VISUAL FIXATIONS, ATTENTIONAL DETECTION, AND SYNTACTIC PERSPECTIVE

AN EXPERIMENTAL INVESTIGATION OF THE THEORETICAL FOUNDATIONS OF RUSSEL S. TOMLIN'S FISH FILM DESIGN

Philip Diderichsen

Dept. of Cognitive Science, Lund University

Kungshuset, Lundagård

S-222 22 Lund

Sweden

philip@alcazar.dk

Abstract: In the so-called *fish film*-experiment designed by the linguist Russell S. Tomlin it is tested whether the assignment of grammatical subject in on-line descriptive discourse production can be controlled by controlling the visual attentional detection of participants in a scene. It is an implicit assumption in the fish film framework that the point of optical focus is a good indicator of focal attention, and that attentional control can thus be gained by making test subjects look at a cued participant. This assumption has not previously been put to the test. In the present paper it is tested by comparing verbal fish film data with visual fixation data from children and adults. This reveals that there is actually no tight correspondence between visual fixations and subject assignment, contrary to what would be expected. It is considered whether an apparent error in the fish film itself can be held responsible for this lacking correspondence. While the error compromises the control of visual fixations, it is still possible to analyze the visual and verbal data for correspondences. Furthermore, it is found that a more traditional interpretation of the data in terms of information structure can be substituted for Tomlin's eye-language correspondence hypothesis in explaining the previous success of the fish film. It is concluded that Tomlin's hypothesis is not directly falsified by the pilot experiment, rather neutralized so that it may still, but not necessarily, be true.

1. INTRODUCTION

In 1995 the linguist Russell S. Tomlin devised an experiment that was to become influential in the field of experimental psycholinguistics, the sub-field of linguistics that takes an empirical-experimental approach to the description of linguistic structure in the minds of speakers and hearers (Tomlin 1995). This experiment, known as the *fish film*, was designed to test the hypothesis that the grammatical choice of sentence subject in English is a linguistic manifestation of attentional focus and can thus be predicted given that focal attention is controlled.

The fish film methodology has been used on many occasions to show that there is indeed a convincing correspondence between the entity in focal attention and grammatical subject, but on closer scrutiny the foundations of the fish film seem to have some minor flaws. In particular one seemingly basic theoretical assumption - that is never really made explicit in Tomlin's writings on the effects of attention in language - is interesting: that attentional focus presupposes visual focus (Tomlin 1984, 1986, 1995, and 1997; Tomlin and Villa 1994).

It is possible to criticize Tomlin's cognitively grounded view on functional discourse for not sufficiently taking into account the possibility that visual perception may at least in part be independent of focal attention. Questions about the visual mechanisms at the base of cognition are simply not addressed. In the present paper, the different mental focusing mechanisms that can be assumed to underly the fish film description task are discussed: visual fixations, attentional detection, and syntactic perspective. Then, different views on how these mechanisms interact will be reviewed. It turns out that these mechanisms are not necessarily connected in a one-to-one fashion, and this forms the theoretical basis for an empirical investigation of the fish film framework.

In particular, it will be investigated whether a one-to-one correspondence between visual fixations and

Lund University Cognitive Studies — LUCS 84 ISSN 1101 — 8453.

attentionally detected information exists or not. This will be accomplished by using an eye-tracker to monitor the eye-movements of test subjects during the fish film experiment. It will be revealed whether the last point of visual fixation before utterance onset is indeed the entity that becomes sentence subject in the ensuing utterance, as would be expected according to Tomlin.

As it turns out, there seems to be an error in the fish film itself, which effectively neutralizes the conclusion reached by Tomlin: that the successful control of attention focus makes it possible to predict the assignment of syntactic subject.

Even if the fish film is thus not suitable to provide valid evidence for the correspondence suggested by Tomlin, it is obviously very suitable to elicit active and passive sentence configurations in a predictable way. This means that Tomlin may still be right, but unfortunately the success of the fish film can also be explained using a more traditional discourse analysis approach.

In an earlier paper (Diderichsen 2001) I have shown how the production of passive constructions in children as old as 7 years of age defies the kind of experimental control of 'visual attention' that can apparently be obtained when the fish film is used with adult subjects. I presented the view that the postulated correspondence between attention and language could still be theoretically defended even though the attempts to control 'visual attention' clearly failed. I further suggested that the development of attentionally determined discourse strategies involves the independent development of selective attention and grammatical means of discourse profiling.

These insights can all be maintained under the assumption that the focus mechanisms of vision are independent of attention and language, at least at the level of individual visual fixations. Thus, in this paper it is argued that the focus mechanisms of vision should be considered a factor in on-line linguistic description of visual material that surely plays an important role, but is not necessarily either determining nor determined by any of these other two factors.

2. THEORETICAL BACKGROUND

One of the most fundamental characteristics of cognition is the ability to focus on some phenomena at the expense of others. This is reflected in different disciplines as theoretical notions of *importance* such as *aboutness*, *topicality*, *relevance*, *salience*, *perspective*, *foregrounding*, the *figure-ground* and *given-new* dichotomies, etc. All these terms have been brought to bear on the selectiveness of human information procesing, and their similarity is highlighted by the fact that they have all occasionally been replaced by a metaphor of *focus* and *periphery* (see, for instance, Chafe 1994). In this section descriptions of the three focusing systems under investigation, namely those of vision, attention, and language, will be reviewed, as well as Tomlin's and other's views on how they interact, providing the theoretical background for the pilot experiment described below.

2.1 Vision

The visual system allowing us to visually perceive a scene is well understood. On the physiological side, the musculature enabling horizontal, vertical, and rolling movement is well-known, and seven distinct eye movement types have been identified, of which the most important one - in this context and otherwise - is the type known as *saccades* (Glenstrup and Engell-Nielsen 1995, Holsánová 2001). This is the kind of quick movement that occurs between visual *fixations*, periods typically ranging from about 100 to about 500 ms where the eyes stand still and allow the *fovea* (a small area at the center of the retina that gives very high visual acuity) to pick up detailed visual information. Thus, it is visual fixations that constitute the focus mechanism of vision.

Figure 1 shows the temporal sequence of saccades and fixations: according to Glenstrup and Engell-Nielsen it takes a minimum of 100 ms from a new stimulus (in this case the large arrow used to attract attention in the fish film) is presented until the actual eye movement starts; it takes around 30 ms to complete the saccade; and the eye stays at its new location for about 200 ms.

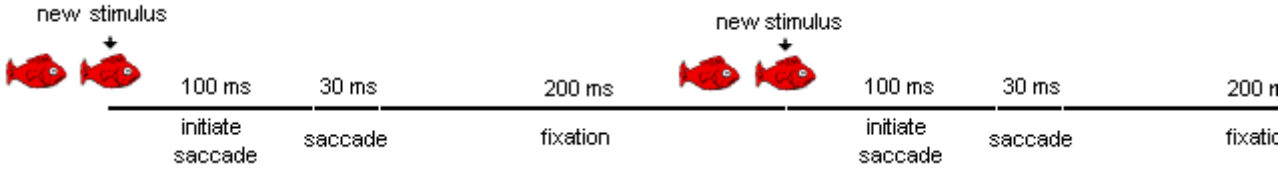

Figure 1. The temporal sequence of eye movements according to Glenstrup and Engell-Nielsen (1995)

Another type of movement relevant to this study is the one known as *pursuit motion*, which is a smooth, slower kind of movement with the function of keeping the eyes focused on moving objects. It may be conceived as a moving fixation; it can only be observed when there is a moving object in the visual field to guide the eyes.

These movement types are the relevant ones for this study as they allow people to orient the eyes toward something, focus, and fixate. The visual fixations can be tracked, and it is thus possible to directly monitor the (purely optic) focus of vision. As will be seen below, tracking gets more difficult when it comes to the foci of attention and language.

2.2 Attention

As long as psychology has existed as a discipline, there has been a notion of attention focus, i.e. the fact that humans can only attend to a limited number of things at a time and that attending to something means more or less ignoring everything else (Gazzaniga, Ivry, and Mangun 1998; Eysenck and Keane 1995).

Most research on attention has been done within the area of *visual attention*, the study of how the information filters of cognition interact with visual perception. While this is presumably because visual orienting can be directly observed, it undoubtedly also has something to do with the fact that the optical systems of vision lend themselves nicely to the metaphorical description of attention. Thus, two of the most popular metaphors for the attentional system are the 'zoom lens' and the 'spotlight' metaphors[1], both of which provide intuitive notions of limitation and focus.

After the finding that it is quite possible to direct spatial attention towards something that is not in optical focus a theoretical distinction between *overt* and *covert* attentional orientation has been introduced, whereby eye movements and head- and body-turning can be distinguished from purely cognitive 'movements' of attention. In 1894, Hermann von Helmholtz described how, during a brief spark of light, he could perceive the letters in an attended location of a screen with a large array of letters without moving his eyes from the center of the screen (Gazzaniga, Ivry, and Mangun 1998: 210). Since then it has been an established notion within cognitive psychology that attentional focus and optical focus are not necessarily the same thing.

Actually, not even the idea of two independent 'spotlights', that of vision and that of attention, interacting in visual attention seems to do the attentional system justice. Eysenck and Keane (1995: 102) cite a study by Juola, Bowhuis, Cooper, and Warner (1991) that showed that attention can be allocated not just as a continuous circular area but also in a pattern of concentric rings. They also cite a study by Neisser and Becklen (1975) that showed that subjects were able to discriminate between two moving scenes that had been superimposed on one another. Thus, the notion of an attentional *focus* in the optic sense of the word should not be taken too seriously. Rather, the attentional spotlight should be considered a metaphor for something like the phenomenon that Chafe (1994) calls the "one new idea constraint": that people can activate only one new idea at a time. In Chafe's own words:

> "The magical number *one* appears to be fundamental to the way the mind handles the flow of information through consciousness and language." (Chafe 1994: 119)

This leaves the question of what 'one idea' is; one thing can be anything. But this might just in itself highlight a relevant point: one can speculate that as soon as something can be seen as one *chunk*, it can be attended to as one unitary idea.

Another important aspect of attention is that its focus can be either exogenously or endogenously cued, i.e. 'controlled' by external stimuli or internal intentionality, respectively. Somewhat misleadingly, some have used the term 'focused attention' for the former and 'selective attention' for the latter (one should think that attention is focused regardless of whether it has been put in that state exo- or endogenously). The fish film, as will be seen, employs both exogenous and endogenous means of keeping the subjects' attention focused.

Furthermore, the state of attentional focus can be divided into substates: *alertness*, *orientation*, and *detection* (Tomlin and Villa 1994). Alertness is a state of general readiness of the attentional system; a person will normally go into an alert state if someone shouts something in the genre of "Look!". Orientation is a state where specific information is expected, a state that might be prompted by an utterance like "Look at that dragonfly!", for instance. Detection is a state where a specific stimulus has been registered for further processing and has one's full attention. Unlike the other substates, this one is a state of such concentration that attention is literally 'locked' to the detected stimulus for a fraction of a second in order to allow the cognitive system to process the information detected. It is *attentional detection* that plays a role in the fish film framework, and the term will be used in the following to refer to the metaphorical-intuitive notion of attentional 'focus'.

2.3 Language

The focus mechanisms of language are notoriously problematic to describe because the terminology in the area has been used somewhat inconsistently by different researchers. Most discussions of linguistic focus fall within the area of *information structure*, a part of the discipline known as *discourse analysis*, which is concerned with how language functions as coherent sequences of sentences, i.e. texts in contexts (Van Dijk 1977; Brown and Yule 1983; Halliday 1985a, 1985b; Chafe 1994; Östman and Virtanen 1995).

[1]. Not to mention conceptual metaphors like Knowing Is Seing (Lakoff and Johnson 1999: 53) or metaphorical expressions like 'the mind's eye'.

The discussions of information structure converge on the point that streams of language can be divided into segments with a dyadic structure. The segments sometimes correspond fairly well, though by no means always, to sentences. Depending on whose kind of analysis is employed the segments are divided into a Given, Theme, or Topic part and a New, Rheme, or Comment part, respectively. These terms have sometimes been used interchangably, resulting in the above-mentioned terminology problems. However, the three term-pairs for this dyadic structure have originally been coined to serve different purposes, and the phenomena they are meant to describe are actually independent of one another, despite the fact that they very often coincide. In the words of Östman and Virtanen:

> "One way of distinguishing between the terms is to view 'theme' and 'rheme' positionally (cf. Halliday 1967, 1985), 'topic' in terms of 'aboutness', i.e. what a proposition is about (cf. van Dijk 1977, 1980), and given and new information cognitively, in terms of the stage of activation assumed by the text producer (cf. Chafe 1976, 1994; Prince 1981)." (Östman and Virtanen 1995: 242)

When the three kinds of structure coincide, the first part (= Theme) of a segment would be a previously established or mentally activated (= Given) entity about which (= Topic) something was asserted. The rest of the segment would then be the interesting one, the 'news' part of the segment containing the *information focus*. This can often be observed in the active-passive distinction, which is the particular linguistic configurational mechanism (among many others) that is relevant in this context. Thus, the analysis of active and passive transitive clauses is likely to show a clustering of Themes, Topics, and Given entities in the subject with Rhemes, Comments, and New material showing up in the rest of the clause, clustered around the information focus.

But what, then, can be considered the linguistic *focus* of a sentence? The Theme or the Rheme, the Topic or the Comment, the Given or the New? An independent definition of 'focus' will not be attempted here, and it will suffice to specify the status of the term in this context. On the one hand there is Halliday's explicit use of the term 'information focus' (1985a: 275), which refers to the linguistic item which has primary stress and marks the culmination of New. This corresponds to Chafe's *intonation unit* and is - in Halliday's terms - "what is made prominent (by the speaker) to the listener; it is 'what you are being invited to attend to'." (Halliday 1985b: 39). On the other hand, there is the Theme, which is described by Halliday as the speaker's point of departure, i.e. "what is prominent for the speaker; it is 'what I am on about'." (Halliday 1985b: 39). "Theme + Rheme is speaker-oriented, while Given + New is listener-oriented." (Halliday 1985a: 278). Furthermore, it is the notion(s) of "theme or topic" that Tomlin wants to replace with 'focal attention' (Tomlin 1995: 517-518).

Could one say, in spite of Chafe's 'focus' terminology, that the linguistic focus of an utterance is the item that the speaker takes as his point of departure? Or that the linguistic focus is what a sentence is about? Perhaps. But it might be even better to abandon the term 'focus' altogether in favour of another term from the same visual domain, namely *perspective*. Literally, a perspective is a point of view that allows one to focus on something, and the term has been used to describe the array of constructions that allow for different linguistic vantage points (Chafe 1994; Clark 1989). This terminology is a nice way of capturing what happens in Tomlin's original scenario, i.e. when one of two possible participants in a transitive event is coded as the subject in an on-line descriptive sentence leading predictably to one or the other grammatical voice. And at the same time it is consistent with the established term 'information focus' that seems to entail some kind of perspective.

2.4 The interaction of vision, attention, and language

Following the considerations above the disussion of the interaction of vision, attention, and language will be in terms of visual fixations, attentional detection and syntactic perspective. As it turns out, it is only vision that has a focus in the optical sense of the word. But granted that the term 'focus' is not taken too literally, it can be seen as a cover term for the above-mentioned, specific mechanisms. Now, how do the focus mechanisms of vision, attention, and language interact? If one thinks in terms of strict correspondences or correlations, the logical possibilities are exhausted by the diagram in figure 2:

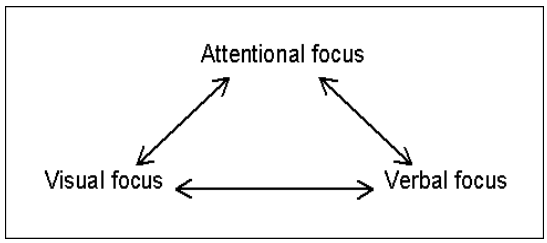

Figure 2. Possible correspondences between visual, attentional, and verbal focus

There have been several suggestions as to the actual occurrence of one or more of these correspondences, some of them more explicit than others. For Tomlin, it is a theoretical pivot that language corresponds to attention in a direct, one-to-one fashion. This can be illustrated with the following central quote:

"If one could independently manipulate the locus of focal attention in speakers about to perform a sentence production task, one should be able to see a dependent assignment of the focally attended referent

to syntactic subject in discourse production. And, if the event were described with a transitive verb, one would expect to see active clauses if and only if the focally attended referent were the agent of the action and one would expect to see passive clauses if and only if the focally attended referent were the patient of the action." (Tomlin 1995: 527-528)

The notion of a close interaction between attention and language is clearly present in this quote. Furthermore, it seems that Tomlin conflates visual and attentional focus when he talks about how "[...] it takes approximately 150 milliseconds to complete the shift of attention and the fovea from one target to another." (Tomlin 1997: 173). This completes the triangle, so to speak: Tomlin seems to assume some kind of correspondence between all three kinds of foci.

An important detail of Tomlin's thinking in this area is the fact that he explicitly differentiates between types of attention focus manifesting themselves in different linguistic phenomena:

> "Within this framework, attentional selection of entire events is associated with the linguistic phenomenon of foregrounding (Hopper 1979; Hopper & Thompson 1980; Tomlin 1985, 1986). The selection of specific parameters is associated with so-called topic or theme (Tomlin 1983, 1992, 1995)." (Tomlin 1997: 173) 2

It is the claim that attentional focus on specific participants ("parameters") manifests itself as thematization in the clause that is under investigation here; thus, the detection and linguistic manifestation of entire events is not so relevant.

Yet, differentiations similar to Tomlin's can be observed elsewhere, particularly in Chafe's 'one new idea constraint' (see above). In Chafe's own words, 'ideas' are "mental representations of perceived, remembered or imagined events, states, and referents." (Chafe 1996: 43). Although Chafe thus shares the view that specific participants ("referents") can show up as linguistically manifested foci of consciousness, he doesn't see the starting point of a sentence - the Theme - as corresponding to one focus of attention. Rather, thematization and subjecthood are seen as expressions of semiactive topic information, i.e. discourse-level topicality or aboutness. But his thoughts are actually not very different from Tomlin's. He observes that sentences are often verbalized as several intonation units and in that case can be considered units of *superfoci* of consciousness, so-called *centers of interest* that contain several foci of consciousness (Chafe 1994: 139ff).

2 Notice the confusion of the terms 'topic' and 'theme'; this is just one of many cases where the terminology is not used consistently. It must be added, however, that Tomlin specifically mentions that he uses both terms to refer to "information the speaker believes is the more central, salient, or important at the moment of speech." (Tomlin 1995: note 1)

Chafe's suggestion that the foci of consciousness constitute the common denominator in the focus mechanisms that can be observed in vision and language have been taken up by Holsánová (2001). In her study of picture viewing and picture descriptions she generates a number of theses of which the following, Hypothesis 1 and Thesis 2, are the relevant ones in this context:

H1. Foci of attention are the comparable units in picture viewing and picture description

T2. There is temporal synchrony between the focus patterns in picture viewing and picture descriptions in individuals. (Holsánová 2001: chapt. IV, section 5)

Both are falsified. The first by the finding that visual fixations by far outnumber the corresponding verbal information foci; thus, language often sums up the information gathered by vision, and a more appropriate unit of comparison turns out to be attentional superfoci - corresponding to sentence-length segments of natural discourse. T2 is falsified in that the verbal description typically lags behind the visual acquisition of the information, even if this lag is not always predictable.

Returning to the diagram in figure 2, the insights presented so far show that a triangle of perfect correspondences between focal properties of vision, attention, and language can hardly be maintained. That eye movements do not necessarily reflect attentional detection of whatever is in visual focus (or vice versa) is proven by the insights about covert attentional orienting as well as by introspective testimonies about the ability to attend to displaced, i.e. remembered or imagined things. (Although the "eye-mind assumption" from reading research states that one can be fairly certain that a tight correspondence exists under some circumstances (Just & Carpenter 1980)).

Holsánová's findings show that Chafe's expectations of a correspondence between visual and verbal foci are not borne out - at least not in the case of her picture description task. Another, but trivial, non-correspondence between vision and language is the case of discourse in the displaced mode (i.e. conversation about things that are not present in the immediate environment), which Chafe himself mentions as the usual type of natural conversation (Chafe 1996: 48). Thus, visual foci do not always entail verbal foci, or vice versa.

Last, there is the possible correspondence between attention and language. Of course, what we attend to is not always linguistically manifested; but does the language we actually produce always correspond perfectly to what we attend to? This is the idea that permeates all the efforts described above; they all in one way or the other assume that language reflects what is in active consciousness at the time of production.

This leaves us with a theoretical toolbox that should come in handy when considering the pilot experiment described below; while all the possible correspondences in figure 2 have been considered in one way or the other by several researchers, the correspondence between language and attention has not yet been challenged. The experiment can not do this either. What it can do is challenge Tomlin's original assumption that attention detection follows from visual focusing and that the assignment of subject can thus be predicted by monitoring 'visual attention'.

3. The Pilot Experiment

The experiment that was piloted in this study is essentially a replication of Tomlin's original fish film experiment with Danish subjects, with the additional feature of direct observationability of the subject's eye movements. The fish film consists of a series of 32 (originally 31) individual computer-animated events showing two fishes with different colours swimming towards each other, one of them swallowing the other as they meet. The events on the screen are to be decribed on-line by the test subjects. Just after the fishes enter the screen a large arrow flashes above one of them. This is an exogenous cue, which is designed to spontaneously attract attention, as well as an endogenous cue in that it is a reminder about the instructions to pay attention to the cued fishes. The arrow cueing occurs once more just before one fish is swallowed by the other in order to make sure that the cued fish is attentionally detected when the subject starts uttering the sentence describing the event. More specifically the second cueing allegedly takes place just 75 ms before the swallowing event is completed (Tomlin 1997: 174) which is well within the period of 150 ms that it takes to foveate a new location in a scene (see figure 3).

(When one looks at figure 2, however, it might be noticed that something is wrong; more specifically, in my depiction of the trial, it takes ca. 75 centiseconds (= 750 milliseconds) from the second cueing to the completion of the swallowing event. My approximate timing of the trials is derived from mean intervals from several time codings of the whole fish film done manually with a stop watch, and thus I believe the error lies with Russ Tomlin who seems to have made the unfortunate mistake of using the wrong order of magnitude. I will come back to this problem below.)

Figure 3. The temporal layout of the events in the fish film

If we ignore for a while the apparent error in the fish film itself, experimental control over visual attention is in theory secured. But as shown above, (visual) attention detection and visual fixations are not necessarily the same thing. In the setup for the pilot experiment an *SMI* eye-tracking device was added to give direct access to the saccades and fixations of the subjects in order to test the effects of the visual arrow-cueing. Furthermore, an additional, sligthly altered version of the fish film was produced by removing the cueing arrow from all 32 trials. This was done in order to allow for the collection of baseline eye movement data from a control group in a non-arrow condition.

The test subjects were divided into two age groups: adults and 5-year-olds. The task of the children was slightly more elaborate than that of the adults to ensure that they mastered certain task-relevant skills. Thus, a child session started with the presentation of eight stills from the fish film showing differently coloured fish facing each other, the cueing arrow hovering above one of them. Here the children were to describe the colours of the fish and identify the cued fish. After this subtask with the purpose of securing the children's ability to refer verbally to the fish, another one followed with the double purpose of dishabituating the children from the fish and rehearsing on-line description. The children were shown two short animated clips with a billard ball bouncing off the cushion, then falling into a hole and a rotating star exploding, respectively. The kinder-garten teacher present throughout the experiment then told the children that they would be watching a short film, that they were to concentrate on the arrow fish in their on-line-descriptions (only in the cued version, of course), and that they were not to guess at the behaviour of the fishes. After this final briefing the fish film started.

A total of 7 children performed the on-line task, and of these 7 most proceeded without anxiety or hesitation, no doubt thanks to the rehearsals prior to the main task, as well as the constant presence of the teacher. Two were excluded from further analysis because of insecure or non-task performance (e.g. production of utterances without verbs or nothing but guesses). 4 adults performed the task without such problems, having been shown only the on-line description rehearsal clips prior to the fish film. The age groups and the two versions of the fish film yielded four different conditions, which were covered

by the 9 subjects analyzed as shown in table 1. Unfortunately, the eye-tracking data from one of the adult subjects in the non-cued condition were lost during the experimental session and thus only one adult subject is analyzed in this condition (which is actually quite regrettable since the subject whose data are missing very unexpectedly uttered only passive constructions).

	Cues	No cues
Children	2	3
Adults	2	2 (1)

Table 1. Distribution of test subjects across condition

One of the adults in the cued condition was selected for detailed analysis because of a performance very close to what could be expected. The performance of this subject was almost exactly as predicted by Tomlin (and confirmed for Danish by Kelstrup (1998)). It is thus justified to treat the data from this subject as a paradigm case in the eye movement analysis.

3.1 Analysis

The utterances of all test subjects were scored as active or passive constructions (or neither) according to the following criteria:

1. There must be an active form of the verb to score it as an active construction

2. There must be a passive form of the verb to score it as a passive construction

3. The verb must describe the transitive swallowing event

4. The sentence must be an online description - that is, at least not a pure guess

The analysis of the eye data was less straight-forward. An eye-tracking analysis program, *iView for Windows*, was employed to trace the path of the subject's eyes. On the basis of traced patterns with a duration of 2 seconds an automatic "order of objects vs. time" analysis was done by the program showing the percentage of fixations relative to total time in predefined regions of the screen (see figure 4).

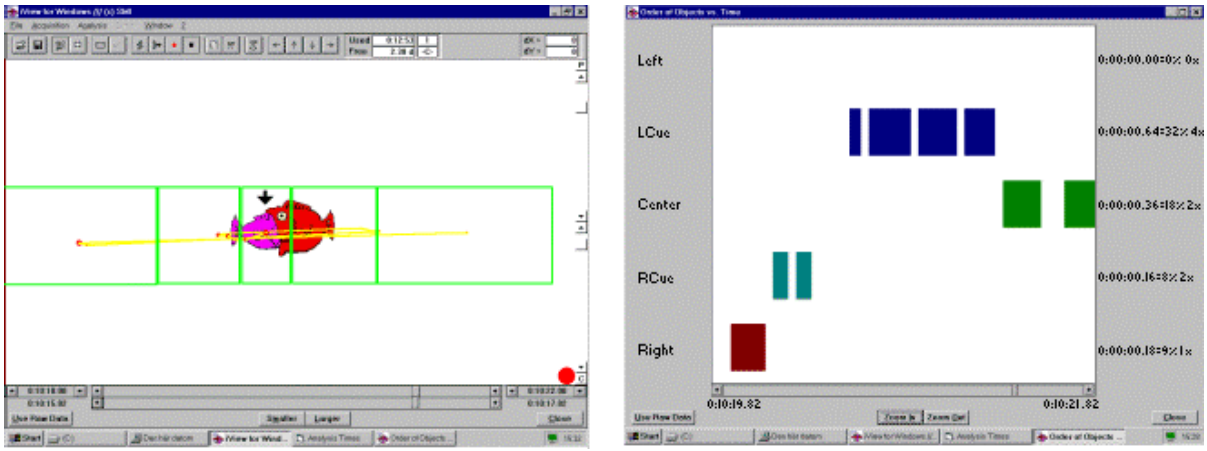

Figure 4. Two different views on the fixations just prior to the swallowing event. The five rectangular areas on the left are called Left, Lcue, Center, Rcue, and Right, respectively.

In this way it could be determined if the last fixation before the swallowing was on the left fish or on the right fish, allowing a comparative analysis of last fixations vs. grammatical voice.

The output from the adult that performed as predicted was scrutinized more closely. Thus, a fixation analysis of the whole session of this subject was produced and qualitatively compared to a transcript of the verbal output from the session.

Unfortunately, there are a number of factors that introduce uncertainty into the analysis. First, it is not possible to know exactly where on the display the subject is looking, due to the less than perfect calibration of the eye-tracker. Each time the subject moves his or her head too much for the camera to stay focused on the eye, and each time the subject blinks his or her eyes, there is a risk that the device falls slightly out of calibration with the result that the represented point of fixation is dislocated a little bit relative to the actual one. This has to be compensated for when one defines the areas to be used in the automatic order of objects versus time analysis, and this can only be done based on an intuitive conception of the true position of the entities on the screen made plausible in large part by the eye-movement traces themselves.

Second, there was a problem concerning the measurement of fixations. The iView software is apparently not designed to handle data collected on the basis of moving pictures and thus a lot of fixations during pursuit motion (see above) are not recorded when a realistic minimal time of, say, 100 ms is set to define a visual fixation. This can be compensated for by setting an unrealistically low minimal time for a

visual fixation, eventually interpreting short fixations in close succession as single fixations in pursuit motion. Based on these considerations, the minimal time defining a fixation was set to 60 ms in my analyses, which is normally too short a time for visual information acquisition to occur. As I did the scoring manually, I was able to interpret each case individually, with the uncertainty that this might lead to.

Third, it is a problem that the eye movement data are not digitally synchronized with the verbal output. It is possible to synchronize the visual and verbal output, but only manually: either the raw visual data in the form of the dot representing the point of fixation and the audio side of the experiment can be manually re-represented in parallel sequence from the video recording of the experiment; or the audio can be time coded (thanks to a beep on the fish film at the moment of swallowing) and manually synchronized with a fixation-vs-time analysis from the iView program. Both methods work best when they take as their point of departure some kind of written representation of the verbal output, since the audio is not available during the necessary frame-by-frame analysis in the first case, or fixation analysis in the second, of the video. It is not advisable to take the visual output as the starting point in the analysis since this is less temporally precise, both in the case of frame-by-frame-analysis - which also does not contain fixation information - and in the case of the iView fixation analyses which are precise unfortu-nately only in principle, not in practice, since it is not possible to specify the exact point and period in time that one would like to look at.

All in all the synchronization problem is a considerable one and introduces both general imprecision and an unneccesary risk of serious human errors in the analysis. With this digression into the technical details of the analyses in mind it will not be necessary to further warn the reader to consider with some caution the data in the next section.

3.2 Results

The results of the pilot experiment are summarized below. Table 2 and 3 show the results that one might expect according to my reading of Tomlin's basic assumptions. There are 16 trials in the fish film where the agent is cued and 16 where the patient is cued, and as the participant roles are counterbalanced in relation to right and left screen entry, there are four different expected points of (last) fixation, namely. on the following location/role-defined participants: left cued agent, right cued agent, left cued patient, and right cued patient.

Hypothetical test subject	Last visual fixation:	Left Fish		Right Fish		Other
Syntactic perspective:		Expected (Cued)	Unexp. (R. cued)	Expected (Cued)	Unexp. (L. cued)	
Agent	expected (agent cued)	8		8		
(active voice)	unexpected (patient cued)					
Patient	expected (patient cued)	8		8		
(passive voice)	unexpected (agent cued)					
Other						

Table 2. Expected data in the cued condition

This is expected on the assumption that the active and passive constructions that are produced in the way predicted by Tomlin actually reflect 'visual attention', i.e. visual fixations. Thus the visual focus is expected to coincide with verbal focus in the form of grammatical voice as outlined in table 2.

Hypothetical test subject	Last visual fixation:	Left fish	Right fish	Other
Syntactic perspective:				
Agent	Left agent	8		
(active voice)	Right agent		8	
Patient	Right patient		8	
(passive voice)	Right patient	8		
Other	Left agent			
	Right agent			

Table 3. Expected data in the non-cued condition

In the non-cued condition the expected pattern is of a similar nature to that in table 2, but in table 3 it is derived from probabilistic considerations instead of visual cues. When there are no manipulative cues there are several factors that one migth be able to use to predict the pattern of visual fixations; one possibility is the colour of the fish: it is not implausible that the eye is attracted by the most vividly coloured fish (although a suggestion to the contrary can be found in Yarbus (1967: 183)). Another possibility is that of a direction bias acquired through reading (in the case of adults). But the effects of these factors are hard to predict in a sensible way and thus table 3 reflects the expectation of a random distribution of last fixations on the left and right fish with a strictly dependent assignment of grammatical voice.

The idealized, expected results in table 2 and 3 are far from what is found in reality. As for the cued-condition child data in tables 4a and 4b there is a deviation that could be expected in that the children almost exclusively utter active constructions.

This confirms a finding from my earlier fish film study of children (Diderichsen 1999 and 2001). When it comes to visual fixations the children seem to perform in a random manner rather than in the systematic way that might have been expected. Only a slight preference for visual fixations on the cued fish can be observed in table 4a.

The results from the adults in table 4c and 4d are not much closer to the ideal in table 3. Table 4c shows results that are as close to Tomlin's predictions as could possibly be hoped for; there is an almost all-exclusive overweight of expected active and passive constructions. But on the visual side there is a substantial number of unexpected last fixations on the left fish, equally non-cued agents and non-cued patients. This is the same as to say that there is a visual bias towards the participant on the left, and this kind of position bias can also be observed elsewhere.

The adult data in table 4d are somewhat more 'messy' than those in table 4c. The expected (agent-cued) active construction is the most frequent one, closely followed by the unexpected (patient-cued) active. These are followed by the expected (patient-cued) passive, and the least frequent construction is the unexpected (agent-cued) passive. On the visual side one can observe a slight bias towards the cued fish. But the pattern most clearly manifested in the table is a strong visual preference for the right fish, no matter if it is cued or not, agent or patient.

Table 4a. Child: ID#5	Last visual fixation:	Left Fish		Right Fish		Other
Syntactic perspective:		Expected (Cued)	Unexp. (R. cued)	Expected (Cued)	Unexp. (L. cued)	
Agent	expected (agent cued)	3	2	3	3	3
(active voice)	unexpected (patient cued)	8	3	5		
Patient	expected (patient cued)					
(passive voice)	unexpected (agent cued)					
Other				1		1

Table 4b. Child: ID#6	Last visual fixation:	Left Fish		Right Fish		Other
Syntactic perspective:		Expected (Cued)	Unexp. (R. cued)	Expected (Cued)	Unexp. (L. cued)	
Agent	expected (agent cued)	5	3	3	3	1
(active voice)	unexpected (patient cued)	3	5	1	5	2
Patient	expected (patient cued)					
(passive voice)	unexpected (agent cued)					
Other		1				

Table 4c. Adult: ID# 10	Last visual fixation:	Left Fish		Right Fish		Other
Syntactic perspective:		Expected (Cued)	Unexp. (R. cued)	Expected (Cued)	Unexp. (L. cued)	
Agent	expected (agent cued)	7	6	2	0	
(active voice)	unexpected (patient cued)					
Patient	expected (patient cued)	7	5	3	1	
(passive voice)	unexpected (agent cued)	1				
Other						

Table 4d. Adult: ID# 11	Last visual fixation:	Left Fish		Right Fish		Other
Syntactic perspective:		Expected (Cued)	Unexp. (R. cued)	Expected (Cued)	Unexp. (L. cued)	
Agent	expected (agent cued)	2	2	4	1	1
(active voice)	unexpected (patient cued)	0	1	3	3	1
Patient	expected (patient cued)	2		3	2	
(passive voice)	unexpected (agent cued)			1	2	1
Other				1	2	

Tables 4a-d. Actual results from the cued condition (expected cells shaded)

Table 5a. Child: ID# 2	Last visual fixation:	Left fish	Right fish	Other
Syntactic perspective:				
Agent	Left agent	7	2	
(active voice)	Right agent	3	4	
Patient	Right patient			
(passive voice)	Right patient			
Other	Left agent	3	4	
	Right agent	5	4	

Table 5b. Child: ID# 3	Last visual fixation:	Left fish	Right fish	Other
Syntactic perspective:				
Agent	Left agent	6	10	
(active voice)	Right agent	5	10	1
Patient	Right patient			
(passive voice)	Right patient			
Other	Left agent			
	Right agent			

Table 5c. Child: ID# 7	Last visual fixation:	Left fish	Right fish	Other
Syntactic perspective:				
Agent	Left agent	6	7	
(active voice)	Right agent	5	8	
Patient	Right patient			
(passive voice)	Right patient			
Other	Left agent	3		
	Right agent	1	2	

Table 5d. Child: ID# 9	Last visual fixation:	Left fish	Right fish	Other
Syntactic perspective:				
Agent	Left agent	1	14	
(active voice)	Right agent	3	13	1
Patient	Right patient			
(passive voice)	Right patient			
Other	Left agent			
	Right agent			

Tables 5a-d. Actual data from the non-cued condition (expected cells shaded)

Also in the non-cued condition the data are quite different from the random pattern that might have been expected. First of all, no passive constructions are observed at all[3], which amounts to saying that there is an overwhelming preference for actives. Second, there seems to be more or less of a visual preference for the left or right fish, at least in tables 5a, 5b, and especially 5d. What distinguishes the adult from the children is the more consistent position bias (towards the right fish), and higher verbal consistency, manifested in the table (5d) by the lack of occurrences in the "Other" row.

So much for the quantitative results. The detailed qualitative analysis of ID# 10 mentioned above yielded the following results. First, those of Holsánová's (2001) observations that are applicable to this experiment generally also hold true for my data. The finding that there are generally more visual fixations than verbal foci can be confirmed, as can the observation that perfect temporal matches between visual fixations and verbal foci are very rare. Thus, Holsánová's rejection of her hypothesis H1 about the unit of comparison for vision and language, as well as thesis T2 about visual/linguistic synchrony (see above) are supported by my data.

Second, some general pattens in the descriptive performance of this subject can be extracted. One thing that is typical is a more or less constant lag of about 300 ms to 1 second between the visual information acquisition and the corresponding verbal output.

Thus, the discription is generally characterized by a pattern that Holsánová calls *series of delays*, although the delays in my data are considerably shorter than the 2-4 second delays reported by Holsánová. The delays are supplemented by what Holsánová calls *triangles*, and thus some of the trials can be described in terms of Holsánovás *series of triangles* where intertwined visual fixations show up as more orderly verbal sequences.

Another pattern that emerges from the analysis of this subject lies in the repeated segmentation of the individual trial descriptions into three intonation units: one containing a presentation of one or both fish, one containing the eating event, and one containing the agent's leaving the screen. A typical example of a trial description is shown in figure 5.

[3] With one big exception; the adult subject whose eye-tracking data are missing uttered only passives.

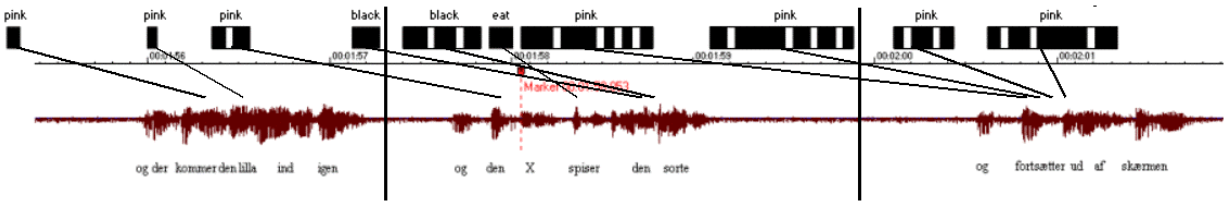

Figure 5. Typical trial divided into three parts (intonation units)

4. DISCUSSION

The data presented above were collected in order to test the assumption that the optical foci, i.e. fixations, of the eyes can be taken as a safe indication of 'visual attention', which would make them suitable to predict the grammatical configuration of the corresponding on-line utterances. All the test subjects whose data are presented in tables 4a-d and 5a-d above performed the task in ways that show that a strict one-to-one correspondence between visual input and linguistic output can not be maintained: each subject has at least some occurrences of non-corresponding last visual fixation and syntactic perspective.

But before this is taken as a direct falsification of one of Tomlin's most basic assumptions, it has to be considered whether the error in the fish film (see above) renders the finding of a less than perfect eye-attention-language correspondence completely meaningless. Tomlin (1997: 173) writes that it takes 75 ms from the second cue to the completion of the swallowing, but the real figure is actually approximately ten times bigger. This means that the fish film, contrary to what Tomlin claims, actually gives the test subjects plenty of time to fixate on something other than the cued fish before the event occurs. Thus, one can not know for sure whether the subject attends to the cued fish at the moment of swallowing by looking at the fixation data at the time of the second cueing, which is about 20 frames before the event. Rather, one has to find the last fixation within the last 2-3 frames before the event to test Tomlin's hypothesis. At this point the fishes are so close to each other that the uncertainty from calibration, poor pursuit motion data handling, and synchronization (see above) makes it unclear which of the two fishes is actually fixated, making such an analysis less than straightforward. But the last-fixation analysis described in this study actually comes close to the ideal attention-controlling timespan of 2-3 frames before the event. Thus, the analysis of the pilot data is justified even though the experimental setup is compromised.

Furthermore, the original experiment has been replicated several times with converging results. But the timing error does neutralize the conclusion that it is the succesful control of visual attentional detection that leads to the predictable assignment of grammatical subject. There might be other factors in the experimental design that are (co-)responsible for its reliability.

What could these other factors be? Of course one could argue on the basis of the demonstrations above of the fact that vision and attention are not the same thing that expected results from test subjects actually do reflect succesful manipulation of attentional detection - temporally independently of visual fixations. But this only recasts the original 'visual attention'-language correspondence hypothesis in terms of a correspondence between temporally vision-independent attention and language and thus adds no new insights to the original fish film research.

An equally speculative, but maybe more integrated line of reasoning brings us back to the notion of information structure. What if the assignment of subject has to be explained in terms of Given-New, Topic-Comment and/or Theme-Rheme after all - in spite of Tomlin's wish to avoid this kind of terms in favour of the extra-linguistic factor of focal attention (Tomlin 1995: 517)? There are reasons to believe that the cued fish in each trial will be at the same time the most cognitively accessible one (Given), the one which the test subject is likely to say something about (Topic), and a natural starting point of the sentence (Theme). When these discourse functions coincide, it is not particularly surprising that they show up as grammatical subject.

The reasons why the cued fish in each trial is likely to be both Given, Topic, and Theme are the following. On the Given-New side, the test subject himself probably often makes the cued fish highly accessible (Given). I have mentioned that there seems to be a tendency to divide the description of each trial into three parts, of which the first part often specifically introduces the cued fish. When this is the case, the cued fish is freshly activated, and must be considered Given information when the next part, describing the swallowing, is uttered. Thus, in addition to 'current attention focus', we have a subject assignment factor of 'recent information focus'. When the cued fish is not specifically introduced, i.e. when there is no first part, or when both fish are introduced, none of the fish are in principle more accessible than

the other, and the Givenness factor loses its power. But then, there is also the topicality factor.

The topicality factor derives from the fact that the cued fish - qua the experimental design, i.e. because of the instruction - has a high aboutness-value. At the beginning of each fish film session, the test subject is told to concentrate on the cued fish. This can be seen as corresponding to a question like "What happens to [the cued fish]?" which in some sense has to remain active in the mind of the test subject throughout the session. WH-questions of this sort in some traditions literally define the notion of Topic. Following van Dijk (1977: 116), the question above is a requests for information about [the cued fish], and [the cued fish] will thus be the Topic of an appropriate ensuing answer. The cueing arrows fit into this scenario as reminders of the request for information about the cued fish.

The Theme factor is not really a factor in its own right. Theme coincides with subject in all relevant respects since Danish is an SOV language with positional case marking; and thus all good reasons to make something the subject of a sentence are also good reasons to make it the Theme of the sentence. This is also reflected in Chafe's 'light subject constraint' (1994: 82ff). However, there is one thing that might somewhat surprisingly be discussed under the heading of Theme: the tight temporal constraints of the fish film. As has been mentioned, there is a shorter vision-language lag of about 300-1000 ms in the on-line fish film task compared to the 2-4 seconds of Holsánová's picture description task, and this may be seen as an effect of the quite fast flow of the unfolding fish film. The temporal constraints on succesful performance calls for efficient managing of thematic structure by the test subjects. It saves time to get the flow of the information one wants to convey right - from the very start of each utterance.

The interpretation above seems to give a plausible explanation for the fairly well-behaved adult data from the various fish film experiments. But what about the child data? In what follows, I will focus on the performance of the children while comparing the pilot child and adult data. A distinct difference between children and adults that can be observed in the data of my earlier study (Diderichsen 2001) as well as the data above is the preference of the children for the active construction. While both children and adults show faint signs of a possible inclination to dedicate their final look before the swallowing to the cued fish, the preference for the active construction in spite of the topicalization of the cued fish is only seen in the children. In (Diderichsen 2001) I concluded that the children's exclusive use of active constructions was probably still a reflection of their attentional focus, but an attentional focus that they could not themselves control yet, rendering them unable to show any consideration for an interlocutor while producing descriptive linguistic output. This thought is not contradicted by the speculations above. On the contrary: it can actually be extended to fit these thoughts. Thus, it is possible that children are not able to keep the request for information about the cued fish in active consciousness while producing their on-line descriptions and instead focus on - i.e. say something about[4] - the most salient participant: the agent. This is the same as to say that both endogenous and exogenous cueing fails when it comes to children. It must be mentioned, though, that the kindergarten teacher repeatedly reminded the children to look at the arrow fish; but this was clearly not enough to topicalize the cued fish[5], which may be seen as another indication that visual fixations do not entail the kind of attention that we are after.

The fact that visual scanning is apparently at least to some degree (i.e. below the level of attentional superfoci) independent of the processing and linguistic rendering of the information is accentuated by the finding of last-look position biases. It turns out that one of the most clearly visible patterns in the data are individual preferences to fixate the left - or the right - fish just before the swallowing. Since these preferences are not found in the children, one might be tempted to explain them by the reading skills possessed only by adults, but this explanation suffers from the fact that both the left and the right fish are preferred by different individuals. There is no obvious explanation for this kind of visual bias, but it is interesting to observe how it is apparently competing with the verbal instructions for the placement of the last fixation before the swallowing. Especially in table 4c one gets the impression that the left uncued fishes have literally 'snatched' the fixations away from the expected destination: the right cued fishes.

Another competing factor, one that actually affects the linguistic output, is present in the entire data set form both children and adults. This is the bias towards the active construction. The active con-struction can uncontroversially be said to be more basic than the passive construction, and this shows up in several ways in the data. First, it manifests itself in the child data, where not a single passive construction is found, no matter if there are visual cues or not. Second, it shows up in the uncued adult session which also only contains active constructions. The other adult whose data were unfortunately lost during the experimental sessions uttered only passives, but considering all the other indications of active bias in the literature, this must be a special case with its own quite specific

[4] Notice this suggestive play on words: could it be that the act of producing a linguistic (or information) focus can be equated to the act of saying something about the participant in attentional focus? This would nicely sythesize Chafe's and Tomlin's thoughts.

[5] - Unlike a proper linguistic elicitation, i.e. a question that topicalizes the relevant participant in the way outlined above. Debriefing sessions were carried out with all the participating children to test their ability to produce passive constructions. Using questions like "What happened to the green fish?" it was possible to reliably elicit passives from all but one of the children.

explanation. Finally, the active bias shows up in cued condition adult subject #11 (table 4d) where one can observe a frequency distribution that arranges the sentences into something that almost resembles a markedness hierarchy: Agent-cued actives: 10 > Patient-cued actives: 8 > Patient-cued passives: 7 > Patient-cued passives: 4. In any case, the active bias in the data is just another indication that the 'visual attention'-language correspondence may not be the whole story; since there is no way of knowing which fish will be the agent before the swallowing has begun, such a bias should not be possible.

5. Conclusion and Directions for Further Research

All in all the above interpretation confirms Holsánovás (2001) finding that there is no strong attentional binding between vision and language on the level of individual foci of attention or consciousness. The basic assumption that I have been attributing to Tomlin, that the control of visual orienting means the control of attentional detection which determines the assignment of syntactic subject can be challenged, not only because of this lacking correspondence between vision and language, but also because of the unfortunate error in the fish film, which means that visual attention can not be controlled in the fish film design in the way outlined by Tomlin.

However, this does not necessarily falsify the finding that attentional detection determines subject assignment; it neutralizes it at worst: even if attention is temporally independent of visual scanning at the sub-second level of individual fixations, it may well be that the attentionally detected participant manifests itself as syntactic subject.

These findings open up for speculation as to what might then cause subject assignment (or possibly the attentional detection that leads to subject assignment), and it turns out that it is possible to show how the cued fish in each trial is likely to be both Given information and Topic, thus making it an obvious candidate for the Theme function and subject position.

The child data with their lack of discourse functions seem to fit into this puzzle if a developmental limit on the ability to keep in mind the topicalization of the cued fish (and possibly an interlocutor's needs in general) during production is assumed.

As this is a pilot study, a couple of suggestions for further research can be added to these concluding remarks. The claims generated in this paper are well suited for further investigations but - if the occasion should arise - it is of great importance to ensure a more detailed data analysis than that presented above. Thus, a proper discourse analysis of each subject's verbal output is strongly recommended, and it is suggested that the automatic analyses by the iView software are supplemented with manual frame-by-frame analyses of all the visual data. Furthermore, it might be interesting to see the effects of placing the second cueing arrow within the last 2-3 frames before the swallowing, thus creating a properly timed exogenous cue.

Tweaked in the ways outlined above, the fish film should still be a useful tool for investigating Tomlin's original thought that language reflects attention during production. This study has hardly but pointed out a few of its minor weaknesses.

References

Brown, Gillian and Yule, George (1983): Discourse analysis. Cambridge etc.: Cambridge University press.

Chafe, Wallace (1994): Discourse, conscious-ness, and time: the flow and displacement of conscious experience in speaking and writing. Chicago & London: The University of Chicago press.

Chafe, Wallace (1996): How consciousness shapes language. Pragmatics & Cognition, 4, pp. 35-54.

Clark, Eve V. (1989): Speaker perspective in acquisition. Linguistics, 28, pp. 1201-1220.

Diderichsen, Philip (1999). Passiven i danske børns sprog: En beskrivelse af den selektive opmærksomheds rolle i tilegnelsen af sproglig perspektivering i dansk, specielt aktiv/passiv-distinktionen. BA-project from the Institute for General and Applied Linguistics, University of Copenhagen.

Diderichsen, Philip (2001): The role of the development of selective attetention in the development of linguistic topicality: a study of the acquisition of the passive construction in Danish. Copenhagen: Netværk for Funktionel Lingvistik.

Van Dijk, Teun A. (1977): Text and Context: Explorations in the semantics and pragmatics of discourse. London and N.Y.: Longman.

Eysenck, Michael W. and Keane, Mark T. (1995): Cognitive Psychology: A Student's Handbook. Hove: Erlbaum (UK) Taylor & Francis.

Gazzaniga, Michael S., Ivry, Richard B., and Mangun, George R. (1998): Cognitive neuroscience: The biology of the mind. N.Y. and London: W. W. Norton & Co.

Glenstrup, Arne John and Engell-Nielsen, Theo (1995): Psychological and Physiological Aspects of Eye-Gaze Tracking. Chapter 3 in Eye Controlled Media: Present and Future State. Thesis for the Partial Fulfilment of the Requirements for a Bachelor's Degree in Information Psychology at the Laboratory of Psychology, University of Copenhagen. URL:http://www.diku.dk/~panic/eyegaze/ article.html.

Halliday, M.A.K. (1985a): An introduction to functional grammar. London etc.: Edward Arnold.

Halliday, M.A.K. (1985b): Dimensions of Discourse Analysis: Grammar. In van Dijk (ed.), Handbook of discourse analysis, vol. 2: Dimensions of discourse, pp 29-56. London etc.: Academic press.

Hoffman, James E (1998): Visual Attention and Eye Movements. In Pashler, Harold (ed.), Attention. Psychology Press.

Holsánová (2001): Picture Viewing and Picture Descriptions: Two Windows to the Mind. Doctoral dissertation from Lund University Cognitive Science.

Juola, J.F., Bowhuis, D.G., Cooper, E.E., and Warner, C.B. (1991): Control of attention around the fovea. Journal of Experimental Psychology: Human Perception and Performance, 15, pp. 315-330. (Cited by Eysenck and Keane 1995.)

Just, M.A. & Carpenter, P.A. (1980): A theory of reading: From eye fixations to comprehension. Psychological Review, 87, pp. 329-354. (Cited by Glenstrup and Engell-Nielsen 1995.)

Kelstrup, Søren (1998): Konceptuel Prominens: En psykolingvistisk tilgang til valget af subjekt i semantisk monotransitive sætninger i dansk. BA thesis, Department of General and Applied Linguistics, University of Copenhagen.

Lakoff, George and Johnson, Mark (1999): Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought. N.Y.: Basic Books.

Neisser, U. and Becklen, P. (1975): Selective looking: Attending to visually super-imposed events. Cognitive Psychology, 7, pp. 480-494. (Cited by Eysenck and Keane 1995.)

Tomlin, Russell S. (1984): Linguistic Reflections of Cognitive Events. In Tomlin, Russell S. (ed.) Coherence and Grounding in Discourse: Outcome of a Symposium, Eugene, Oregon, June 1984, pp. 455-479. Amsterdam and Philadelphia: John Benjamins Publishing Co.

Tomlin, Russell S. (1986): The Identification of Foreground-Background Information in Oral Descriptive Discourse. Papers in Linguistics, 19, pp. 465-494.

Tomlin, Russell S. (1995): Focal Attention, Voice, and Word Order: An Experimental. Cross-Linguistic Study. In Downing, Pamela and Noonan, Michael (Eds.), Word Order in Discourse, pp. 517-554. Amsterdam and Philadelphia: John Benjamins Publishing Co.

Tomlin, Russell S. (1997): Mapping Conceptual Representations into Linguistic Representations: The Role of Attention in Grammar. In Nyuts, Jan and Pederson, Eric (Eds.), Language and Conceptualization, pp. 162-189. Cambridge: Cambridge University Press.

Tomlin, Russell S. and Villa, Victor (1994): Attention in Cognitive Science and Second Language Acquisition. Studies in Second Language Acquisition, 16, pp. 183-203).

Yarbus, A.L. (1967): Eye movements during perception of complex objects. In Riggs, L.A. (Ed.), Eye Movements and Vision, pp. 171-196. New York: Plenum Press.

Östman, Jan-Ola and Virtanen, Tuija (1995): Discourse analysis. In Handbook of pragmatics: manual, pp 239-253. Amsterdam/Philadelphia: John Benjamins Publishing Company.