# **Optimal Design in Language - A research proposal**

Tanya Reinhart and Kriszta Szendrői

(Submitted to NWO, Holland, in September 2003. Not funded)

## 9. Description of the proposed research

## **1 BACKGROUND**

## 1.1 Optimal design: language competence and use

The introduction of the competence/performance distinction at the early stages of generative grammar (Chomsky 1965) inadvertently led to a trend in theoretical linguistics that ignored complexity considerations in the grammar. On this view, since the grammar is distinct from the parser, complexity problems which may slow down parsing are irrelevant for the research of linguistic competence, because the parser operates by an independent set of rules, which may bypass the computations required by the grammar. Complex grammars preclude the possibility of a *'transparent' parser*, where sentence processing simply follows the rules of the grammar, because performing the complex rules of the grammar on-line exceeds the limited capacity of the human working memory. Along with conceptual problems, this is methodologically problematic, as it means that we cannot learn anything about human linguistic competence through processing. If the parser is non-transparent, information about actual language use only gives insight into the nature of the parser.

Importantly, the problem is not with the competence/performance distinction itself, which is necessary (see Crain et al 2001). Rather, the false logical step was to conclude that processing complexity is irrelevant for competence models.

A new perspective on the relation between competence and parsing was presented in Chomsky's (2000) hypothesis of *optimal design*. The basic idea is that an optimally designed computational system (grammar), namely a system that optimally connects sound and meaning (thus satisfying the interface conditions), would turn out also optimal for the processing systems. If human language is optimally designed, then its parser is essentially transparent, with minimum parser specific rules or algorithms. This entails that any complexity in the grammar should be observable by increased processing cost—the topic of the present investigation.

#### 1.2. Reference set computation

An important idea of the nineties was that grammaticality of sentences is not always determined by absolute conditions, but it may be determined by a selection of the optimal competitor from a relevant reference set. A restricted version of this was assumed at early stages of the minimalist program (Chomsky 1993), and simultaneously, it was the central notion developed in Optimality Theory (henceforth OT) (Prince and Smolensky 1993). In the minimalist program the notion was gradually adandoned, as there is no evidence for the need of such computation at the syntax level (Chomsky, 1995, Chap.4). However, it was later found that instances of this computation must be operative at the level of the interface between syntax, semantics and context. (MP: Fox 2000; Reinhart 1998, to appear; OT: De Hoop and De Swaart 1998; Blutner 2000, Hendriks and De Hoop 2001).

In schematic terms, reference set computation (henceforth <u>RS-computation</u>) is found when a particular linguistic form  $FORM_1$  is potentially ambiguous; the rules of the interpretative system can assign to it both INTERPRETATION<sub>1</sub> and INTERPRETATION<sub>2</sub>, as in (1). However, there exists an alternative linguistic form in the language,  $FORM_2$ , which also has INTERPRETATION<sub>2</sub>, and is a more optimal way to express it. In many such cases, the association of  $FORM_1$  with INTERPRETATION<sub>2</sub> is blocked, because of the existence of the alternative  $FORM_2$ .



In terms of processing, the computation of RS-economy involves three crucial steps: FORM<sub>1</sub> is stored in working memory, then an alternative derivation is selected and formed (FORM<sub>2</sub>), and finally the two are compared under the same interpretation, INTERPRETATION<sub>2</sub>. The characteristic property of this type of RS-computation is that it's "global", namely it requires comparing two full derivations. (For details, see Collins 1997.)

This is difficult for the human parser, because it operates with limited working memory (Frazier 1978; Ferreira and Clifton 1986; Crain & Hamburger 1992; Trueswell et al 1994). The assumption shared by all processing studies (since, at least Fodor et al 1970) is that given these limitations, the human processor attempts to close constituents as soon as possible. Chunks of closed derivations are assigned some abstract representation. Opening a closed constituent to access its subparts is possible, but costly, leading (in some contexts) to a garden-path effect.

The complexity of RS-computation lies in the fact that the derivation cannot be closed and assigned an interpretation until an alternative is computed and the two are compared. We should note that RS-computation is only difficult, but not unprocessable. If there are isolated instances where it needs apply, it's possible that processing is simply slowed down in these cases. But if global RS-computation is required all around, then either nothing gets closed and eventually the overload is too big for processing, or constituents constantly close and reopen. Neither is consistent with the fact that in actual language use, sentences usually get processed smoothly.

A line developed to address this problem in the OT-framework is that the actual processing of derivations need not literally compute RS-s, rather, some heuristic strategies, or algorithms, are developed by speakers for quick assessment. (for acquisition see Pulleyblank and Turkel 1998; Tesar 1998) This, however, brings us back to the question of the transparent parser. If there are isolated instances of RS-computation, then the parser could still be essentially transparent, with minimum parsing-specific algorithms and adjustments. But if it's a wide-spread computation, then essentially we are forced to assume that the parser operates by its own algorithms, completely independent of the computational system; namely, language is not optimally designed for use.

The aim of the proposed project is to show that this interface computation does not require departure from the concept of a transparent parser. RS-computation exists in isolated areas, and in these areas there is evidence for increased processing cost. More broadly, it's often debated whether a particular interface problem requires RS-computation or not. Within the

hypothesis of optimal design and maximally transparent parser, if such computation is proposed for a given problem, then either a simple, inferable local algorithm can be defined to allow bypassing the computation, or there should be visible evidence for processing difficulties. A strong theoretical implication (to which we are committed) is that if none of these can be established, then our hypothesis that the given problem requires RS-computation is wrong.

We should note that the scope of this claim concerns only global RS-computation. Many of the computations assumed in OT are, in fact local (applying at the level of a constituent or during the derivation). Similarly, Fox (2000) argues that the minimal link condition, regulating movement, is, in fact, operative in both syntax and the interface, but it's local. There is no (known) reason to assume that local RS-computation is costly.

## 1.3. The computation of focus and quantifier scope

In Reinhart (to appear) she argues that RS-computation is involved in areas where the output of the computational system are not sufficient for the interface needs. There are four instances where theoretical investigations give substantial evidence for assuming that such computation is indeed present: Quantifier raising (Fox 2000); stress-shift for focus construal (Cinque 1993, Reinhart 1998, Szendrői 2001); pronominal coreference (Reinhart 1983), and scalar implicatures (Chierchia 2001). In the last two areas, acquisition studies already established that children have difficulties, which, it has been argued, can be interpreted as reflecting difficulties in processing the required RS-computation (Grodzinsky and Reinhart 1993; Chierchia et al 2001). The focus of the present project is the first two areas.

The view of focus identification we assume, following Cinque (1993) (and Chomsky 1971) is that the (neutral) main stress, which is assigned independently during phonetic realization, also determines directly the set of possible foci of a given derivation. The focus set associated with each derivation includes all the constituents that contain the main stress. E.g. (2a) is associated with the focus set (2b). (**bold**=main stress) In actual use, the context determines which of these possible foci is relevant.

- (2) a. [IP My neighbor [VP is building a desk ]].
  b. Focus set: {IP, VP, Object}
- (3) My **neighbor** is building a desk.

However, the independent mechanism of main-stress assignment is not always sufficient for the needs of the context-interface. E.g. the subject is not in the focus set of (2a) For this interpretation, a special mechanism, stress-shift, needs to apply, displacing main stress, and deriving (3).

My general claim is that such adjustment operations are permitted only when there is no other way to obtain the intended interpretation Deciding this requires the construction of a RS, which, for (3) includes (2). So, (3) cannot have IP-focus, because this interpretation is available also in (2a), without stress-shift.

There is an interesting similarity between stress-shift and Quantifier Raising (henceforth QR). To generate wide scope interpretation for the object in (4), QR must apply covertly (5).

- (4) A table cloth covers every table
- (5) [every table]<sub>1</sub> [a table cloth covers  $e_1$ ]

Like stress-shift, QR is superfluous for the needs of the computational system and it's only enforced by needs of the interpretation interface. Fox (2000) provides impressive evidence that it's permitted only when the desired interpretation could not be obtained otherwise: its application requires RS-computation.

The hypothesis of a transparent parser entails that there should be some visible processing cost when stress-shift an QR apply. The aim of the project is to test whether this entailment holds.

#### **2** THE DESIGN OF THE PROJECT

Evidence for processing cost of a given computation can come from two sources: direct experimentation of adult processing, where processing difficulties are witnessed by slower reaction time or e.g. eye-tracking, and experiments on children's performance. As mentioned, the specific difficulty in (global) RS-computation is the load it poses on working memory. As children's working memory is not yet fully developed (Gathercole and Hitch 1993), a costly task for adults may simply be impossible for children. Grodzinsky and Reinhart (1993) argue that the established 50%-pattern in children's performance on coreference (condition B) reflects their inability to execute the required computation. Children know (innately) what they have to do to determine the interpretation and answer the experiment's question, but since the task exceeds their working memory capacity, they give up and offer a guess.

Correspondingly, the project has two sub-projects: adult processing of stress-shift and scopeshift, and children's processing. The second is more difficult, for reasons explained below, so it will be carried out by a post-doc; the first is a dissertation project (requiring also theoretical research on QR).

As an added value, the combined results will shed light on the development of the human ability to handle RS-computation.

Both subprojects will examine both stress-shift (focus) and scope-shift (QR). The theoretical interest in combining these two is not just their similarity, but also their potential difference: As mentioned in 1.2., RS-computation can either be bypassed in case a simple algorithm exists, or it applies with visible processing cost.

In the case of stress-shift, it's possible to define a bypassing algorithm. In an overwhelming majority of cases (e.g. in 3), the outcome of RS-computation is that the only possible focus is the element bearing shifted-stress (because all other alternatives can be obtained also without stress-shift). The algorithm (6) is therefore statistically motivated.

(6) If stress-shift applied, then the terminal bearing main stress is the focus of the utterance.

But it will sometimes fail: RS-computation allows focus=DP-subject interpretation (underlined) in (7b), because this focus cannot be obtained without stress-shift. The algorithm (6) fails to account for this, assigning focus to *hat*.

- (7) a. Who committed the murder?
  - b. <u>The man in the red **hat**</u> committed the murder.

In principle, then, it's possible that no processing cost will be found for cases like (3), but only in (7b), where the algorithm fails and RS-computation must apply. If the experiments show no processing cost even in (7b), we will have to conclude that my analysis of stress-shift is mistaken, and it does not involve RS-computation.

With QR, by contrast, it's not possible to define an algorithm that can bypass processing the question whether the same interpretation could not be obtained without QR. If the experiments do not find processing cost for derivations with QR, this would falsify my analysis of QR, and will confirm Fox's (2000) analysis, in which QR involves local, rather than global RS-computation.

For the study of stress-shift, the basic language for the experiments will be Dutch. For quantifier scope, it must be English, because it's known that specific restrictions apply in Dutch, e.g. singular indefinite subjects are always specific and thus resist narrow scope (Reuland 1988, Rullman 1989; vanden Wyngaerd 1992).

#### **SUBPROJECT 1: Adult processing of reference set computation**

### Objective 1: Focus

**Experiment 1** provides a baseline for further experiments.

The method is a <u>bi-modal truth-value judgement task</u>. The subjects see a picture on a screen and hear an utterance involving stress-shift (cf.9) or neutral stress (cf.8) describing the picture correctly or incorrectly depending on the stressing of the utterance. The subject has to decide whether the utterance is a true statement about the picture or not. <u>Response time</u> will be measured. (All experimental materials are given in Table 1, Appendix.)

In the same session, <u>eye fixations</u> on the visual stimuli are measured (following technique of Cooper 1974; Eberhard et al 1995). Further evidence for the presence of RS-computation in parsing comes form characteristic eye-fixations revealing the computation of the alternative derivation with neutral stress. UiL OTS has adequate equipment and technical expertise.

**Experiment 2**: We employ the <u>dual task method</u> (Shapiro et al 1987, 1989): subjects get an additional task (Task A) that is known to use the same resources. Task B involves either neutral or shifted stress. Such a method can detect more subtle processing differences by 'magnifying' the effect by the presence of the additional task.

#### **Objective 2: Quantifier raising**

The task under consideration is to determine whether the object in (I)-(III) can take scope over the subject by QR. The intended reading is given in parentheses, with its implication for

the topmost number allowed for the set denoted by the subject. (In the experiments, the pictures illustrate this reading).

(I)	a.	A tablecloth covers every table.	(Up to as many tablecloths as tables)
	b.	A doctor will examine every patient.	(Up to as many doctors as patients).
(II)	a.	A tablecloth covers two tables.	(Up to two tablecloths)
	b.	A doctor will examine ten patients.	(Up to ten doctors)
(III)	a.	Two doctors will examine ten patient	ts. (Up to twenty doctors)
	b.	Three men lifted two tables.	(Up to six men)

In the semantic literature, (I) was judged as fully possible, (III) as utterly impossible, and (II) is subject to judgment-debates. Reinhart (to appear) argued that this grading corresponds to the size of the RS involved in the computation. To determine whether the object is allowed to undergo QR, it's necessary to check whether the same interpretation could not be derived without QR. This requires listing in the RS **all** possible scope construals of the same derivation without QR. Applying rigorous counting, (I) has two members in the RS, (II) has three, and (III) has five. Assuming a transparent parser, more complex computation comes with slower processing, up to the point when the computation is so complex that it exceeds human processing capability.

There are other factors influencing the availability of the inverse reading including pragmatics, world knowledge, lexical differences (Ioup 1975), discourse considerations (Krifka 1998), and even syntactic differences. Crucially however, we can assume that these factors have a similar effect on judgements of examples from each levels of the scale.

The availability of a three-level complexity scale for RS-computation allows for robust, unambiguous predictions. Examples higher on the scale should result in longer response time. In addition, increase in processing load should manifest itself in a greater number of incorrect NO judgements. On the hypothesis (based on off-line judgements) that (III) is unprocessable with wide scope, a NO answer is expected in this case after the computation is attempted and aborted. No prediction for response time is made for (III): subjects may take a long time to abort the computation or abort it straightaway when they feel overwhelmed. The expected result for the experiments is depicted in Figures 1-2 (0=example without RS-computation).<sup>1</sup>



Figure 1: Expected results 1

Figure 2: Expected results 2

<sup>&</sup>lt;sup>1</sup> As for **Experiment 1**, cases with stress-shift correspond to level (I) in the QR-experiment, while neutral stress involves no RS-computation (i.e. 0).

As working memory capacity is known to vary between people, individual subjects' data will also be analysed. <u>Independent cognitive measures of working memory capacity</u> will be applied for each subject (Daneman and Carpenter 1980; Stowe et al. 1998). It's a formidable advantage of the existence of the scale that results of individuals with working memory differences can be predicted. Figures 1-2 depict the expected results of adults with average (or higher) working memory capacity. Individuals with lower capacity will attempt but fail to carry out the computation even in easier cases i.e. (II), sometimes even (I), and are thus predicted to exhibit computational failure in these cases too: a NO-judgement after an unpredictable length of time.

**Pretests:** A <u>series of grammaticality judgement surveys</u>, with English native speakers, will be conducted with many sentences of levels (I)-(III), to determine the effects of scale-independent factors.

**Experiment 3**: A <u>bi-modal truth value judgement task</u> measuring response time will be performed (cf. Experiment 1), with test items from **Pretests**. The inverse scope reading of the test stimulus is depicted in a picture presented on a screen while test stimulus is simultaneously presented auditorily. (The picture is incompatible with the overt scope reading.) The controls are the same examples with overt scope. The <u>eye-tracking patterns</u> will also be analyzed, to determine any preferential scanning patterns while subjects perform the truth-value judgement task.

Due to the high number of factors involved, a large number of subjects will be tested.

#### SUBPROJECT 2: CHILD PROCESSING OF REFERENCE SET COMPUTATION.

Following the literature, children aged 4-8 will be tested. The focus experiments will involve Dutch native speakers, but the scope-shift experiments will be primarily with English speakers for reasons given in Section 2.

#### Objective 1: Focus

(8)-(9) were tested in a truth-value judgement task (Gualmini et al 2002). (8) was false in the context, where Barney also sold a cake to Winnie. (9) was true: the only thing Barney sold to Snow-White was a banana, but he also sold a banana and a cake to Grumpy.

- (8) Barney only sold a cake to SNOW-WHITE.
- (9) The farmer only sold a BANANA to Snow-White.

The neutral stress results were adult-like. The stress-shift condition results appear consistent with the expected guessing pattern. So, initial results show degraded comprehension of stress-shift cases with *only* (Halbert et al 1995, Gualmini et al 2002) and Dutch *alleen* (Szendrői in preparation) in first language acquisition. However, interpreting these results is difficult, because of the interaction with a VP-default interpretation found independently in the acquisition of *only* (Crain et al 1994). A NO answer to (9) (also to 8) is consistent with a VP-default: Selling a banana to Snow-White is not the only thing the farmer did.

To decide whether the results indicate guessing or application of the default we need to study individual performance. **Experiment 4** will replicate the Gualmini-study using *alleen* 'only'. The alternative to guessing is that children with adult-like answers have a larger working memory than children who resort to the VP-default. So, standardised working memory tests will be applied to determine the working memory capacity of the individual children (Gathercole 1999; Adams and Gathercole 2000; Gathercole and Pickering 2000; also work by *Dyslexia*-Group, UiL OTS e.g. De Bree et al 2003).

However, establishing the presence of RS-computation in stress-shift requires checking it in other contexts, and, thus, abstracting away from the VP-default interpretation of *only*. Most experiments will use <u>negation contexts for stress-shifted foci</u>: negation reverses subset-superset relations, so in contrast to **Experiment 4**, in **Experiments 5a-d** children's default focus interpretation is predicted to be (in)direct object (narrowest possible), rather than VP (widest possible).

As Dutch negation and indefinites 'merge' (i.e. geen≅niet+een), the scope of negation in this case is determined independently from prosody. To avoid this, definites will be used. This introduces a new factor of word order: scrambling (Schaeffer 1995). The test items will include both scrambling and non-scrambling orders and PP objects that resist scrambling.

Two techniques will be employed: <u>picture selection task</u>, which determines the preferred interpretation that children assign to test items, and <u>truth-value judgement task</u>, determining all interpretations children allow.

### Objective 2 - Resolving scope shift

Experiment 6, will duplicate Experiment 3, adopted to acquisition experimental setting.

**Experiment 3** asks for truth/falsity judgments (NO: overt scope; YES: inverse scope). It's widely accepted that children have a bias for YES unless the preceding discourse clarifies why a truth/falsity judgement is expected of them (condition of plausible dissent Crain et al 1996). So, the test utterance has to be preceded by an appropriate context story. We will also analyze the eye-tracking patterns, to determine any preferential scanning patterns during performance of the task.

Note that QR is more difficult to test for at least two reasons. First, children exhibit what is known as quantifier spreading. They reject utterances such as *Every boy is riding an elephant* in contexts where adults would judge it to be true, e.g. three boys are riding three elephants and there is an additional elephant (Inhelder and Piaget 1964; Philip 1995; cf. Crain et al 1996). The different suggestions from the literature for the reasons of quantifier spreading will be evaluated, the experimental designs will take these into consideration.

Next, many potentially intervening factors have to be adequately controlled for. Alongside the factors mentioned in **Experiment 3**, child-specific factors include aversion to wide scope indefinites (Krämer 2001, cf. Musolino 1998); lexical differences (Brooks and Braine 1996); syntactic differences (Schaeffer 1995). These necessitate greater care (and theoretical research) in selecting the experimental sentences, and **pretests**, designed to isolate interfering factors.

Individual data will be analysed, and standardised working memory tests will be applied to determine the working memory capacity of the children (see objective 1). The prediction is that children will follow the same pattern as adults (Figures 1-2), but overall do worse than adults (i.e. slower response time, more incorrect answers).

APPENDIX: REFERENCES AND TABLE 1: EXAMPLES OF EXPERIMENTAL MATERIALS

- Adams, A.-M.and S.E. Gathercole (2000). Limitations in working memory: implications for language development. *International Journal of Language Communication Disorders* 35 (1), 95-116.
- De Bree, Elise, Carien Wilsenach, Ellen Gerrits & Frank Wijnen (2003). In search for precursors of dyslexia. Phonological processing and working memory in children at risk for dyslexia and children with SLI. Poster presented at the International Child Phonology Conference, Vancouver, Canada 1-4 July, 2003.
- Blutner, R. (2001) Some aspects of optimality in natural language interpretation. *Journal of Semantics* 17, 189-216.
- Brooks, P.J. and M. Braine (1996). What do children know about the universal quantifiers *all* and *each? Cognition* 60, 235-268.
- Chierchia, Gennaro (2001). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. Ms. University of Milano.
- Chierchia, G., S. Crain, M.T. Guasti, A Gualmini, and L. Meroni (2001). The Acquisition of Disjunction: Evidence for a Grammatical View of Scalar Implicatures. *BUCLD* 25. Cascadilla Press: Sommerville.
- Chomsky, N. (1965). Aspects of the theory of syntax. Cambridge, Mass.: MIT Press.
- Chomksy, N. (1971). Deep structure, surface structure and semantic interpretation. In Semantics: An interdisciplinary reader, eds. D.D. Steinberg and L.A. Jakobovits, 183-216. Cambridge: Cambridge University Press.
- Chomsky, Noam (1993) "A Minimalist Program For Linguistic Theory" in K. Hale & S. J. Keyser (eds) The View from Building 20, p. 1-52. Cambridge: The MIT Press,
- Chomsky, N. (1995). The Minimalist Program. Cambridge, Mass.: MIT Press.
- Chomsky, N. (2000). Minimalist inquiries: the framework, In R. Martin, D. Michaels and J. Uriagereka (eds.) *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*. Cambridge: MIT Press.
- Cinque, Guglielmo. 1993. A null theory of phrasal and compound stress. *Linguistic Inquiry* 24: 239-267.
- Collins, C. (1997). Local economy. Cambridge, Mass.: MIT Press.
- Cooper, R.M. (1974). The control of eye-fixation by the meaning of spoken language. *Journal of Cognitive Psychology* 6, 84-107.
- Crain, S. & Hamburger, H. (1992). Semantics, knowledge and NP modification. In R.Levine ed. *Formal grammar, theory and implementation*. U. of British Columbia Press, 372-401;
- Crain, S., W. Ni & L. Conway (1994). Learning, parsing and modularity. In C.Clifton, L.Frazier & K.Rayner eds. *Perspectives on sentene processing*. Lawrence Erlbaum, 443-467.
- Crain, S., R. Thornton, C. Boster, L. Conway, D. Lillo-Martin and E. Woodams (1996). Quantification without qualification. *Language Acquisition* 5, 83-153.
- Crain, S., W. Ni and D. Shankweiler (2001). Grammatism. Brain and Language 77, 294-304.
- Daneman, M. and P.A. Carpenter (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19, 450-466.
- Eberhard, K., S. Spivey-Knowlton, J. Sedivy and M. Tanenhaus (1995). Eye movements as a window into real-time spoken language processing innatural contexts. *Journal of Psycholinguistic Research* 24, 409-436.
- Ferreira, F., and C. Clifton (1986). The independence of syntactic processing. *Journal of Memory & Language*, 25(3), 348-368.
- Fodor, J., T. Bever and M. Garrett (1974) The psychology of language, McGraw Hill.

Fox, Danny (2000) Economy and semantic interpretation. Cambridge, Mass.: MIT Press.

- Frazier, L. (1978) On comprehending sentences: Syntactic parsing strategies. PhD Dissertation, University of Connecticut.
- Gathercole, S. (1999) Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences* 3(11)410-19.
- Gathercole, S., G. Hitch, (1993). Developmental changes in short-term memory: a revised working memory perspective. In Collins, A. F., Gathercole, S. E., Conway, M. A., Morris, P. E. (eds.), *Theories of Memory*. Hove: Lawrence Erlbaum, 189-211.
- Gathercole, S.; S. Pickering, (2000). Assessment of working memory in six- and seven-year-old children. *Journal of Educational Psychology* 92(2), 377-390.
- Gennari, S., A. Gualmini, S. Crain, L. Meroni, and S. Maciukaite (2001). How Adults and Children Manage Stress in Ambiguous Contexts. Proceedings of 1st Workshop on Cognitive Models of Semantic Processing. University of Edinburgh, Edinburgh, UK.
- Grodzinsky, Yosef. and Tanya Reinhart (1993), 'The innateness of binding and coreference' *Linguistic Inquiry* 24(1): 69-101.
- Gualmini, A., S. Maciukaite and S. Crain (2002). Children's insensitivity with contrastive stress with *only. Penn Working Papers in Linguistics* 9(1). University of Pennsylvania.
- Halbert, A.M., S. Crain, D. Shankweiler and E. Woodams (1995). Interpretative use of contrastive stress by children. Poster presented at CUNY 8, March 16-18, 1995, Tucson, AZ.
- Hendriks, Petra and Helen de Hoop (2001). Optimality theoretic semantics. *Linguistics and Philosophy*, 24(1): 1-32.
- de Hoop, Helen and Henriette De Swaart (1998). *Temporal adjunct clauses in Optimality Theory*. Ms. UU and KUN
- Inhelder, B. and J. Piaget (1964). *The early growth of logic in the child*. London: Routledge and Kegan Paul.
- Ioup, G.L. (1975). The treatment of quantifier scope in a transformational grammar. PhD dissertation. CUNY.
- Krämer, I. (2001). Interpreting indefinites. Phd dissertation, MPI Nijmegen.
- Musolino, J. (1998). Universal grammar and the acquisition of semantic knowledge: an experimental investigation into the acquisition of quantifier-negation interaction in English. University of Maryland, College Park.
- Krifka, M. (1998). Scope inversion under the rise-fall pattern in German. *Linguistic Inquiry* 29.1, 75-112.
- Philip, William (1995). Event quantification in the acquisition of universal quantification. PhD dissertation. UMass, Amherst.
- Prince, Alan and Paul Smolensky (1993). Optimality Theory: Constraint interaction in generative grammar. Technical Report RuCCS Technical Report #2, Center for Cognitive Science, Rutgers University, Piscataway.
- Pulleyblank, Douglas and William J. Turkel. 1998. The logical problem of language acquisition in optimality theory. In Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis, and David Pesetsky (eds.) *Is the best good enough? Optimality and competition in syntax*, Cambridge, MA: MIT Press and MIT Working Papers in Linguistics, 399-420.
- Reinhart, T. (1983). Anaphora and semantic interpretation. London, Sydney: Croom helm.
- Reinhart, T. (1998). Wh-in-situ in the framework of the minimalist program, *Natural Language Semantics*.
- Reinhart, T. (to appear). *Interface strategies*. Linguistic Inquiry Monographs. Cambridge, Mass.: MIT Press.
- Reuland, E. (1988). Indefinite subjects. NELS 18, 375-394. Amherst: GLSA.
- Rullmann, H. (1989). Indefinite subjects in Dutch. In E.Bach, A. Kratzer and B. Partee (eds.), Papers on quantification. Amherst, GLSA, 307-328.
- Schaeffer, J. (1995). On the acquisition of scrambling in Dutch. BUCLD 19, Cascadilla Press: Sommerville, 521-532.
- Shapiro, L.P., E. Zurif, & J. Grimshaw. (1987). Sentence processing and the mental representation of verbs. Cognition, 3, 219-246.

- Shapiro, L.P., E. Zurif, & J. Grimshaw. (1989). Verb processing during sentence comprehension: Contextual impenetrability. *Journal of Psycholinguistic Research*, 18, 223-243.
- Stowe, L.A., R. Withaar, A. Wijers, C. Broere and A. Paans (1998). *Encoding and storage in working memory during sentence comprehension*. CUNY 11, New Brunswick, NJ.
- Szendrői, K. (2001). Focus and the syntax-phonology interface. PhD dissertation, University College London.
- Szendrői, K. (in preparation). Narrow and wide focus interpretation in the acquisition of onlysentences. Ms. UiL OTS, Utrecht.
- Tesar, Bruce (1995). *Computational Optimality Theory*. PhD dissertation. University of Colorado: Boulder.
- Trueswell J. C., Tanenhaus M. K. and Garnsey S. M. (1994). Semantic Influences On Parsing: Use of Thematic Role Information in Syntactic Ambiguity Resolution. *Journal of Memory and Language* 33(3), 285-318
- vanden Wyngared, G. (1992). Een reviewer zal elk abstract nalezen. [A reviewer will read each abstract.] *Tabu* 22(1).

Table 1	EXAMPLES OF EXPERIMENTAL MATERIAL					
	eutral stress sentences (control) Stress-shift sentences					
Experiment 1a:	Subjects: adults; Language: Dutch; Material: alleen 'only'; Method: RT+eye-tracking					
	Hij heeft alleen een taart aan MICKEY	Hij heeft alleen een TAART aan Mickey				
Experiment 2a:	verkocht.	verkocht.				
(dual-task)	he has only a cake to Mickey sold	he has only a cake to Mickey sold				
Experiment 1b:	Subjects: adults; Language: English; Material: only+long stress-shift; Method: RT+eye-tra					
Experiment 2b:	The man with the HAT committed the	The man with <u>the HAT</u> committed the murder.				
(dual task)	murder. (i.e. not the woman with the poodle)	(i.e. not the man with the wig)				
Experiment 4:	Subjects: children; Language: Dutch; Material:	alleen 'only'; Method: truth-value judgement				
	Hij heeft alleen een taart aan MICKEY	Hij heeft alleen een TAART aan Mickey				
	verkocht.	verkocht.				
	he has only a cake to Mickey sold	he has only a cake to Mickey sold				
Experiment 5a:	Subjects: children; Language: Dutch; Material: niet 'not' +PP; Method: truth-value judgement					
	Bert wil niet aan de LERAAR denken.	Bert wil niet aan de leraar DENKEN.				
	Bert wants not of the teacher think-to	Bert wants not of the teacher think-to				
Experiment 5b:	Subjects: children; Language: Dutch; Material: niet 'not' +no scrambling; Method: TV jud					
	Hij wil niet het boek aan TOMMIE geven.	Hij wil niet het BOEK aan Tommie geven.				
	he wants not the book to Tommie give-to	he wants not the book to Tommie give-to				
Experiment 5c:	Subjects: children; Language: English; Material: not; Method: truth-value judgement					
	He doesn't want to give a book to TOMMIE.	He doesn't want to give a BOOK to Tommie.				
Experiment 5d:	Subjects: children; Language: Ducth; Material: niet 'not' +scrambling; Method: TV judgement					
	Hij wil het boek niet aan TOMMIE geven.	Hij wil het BOEK niet aan Tommie geven.				
	he wants the book not to Tommie give-to	he wants the book not to Tommie give-to				
	Overt scope interpretation (control)	Scope shift interpretation (test item)				
Experiment 3:	Subjects: adults; Language: English; Material: (1), (2), (3); Method: RT+ eye-tracking					
	A tablecloth covers every table. (separate	A tablecloth covers every table. (one cloth)				
	cloth for each table)					
	A tablecloth covers two tables. (2 cloths)	A tablecloth covers two tables. (one cloth)				
Two men lifted three tables. (6 men)		Two men lifted three tables. (2 men)				
Experiment 6:	Subjects: children; Language: English; Material: (1), (2), (3); Method: RT+ eye-tracking					
	A tablecloth covers every table. (separate	A tablecloth covers every table. (one cloth)				
	cloth for each table)					
	A tablecloth covers two tables. (2 cloths)	A tablecloth covers two tables. (one cloth)				
	Two men lifted three tables. (6 men)	Two men lifted three tables. (2 men)				

#### REPLIES TO REVIEWERS COMMENTS (OCT 2003)

We are grateful for all the reviewers for the useful and stimulating comments. In this reply we will concentrate on the comments of the two somewhat critical reviews. Those of reviewer 3 and 4 (henceforth R3 and R4).

The proposal lies at the heart of a current theoretical debate between the Chomskvan program and Optimality Theory (OT), regarding the nature of syntactic computations and the relations of syntax and the parser. This is reflected in the comment of reviewer 3 (R3) that "the transparency of the parser claim... flies in the face of significant and increasing evidence from wide-ranging studies by so-called OT syntacticians both in the US and Europe that syntax is indeed an optimizing system. The proposal simply dismisses these as irrelevant..." Quite to the contrary, rather than dismissing OT claims, our aim in the proposed research is to switch mode from a purely theoretical debate, where each side accuses the other of ignoring their theoretical evidence, into an empirical question. In OT, all syntactic computations involve the type of reference-sets exemplified in the proposal (and which we view as a restricted interface phenomenon). If processing cost is indeed found in the planned experiments, it would be extremely difficult to explain why they are lacking in all other instances where OT assumes equally complex computations. If, however, no processing cost is found, this provides support to the OT claim that the parser operates by its own algorithms, bypassing the computational complexity of syntax.

R4's concrete comments are directed at the experiments of Subproject 1- adult processing... R4 points out correctly that the description of the experiments is not very detailed, with "just over 3 lines" devoted to one of them. This, alas, is an unavoidable consequence of the strict restriction on the size of the proposal. Given more space now, we can provide more details, which answer R4's queries.

A fully designed pilot for experiment 1 has been prepared by Dr Kriszta Szendroi, Dr Iris Mulders (eye-tracking technician, UiL OTS), Dr Ignace Hooge (cognitive scientist, Psychometrie, UU), with Dr Frank Wijnen. The experiment requires a true/false decision from the subjects when they hear a sentence and look at a picture on a screen. The experiment simultaneously measures response time and eye movement–a standard procedure.<sup>2</sup>

R4 argues that "it is not clear how the semantic implications of specific focus interpretations will be presented in a static picture" (p1 final paragraph). We follow the method of Gennari et al, who also used static pictures. In addition, like in Gualmini et al's study, each picture is preceded by a short explanatory story, e.g. '*The waiter has just delivered some food for the man, the girl and the boy. Then he went to look after the other customers. Look what he brought!*' The picture, then, represents the set of conditions (or semantic implications) needed for the focus computation.

The team designed a concept for the visual stimuli suitable for adults, illustrated in Figure 1 (see overleaf). Examples for the auditory stimuli corresponding to Figure 1 (CAPS represent main stress) are given in (1) and (2):

(1)	Hij	heeft	alleen	appels	aan	de JONGEN	gegeven.		
(2)	Hij	heeft	alleen	APPELS	aan	de jongen	gegeven.		
	he	has	only	apples	to	the boy	given		
	'He only gave apples to the boy.'								

<sup>&</sup>lt;sup>2</sup> See e.g.: Sedivy, J., M. Tanenhaus, C. Chambers, G. Carlson (1999). Acieving incremental semantic interpretation through contextual representation. Cognition 71, 109-147

Sentence (1) is false, given the situation in Figure 1, and sentence (2) (with stress shift) is true. We expect adults (unlike children) to reach the correct answer, but the experiment measures the amount of processing involved, with the theoretical hypothesis being that (2) involves more steps, and thus takes longer to decide.



Figure 1: Visual stimulus reproduced from pilot for Experiment 1

Of the two measurement methods we use, eye-tracking is relatively new. Along with R4, R1 points out as well that this is the most difficult aspect of the experiments (P2, paragraph 1). Nevertheless, our pilot explicitly defines experimental expectations (taking into account Gennari et al's considerations, mentioned by R4, but we believe, also overcoming serious short-comings of that initial study). It has been previously established that in eye tracking experiments subjects fixate on the images corresponding to the referents,<sup>3</sup> and the experiments regard the order and duration of fixation. Our starting assumption is that the referents selected reflect the elements considered in the focus computation. Subjects will fixate on: A. the image corresponding to the *stressed word* (e.g. the boy, for sentence (1)); B. the second referent mentioned, which provides the anchor for search (the boy's apple, for sentence (1)); C. the *falsifying entity* (For (1) this is the man's apple in Figure 1, whose presence falsifies this sentence in this context.). For sentence (2), the stressed-word referent is the boy's apple, the second referent is the boy, and the falsifying entity is the boy's empty tray (because if it were not empty, this would falsify (2)). While this is expected to be the pattern in all focus tasks, the specific prediction of the proposed research (and, thus, the gist of the experiment) is that when reference set computation is involved, as in (2), this will be reflected in the choice of referents fixated on. In processing input (2), with a stress shift, we

<sup>&</sup>lt;sup>3</sup> Cooper, R.M. (1974). The control of eye-fixation by the meaning of spoken language. J. of Cognitive Psych. 6, 84-107.

Eberhard, K., S. Spivey-Knowlton, J. Sedivy and M. Tanenhaus (1995). Eye movements as a window into real-time spoken language processing innatural contexts. *Journal of Psycholinguistic Research* 24, 409-436.

expect to find also fixation on the falsifying entity of sentence (1), namely, the man's apple, which is, in fact, irrelevant for the truth assessment of (2). This is because it is my hypothesis that the interpretation of (2) involves the construction of (1) and a comparison of (1) and (2). By contrast, we do not expect to find, during the processing of (1), fixations that are characteristic to (2), e.g. on the boy's empty tray. This is because, according to my hypothesis, the interpretation if (1) does not require comparison with (2).

R4 also expects clarifications regarding the coordination of the picture and the sound stimulus (p1 last line). Here we follow Gennari et al and present first the picture (e.g. Figure 1) for 800 ms and then, after a 500ms pause, the picture and the sound simultaneously (e.g. Figure 1 and sentence (1)). The main reason for showing the picture alone first is that if subjects are unfamiliar with the picture, they may start examining it without paying attention to the sound stimulus. Nevertheless, in the pilot, we measure the eye-tracking patterns from the first onset of the picture, to reveal if a picture contains an item drawing specific attention.

R4 is rightly concerned about "where the timing of response times will begin" (p1 last line). In fact, however, his concern here is based on what we believe is a mistake in Gennari et al's study. They measured RT from the offset of the utterance, thereby biasing the results in favour of the null hypothesis (a Type II error)). In contrast, in our experiment, RT is measured from the onset of the stressed word, i.e. *JON* in (1), *APP* in (2), because it is only at this point that focus processing can start. Note, however, that care should be taken that this does not introduce a counter-bias. For this reason, the crucial stimuli will involve subject foci of the type listed in experiment 1b of the proposal's Appendix, namely, comparing neutral stress within the subject (*Only the tall MAN...*), with stress shift (*Only the TALL man...*). Such stimuli involve a minimal difference (only one syllable) between the positions of the main stress in the sentences.