Scandinavian Journal of Psychology, 2005, 46, 229–238

# Antecedent reactivation by surface and deep anaphora in Norwegian

ARILD HESTVIK,[1,2] HELGE NORDBY[3] and GEIR KARLSEN[3]

[1]*PhD Program in Speech and Hearing Sciences, City University of New York Graduate Center, USA*
[2]*Section for Linguistic Studies, Department of Linguistics and Comparative Literature, University of Bergen, Norway*
[3]*Division of Cognitive Neuroscience, Department of Medical and Biological Psychology, University of Bergen, Norway*

Hestvik, A., Nordby, H. & Karlsen, G. (2005). Antecedent reactivation by surface and deep anaphora in Norwegian. *Scandinavian Journal of Psychology*, *46*, 229–238.

Anaphora are expressions in language that depend on other linguistic entities for their full meaning. They can furthermore be divided into two types according to the level of representation where they find their antecedents: *Surface anaphora*, which resolve their reference at the sentence representation level, and *deep anaphora*, which resolve their reference at the non-grammatical level of discourse representation. The linguistic theory of these two anaphor types, and recent findings about processing differences at these two levels, combine to predict that surface anaphora should show fast and immediate reactivation of their antecedents, whereas deep anaphora should have a slower time course of antecedent reaccess. These predictions were confirmed with two lexical decision task experiments with Norwegian stimuli.

*Key words:* Psycholinguistics, syntax, anaphora, sentence processing, reactivation, semantic priming.

*Dr Arild Hestvik, PhD Program in Speech and Hearing Sciences, CUNY Graduate Center, 365 Fifth Avenue, New York, New York 10016, USA. Tel: +1-212-817-8806; fax: +1-212-817-1537*; e-mail: ahestvik@gc.cuny.edu.

## INTRODUCTION

Linguistic theory can be regarded as a theory about mental states and tacit knowledge, embodied by linguistic competence and played out as linguistic performance (Hauser, Chomsky & Fitch, 2002). But to what degree does the mind/brain process language as if it is computing the relations and constructing the representations postulated by linguistic theory? A general project is to search for distinctions in language processing that match the distinctions in linguistic models. If such matches are found, one can conclude that the language performance system is an implementation of the constructs set forth by language competence theory (theoretical linguistics). In this article, such distinctions are sought in the processing of *anaphora*.

An anaphor is a linguistic expression that does not carry meaning by itself, but inherits its meaning from another expression. According to Hankamer and Sag (1976), anaphora can be subdivided into two types according to how they relate to this other expression. *Surface anaphora* get their meaning as a function of the meaning of another syntactic expression, and require that this other expression is contained in the same syntactic representation (of a sentence) that contains the anaphor. This is so because the constraints on the relation between the anaphor and the antecedent are only statable by rules that refer to syntactic structure. *Deep anaphora*, on the other hand, establish their meaning by searching for an antecedent among elements in the non-linguistic and non-syntactic representation of the current discourse. Hankamer and Sag showed that this is so because the units that make up the antecedent of a deep anaphor do not have to correspond to a syntactic unit.

To illustrate this difference, consider reflexive pronouns, which are surface anaphora by the criteria established by Hankamer and Sag. In the sentence "John likes himself", the word "himself" is an anaphor because it cannot establish reference by itself, but is dependent upon the reference already established by the word "John". The non-reflexive pronoun "he", on the other hand, can be dependent on an expression outside the sentence, or can get its meaning from an entity that has no linguistic expression, as in the following example: Suppose two men are watching a third man scaling a wall, and one man says to the other out of the blue: "He's really good!" In this case, "he" refers directly to the object in the mental representation of the visual scene induced by purely visual and cognitive (but not language-) perception. This is typical of deep anaphora.

Surface anaphora are most typically instantiated by elided syntactic material, as in VP-ellipsis ("I like apples, and you do [. . .], too"), and gapping ("I like apples, and you [. . .] oranges."). In these two cases, the elided part of the sentence (indicated by "[. . .]") gets its value from the verb phrase and the verb, respectively, in the previous sentence. Deep anaphora are most often overt pro-forms, as in "Do it"–anaphora ("I often cheat, and you do it, too!"), where *it* is evaluated directly as having the same *meaning* as the first sentence.[1]

In a subsequent paper, Sag and Hankamer (1984) made further claims about the psychological status of deep and surface anaphora. In their model, both syntactic representations and pure meaning representations are psychological objects, and they claimed that "the differences between the two kinds of anaphora . . . can only be understood in terms of a performance model: a model of how discourses are represented, produced, and comprehended" (p. 329). Drawing

on Johnson-Laird (1983), their model entailed that the process of going from sound to meaning works as follows:

As a new piece of discourse is produced, what is comprehended first is a propositional representation. . . . Thus at any given point in a discourse, what is present in the mind of a comprehender is a pair: a model of the discourse to date, and a propositional representation of the immediately preceding discourse. As the discourse proceeds, the content of the propositional representation, which is held in short-term memory, is integrated into the model. At some point the propositional representation is discarded, making room for new propositional representations . . .

The gradual breakdown of these (linguistic) representations into a "model of the discourse" matches what one finds in current discourse-semantic theories such as the Discourse Representation Theory of Kamp and Reyle (1993).

In summary, the psychological content of Hankamer and Sag's theory is the claim that surface anaphora look for antecedents at a syntactic level of mental representation, whereas deep anaphora look for antecedents at a discourse-semantic level of mental representation. Note that Hankamer and Sag themselves did not attempt to derive specific *processing* predictions from this hypothesis; instead, they focused on grammaticality predictions. However, the deep/surface distinction did initially receive some attention in the psycholinguistic literature, by investigators who examined whether violations of the various *parallelism requirements* between antecedent and anaphor differentially affected processing of surface versus deep anaphora (see Tanenhaus & Carlson, 1990; Mauner, Tanenhaus & Carlson, 1995). In the work reported here, the time course of the anaphor resolution process is examined. The background for this is that when recent findings about time course differences between purely syntactic versus discourse processing are combined with Hankamer and Sag's theoretical claims, the prediction arises that surface anaphor resolution should be faster than deep anaphor resolution. This is borne out by our experimental results, and we conclude that the findings are consistent with Hankamer and Sag's claim that the two anaphor types are relegated to different psychological representation and processing mechanisms.

## SYNTACTIC VERSUS DISCOURSE PROCESSING

Several recent theoretical and empirical studies have concluded that processing of formal syntactic representations must be different from processing of discourse-related and pragmatic information. First, experimental studies of gap-filling processes have shown that this process is fast, automatic and reflex like (see, for example, Love & Swinney, 1996; Nicol, Fodor & Swinney, 1994; Nakano, Felser & Clahsen, 2002; Osterhout & Swinney, 1993; Stowe, 1986), and gaps are surface anaphors. Grodzinsky & Reinhart (1993) interpret classical behavioral findings on pre-school children's understanding of anaphor relations (see Chien & Wexler, 1990), as showing that children automatically and correctly process pronouns that function as bound variables (which must find a sentence internal antecedent and are surface anaphora), whereas pronouns that may refer deictically (and which are deep anaphora) involve computations over discourse representations, which require more cognitive resources, and for this reason leads to errors in pre-school children. Avrutin (2000) report experimental findings that shows that aphasics and children make significantly more comprehension errors with discourse-linked versus non-discourse linked Wh-phrases. According to Avrutin, the computations required for discourse-linked Wh-phrases involve "the integration of syntactic and discourse-related knowledge"; speakers are required to "introduce a set of presupposed" entities to which the discourse linked phrase can refer. Operations that obligatorily relate to discourse representations require more processing resources and therefore are harder to "learn" (for children) and take a longer time to process (for both adults and children), according to Avrutin. Similarly, Shapiro (2000) argues, based on reaction time studies, that reaccess of the antecedent of discourse-linked Wh-phrases are delayed relative to non-discourse linked Wh-phrases. Finally, Friederici (1995) adds electrophysiological evidence for the claim that language processes related to purely syntactic form has an early time course of processing relative to other processes.

Whereas there is no *a priori* theoretical reason why discourse processing should have a different time course than syntactic processing, the empirical findings suggest that it does. We will follow the lead of Avrutin (2000) and take these experimental findings to suggest a model of anaphor processing where mental computations that require inspection of non-linguistic representations of discourse are more resource demanding, less automatic, and therefore has a relatively slower time course. On the other hand, form-driven, syntactic processing is reflex-like, automatic and fast. This predicts that deep anaphora should in general have a slower time course of accessing their antecedent than surface anaphora.

In order to test these predictions, two experiments were designed to measure the time course of antecedent reactivation for deep and surface anaphora, and to compare the time course of antecedent reactivation for the two anaphor types.

## EXPERIMENT 1: SURFACE ANAPHORA

An all-visual lexical decision reaction time (RT) experiment was first conducted to replicate the finding by Nicol & Swinney (1989) that reflexive pronouns immediately reactivate their antecedents.

### *Method*

*Subjects.* 29 Norwegian students were paid NOK 50 for participation.

*Materials.* Twenty experimental sentences in Norwegian were constructed that contained the reflexive pronoun "seg selv". All sentences had the structure "[$_{NP}$ N PP] [V (P) seg selv + Adjunct]" (see Appendix A for the full set). In addition, 40 filler sentences were constructed. Ten of these were presentational constructions of the form "It was X who V'ed XP"; 10 were Yes/No-questions, 10 were declarative sentences introduced by an object relative clause NP, and 10 were declaratives introduced by a subject relative clause NP.

The subject noun of each sentence was matched with two words (probes): One probe was semantically related to the subject noun (as determined by an independent rating study, see Hestvik, Nordby, Karlsen & Myklebust, 2002), and the other probe was semantically unrelated to the prime noun or any other noun in the sentence. The two probes in each pair were matched for length and phonological structure, and did not differ significantly in reaction time in an independent lexical decision task (see Hestvik *et al*., 2002). The probe words were mostly nouns, but in some cases verbs. The 40 fillers were matched with 20 words and 60 non-words as probes. In this way, the proportion of words to non-words in the entire stimuli set was 1:1.

Each experimental sentence had two probe positions: Position 1 was 1,000 ms after the prime noun, but still temporally before the reflexive pronoun. Position 2 occurred 500 ms after the reflexive pronoun was shown. The following illustrates by example:

(1) Brannmannen-i terapigruppen-[1]-kritiserte-seg selv-[2]-etter redningsaksjonen. "The fireman in the therapy group criticized himself after the emergency operation."

The pair of probe words for this stimulus was ILD "fire" vs. DILL (the spice), where "fire" is semantically related to "fireman" and "dill" is semantically unrelated.

Probe positions in the fillers were in 4 different positions: In set 1, probes were placed before the last word chunk (see below), in set 2 before the penultimate chunk, in set 3 before the antepenultimate chunk, and in set 4 at the end of the sentence. In this way, across the entire set of stimuli, probe positions varied considerably, lowering subject's expectations as to where they would appear.

*Design.* Subjects were randomly assigned to two experimental groups, so that one group saw probes in position 1, and one group saw probes in position 2. Position thus constituted the between subject factor. Each subject was presented with 20 trials of related probes and 20 trials of unrelated probes, by presenting each of the 20 experimental sentences twice, once with a related probe and once with an unrelated probe. Probe type thus constituted the repeated measures within-subject factor, yielding a 2 (position) × 2 (probe type) mixed factorial ANOVA design.

*Procedure.* The experiment was programmed using MEL2 software running on a Fujitsu laptop computer. The 60 sentences were visually presented on the computer screen (cf. Fig. 1) in a random order in two immediately succeeding sessions, one session presented only semantically related probes to experimental sentences, and the other session presented only semantically unrelated probes (or word/non-word in the filler sentences). The order of presentation was counterbalanced within each group, so that half the subjects saw related probes in the first session and half the subjects saw the related probes in the second session.

Subjects were instructed to read the sentences for comprehension. When the probe appeared, subjects made a speeded lexical decision (word/non-word) by pressing a "yes" or "no" button. The response terminated the probe presentation and initiated presentation of the rest of the sentence. To ensure attention to the stimuli, subjects were instructed to paraphrase each sentence after it was finished.

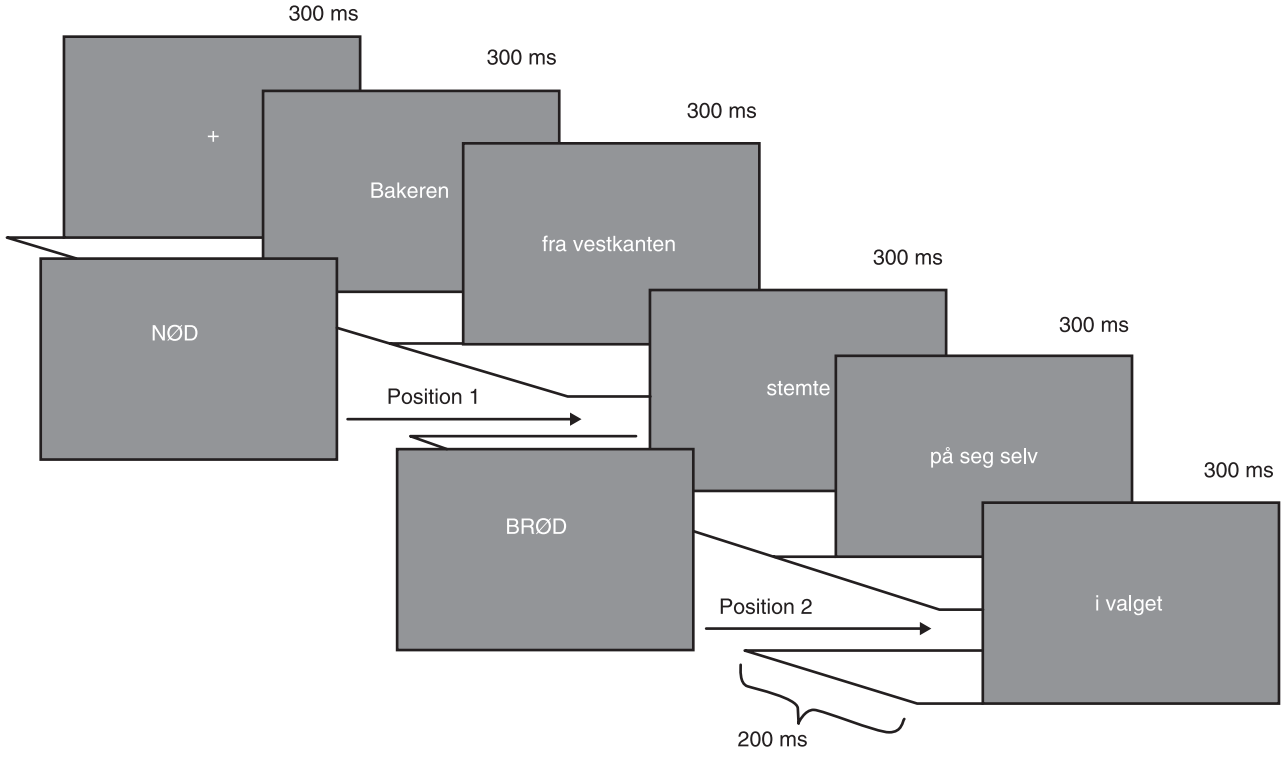


*Fig. 1.* Sentence presentation. Word chunks were presented centrally on the screen in white letters on a dark background. Probe words were presented in green uppercase letters 1° above the sentence presentation text.

Table 1. *Reaction times to related and unrelated probes by position, Experiment 1. Numbers in parentheses indicate standard errors*

| Position | N | Unrelated | Related |
|---|---|---|---|
| 1 | 13 | 867 (39) | 853 (37) |
| 2 | 15 | 871 (52) | 779 (45) |

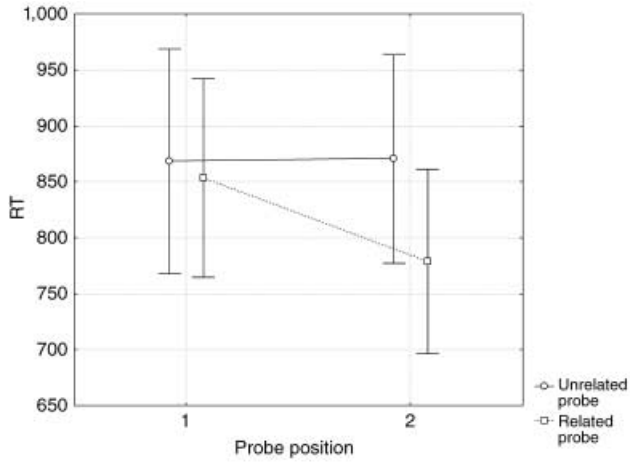


*Fig. 2.* Reaction times by probe type and position, Experiment 1. Vertical bars denote 0.95 confidence intervals.

*Results*

Because the lexical decision response was either incorrect or missing, 6.3% of the trials were removed from analysis. An additional 3.8% of the total trials were removed because the reaction times constituted outliers. The non-outliers range within a subject and within a condition was defined as the interquartile range (the *H*-spread) ± 1.5 times the *H*-spread (cf. Tukey, 1977); reaction times above or below this range were removed. One subject pressed the same button at every trial and was removed from the analysis. The overall average means per condition (with standard errors in parentheses) are given in Table 1.

Mean reaction time per condition was calculated per subject; these means were then used in a mixed factorial repeated measures ANOVA with Position (before vs. after the reflexive) as between factor and Probe Type (related vs. unrelated probe) as within factor. This ANOVA yielded a main effect for probe type, $F(1, 26) = 9.21$, $p < 0.006$, and a significant interaction between probe type and probe position, $F(1, 26) = 4.99$, $p < 0.04$ (see Fig. 2).

Planned orthogonal contrasts comparing related and unrelated probes in position 2 showed that the difference immediately after the reflexive (91 ms) was highly significant, $F(1, 26) = 14.95$, $p < 0.0007$. The difference between related and unrelated probes in position 1 was not significant, $F(1, 26) = 0.29$, $p = 0.58$.

*Discussion*

The results show that related probe words immediately following the reflexive are indirectly primed by the antecedent of the reflexive, via the relation it bears to the reflexive. This shows that the reflexive virtually instantaneously reactivates its antecedent and the semantic associative network of this antecedent. No such effect was observed in position 1, even though position 1 is in fact temporally closer to the antecedent (occurring exactly 1,000 ms after it). This shows that any recency priming (or perhaps more aptly termed, repetition priming) has dissipated after 1,000 ms. The priming caused by the reflexive must therefore be due to immediate reactivation of the antecedent. In effect, the reflexive is interpreted by the language processing system as corresponding to the lexical item of the grammatical antecedent, and the lexical semantic information of that item is what primes the decision task for the probes. Note that the significance of this is not the priming itself, but rather that the priming shows that the reflexive reactivates the semantic information associated with its antecedent, replicating for Norwegian the finding based on English language stimuli by Nicol & Swinney (1989).

## EXPERIMENT 2: DEEP ANAPHORA

We next investigated whether deep anaphora of the "Do-It"-type reaccess their antecedents in the same way as reflexives. A lexical decision task with the same design as Experiment 1 was run with the Norwegian deep anaphor "gjøre det" ("do it"). This expression was determined to be a deep anaphor by the tests in Hankamer and Sag (1976).

Lexical decision probes were placed in a control position before the anaphor, immediately after the anaphor and in two downstream positions to test for late reactivation. We here predicted a slower timecourse of reaccess, since the resolution of the anaphor is not mediated by fast automatic processing.

*Method*

*Subjects.* 43 Norwegian students were paid NOK 50 for participation.

*Materials.* Twenty stimuli sentences with the Norwegian expression "gjorde også det" ("did it too") were constructed as follows: each stimuli sentence from Experiment 1 was turned into a sentence where the subject in Experiment 1 now was a direct object. E.g.

(2) Brannmannen i terapigruppen kritiserte seg selv – Lederen kritiserte brannmannen. "The fireman in the therapy group criticized himself" – "the leader criticized the fireman."

This sentence was then conjoined with another clause containing a modified subject and the deep anaphor VP:

(3) Lederen kritiserte *brannmannen* og kvinnene i terapigruppen gjorde også *det* uten at noen brydde seg. "The leader criticized the fireman and the women in the therapy group did it too, without anyone caring."

The same probe pairs matched with the subject in Experiment 1 were now matched with the object of the first clause in Experiment 2. In this way, the materials were kept fairly constant across the two experiments. In addition, 40 filler sentences matched with 20 words and 60 non-word probes were constructed. The fillers were varied in structure along the same lines as in Experiment 1, to lessen subject expectations about the nature of the stimuli.

Probes were placed in four different syntactic positions, between word chunks as in Experiment 1. Probes in position [1] were shown immediately after the subject noun phrase of the second sentential conjunct. This position was always 2,000 ms (corresponding to 4 word chunks) after the "priming" noun phrase ("brannmannen" in the example below), and 700 ms before the deep anaphor. No "repetition" priming by the subject of the second clause, as a function of mere temporal vicinity as opposed to reactivation, is expected at this long latency. Indeed, the lack of probe type effect at the corresponding control position in Experiment 1 shows that any "recency" or repetition priming has dissipated as early as 1,000 ms after presentation of the antecedent noun phrase. This probe position therefore serves as a control condition for testing priming at the next probe position, position [2]. Probes in position [2] always occurred 500 ms after the onset of the anaphor word chunk. If deep anaphora reactivate their antecedents in the same manner as surface anaphora, a probe type effect (i.e. priming) is expected in this position, as opposed to in position [1].

Finally, probes in position [3] occurred 1,000 ms after the deep anaphor and probes in Position [4] were presented 1,500 ms after the onset of the anaphor. The purpose of testing for a probe effect in these positions were exploratory: Even if we don't observe reactivation priming immediately after the deep anaphor, it is conceivable that reactivation in this case has a slower time course than that observed with surface anaphora. If so, we might observe a priming effect in one or both of these "downstream" positions. The four probe positions are illustrated below with a sample stimulus from Appendix B:

(4) Lederen-kritiserte-brannmannen-og-kvinnene-i terapigruppen-[1]-gjorde-også det-[2]-uten at-[3]-noen-[4]-brydde seg.

As in Experiment 1, a pair of probes was associated with each stimulus sentence. For example, the probe pair for the stimulus sentence in (4) was ILD ("fire") versus DILL (the spice). Specifically, the same prime word-probe pair match-ups were used as in Experiment 1. For example, in Experiment 1, *brannmann* ("fireman") would be a prime word and the probes for the sentence containing this prime word would be ILD "fire" or DILL (the spice). Since we observed reactivation priming with the same prime word-probe pairs in Experiment 1, these stimuli should be well suited for detecting reactivation priming in Experiment 2 as well, if it indeed does occur.

*Procedure.* The procedure was identical to that of Experiment 1, except that half the stimuli sentences were matched with a related probe and half with an unrelated probe in session 1. In session 2, the stimuli that received a related probe in session 1 now received an unrelated probe, and vice versa.

*Design.* The subjects were randomly assigned to four experimental groups based on probe position. Group 1 saw probes in position 1, group 2 saw probes in position 2, etc. Position thus constituted the between-subjects factor, and probe type (related/unrelated) the within-subjects factor.

### *Results*

Only correct responses were included for analysis. In addition, outliers (defined in the same way as in Experiment 1) within condition per subject were removed. The average means per condition (standard errors in parentheses) are given in Table 2.

Table 2. *Reaction times to related and unrelated probes by position, Experiment 2. Numbers in parentheses indicate standard errors*

| Position | N | Unrelated | Related |
|---|---|---|---|
| 1 | 10 | 806 (31) | 792 (35) |
| 2 | 10 | 818 (41) | 829 (42) |
| 3 | 11 | 852 (41) | 843 (31) |
| 4 | 11 | 824 (43) | 814 (39) |

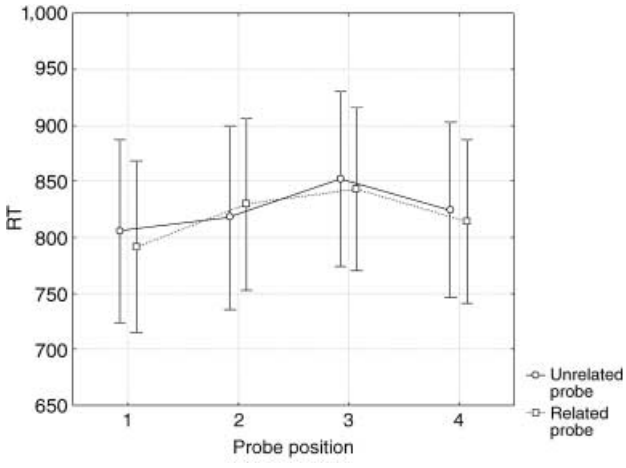


*Fig. 3.* Reaction times by probe type and position, Experiment 2. Vertical bars denote 0.95 confidence intervals.

A mixed factorial repeated measures ANOVA with probe position as between-subjects factor and probe type as within-subject factor showed no main effect of position ($F(3, 38) = 0.29$, $p = 0.83$) nor of probe type ($F(1, 38) = 0.44$, $p = 0.51$), and the interaction between probe and probe position was not significant ($F(3, 38) = 0.49$, $p = 0.69$).

### *Discussion*

This experiment yielded no significant effect of probe type in either probe position. We interpret the lack of priming in position [2] as evidence that the deep anaphor does not immediately reaccess the semantic information associated with its antecedent. Furthermore, the lack of priming in the two downstream positions shows that the resolution of the anaphor is probably delayed until after the entire sentence is processed.

## GENERAL DISCUSSION

Using Norwegian stimuli we first replicated previous findings based on English stimuli that reflexive pronouns immediately reactivate their antecedents. This was as expected, since reflexive pronouns are surface anaphora, and surface anaphora establish their antecedents at the syntactic level of

representation. Based on the model that says that syntactic representations are computed quickly and early in the sentence understanding process, this is an expected result.

We then constructed a set of sentences containing deep anaphora that referred back to an entire VP. This VP contained as grammatical objects the exact same prime nouns used in the surface anaphor experiment. The exact same probe words were used to test activation as in Experiment 1. The fact that the deep anaphora stimuli did not yield priming effects, despite the identity of materials and method across the two experiments, is consistent with the interpretation that a deep anaphor of the "Do it"-type does not immediately reactivate its antecedent during online processing, and in fact has a slower time course of processing. If the "Do it"-type anaphor must wait until a discourse representation has been constructed, and only then can establish its reference via anaphoric linkage to an object at this level of representation, it is expected that we should not observe any priming effect immediately after the presentation of the anaphor, unlike the finding with reflexives.

This is consistent with the predictions we derived for deep anaphora versus surface anaphora based on the Hankamer–Sag theory, and thus supports the claims made by Sag and Hankamer that the two anaphor types are processed differently by the language processing system. The results also support the proposal by Avrutin (2000) that form-driven linguistic processing (as in surface anaphora) has a faster time course of mental computation than grammatical processing that requires access to pragmatics and discourse-related information, as with deep anaphora.

A weakness of the current finding is that we did not observe antecedent reactivation for deep anaphora even in the downstream positions. One would expect that even if the anaphora resolution process for deep anaphora had a slower time course, a meaning is nevertheless eventually assigned to it, and that this meaning assignment should have been reflected in priming at a late probe position. The current finding could therefore also be interpreted as if Experiment 2 simply failed to detect anything. There were also other differences in the stimuli presentations across the two experiments that confound the interpretation of the two experiments. For example, the prime word in Experiment 2 was a grammatical object whereas the prime word in Experiment 1 was a grammatical subject, and the sentences in Experiment 1 were simple clauses whereas Experiment 2 used coordinated structures. In principle, these differences confound the interpretation of the results, but there is no theoretical reason or model that would entail lack of immediate reactivation for these reasons. For lack of a better explanation, the deep versus surface anaphor resolution mechanisms therefore seems to be the best explanation.

Another recent finding based on English stimuli also bolsters the current interpretation. In Experiment 3 in Shapiro, Hestvik, Lesan & Garcia (2003), we used stimuli and methods that were essentially identical to those used in Experiment 2, except VP-ellipsis was used instead of "Do it"-anaphora. The following is an example, with probe positions indicated:

(5) The mailman bought a tie for Easter, and his brother, who was [1] playing volleyball, did [2] too, according to the sales clerk.

As shown in that paper, the English surface anaphor "Did too", which differs minimally from "Did it", did indeed generate immediate reactivation priming by the grammatical object in the elided VP. Since the method was the same in both experiments, and the stimuli only differed, from a theoretical perspective, in whether the first clause VP contained an "it" or not, the fact that there was antecedent reactivation with the English surface anaphor VP-ellipsis stimuli but not with the Norwegian deep anaphor "Do it"-stimuli can be explained if the latter is a deep anaphor and the former a surface anaphor, and if syntactic processing precedes discourse-semantic processing. (This direct comparison could not be carried out with Norwegian stimuli since English is the only Germanic language with true VP-ellipsis). In conclusion, we interpret our results as showing that surface anaphora establish their anaphoric relation immediately, as shown by antecedent reactivation priming, whereas deep anaphora do not. These findings can therefore be viewed as experimental data supporting the theoretical distinctions proposed by Hankamer and Sag.

This research was supported by a Norwegian Research Council grant to the first two authors during 1997–99 (project #10094, "Language processing indicated by event-related brain potentials"); by a Meltzer Foundation grant to the first two authors during 1997, and by a NIH/NIDCD training grant fellowship to the first author during 2002 (PHS grant #5T32DC0000-39). The authors also gratefully acknowledge the help of Ninni Pedersen for contributing to this project.

## NOTES

[1] However, the distinction cuts across the overt-covert division: some anaphors that are surface anaphors by the tests in Hankamer and Sag are overt (as reflexive pronouns), and some deep anaphors are covert, as Null Complement Anaphora (NCA) ("I know what to say, but just don't know how!"). The reader is referred to Hankamer and Sag's original papers for a full overview of the linguistic arguments and criteria for placing an anaphor in one category or the other.

## REFERENCES

Avrutin, S. (2000). Comprehension of discourse-linked and non-discourse-linked questions by children and Broca's aphasics. In Y. Grodzinsky, L. P. Shapiro & D. Swinney (Eds.), *Language and the brain: representation and processing*, Academic Press.

Chien, Y.-C. & Wexler, K. (1990). Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition*, *1*, 225–295.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: a model based on neuropsychological and neurophysiological data. *Brain and Language*, *50*, 259–281.

Grodzinsky, Y. & Reinhart, T. (1993). The innateness of binding and coreference: a reply to Grimshaw and Rosen. *Linguistic Inquiry*, *24*, 69–101.

Hankamer, J. & Sag, I. (1976). Deep and surface anaphora. *Linguistic Inquiry*, *7*, 391–426.

Hauser, M. D., Chomsky, N. & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, *298*(*5598*), 1569–1579.

Hestvik, A., Nordby, H., Karlsen, G. & Myklebust, E. (2002). *A Norwegian stimuli set for semantic priming experiments* (Technical report). Bergen: University of Bergen, Department of Linguistics and Comparative Literature.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

Kamp, H. & Reyle, U. (1993). *From discourse to logic*: Kluwer Academic Publishers.

Love, T. & Swinney, D. (1996). Coreference processing and levels of analysis in object-relative constructions: Demonstration of antecedent reactivation with the cross-modal priming paradigm. *Journal of Psycholinguistic Research*, *25*(1), 5–24.

Mauner, G., Tanenhaus, M. K. & Carlson, G. (1995). A note on parallelism effects in processing deep and surface verb-phrase anaphors. *Language and Cognitive Processes*, *10*, 1–12.

Nakano, Y., Felser, C. & Clahsen, H. (2002). Antecedent priming at trace positions in Japanese long-distance scrambling. *Journal of Psycholinguistic Research*, *31*(5), 531–571.

Nicol, J. & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, *18*(1), 5–19.

Nicol, J. L., Fodor, J. D. & Swinney, D. (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*(5), 1229–1238.

Osterhout, L. & Swinney, D. A. (1993). On the temporal course of gap-filling during comprehension of verbal passives. *Journal of Psycholinguistic Research*, *22*(2), 273–286.

Sag, I. & Hankamer, J. (1984). Toward a theory of anaphoric processing. *Linguistics and Philosophy*, *7*, 325–345.

Shapiro, L. P. (2000). Some recent investigations of gap-filling in normal listeners. In Y. Grodzinsky, L. P. Shapiro & D. Swinney (Eds.), *Language and the brain: representation and processing*: Academic Press.

Shapiro, L. P., Hestvik, A., Lesan, L. & Garcia, R. (2003). Charting the time-course of sentence processing: reconstructing missing arguments in VP-ellipsis constructions. *Journal of Memory and Language*, *49*(1), 1–19.

Stowe, L. A. (1986). Parsing WH-constructions: evidence for on-line gap location. *Language and Cognitive Processes*, *1*(3), 227–245.

Tanenhaus, M. K. & Carlson, G. N. (1990). Comprehension of deep and surface verb phrase anaphors. *Language and Cognitive Processes*, *5*, 257–280.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Received 19 September 2003, accepted 21 June 2004

## APPENDIX A: STIMULI SENTENCES AND PROBE WORDS, EXPERIMENT 1

Word chunks are indicated by dashes; probe positions between square brackets; probe words are set in capitals.

*Experimental sentences*

1. Journalisten-i rommet-[1]-hentet vann-til seg selv-[2]-fordi-han var tørst. AVIS-LAKRIS
2. Tannlegen-i restauranten-[1]-skyldte-på seg selv-[2]-for -den dårlige vinen. BOR-KOR
3. Politimannen-i salen-[1]-forsvarte-seg selv-[2]-mot anklagene. BOT-ROT
4. Vaskekonen-med astma-[1]-snakket-med seg selv-[2]-i gangen. BØTTE-STØTTE
5. Postbudet-på ferie-[1]-kjøpte billetter-til seg selv-[2]-til konserten. BREV-VEV
6. Bakeren-fra vestkanten-[1]-stemte-på seg selv-[2]-i valget. BRØD-NØD
7. Veterinæren-med utstyret-[1]-siktet-på seg selv-[2]-med kameraet. DYR-MYR
8. Frisøren-med kappen-[1]-klippet-seg selv-[2]-foran speilet. HÅR-KÅR
9. Brannmannen-i terapigruppen-[1]-kritiserte-seg selv-[2]-etter redningsaksjonen. ILD-DILL
10. Presten-fra Frankrike-[1]-beskrev-seg selv-[2]-i en artikkel. KIRKE-KURV
11. Soldaten-med langt hår-[1]-refererte-til seg selv-[2]-i rapporten. KRIG-KVELD
12. Gruvearbeideren-på besøk-[1]-tenkte bare-på seg selv-[2]-ifølge -vennene hans. KULL-MUGG
13. Joggeren-fra Askøy-[1]-foretrakk-seg selv-[2]-fremfor andre. LØPE-DØPE
14. Kokken-på møtet-[1]-bebreidet-seg selv-[2]-for det -som hadde hendt. MAT- HAT
15. Turisten-i butikken-[1]-åpnet et brev-til seg selv-[2]-i full fart. REISE-MALE
16. Mureren-i banken-[1]-skjøv-seg selv -fremover -i køen. SEMENT-SKRIBENT
17. Polfareren-fra Canada-[1]-roste-seg selv-[2]-i boken sin. SKI-STI
18. Jegeren-på vei til Oslo-[1]-leide en bil-til seg selv-[2]-og kameraten sin. SKOG-TOG
19. Elektrikeren-med skrujernet-[1]-studerte-seg selv-[2]-nøye i speilet. STRØM-DRØM
20. Fjellklatreren-i rommet-[1]-bestilte middag-til seg selv-[2]-og de andre. TAU-SAU

*Fillers: (probe positions indicated by [*])*

21. Det-var-snekkeren-som la ut-om problemene-[*]-han hadde. OLJE-EKJELIG
22. Det-var-statsministeren-som støttet -forslaget-[*]-fra departementet. KIOSK-FLIASKO

23. Det-var-svømmeren-som hvisket-nedsettende-[*]-om konkurrentene. KAVIAR-HJERDE
24. Det-var-indianeren-som spikket-fine buer-[*]-til barna sine. PLYSTRE-ITLEN
25. Det-var-husmoren-som kom -ikke-[*]-overens. BILDE-KESPET
26. Det-var-kiropraktoren-som holdt på-å besvime-[*]-etter behandlingen. RADIO-KRUNN
27. Det-var-birøkteren-som betraktet-kunden-[*]-som en plage. STOL-NAKTE
28. Det-var-eiendomsmegleren-som plukket ut-teltplassen-[*]-for turen. BOK-SAN
29. Det-var-legen-som deklamerte-et vakkert dikt-[*]-for sykesøsteren. STILLAS-SLAMT
30. Det-var-mekanikeren-som fikk seg -en god latter-[*]-etter bilsalget. KIKKE-STROR
31. Knyttet-filmstjernene-[*]-silkeslips-på hverandre? KONFUMENT-GJERNE
32. Klorte-vinkjennerne-[*]-seg fast-på fjellhyllen? KONLUTT-STOR
33. Minte-astronautene-[*]-sjefen-på å slutte -tidlig? KØKKJEN-BESKRIVE
34. Tok-alkoholikerne-[*]-testen-på seg selv? KLION-FIASKO
35. La-bartenderne-[*]-skylden-på kollegaene -for ranet? KRETRONE-NAKEN
36. Sammenlignet-skipsrederne-[*]-klienten-med Marilyn Monroe? LYKVOM-APPELSIN
37. Lagde-glassblåserne-[*]-en gjenstand-for turistene? KENN-BILMERKE
38. Smørte-bøndene-[*]-brødskiver-med gjetost? TORFAST-REGNFRAKK
39. Skrev-munkene-[*]-et dikt-på pergamentet? PROR-TAK
40. Øvde-bokserne-[*]-lenge-på fotarbeidet? KOMER-KVELD
41. Generalen-som alle -trodde på-[*]-overbeviste-juryen-om sin uskyld. UNTATT-NELDIG
42. Buddhisten-som ingen -likte-[*]-ba-ofte-om penger. MIPE-SPITSE
43. Forskeren-som innvandreren -snakket med-[*]-anså-seg selv -som kompetent. ALIS-FIGNER
44. Feieren -som naboen -hadde hentet-[*]-skadet-armen-under trening. LABRYNT-SKYRTE
45. Syklisten-som vennene -besøkte-[*]-sølte-på jakka-under skålingen. ROLD-NEFIN
46. Gartneren-som alle -ventet på-[*]-holdt-en lang tale-om prydbusker. STÆVE-ØDEGLATT
47. Flykapteinen-som butikksjefen -kjente-[*]-prøvde-et par bukser-og en jakke. GIRS-FOSK
48. Matrosen-som jenta -hadde invitert-[*]-ba-om en pils. VOSKE-DORFER
49. Psykologen-som universitetet -hadde ansatt-[*]-beordret-assistenten-til å slutte. SMØT-GRENI
50. Vakten- som fabrikken -alltid brukte-[*]-frakoblet-alarmsystemet. KRELLE-NUNNEN
51. Veiarbeideren-som alltid- jobbet sent-ringte-for å avlyse-middagen.[*] KÅLSUR-FYLFET
52. Filosofen-som var -veldig populær -foreleste-lenge-for studentene.[*] REFFEM-SÆN
53. Ornitologene-som observerte-terner-var-ikke-sultne.[*] KATRE-KRØY
54. Spionen-som var stolt-av sitt fag-klaget-over den dårlige-forberedelsen.[*] GULENE-GÅRTE
55. Smeden-som hadde-mange oppgaver-bearbeidet-metallet-nøye.[*] SUBB-SKERV
56. Uteliggeren-som hadde-kollidert-beskyldte-sjåføren-for ulykken.[*] ENNIP-VAT
57. Urmakeren-som var-meget pertentlig-spiste-en stor middag-uten dessert.[*] HJOL-SEPO
58. Kirurgen-som var-på handletur-skaffet-parfyme-til sin kone.[*] KURRES-KLADE
59. Musikeren-som hadde-kommet sent-protesterte-mot-vedtaket.[*] GORT-GARIN
60. Servitøren-som var-fra byen-vrikket-foten-under kampen.[*] STYKRE-SNALB

## APPENDIX B: STIMULI SENTENCES AND PROBE WORDS, EXPERIMENT 2

*Experimental items*

1. Vitnet-hentet-en journalist-og-offeret-i ulykken-[1]-gjorde-også det-[2]-etter at-[3]-alt-[4]-hadde-roet seg. AVIS-LAKRIS
2. Kunden-skyldte-på tannlegen-og-eieren-av restauranten-[1]-gjorde-også det-[2]-ut ifra-[3]-papirene-[4]-i journalen. BOR-KOR
3. Jens-forsvarte-politimannen-og-Marit-som var der-[1]-gjorde-også det-[2]-ifølge-[3]-programmet-[4]-som ble sendt. BOT-ROT
4. Studenten-snakket-med vaskekonen-og-gutten-med astma-[1]-gjorde-også det-[2]-på-[3]-vei-[4]-ut-av bygget-etter skoletid. BØTTE-STØTTE
5. Den gamle damen-hjalp-postbudet-og-en tilfeldig-forbipasserende-[1]-gjorde-også det-[2]-ifølge-[3]-barna-[4]-på lekeplassen. BREV-VEV
6. Mange-stemte-på bakeren-og-de-fra vestkanten-[1]-gjorde-også det-[2]-ifølge-[3]-opptellingen-[4]-etter valget. BRØD-NØD
7. Den unge kvinnen-ringte til-veterinæren-og-tyskeren-på ferie-[1]-gjorde-også det-[2]-før-[3]-hele dagen-[4]-hadde-gått med. DYR-MYR
8. Fru Hansen-foretrakk-frisøren-og-Fru Nilsen-over gaten-[1]-gjorde-også det-[2]-om-[3]-en-[4]-tar ryktene-alvorlig. HÅR-KÅR
9. Lederen-kritiserte-brannmannen-og-kvinnene-i terapigruppen-[1]-gjorde-også det-[2]-uten at-[3]-noen-[4]-brydde seg. ILD-DILL
10. Ola-oppsøkte-presten-og-slektningen-fra Frankrike-[1]-gjorde-også det-[2]-fordi-[3]-tidspunktet-[4]-passet slik. KIRKE-KURV

11. Rapporten-refererte-til soldaten-og-en bok-som kom ut-[1]-gjorde-også det-[2]-uten at-[3]-noen-[4]-hadde reagert. KRIG-KVELD
12. Gjestene-roste-kongen-og-de andre-som så på-[1]-gjorde-også det-[2]-hvis-[3]-ukebladet-[4]-har rett. KRONE-GARDIN
13. Konene-tenkte-på gruvearbeiderne-og-funksjonæren-med bart-[1]-gjorde-også det-[2]-fordi-[3]-det-[4]-var sent-allerede. KULL-MUGG
14. Bilen-passerte-joggeren-og-bussen-fra Askøy-[1]-gjorde-også det-[2]-ifølge-[3]-mannen-[4]-som så alt. LØPE-DØPE
15. Leverandøren-bebreidet-kokken-og-selgeren-på møtet-[1]-gjorde-også det-[2]-ettersom-[3]-saken-[4]-utviklet seg. MAT-HAT
16. Gateselgeren-så-en turist-og-mannen-i butikken-[1]-gjorde-også det-[2]-ifølge-[3]-folkene-[4]-som ble spurt. REISE-MALE
17. Utbyggeren-leide-en murer-og-sjefen-i banken-[1]-gjorde-også det-[2]-før-[3]-dagen-[4]-var omme. SEMENT-SKRIBENT
18. Barna-møtte-en jeger-og-de voksne-som passet dem-[1]-gjorde-også det-[2]-innen-[3]-alle-[4]-hadde gått-hjem. SKOG-TOG
19. Vi-trengte-en elektriker-og-de som var-i nabohuset-[1]-gjorde-også det-[2]-ettersom-[3]-hele gata-[4]-var mørkelagt. STRØM-DRØM
20. Tilskueren-beskrev-fjellklatreren-og-hytteieren-nedenfor-[1]-gjorde-også det-[2]-i-[3]-løpet-[4]-av samtalen. TAU-SAU

*Fillers*

21. Det-var-læreren-[*]-i fransk-som vurderte-oppgavene-til studentene. GÆRTE-GARIN
22. Det-var-kassereren-[*]-i banken-som ødela-skrivebordet. SÆN-SEPO
23. Det-var-piken-[*]-i kafeteriaen-som smakte-på desserten. FYLFET-VAT
24. Det-var-en kineser-[*]-fra Taiwan-som oppfant-kruttet. SKERV-SNALB
25. Det-var-pensjonisten-[*]-som malte-akvareller-i fritiden. KRØY-KLADE
26. Burde-offiserene-[*]-i Kosovo-hilse-på flyktningene? SELIE-VAMDEL
27. Måtte-sangerinnen-[*]-alltid-neie-for publikum-etter forestillingen? SAPPELIG-DARM
28. Prøvde-barnebarna-[*]-å finne-brillene-til tanten? GREDNA-LØTFE
29. Skulle-eieren-[*]-av huset-filme-datterens-bryllup? FETMI-FANLE
30. Ville-Hitler-[*]-under krigen-overta-nabolandene? SLIRTE-BODRET
31. Det-var-snekkeren-som la ut-om problemene-til de andre.-[*] OLJE-EKJELIG
32. Det-var-indianeren-som spikket-en fin bue-til barna.-[*] PLYSTRE-ITLEN
33. Det-var-mekanikeren-fra Nestun-som lo-av opptrinnet-etter bilsalget.-[*] KIKKE-STROR
34. Det-var-ministeren-som støttet-forslaget-på møtet.-[*] KIOSK-FLIASKO
35. Det-var-eiendomsmegleren-som plukket ut-tomten-uten tillatelse.-[*] BOK-SAN
36. Det-var-mange-som betraktet-arbeidet-som en plage.-[*] STOL-NAKTE
37. Det-var-kiropraktoren-som ville-massere meg-på ryggen.-[*] RADIO-KRUNN
38. Det-var-innvandreren-som vant-gullmedaljen-i konkurransen.-[*] KAVIAR-HJERDE
39. Det-var-husmoren-som ikke-ville veksle-pengene.-[*] BILDE-KESPET
40. Det-var-noen-som deklamerte-et vakkert dikt-midt-på natten.-[*] STILLAS-SLAMT
41. Passasjeren-som satt-[*]-i kabinen-betalte-ikke forturen-og-ble deretter-kastet av. SPITLE-KVUTRE
42. Polfareren-som ville-[*]-komme i avisen-badet-i isvannet-og-endte opp-med koldbrann. SALGNE-MOLEDD
43. Satanisten-som alle-[*]-unngikk-hermet-etter-de store guttene-og-ble ansett-som en dust. BILK-TIGNE
44. Sjakkspilleren-som hadde-[*]-vunnet gull-luftet-hunden-til treneren-og-var bare-glad til. FOKLE-SUKLE
45. Sjarmøren-som alle-[*]-ville varte opp-elsket-sjokoladepudding-og-bestilte-mer champagne-til damene. STITLE-MAKREN
46. Skomakeren-som byen-[*]-satte pris på-beilet-til en kvinne-og-ville gjerne-gifte seg-snarest mulig. BURK-KANRE
47. Skredderen-som alle-[*]-gikk til-modellerte-en smoking-for brudgommen-og-en kjole-for bruden. SURIP-KJENRE
48. Skribenten-som de fleste-[*]-kritiserte-tenkte-på livet-og-skrev deretter-en bok-som ingen-ville kjøpe. RESDEL-SKJEVLE
49. Skurken-som barna-[*]-var redd for-fikk-fullt hus-i pokerspillet-og-vant-mange penger. LATNE-TRAPME
50. Vaktmesteren-som hadde-[*]-god tid-kostet-opp alt-søppelet-og-hev det-utenfor skolen. BLOSKDUTT-SETRERETT
51. Burde-astronauten-fra Russland-påminne-de andre-[*]-astronautene-om rutinene-eller-visste de-alt-fra før? KØKKJEN-BESKRIVE
52. Hadde-kameratene-på fisketur-laget seg-brødskiver-[*]-på forhånd-eller-kjøpte de-lunsjen-i butikken-som vanlig? TORFAST-REGNFRAKK
53. Hadde-guttungen-lagt-skylden-for fadesen-[*]-på broren-selv om-det-sannsynligvis-var-hans egen skyld? KRETRONE-NAKEN
54. Hadde-vinkjenneren-fra Tyskland-tilfeldigvis-spyttet ut-[*]-vinen-som ble-servert-eller-likte han-den ikke? KONLUTT-STOR
55. Hadde-filmstjernen-knyttet-sløyfe-før middagen-[*]-men helt-glemt-å ta-av seg-hatten-før forestillingen. KONFUMENT-GJERNE

56. Hadde-munken-skrevet-et dikt-om dyrene-[*]-eller-var det-en salme-som han-ville-skrive? PROR-TAK
57. Hadde-alkoholikeren-tatt-testen-på seg selv-[*]-fordi han-var redd-eller-var det-legen som-tok feil? KLION-FIASKO
58. Måtte-glassblåseren-på verkstedet-prøve-å lage-[*]-et glass-som ikke-kunne knuses-eller-var det-umulig? KENN-BILMERKE
59. Ville-bokseren-absolutt-trene-i fotarbeid-[*]-hele dagen-eller-ville han-heller-bare jogge-en tur? KOMER-KVELD
60. Ville-budet-i firmaet-stadig-plage-[*]-sjefen-selv om-han ikke-turde-innrømme det-til de andre? LYKVOM-APPELSIN