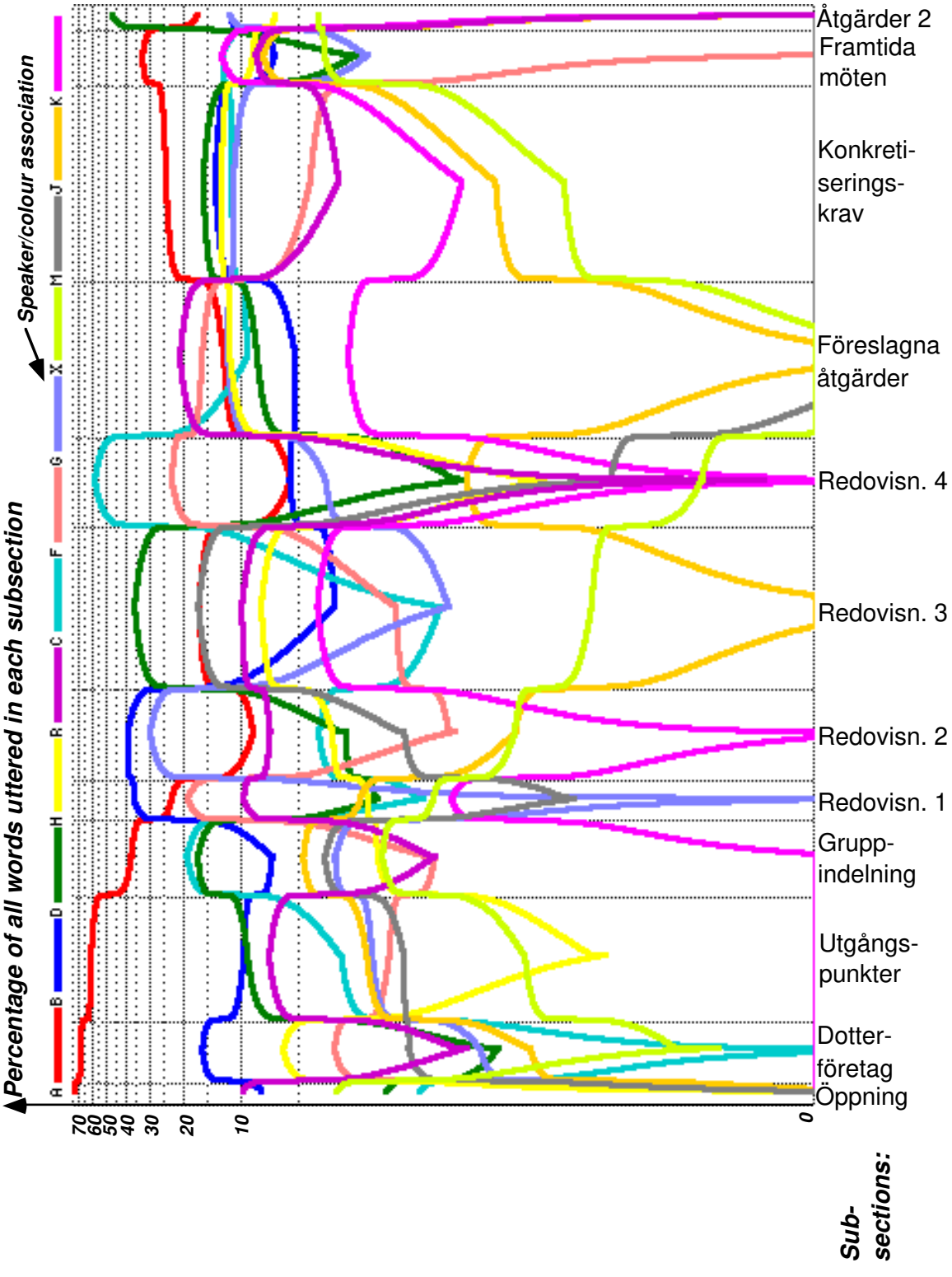


FIGURE 1 Logarithmic diagram showing the share of all uttered words for each speaker, sectionwise

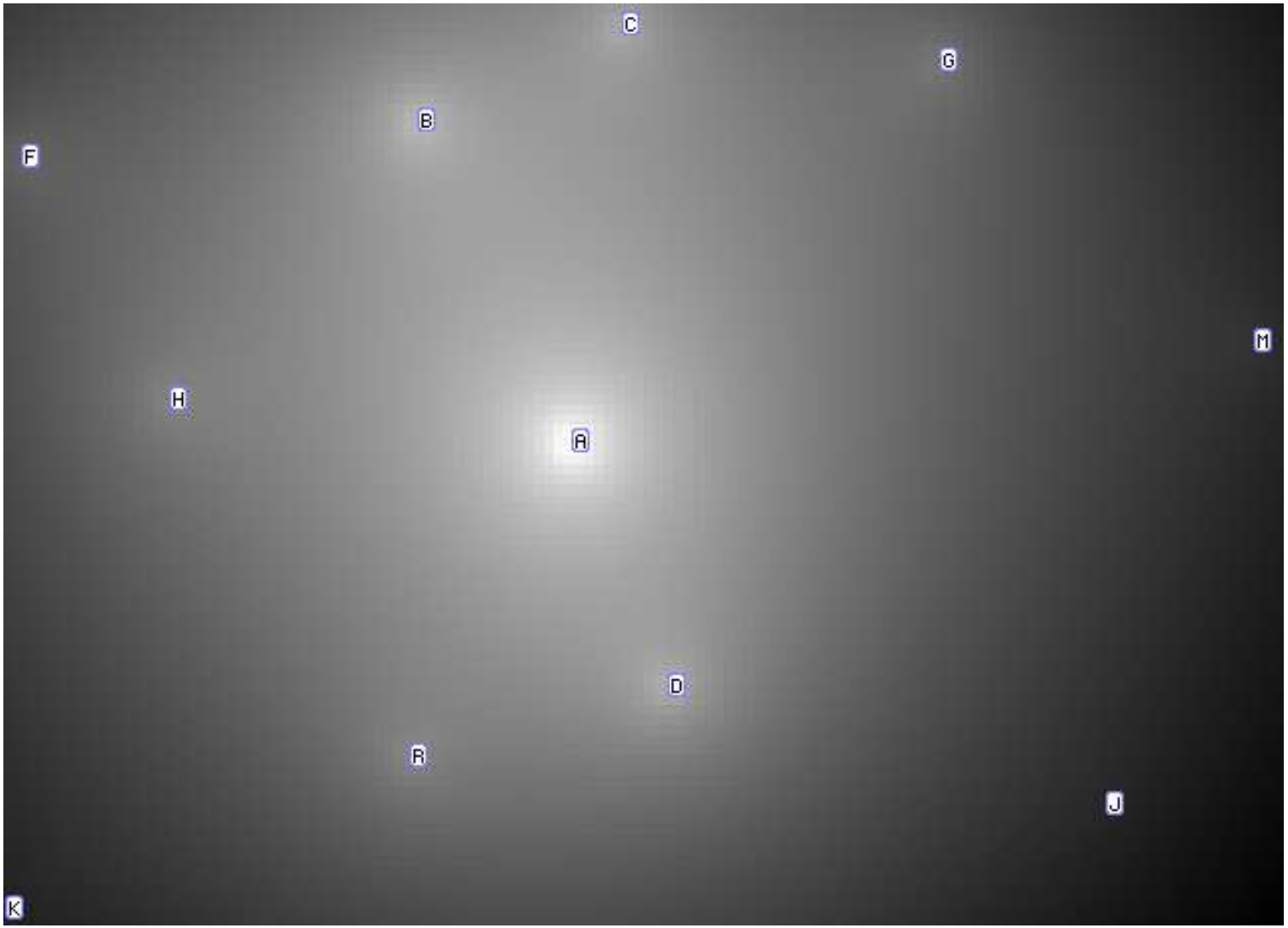


FIGURE 4 Configuration of speakers obtained by cluster analysis of the values in TABLE 10. The actual numbers of times when e.g. A follows B and B follows A were added and treated as an index of attraction. These indeces for every possible pair of speakers then contributed to the cluster formation shown here. The background's gradually changing shades of gray mirror the nearest speakers' shares of words in the Company Meeting.