FOCUS MOVEMENTS AND THE INTERNAL IMAGES OF SPOKEN DISCOURSE

Kenneth Holmqvist
Jana Hol ánová

Lund University Cognitive Science
Kungshuset, Lundagård
S–222 22 LUND

kenneth.holmqvist@fil.lu.se
jana.holsanova@fil.lu.se

Abstract: When a speaker spontaneously draws, what is the correlation between visual focus movements over the drawing and focus movements in the discourse? How is the listener's internal image constructed from the spoken discourse? We studied these questions empirically by transcribing spoken language data into intonation units and comparing them to events in the drawing. In our analysis, we employed an image oriented semantics. Several visual processes could then be identified in spoken conversation. A number of small markers, which we named attention movers, played a fundamental role in marking the transition between focuses of attention. And, when building his/her internal image, the listener simply superimposes the lexical images of words and intonation units onto a progressing internal image of the discourse.

1. INTRODUCTION[1]

Chafe (1980, 1994) formulated the hypothesis that similar mechanisms are used a) when we perceive information visually, b) when we recall it from memory, and c) when we verbalise it. "All three processes may be guided by a single executive mechanism which determines what is focused on, for how long, and in what sequence." Chafe (1980:16).

The analyses we present below stem from two projects studying the correlations between units in vision and in discourse. The first of these projects (Kenneth Holmqvist) attempts to describe focus movements to an extent that will allow a future computational model of discourse understanding, in terms of an evolving mental image. The other project (Jana Hol ánová) is an empirical investigation of the correlation between eye movements when watching a narrative film, and focus movements in the spontaneous retellings of such dynamic visual events.

In this paper, we will concentrate on the reconstruction of the participants' internal images during spoken discourse. Our first question is: What happens if the speaker draws to illustrate something in the discourse? The focus movements should then leave traces in the drawing, that could tell us how the focus of attention moves over the undrawn internal images of speaker and listener. The first part of this paper discusses the correlation between spoken discourse and focus movements, exemplified by an analysis where a speaker uses an abstract drawing as a complement to a descriptive explanation.

[1]This research has been financed by the Swedish Council for Research in the Humanities and Social Sciences.

The second part of this paper discusses how we can construct a close representation of the internal images and the focus movements over it. We use Langacker (1987) and Holmqvist (1993) on an authentic example to build a spatial layout, which can then be filled in with complete lexical images.

2. FOCUS OF ATTENTION AND INTONATION UNITS

When we listen to spoken discourse in an authentic communicative situation, we filter out all interruptions and odd repetitions, and perceive it as a continuous stream of ideas. The discourse seems coherent and fluent. Not until we transcribe and analyse a spoken text, do we realise that it consists of small units and includes small words (discourse markers) that reflect the planning and production process of the speaker. Studies made on reports of an event, on spoken descriptions, and on retellings of a film (Chafe 1980, Redeker 1990), show that we formulate in brief spurts, focus on small units, jump between different sequences, and add digressions and comments.

How can we explain this normally imperceptible fragmentation of spoken language? Chafe (1980) decided to base the explanation on the properties of our attention: We can concentrate only on parts of the vast amount of information we get from our perceptual system, our emotions, and our memory. We choose a specific part at a time, according to our needs, interests, and current goals. Our attentional system can be conceived of as a limited capacity resource which selects limited information for further specialised processing (cf. Tomlin, in press). Therefore, when we retell something from memory, our attention moves sequentially from focus to focus as the retelling progresses.

In spoken language, these foci are verbalised in what Chafe (1994) calls *intonation units*. In our material, intonation units vary in length between 0.3 and 2.5 seconds. There are often pauses between the units, but they are rather given a prosodic characterisation (as in Chafe 1994:58): "The features that characterise intonation units may involve any or all of the following: changes in fundamental frequency (perceived as pitch), changes in duration (perceived as the shortening or lengthening of syllables or words), changes in intensity (perceived as loudness), changes in voice quality of various kinds, and sometimes changes of turn." Intonation units may coincide with syntactic units (such as sentences), but more often are there mismatches between syntactic and intonation units in spoken discourse. A sentence is often spread over a number of intonation units.

Intonation units group to larger units in spoken discourse. According to Chafe, foci combine to *centres of interest* or *super foci*, i.e. cognitive units based on experience, intellect, and judgement, which intermediate between intonation units and *basic-level topics*, which are units at an even higher level (cf. Chafe 1994: 137-138).

3. ANALYSES OF TWO DISCOURSE PASSAGES

Below we present focus movement analyses from two discourse passages. Both have the same main speaker, a young Swedish-Canadian man, whom we call A. During his short visit to Sweden, he is having an evening of small talk with his friends. The other two participants, B and C, leave the initiative to A and usually speak only in the form of questions or support signals such as humming and laughter.

Focus in the drawings and in the figures is shown with a white spot-light. Those parts that are currently not focused are shadowed (Chafe uses the term *periphery* for these attentional outskirts). We show the English translation to the right of the corresponding drawing. The Swedish original is found in appendix II. The numbering indicates intonation unit number in a larger transcript, from which both examples (312–369 and 1–9) are taken. The symbols used in the transcriptions are explained in appendix I.

4. FIRST ANALYSIS: FOCUS MOVEMENTS OVER AN ABSTRACT DRAWING

In the first extract, 1.13 minutes long, the speaker spontaneously produced an abstract drawing as a part of his explanation to the low road quality in Canada. As he moved through his explanation, A used a combination of physical pointing (with the pencil) and linguistic means, to help the listeners identify the current focus within the picture. Our first analysis will therefore mainly concern the correlation between visual focus movements over the drawing and focus movements in the discourse.

After having talked a while about his experiences of poor Canadian road quality, speaker A starts drawing the one-dimensional spectrum in Figure 1. From preceding context, we can guess that the

spectrum content is road quality or road builder quality or something similar.

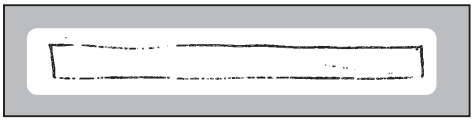


Figure 1. Drawing focus in 312.

312(A) here is the whòle **spéctrum**.

In intonation units 313–337, the empty spectrum is filled in from discourse by the superimposition of at least three polary contrast dimensions onto it. The description of the contrast dimensions is given by focusing on their end points. In 313, focus is placed on the right hand side of the drawing. Intonation units 314 and 315 then superimpose the *much money* end point and the *high quality* end point of two contrast dimensions onto the focused pole in the drawing.

Speaker B signals understanding in 316, and A then elaborates the relations between the two dimension end points, focusing back and forth between them. In 320, B expects a continuation, signalling understanding even before A has ended. Since contrast dimensions have two poles, a natural expectation is that the unfilled poles of the contrast dimensions (see Figure 2b) are to be elaborated next.

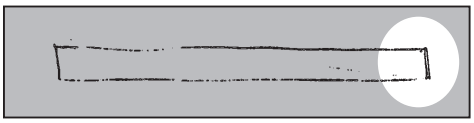


Figure 2a. Drawing focus in 313–320.

313(A) …$_{1.40}$ here 's
314(A) …$_{0.79}$ much money
315(A) and vèry gòod quálity.
316(B) ..$_{0.35}$ mhm
317(A) …$_{0.55}$ they= do a góod job,
318(A) but they know it costs a little more
319(A) to dó a [gòod jòb].
320(B) [mhm]

Initially this also seems to happen. In 321, speaker A moves to the opposite pole of the drawing. However, he then unexpectedly introduces the negative pole of an ethnic contrast dimension and superimposes it onto the spectrum (322–324).

To complete the spectral contrast, the two unfilled dimension poles from Fig 2b must still be filled in, and A does this by moving first to the money dimension in 325 and then to the quality dimension in 326. All these movements are internal. In the drawing, focus stays on the left pole.

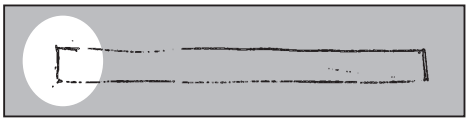
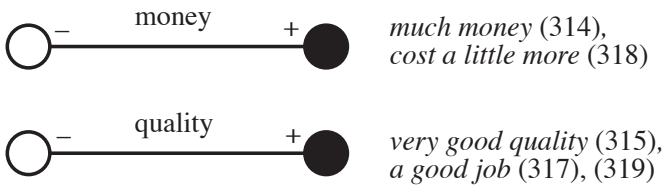


Figure 3a. Drawing focus in 321–327.

321(A) …$_{0.62}$ then we have down here
322(A) we have ..$_{0.27}$ the fellows who
323(A) …$_{1.50}$ come from ítaly
324(A) an all those countries
325(A) they spend quickly the mòney
326(A) an they ..$_{0.36}$ **dónt** càre.
327(B) ..$_{0.42}$ mhm

Figure 3b shows the filling state of the contrast dimensions at intonation unit 327. There now remains one dimension pole unfilled, and A immediately undertakes the task. Since it means a move in the drawing, A uses a clear attention mover *we have … up here*, contrasting against the *down here* mover of 321. This is an implicit fourth contrast dimension, which is verbalised again by *at the bottom* in 336.



much money (314),
cost a little more (318)

very good quality (315),
a good job (317), (319)

Figure 2b. Contrast dimensions, filling of right pole.

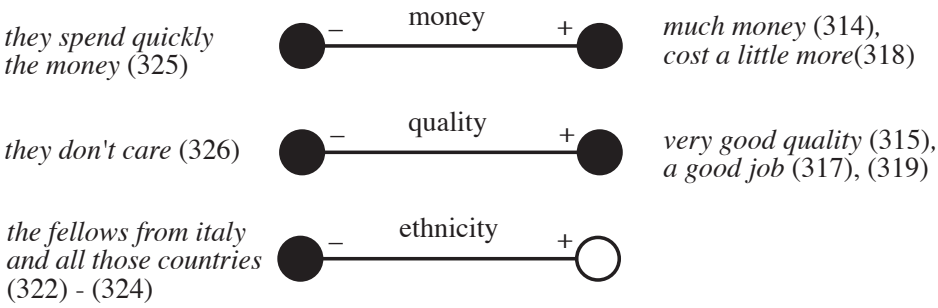


Figure 3b. Contrast dimensions, filling of left pole.

Continuing the summary announced by the marker *so* in 328, A uses the mover *then* to go back down to the negative pole of the ethnic dimension, and dwell a little on it in 330–336, thereby entertaining his partners. The three dimensions (money, quality, ethnicity) that have been superimposed onto the drawn spectrum, are by now well understood by all partners. The introductory description of the spectrum is complete. When looking at the drawing on the paper, they now see these contrasts and not the initially empty bar.

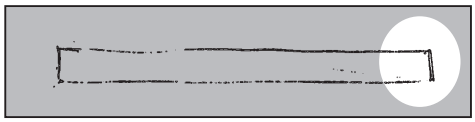


Figure 4. Drawing focus in 329.

328(A) so we have mòre or less
329(A) scandinavians and scots up here

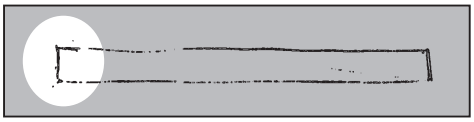


Figure 5. Drawing focus in 330–337.

330(A) then we have the italians and the portu[guese]
331(B) [hn]
332(A) and other..$_{0.23}$ [tr/]
333(B) [haha]ha
334(B) ha[haha]
335(A) [sílly] <LAUGHING>
336(A) …$_{0.64}$ o other trásh at the bòttom <LAUGHING>
337(B) ..$_{0.30}$ hnhnhn

The partners are thus prepared for the causal part of the explanation, in which the spectrum is to be the main participant. The change from spectrum description to causal description is marked by 338, a regulative intonation unit, including the mover *an now,* the deictic *like this* and a very clear voice-quality change from laughter to seriousness.

In 339, the deictic attention mover *all these* makes us focus on the entire spectrum. A continues by describing the process of the road builders giving offers to districts etc, and even quotes the offers in 342/343.

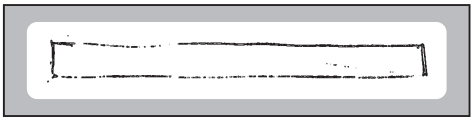


Figure 6. Drawing focus in 339.

338(A) …$_{0.52}$ an now its like this <SERIOUSLY>
339(A) ..$_{0.42}$ àll these give **óffers**
340(A) to districts,
341(A) an an counties an such.
342(A) **nów**
343(A) now we can build róads an all this stuff.
344(A) ..$_{0.17}$ hnhn
345(A) ..$_{0.35}$ an the district looks at
346(A) ..$_{0.14}$ at all the **óffers,**
347(A) an then they look at their **fúnds,**

In 345–350, A demonstrates the decision process of the authorities. There are several attention movers here. The *an* in 345 simply signals progression in time and discourse. The *an then* in 347 and 348 also signals movement of the (visual) attention of the authorities.

During 340–347, there are no focus changes in the drawing. The focused entities, perhaps especially the district, are judged to be efficiently communicated by language, and would be messy if drawn. Our focus of attention reappears in the drawing when the deictic attention mover *like this* in 348 coincides with A drawing the line across the spectrum.

In 350, *down here* takes us to the cheap and low quality side of the spectrum (the ethnic dimension is now forgotten). The point being made is that authorities sacrifice quality for a low price. In the

following, A will describe how this effects the spectrum of road builders in the long run.

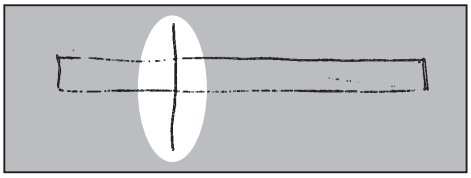


Figure 7. Drawing focus in 348–349.

348(A) an then they do like **thís.**
349(B) ..0.12 aha

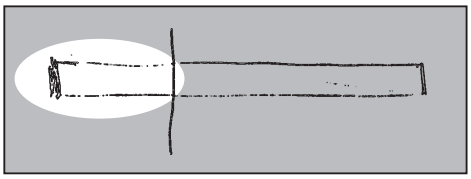


Figure 8. Drawing focus in 350 and 353.

350(A) ..0.18 then they trade down **hére.**
351(B) …1.01 mhm.
352(A) …0.62 an of course,
353(A) when ..0.41 thése have been given work
354(A) enough tímes,

In 355, *then* signals causal and temporal progression, and *these* moves us to the good side of the spectrum. In 357, *here* moves focus back to the cutting line.

The *an* in 356 and 357 and the *an then* in 358 continue to signal progression. The stressed *this* takes us to the bad side of the spectrum.

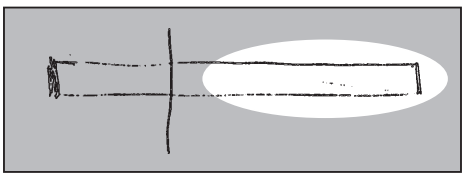


Figure 9. Drawing focus in 355.

355(A) then …0.54 these wind up bankrúpt.
356(A) …0.93 an disappeár.

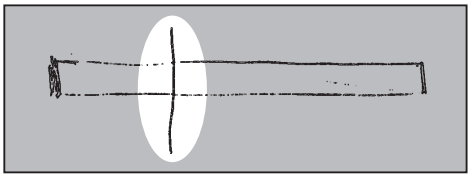


Figure 10. Drawing focus in 357.

357(A) …0.50 an it is cut óff here.

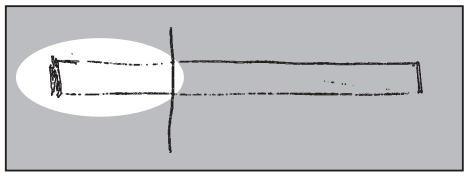


Figure 11. Drawing focus in 358.

358(A) …1.09 an then **thís** thing spreads out.

In 358–361, A gives the idea of the bad part of the spectrum growing over time, taking up the space from the bankrupt good side. The focused magnification of the bad section of the spectrum, is drawn as two lines leading downward (ahead in time) in the drawing. It is interesting to notice that originally, A's three dimensions were polary contrast dimensions, but in order for this magnification process to work, the dimensions have to be continuous.

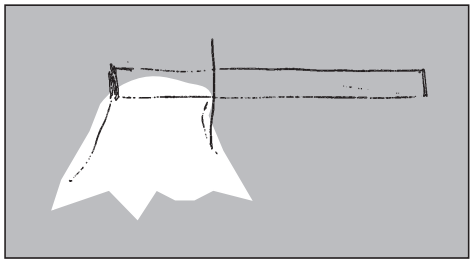


Figure 12. Drawing focus in 358–362.

359(A) …0.97 with the sàme quálity,
360(A) only
361(A) its a **broáder** spèctru[m]
362(B) [m]hm

In 363, we have the situation a few years ahead in time, after the development above has taken place. The new spectrum in Figure 13 is immediately filled by the dimensional content from the bad part of the old spectrum.

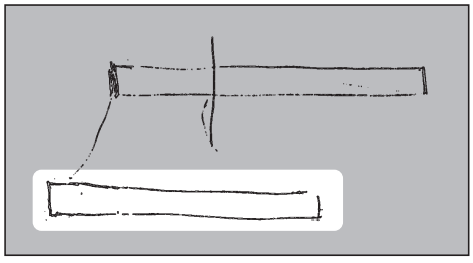


Figure 13. Drawing focus in 363–365.

363(A) …1.34 then we have a **néw** group here
364(A) who
365(A) the **néxt** generation of road builders.

In 366, when the new cutting line is drawn and probably focused upon, A quickly mentions both sides of the new spectrum (referring to the money dimension). This does probably not mean that the focus of attention moves to both poles of the spectrum. Rather, the cutting line metonymically refers to the two poles, which give meaning to the cutting line when mentioned, because it links the new cut back to what the authorities did to the original spectrum.

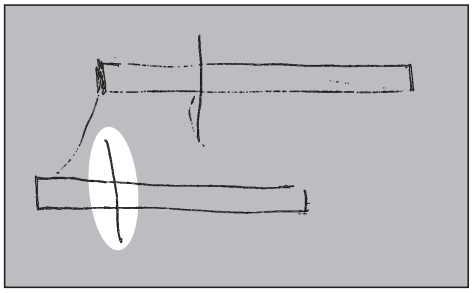


Figure 14. Probable drawing focus in 366.

366(A) …$_{1.38}$ most **expénsive** an **cheápest.**

The long awaited final in 367/368 takes us (by *an* and *these*) to the bad side of the new spectrum, which is marked very clearly in the drawing. We are suddenly placed in the present, and the previous discourse becomes a historical explanation to current Canadian road quality.

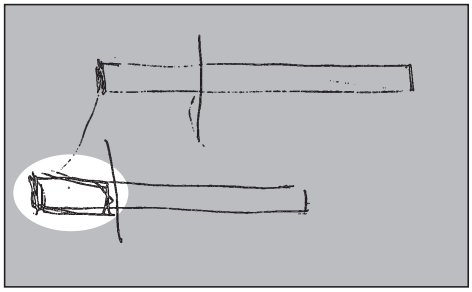


Figure 15. Drawing focus in 367–369.

367(A) …$_{1.76}$ an its **thése**
368(A) who now build the roads in **cánada.**
369(B) …$_{2.61}$ mhm

5. FOCUS MOVEMENTS

The material used here was recorded on audio tape for another purpose (a study of spontaneous descriptions of national identity, Holmqvist and Hol ánová (1996)). We reconstructed the drawing process in Figures 1 to 15, by erasing pieces of the final drawing (Figure 15 except for focus marking), according to what was said in the discourse. In doing so, we have paid attention to what is being focused at each point in the discourse, and related that to elements in the drawing, as was indicated above.

For the speaker, the purpose of focus movements is to guide the listener, to let the listener build her own internal image the way the speaker wants her to. What should count as a focus movement depends on what the focus movements are supposed to move over. Chafe (1994: 63–65) points out that every substantive intonation unit verbalises an idea, and gives examples of types of ideas. The quoted offers in 342/343, the magnification in 358–361 and the contrast dimensions are all ideas, to which, from which, along and between which our focus of attention can be made to move.

When A fills up the original spectrum in 313–337, there are a number of focus movements in and between the three dimensions (money, quality and ethnicity) that are not marked, neither in the drawing nor by any markers in the discourse. The focus moves between 314 and 315, simply because A mentions one dimension (quality) after the other (money). In this sense, virtually every new substantive intonation unit means a focus movement in some domain.

Some of these domains are not easy to integrate in a drawing. The causal relations in 317–319 cannot easily be mapped onto two static dimensions. The district in 340 has too many domains to be able to put it on paper: Which aspects of the district should be drawn? Speaker A chooses to draw the contrast dimensions and the development over time, and those two mappings occupy the two dimensions of the paper. Mapping further dimensions onto it would be messy. Most of the image remains internal and consequently most of the focus movements are internal.

6. LINGUISTIC MARKERS FOR TRANSITIONS BETWEEN ATTENTION FOCUSES

As we have seen in the analysis, transitions between the foci in discourse were often accompanied by hesitations, pauses, and linguistic markers. Focus movements that are difficult for the listener to anticipate, and to predict the direction of, are probably the ones that are lexically marked in the discourse. The many *then we have down here* and similar markers mean: Now we are going to move the focus (regulation), and we are moving it to this place (direction/deixis), and we are going to stay in this

neighbourhood for some time (planning/prediction). These markers give the speaker the time needed for careful description of the upcoming centres of interest, such as the poles of the spectrum (313, 321, 328, 330), the process of sending offers (338), and the second spectrum in analogy to the first one (363).

The marked focus movements therefore also influence pronominal reference (which Redeker 1993a, b studied closer). The pronoun referent is primarily found within the focused parts of the image, so establishing these centres of interest makes it simpler for both speaker and listener: There are fewer possible referents, and they can be referred to by short and efficient pronouns. (Grosz and Sidner 1986 implemented a stack of lisp-structures to simulate levels of discourse focus, for the very reason of making nominal and pronominal resolution more efficient).

When there is no need to establish a new centre of interest, markers such as *an* and *an then* were used (347, 348, 355, 357, 358). The function is one of progression, in time or in causality, within the same centre of interest.

When moving the focus to a new centre of interest, *an* (345), *an now* (338) and *then* (321, 363) were used.

The marker *then* can also mark cases when the focus is moved back to a place already described, repeating the description, as in 330 (we will see the same in our second analysis, in Section 7.2, unit 6). A clearer marker of refocusing is *so*, as in 328. *So* prepares the listener for a general summary, a new center of interest, in which some things may already have been described and others not.

The place to which the focus moves is sometimes given by deixis (*here* in 357, *down here* in 350, *these* in 355, 367, *all these* in 339, *this* in 358), but more often simply by mentioning the new focus.

In the literature, these linguistic markers appear under many names, depending on the perspectives and goals of the respective researcher: Gülich (1970) speaks of *Gliederungssignale*, Quasthoff (1979) of *Verknüpfungs- und Gliederungssignale*, Schiffrin (1987) of *discourse markers*, Aijmer (1988) of *discourse particles*, Rudolph (1989) and Weydt (1989) of *Partikeln*, Stenström (1989) of *discourse signals*, Redeker (1990) of *discourse markers* and Redeker (1991) of *discourse operators*. There is only a partial agreement about the classification of these signals. They are often divided into different classes and attributed different functions by different authors.

Discourse markers appear as conjunctions, adverbials, interjections, particles, final tags, or lexical clauses. Examples are the English *well, but, cause, so, y'know, I mean, now, anyway, and so, okay, by the way*.

The transitions are not only marked in monologues, but also in dialogical situations as the following example from Schiffrin (1987:199) shows.

(1) Sally You said your teachers were old fashioned. Did they ever hit kids, or
(2) Irene Yeah. I had one teacher, her name was Frank,
(3) Irene we used t' call her Frankenstein.
(4) Irene **So**, yeh, she would hit kids with a ruler.

Sally puts a question to Irene, who starts her answer with a description of a person. This beginning could have developed into an entire story, but Irene interrupts herself, using *so* as a marker, and goes back to the question.

What functions do such markers fulfil? To us the most interesting definitions of discourse markers are: "A discourse marker is a linguistic expression that is used to signal the relation of the utterance to the immediate context." (Redeker 1990: 372). "A discourse operator is a word or phrase … that is uttered with the primary function of bringing to the listeners' attention a particular kind of linkage of the upcoming utterance with the immediate discourse context." (Redeker 1991:1168). Schiffrin characterises discourse markers as "sequentially dependent elements which bracket units of talk" (Schiffrin 1987:31).

We want to propose a broader definition than the ones above. These markers should include not only the movements internal to discourse, as in the definitions above. We also want to include the external focus movements that take place in the speech situation. If, for instance, we make a drawing or point at a painting while speaking, deictic expressions such as *here* (excluded by Redeker 1991) are necessary to guide our focus movements over the drawing or painting. The deictic *here* signals that a new focus of attention is to be pointed out in the participant's immediate perceptual space. Stressed deictic referents like *here* in 350, 358 and 367 often correspond to stress in the drawing, see Figures 8 and 15 above.

The similarity between discourse markers and deictic expressions is closer than it may seem: Neither can by itself specify the location of the next focus.

They both need either the succeeding intonation unit or the physical pointing with a finger to locate the new focus. There are also combinations of discourse markers with deictic markers which have a regulative function. The phrases *an now its like this* (338), *an then they do like this* (348) serve as a linkage to the following context and prepare the listener for a more complex explication that will follow. But primarily, both discourse markers and deictic expressions mark an immanent move to a new focus of attention, and therefore we have chosen to call them *attention movers*.

7. SECOND ANALYSIS: FOCUS MOVEMENTS AND THE RECONSTRUCTION OF THE LISTENERS INTERNAL IMAGE

We will now discuss in detail how a valence analysis based on the works of Langacker (1987) and Holmqvist (1993), can help us understand how the listener's internal image is built up during the progression of discourse, even when there are no corresponding external images. The valence analysis is presented as a spatial layout, which has the same topology as the image, but in which we have not filled in any other image content from individual words, than the relative locations of entities and relations.

This second analysis makes use of an excerpt from the same discussion as above. The passage is only 11.8 seconds long, and there is no corresponding drawing. Still, there are obvious focus movements, again across a polar contrast.

1(A)	…0.8 the amèricans have a small fall.
2(A)	..0.15 which is ..0.19 sèparate fro=m the real fall,
3(B)	..0.35 hm
4(A)	…0.33 the **bíg** fall is on the canadian side.
5(B)	..0.30 mhm.
6(A)	…0.71 then the americans have a **smáll** fall
7(A)	and they make a great hullabaloo about their little fall.
8(A)	**å´h** ja mej [ja **ja**!] <A IS VENTING OR QUOTING>
9(B)	[hn hn]

To the listener, this excerpt appears to be a delimited passage, combined from several intonation units, and it may therefore constitute a super focus. We will reconstruct the listener's process of understanding language, using a valence analysis, in which we assume that the semantic pole of words is schematic and can be incorporated into the discourse image.

We take valence relations to be all semantic relations between the different words that appear in discourse. Thus the relation between *small* and *fall* is a valence relation, just as much as the relation between *the americans* and *have*.

In intonation unit (1), *the americans have a small fall*, *the* obviously connects to *americans*. We know this because we know the word order in English (originally Swedish). There is a corresponding connection in the semantic pole, between the schemata of *the* and of *americans*. In the terminology of cognitive linguistics, we say that *the trajector* (TR) *part of the* [THE] *schema is placed in a valence relation with the schema* [AMERICANS].

The TR is the most salient part in relational schemata (adjectives, prepositions, cases, verbs etc.). In verbs the TR normally appears as the unelaborated agent, such as [HAVE].TR. Other parts of relational schemata are the landmarks (LMs). LMs appear as unelaborated objects, patients, reference points etc., such as [HAVE].LM and [SMALL].REFP. The most common type of valence relations are found between schema parts such as TRs and LMs on the one hand, and an entire schema on the other hand. Such is the case with both the valence relations [AMERICANS]–[HAVE].TR and [SMALL].REFP–[NIAGARA FALL(S)]

7.1. Intonation unit (1)

In the intonation units preceding excerpt (1)–(9), the speaker had described the Canadian *Niagara fall* (in Swedish: *niagarafallet*, in singular) in general. The main topic had been how fast the erosion of the cliff moves the fall backwards. Just before the passage (1)–(9) appears, we therefore had a fairly detailed image in mind of A's Niagara fall, combined with whatever pictures or films of it that we could remember.

We then receive intonation units (1)–(9). Each of these units evoke a more or less drawable image. We place each image, one by one, upon our semiactive image of the Niagara fall(s). The superimposition of a new unit image onto the previous discourse image results in the next discourse image.

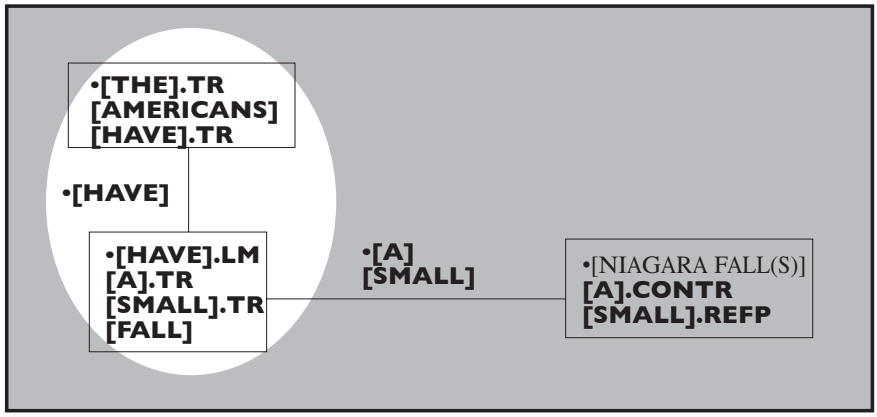


Figure 16. Entities and relations after intonation unit (1): *the americans have a small fall.*

The first valence relation in intonation unit (1) is between [THE].TR and [AMERICANS]. The essence of this valence relation is that the determined thing and the Americans are the same. Protoformalistically, we can say that in this valence relation, [THE].TR and the [AMERICANS] schemata have been made identical. In Figure 16, we represent this identity by placing [AMERICANS] and [THE].TR in the same entity box.

In Figures 16–20, boxes are entities and lines are relations. A name with a dot shows that it is the first occurrence of this entity or relation. The white spotlight marks the focused entities and relations. Bold schema names in the sans font mark schemata evoked by the last intonation unit.

The next schema in intonation unit (1), [HAVE], is an ownership or spatial proximity relation between the [HAVE].TR (owner) and the [HAVE].LM (owned object). On the basis of word order, it is clear that [HAVE].TR is in a valence relation with [THE].TR and [AMERICANS]. Conversely, [HAVE].LM is in a valence relation with [A].TR, [SMALL].TR and [FALL].

With this, we have the left-hand side of Figure 16: The Americans, the fall, and the [HAVE] relation that connects the two entities. We have taken the ownership relation to mean that the [AMERICANS] entity and the [FALL] entity should be close to one another in the image.

There is however more to intonation unit (1). The [A] and [SMALL] schemata not only express the indefiniteness and smallness of their respective TRs. They also contrast their TRs, the American fall, to a definite and well-known [A].CONTR, and to a not so small reference point [SMALL].REFP, both of which are identical to the Niagara fall(s). In Figure 16, the contrasted entities have been placed with a relatively longer distance between them.

It should be obvious that it is the left-hand side of Figure 16 that is focused. After hearing (1), we do not direct our attention to the Niagara fall(s) anymore, but to the small American fall. The reasons for this are:

First, the [AMERICANS], [FALL] and [HAVE] schemata all have a rich lexical content. This makes them the most important schemata of (1), and therefore draws focus to the parts of the image they are being built into. Second, the TRs of [HAVE], [A] and [SMALL] are focused because they are the most salient parts of their respective schema. Third, the CONTR and REFP of [A] and [SMALL] are not focused, because they are not salient parts of their schemata.

How do we know that [A] and [SMALL] contrast the American fall to the Niagara fall(s)? Simple word order tells us that their TRs are in valence relations with the American fall. Therefore the [A] and [SMALL] schemata are about falls. Therefore the CONTR and REFP parts must also be in valence relations with falls. In other words, CONTR and REFP expect there to be another fall more or less focused in the discourse. What other falls are there in the discourse? The Niagara fall(s).

The semantic mechanism establishing this connection is what Holmqvist 1993 calls *semantic expectations*. The hypothesis is that whenever two entities (such as the CONTR and the Niagara falls) have coinciding semantic properties, they are set in a valence relation to one another (unless there are explicit identity differentiators, as in *the other fall*).

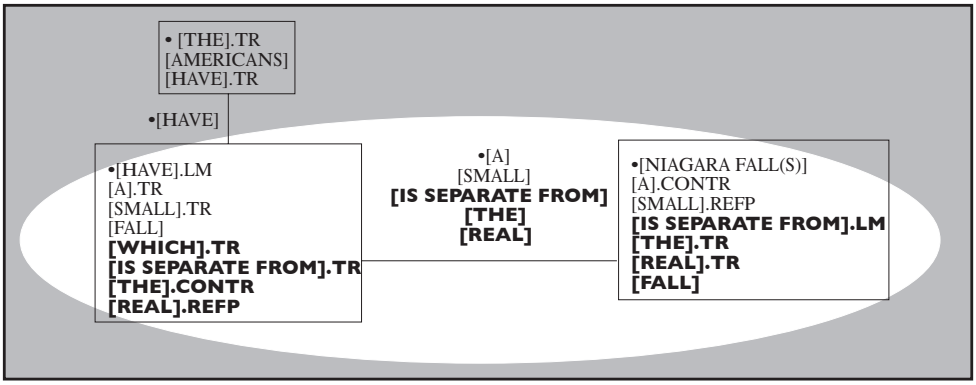


Figure 17. Entities and relations during intonation unit (2): *which is sèparate from the real fall*

7.2 Intonation units (2)–(9)

Intonation unit (2), *which is sèparate from the real fall*, starts off from the small American fall ([WHICH].TR), which was in focus after intonation unit (1). Then three relations, [IS SEPARATE FROM], [THE] and [REAL], contrast this small fall to the real fall. The *fall* in (2) is identified with the Niagara fall(s), because they are both *real* falls and separate from the American fall (again semantic expectations). The three relations therefore describe the contrast between the two falls, together with the contrasting relations from intonation unit (1).

Focus has now changed to the contrast relation and to the two fall entities. The Americans are disappearing into the periphery. In addition to showing us again the same focus mechanisms as in (1), intonation unit (2) is also interesting because of the [IS SEPARATE FROM]. Many schemata that correspond to verbs, especially so-called events, contain change over time in some other domain. This change often follows a path, as in the example *She went up the stairs, walked along the corridor and entered room no 12*. The focus of attention simply follows this path. [IS SEPARATE FROM] also has a path in it, and our focus follows that path. In the beginning of (2), the small American fall is focused. Towards the end, focus is on the real Niagara fall(s).

In intonation unit (4), *the **bíg** fall is on the canadian side*, focus is completely on the Niagara fall(s). The new [BIG] and [THE CANADIAN SIDE] schemata provide not only spatial information, but also more contrast.

The stress on ***bíg*** is of course another focus marker. Not only does the stress say that the difference in size between the Niagara fall(s) ([BIG].TR) and the American fall ([BIG].REFP) is considerable. The salient part [BIG].TR is also given extra focus.

Having presented both the [SEPARATE] and the [CANADIAN] schemata, the speaker has conveyed most of his image to the listener. Figure 18 is not only a valence analysis of intonation units (1)–(4). It is the *framework for a drawing* of the mental image that (1)–(4) evokes: It includes the Niagara fall(s), on the Canadian side. This is the big, real fall. It is separate from the small fall, which the Americans have. The Canadian-American border can be drawn straight through Figure 18.

In order to draw this image, we must not only have images for the lexical schemata in Figure 18, but also a mechanism for conjoining or superimposing the separate images onto one another. Finally, we have the focus on the Canadian side, which tells us that now the focus of attention should be directed to that part of the image.

However, during intonation unit (4), the speaker notices that he has been lead off from the discourse path to his punch line. The focus is now on the Canadian side, but his punch line will be about the Americans. He therefore has to move the focus back to the American side of the border. In order to notify us listeners of his focus move, he starts intonation unit (6) with the attention mover [THEN]: *then the americans have a **smáll** fall.*

The rest of intonation unit (6) is a repetition of intonation unit (1), with added stress on *small*. Repeating intonation unit (1) may seem superfluous, but the speaker must ensure that our focus is moved to the proper parts of the image: The Americans and their little fall. He moves the focus by mentioning *the americans* and by stressing *small*, which contrasts against the *big* in intonation unit (4). Since this is a big move to make in only one intonation unit, the attention mover [THEN] is inserted to notify us.

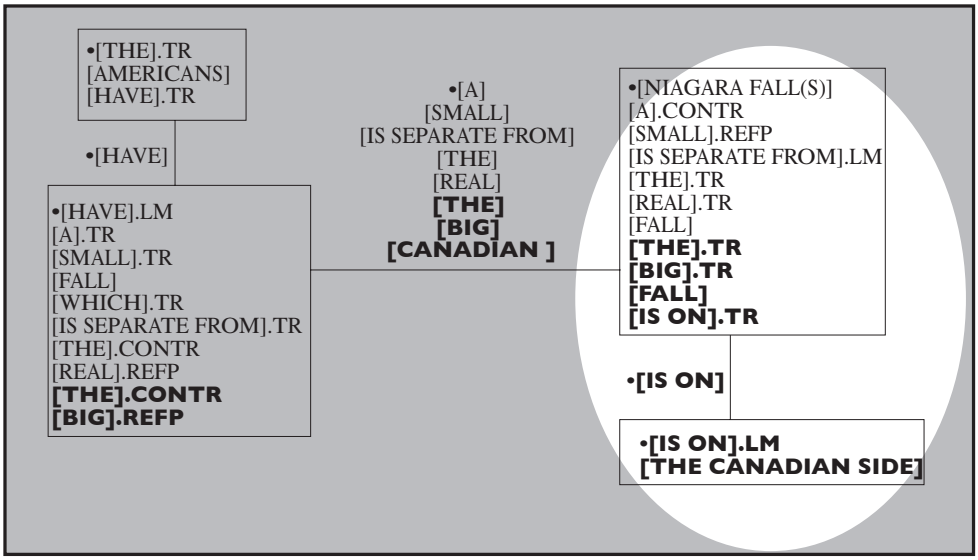


Figure 18. Entities and relations after intonation unit (4): *the **bíg** fall is on the canadian side*.

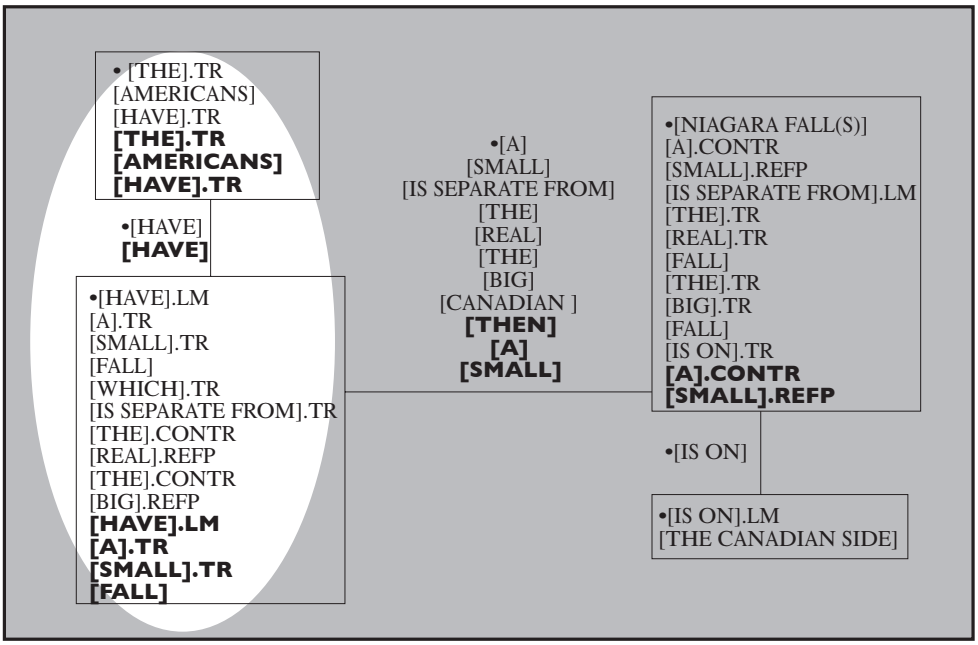


Figure 19: Entities and relations after intonation unit (6): *then the americans have a **smáll** fall*.

Once the focus has been moved over to the American side, the speaker can get to his punch line. Intonation unit (7) is *and they make a great hullabaloo about their little fall*. In his evaluation of the current image, A uses salient schemata like [A], [GREAT] and [HULLABALOO], [THEIR], [LITTLE] and [FALL].

Intonation unit (6) moved the listener's focus of attention, and made her prepared for the punch line. Without this move, the punch line could not refer to the Americans with a simple pronoun, and the hullabaloo idea would have got less focus than intended.

The most important goal of the punch line is to describe the American attitude to their little fall as exaggerated ([HULLABALOO]), childish and silly ([THEIR LITTLE FALL]). It is unclear what [HULLABALOO] exactly involves. We have taken the Swedish [HALLÅ] to mean that the Americans are making exaggerated sounds of praise over their fall, and this interpretation is used in Figure 20.

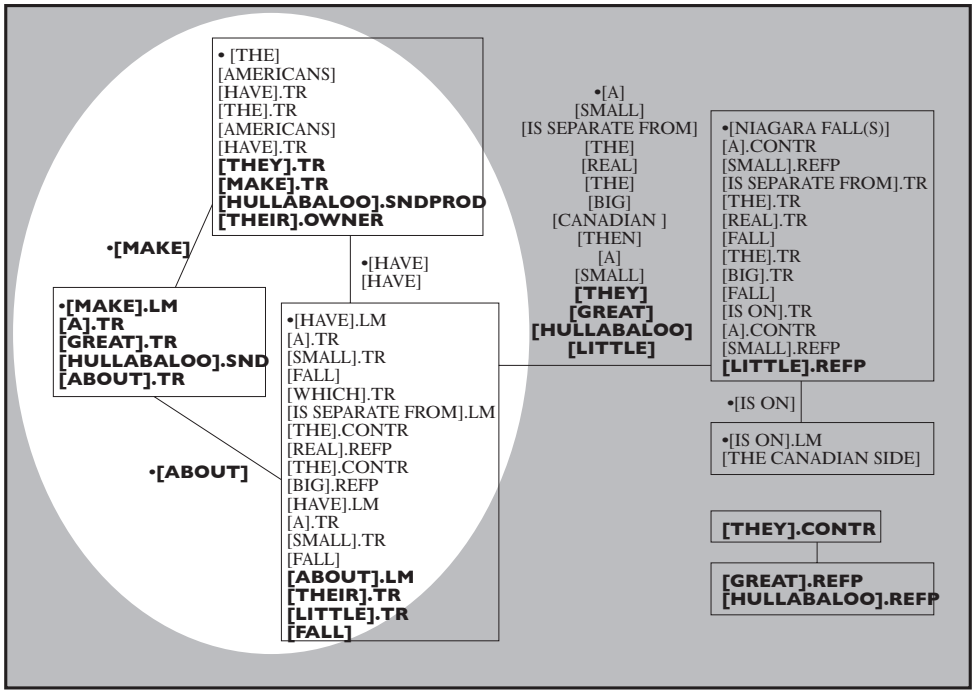


Figure 20. Entities and relations after intonation unit (7): and they make a great hullabaloo about their little fall.

Intonation unit (8), ***å´h*** *ja mej [ja* ***ja****!]* is very difficult to translate. It may be a quote, in which the speaker tries to give a condensed abridgement of how he thinks that the Americans sound when they make their hullabaloo in praise over their fall. But it may also be the speaker just expressing his own feelings when thinking about the Americans.

This second analysis shows how we can reconstruct the internal images and the focus movements over them. The first step of the reconstruction process results in predrawings such as Figures 16–20, in which only the relative positions of entities have been used from the word content. The next step is to evoke the remaining image or image schema structure of singular words and constructions. Each individual lexical image contributes to the discourse image, and is added to it. Such an addition of an image onto another is sketched in Holmqvist (1993) under the general term *superimposition*.

However, before we print out the drawing of the image, we have to choose what *domains* are to be visible. For instance, the spatial, geographical domain is easy to include. The size and other intensity domains in Figure 20 are not all easy to map onto a 2D surface. The sound domain is even more complicated to map onto paper.

The text also includes a *perspective*. We are together with the speaker on the Canadian side in the image. The Americans are the *them*-characters in the *us – them* contrast. One way of including that perspective into the image would be to impose a spatial perspective: Turn the spatial, geographical domain so as to make the Canadian side appear close and the American side further away.

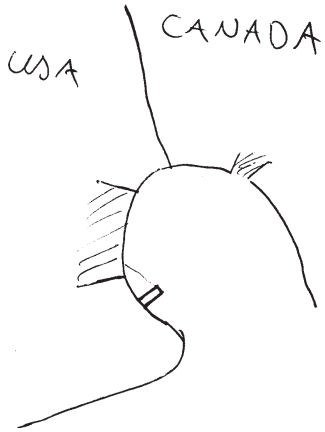


Figure 21. A drawing produced as part of a listener's explanation of speaker As allegedly twisted truths. Compare to Figures 18–20.

Figure 21 was produced by a person who at a seminar had listened to the excerpt (1)–(9) and heard the above analysis. He protested, not against the analysis, but against what he meant were the twisted truths in (1)–(9). He argued that the [HAVE] relation probably meant "have visual access to", and that the Americans have the big fall on their side, but bad visual access to it (only from a

bridge to the right of the fall). The Canadians instead have excellent visual access to the American fall, but only a small fall on their own side.

8. DISCUSSION

It is sometimes argued that we do not visualise, at least not generally, when we understand language. As long as the answer to this question depends on introspective observation, the matter cannot be objectively settled. There is however ample indirect proof in favour of visualisation of language: Speakers who spontaneously draw and listeners who can draw what they think a drawing speaker is drawing, even in nonspatial domains. Above all, the traces of the restless wandering of the speaker's attention over a reconstructable internal image appear in all of language, even if the speaker denies conscious access to those images.

It is interesting in itself that these visual processes can be revealed in linguistic communication. For the computational linguist it might also be worthwhile to consider that the use of an image oriented semantics for natural language processing promises to simplify traditional NLP problems considerably, at the expense of refocusing part of the research towards visual computation.

REFERENCES

Aijmer, Karin (1988): "'Now may we have a word on this.' The Use of 'now' as a discourse particle." In: *Corpus Linguistics, Hard and Soft*. Proceedings of the Eighth International Conference on English Language Research on Computerized Corpora. M. Kytö/ O. Ihalainen/ M. Rissanen (eds.), pp. 15–33.

Chafe, Wallace L. (1980): "The Deployment of Consciousness in the Production of a Narrative." In: W. L. Chafe (ed.) *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. Norwood, NJ: Ablex.

Chafe, Wallace L. (1994): *Discourse, Consciousness, and Time*. Chicago: Chicago University Press.

Grosz, Barbara and Sidner, Candace (1986): "Attention, Intention and the Structure of Discourse", in *Computational Linguistics*, volume 12, number 3 1996, pp. 175–204.

Gülich, Elisabeth (1970): *Makrosyntax der Gliederungsignale im gesprochenen Französisch*. Wilhelm Fink Verlag München 1970.

Holmqvist, Kenneth (1993): *Implementing Cognitive Semantics*. Lund University Cognitive Studies 17, Lund.

Holmqvist, Kenneth, Hol ánová, Jana (1996): *"Jag längtar efter svensk kvalité och tänkande…" Hur vi använder kategoriseringar i samtalet*. Lund University Cognitive Studies 49, Lund.

Quasthoff, Uta (1979): "Verzögerungsphänomene, Verknüpfungs- und Gliederungssignale in Alltagsargumentationen und Alltagserzählungen," In: *Die Partikel der deutschen Sprache*. H. Weydt (Ed) Walter de Gruyter: Berlin, New York 1979, pp 39–57.

Redeker, Gisela (1990): "Ideational and pragmatic markers of discourse structure." *Journal of Pragmatics* 14 (1990), pp. 367–381.

Redeker, Gisela (1991): "Linguistic markers of discourse structure." Review article. Linguistics 29 (1991), pp. 1139–1172.

Redeker, Gisela (1993a): "Discourse Markers as Attentional Cues to Discourse Structure." Poster presented at NATO Advanced Research Workshop "Burning issues in Discourse", Maratea, Italy, April 1993.

Redeker, Gisela (1993b): "The Attentional Structure of Spoken Discourse." Presentation at the 3rd International Cognitive Linguistics Conference in Leuven, Juli 1993.

Rudolph, Elisabeth (1989): "Partikeln in der Textorganisation." In: H. Weydt (1989): pp. 489–498.

Schiffrin, Deborah (1987): *Discourse markers*. Cambridge: Cambridge University Press.

Stenström, Anna-Brita (1989): "Discourse Signals: Towards a Model of Analysis." In: H. Weydt (1989): pp. 561–574.

Tomlin, Russell (in press): "Mapping Conceptual Representations into Linguistic Representations: The Role of Attention in Grammar." In: J. Nuyts and E. Pederson, *With Language in Mind.* Cambridge: CUP.

Weydt, Harald (ed.) (1989): *Sprechen mit Partikeln*. Walter de Gruyter: Berlin, New York .

APPENDIX I

Symbols used in our transcription of speech, adopted from Chafe (1994).

skílt	primary accent (a pitch deviation accompanied by loudness or lengthening)
frå`n	secondary accent (a pitch deviation without loudness or lengthening), slightly displaced for typographical reasons
··0.31	a brief pause, shorter than 0.5 seconds
···1.25	a pause longer than 0.5 seconds
=	lengthening of the preceding vowel or consonant
,	a terminal contour which is not sentence-final (speaker proceeds)
.	a sentence-final falling pitch
!	exclamational terminal contour
big	loudness, stress

<SHOWS>	non-verbal action; the only use of capital letters
[	overlapping speech

Capital letters are never used for speech. Border pauses between intonation units are always written in the beginning of the second intonation unit.

APPENDIX II

The Swedish original for analysis 1.

312(A) här har vi hèla **spéktrat**.
313(A) …1.40 här har vi
314(A) …0.79 många pengar
315(A) och my`cket gòd kvalité.
316(B) ..0.35 mhm
317(A) …0.55 do=m gör ett gótt jobb,
318(A) men dom vet att det kostar lite mer
319(A) å gö´ra ett [brà jòbb].
320(B) [mhm]
321(A) …0.62 så då har vi här nere
322(A) har vi ..0.27 nissarna som
323(A) …1.50 kommer från itàlien
324(A) å alla dom där länderna
325(A) dom gör av med snabbt pèngarna
326(A) å dom ..0.36 **bry´r** sig ìnte.
327(B) ..0.42 mhm
328(A) så vi har mèr eller mindre
329(A) skandinaver och skottar här uppe
330(A) så har vi italienare och portugis[er]
331(B) [hn]
332(A) och annat ..0.23 [sl/]
333(B) [haha]ha
334(B) ha[haha]
335(A) [tráms] <SKRATTANDE>
336(A) …0.64 a annat slö´dder i bòtten <SKRATTANDE>
337(B) ..0.30 hnhnhn
338(A) …0.52 å nu e de så att <ALLVARLIGARE>
339(A) ..0.42 àlla dom här kommer in med **búd**
340(A) till kommuner,
341(A) å å provinser å sådana.
342(A) **nú**
343(A) nu kan vi bygga vä´g och allt sånt här.
344(A) ..0.17 hnhn
345(A) ..0.35 å kommunen tittar på
346(A) ..0.14 på alla **búden,**
347(A) å så tittar dom i sin **kássa,**
348(A) å så gör dom så **hä´r**.
349(B) ..0.12 jaha
350(A) ..0.18 så handlar dom **hä´r** nere.
351(B) …1.01 mhm.
352(A) …0.62 å naturligtvis,
353(A) när ..0.41 dom hä´r har fått arbete
354(A) tìllräckligt många gå´nger,
355(A) så …0.54 går dom här i konkúrs.
356(A) …0.93 och försvínner.
357(A) …0.50 å så kapas de áv här.
358(A) …1.09 å så sprids den **hä´r** ut.
359(A) …0.97 med sàmma kvalité,
360(A) de e bara de att
361(A) de e ett **vídare** spèktru[m],
362(B) [m]hm
363(A) …1.34 så har vi då en **ny´** grupp här
364(A) som
365(A) **nä´sta** generation vägbyggare.
366(A) …1.38 **dy´rast** å **bílligast**.
367(A) …1.76 å de e dom **hä´r**
368(A) som nu bygger vägarna i **kánada**.
369(B) …2.61 mhm

The Swedish original for analysis 2.

1(A) …0.8 amerikànerna har ett litet fall.
2(A) ..0.15 som e ..0.19 skìlt frå=n det riktiga fallet,
3(B) ..0.35 hm
4(A) …0.33 det **stóra** fallet e på den kandensiska sidan.
5(B) ..0.30 mhm.
6(A) …0.71 sen har amerikanerna ett **lítet** fall
7(A) å dom har ett himla hallå om sitt lilla fall.
8(A) **å´h** ja mej [ja **ja**!]
9(B) [hn hn]