Linguistik Aktuell Linguistics Today 175

Experimental Pragmatics/ Semantics

Edited by Jörg Meibauer Markus Steinbach

John Benjamins Publishing Company

Experimental Pragmatics/Semantics

Linguistik Aktuell/Linguistics Today (LA)

Linguistik Aktuell/Linguistics Today (LA) provides a platform for original monograph studies into synchronic and diachronic linguistics. Studies in LA confront empirical and theoretical problems as these are currently discussed in syntax, semantics, morphology, phonology, and systematic pragmatics with the aim to establish robust empirical generalizations within a universalistic perspective.

General Editors

Werner Abraham University of Vienna / Rijksuniversiteit Groningen Elly van Gelderen Arizona State University

Advisory Editorial Board

Josef Bayer University of Konstanz

Cedric Boeckx ICREA/Universitat Autònoma de Barcelona

Guglielmo Cinque University of Venice

Liliane Haegeman University of Ghent

Hubert Haider University of Salzburg

Terje Lohndal University of Maryland Christer Platzack University of Lund

Ian Roberts Cambridge University

Lisa deMena Travis McGill University

Sten Vikner University of Aarhus

C. Jan-Wouter Zwart University of Groningen

Volume 175

Experimental Pragmatics/Semantics Edited by Jörg Meibauer and Markus Steinbach

Experimental Pragmatics/Semantics

Edited by

Jörg Meibauer Gutenberg University Mainz

Markus Steinbach Georg-August University Göttingen

John Benjamins Publishing Company Amsterdam/Philadelphia



The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Experimental pragmatics/semantics / edited by Jörg Meibauer, Markus Steinbach.

p. cm. (Linguistik Aktuell/Linguistics Today, ISSN 0166-0829 ; v. 175) Includes bibliographical references and index.

1. Pragmatics. 2. Semantics. 3. Reference (Linguistics) I. Meibauer, Jörg. II. Steinbach, Markus.

P99.4.P72.E95 2011 401².45--dc22 ISBN 978 90 272 5558 7 (Hb ; alk. paper) ISBN 978 90 272 8715 1 (Eb)

2011000205

© 2011 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ме Amsterdam · The Netherlands John Benjamins North America · P.O. Box 27519 · Philadelphia ра 19118-0519 · USA

Table of contents

Preface	VII
List of contributors	IX
Introduction: Experimental research at the pragmatics/semantics interface Jörg Meibauer & Markus Steinbach	1
The development of conversational competence in children with Specific Language Impairment <i>Robert M. Kurtz & Ronnie B. Wilbur</i>	19
The impact of literal meaning on what-is-said Frank Liedtke	43
Discourse under control in ambiguous sentences Vincenzo Moscati	63
Pragmatic children: How German children interpret sentences with and without the focus particle <i>only</i> <i>Anja Müller, Petra Schulz & Barbara Höhle</i>	79
Adult response uniformity distinguishes semantics from pragmatics: Implications for child language <i>Leah R. Paltiel-Gedalyovich</i>	101
Numerals and scalar implicatures Daniele Panizza & Gennaro Chierchia	129
Meaning in the objects <i>Katharina J. Rohlfing</i>	151
Blocking modal enrichment (<i>tatsächlich</i>) Hans-Christian Schmitz	177
<i>The hepatitis called</i> : Electrophysiological evidence for enriched composition <i>Petra B. Schumacher</i>	199
The role of QUD and focus on the scalar implicature of <i>most Arjen Zondervan</i>	221
Index	239

Preface

This volume is a collection of original papers on topical issues concerning various empirical and theoretical aspects of linguistic meaning. Its aim is to advance the current debate among theoretical and experimental linguists on the interface between pragmatics and semantics. All articles are versions of presentations given at the workshop "Experimental Pragmatics/Semantics", held in February 2008 at the University of Bamberg, Germany. The workshop was part of the Annual Meeting of the German Linguistic Society (DGfS).

We would like to thank the authors for their contributions to this volume and for their cooperativeness during the reviewing and publication process. We would also like to thank Werner Abraham and Elly van Gelderen for their helpful comments and for the opportunity to publish this volume in their Linguistik Aktuell/ Linguistic Today series. We are grateful to Sina Schade and Sven Müller in helping us with the preparation of the typescript. And last but not least we thank Kees Vaes for his patience and his professional editorial assistance.

> Mainz & Göttingen, November 2010 Jörg Meibauer & Markus Steinbach

List of contributors

Jörg Meibauer

Johannes Gutenberg-Universität Deutsches Institut 55099 Mainz Germany E-mail: meibauer@uni-mainz.de

Markus Steinbach

Georg-August-Universität Seminar für Deutsche Philologie Käte-Hamburger-Weg 3 37073 Göttingen Germany E-mail: markus.steinbach@phil.uni-goettingen.de

Robert M. Kurtz and Ronnie B. Wilbur Purdue University Linguistics Program 500 Oval Drive West Lafayette, IN 47907-2038 United States E-mail: kurtz@purdue.edu E-mail: wilburg@purdue.edu

Frank Liedtke Universität Leipzig Institut für Germanistik Beethovenstr. 15 04107 Leipzig Germany E-mail: liedtke@uni-leipzig.de

Vincenzo Moscati

University of Siena CISCL - Centro Interdipartimentale di Studi Cognitivi sul Linguaggio III piano, Complesso S. Niccolò Via Roma, 56 53100 Siena Italy E-mail: moscati@unisi.it

Anja Müller and Petra Schulz

Johann Wolfgang Goethe-Universität Institut für Psycholinguistik und Didaktik der deutsche Sprache Am Grüneburgplatz 1 60323 Frankfurt am Main Germany E-mail: P.Schulz@em.uni-frankfurt.de E-mail:AnjaMueller@em.uni-frankfurt.de

Barbara Höhle University of Potsdam

Department of Linguistics Karl-Liebknecht-Str. 24-25 14415 Potsdam Germany E-mail: hoehle@rz.uni-potsdam.de

Leah R. Paltiel-Gedalyovich Ben-Gurion University of the Negev Derech Ramon 14/6 Mizpeh Ramon 80600 Israel E-mail: leah.gedalyovich@gmail.com

Daniele Panizza

Department of Cognitive Science University of Trento Corso Bettini n. 31 38068 Rovereto (TN) Italy E-mail: daniele.panizza@gmail.com

Gennaro Chierchia

Department of Linguistics 316 Boylston Hall Harvard University Cambridge, MA 02138 United States E-mail: chierch@fas.harvard.edu

Katharina J. Rohlfing

Technische Fakultät Universität Bielefeld AG Angewandte Informatik Postfach 100 131 33501 Bielefeld Germany E-mail: rohlfing@techfak.uni-bielefeld.de

Hans-Christian Schmitz

Frauenhofer FIT.ICON Schloss Birlinghoven 53757 St. Augustin Germany E-mail: hans-christian.schmitz@fit.fraunhofer.de

Petra B. Schumacher

University of Mainz Department of English and Linguistics Jakob-Welder-Weg 18 55099 Mainz Germany E-mail: petra.schumacher@uni-mainz.de

Arjen Zondervan

Utrecht Institute of Linguistics OTS (UiL-OTS) Janskerkhof 13A room 0.05 3512 BL Utrecht Netherlands E-mail: arjen.zondervan@let.uu.nl

Introduction

Experimental research at the pragmatics/semantics interface

Jörg Meibauer & Markus Steinbach Mainz University/Göttingen University

Focusing on the semantics/pragmatics divide, the contribution of experimental pragmatics to pragmatic theory is discussed from a number of angles, ranging from implicature theory and theories of pragmatic enrichment to pragmatic acquisition, pragmatic impairment, and pragmatic processing. In addition, methodological issues are touched upon, and finally, the single contributions to this volume are introduced.

1. Introduction

Drawing the boundary between pragmatics and semantics - both being disciplines that deal with linguistic meaning - belongs to the most basic problems of modern linguistics. In recent years, a lively debate has emerged about that problem (see Bianchi 2004; Szabó 2005; Jaszczolt, to appear). Most researchers engaged in the debate relate their approaches to the fundamental work of Paul Grice, who made the by now classical distinction between 'what is said' and 'what is implicated' (Grice 1989). Numerous sophisticated approaches to the question of how that distinction should be spelled out have been put forward. However, the rough picture of the major camps involved in the debate is that we have the so-called Neogriceans on the one hand, and the Relevance theorists on the other. Neogriceans by and large tend to defend the conceptual value of Gricean maxims or principles (Levinson 2000; Horn 2004; Atlas 2005), while Relevance theorists argue against such maxims or principles and refer to general cognitive principles such as the Principle of Optimal Relevance (see Sperber & Wilson 1995; Carston 2002; Wilson & Sperber 2004). A more fine-grained sketch would include further important theoretical approaches provided, for instance, by Bach (1999); Recanati (2004); Jaszczolt (2005); Ariel (2008); Potts (2005).

Indeed, there are many differences in the pragmatic architecture of the rivalling camp's approaches as well as in the coverage of empirical phenomena. But, quite surprisingly, Neogriceans as well as Relevance theorists go for the assumption that propositional structures are systematically underdetermined and therefore are in need of enrichment. These processes of enrichment are of an essentially pragmatic nature, and hence pragmatics is conceived of as being able to influence semantics. Several terminological proposals are on the market to fix the phenomenon of pragmatically steered propositional enrichment: **explicature** (Carston 2002), **impliciture** (Bach 1999; Garrett & Harnish 2007), **pragmatic intrusion** (Levinson 2000), or **intuitive content** (Recanati 2004). The detailed comparison of these proposals, or, more generally, the explicit/implicit distinction, is of course on the pragmaticist's research agenda. Insofar as **underdeterminacy** is taken as a serious linguistic phenomenon, there is an in-built tendency of the approaches mentioned to restrict the realm of truth-conditional semantics, or to downplay its importance. According to those approaches, there are truthconditions or logical forms for sentences, but they occur only at a certain (intermediate) stage in the generation of the comprehensive meaning of an utterance.

The tendency sketched has of course provoked conjectures. For instance, some researchers defend a classical, minimalist approach to the truth conditions of a sentence, and consequently assume a more powerful apparatus for pragmatic interpretation (cf. Borg 2004; Cappelen & Lepore 2004; Bach 2005). Minimalism is thus opposed to Contextualism, understood as the persuasion that context influences semantics, not only in the case of indexicals but also in numerous further aspects. Regrettably, contributions to these debates are not always founded in large-scale empirical research. Very often, a rather restricted phenomenon serves as the 'evidence' for a certain view of the explicit/implicit distinction or a particular concept of the 'proper' semantics/pragmatics divide. It is obvious that the field will profit from more truly empirical exploration of the data.

With the advent of **experimental pragmatics** the scene has changed. Experimental pragmatics, as we understand it, is the application of psycholinguistic and neurolinguistic methods to the exploration of pragmatic phenomena (cf. also Katsos & Cummins 2010; Bezuidenhout 2010). When the focus is on the experimental research into the semantics/pragmatics distinction, we may very well speak of 'experimental pragmatics/semantics', thus alluding to the fact that research into experimental pragmatics necessarily is involved in reflecting semantics, too. As a quick glance into the seminal collection edited by Noveck & Sperber (2004) and the more recent volume edited by Sauerland & Yatsushiro (2009) shows, experiments have been run with regard to pragmatic phenomena as diverse as reference, felicity conditions, scalar implicatures, presuppositions, negation, irony and metaphor. But another view into the *Handbook of Pragmatics* (Horn & Ward 2004) and the handbook on semantics (von Heusinger et al., to appear) shows that there is still much experimental ground to cover.

It is not by mere chance that experimental pragmatics profits from experimental **psycholinguistics**, especially with regard to **language acquisition**. Research into language acquisition has a strong focus on elicitation studies, and it is in the study of research into pragmatic development (and impairment) where the two strands of research, experimental pragmatics and research into pragmatic development, meet. Indeed, it appears as if the appeal of the new research paradigm benefits from this particular constellation, as the numerous studies on the acquisition of scalar implicatures impressively show. Many of the experimental designs used in experimental pragmatics already have a history. For example, the experimental design in today's research into scalar implicatures goes back to Smith (1980). It goes without saying that there is reinterpretation and replication of classical psycholinguistic experiments, too. Still rare are production tasks, and the very promising neurolinguistic research into pragmatics is still in its beginnings (Paradis 1998 a,b; Stemmer & Schönle 2000).

The next section is concerned with one of the most important debates in theoretical and experimental pragmatics, i.e. the analysis of conversational implicatures. The discussion focuses on scalar implicatures and the distinction between generalized and particularized conversational implicatures. Section 3 briefly addresses some varieties of pragmatic enrichment. The acquisition of pragmatics, semantics and the interaction of these two linguistic modules are discussed in Section 4. Here, the discussion focuses again on scalar implicatures. Section 5 deals with the deficits of pragmatic impaired children. Sections 6 and 7 finally briefly touch aspects of pragmatic processing and some methodological issues. The last section gives a brief summary of the contributions to this volume, which deal with various theoretical and experimental aspects of the interface between pragmatics and semantics addressed in this introduction.

2. Testing for scalar implicatures

Scalar implicatures are conversational implicatures due to the observation of the maxims of Quantity. For instance, in a scale (all, some), *some* is an informationally weaker term than *all*. From the assertion of the weaker term the negation of the stronger may be inferred. Thus, if (1a) is asserted, (1b) may be inferred. If the speaker had known for sure that all of the guests were drunk, he should have – observing the maxims of Quantity – said so. Since he didn't, the hearer is licensed to conclude that the speaker intended to convey that not all of the guests were drunk. That we have to do with a type of conversational implicature here, is usually shown with a hint towards cancellability as is illustrated in (1c).

- (1) a. Some of the guests were drunk.
 - b. Not all of the guests were drunk.
 - c. Some of the guests were drunk, indeed all of them.

For Levinson (2000), scalar implicatures are generalized conversational implicatures (GCIs). GCIs arise due to pragmatic principles like the Q-principle, the I-principle, and the M-principle and are distinguished from particularized conversational implicatures (PCIs). GCIs are close to the grammar, they are implicatures that may directly influence the truth conditions of a sentence. This phenomenon of 'pragmatic intrusion' leads to the idea of a pre-semantic pragmatics. Similar ideas are pursued by a number of scholars, most notably by Relevance theorists and François Recanati ('truth conditional pragmatics').

There are, however, many differences with regard to terminology, as well as with regard to the broader conception of the semantics/pragmatics divide, and the distinction between 'what is said' and 'what is implicated' (cf. Gibbs 1999, 2002; Nicolle & Clark 1999; Liedtke, this volume). Relevance theorists like Carston (2002) create a separate pragmatic category, called 'explicature', which is portrayed as a development of a logical form. The usefulness or empirical adequacy of the PCI/GCI-distinction is denied by them, as is the distinction between sorts of pragmatic principles like the Q-principle, the I-principle and the M-principle. Instead, Relevance theorists postulate the operation of general cognitive principles like the Cognitive Principle of Relevance, the Communicative Principle of Relevance, and the Presumption of Optimal Relevance, all of which play some role in guiding pragmatic inferencing.

As far as scalar implicatures are concerned, recent research in experimental pragmatics is dedicated to the question whether there is evidence for GCI-theory versus Relevance Theory. The alternatives are nicely sketched by Noveck & Sperber (2007: 196) with regard to the computational factor 'speed of interpretation'. If a scalar term is interpreted literally (e.g. 'some, maybe all'), so they demonstrate, the GCI theory would nevertheless predict local default enrichment, because of the generalized and relatively context-independent nature of GCIs. If a context is taken into account that is not compatible with default enrichment, the implicature has to be cancelled in order to arrive at the literal interpretation, a process that obviously is costly measured in time. Relevance theory, in contrast, doesn't assume enrichment in the first place (Relevance principles being operating "fast and automatic"), and therefore predicts fast arrival at the correct interpretation.

Conversely, if a scalar term needs enrichment (e.g. 'some, but not all'), GCI theory predicts default enrichment, whereas Relevance theory demands calculation of the context (background knowledge has to be considered), so GCI theory should predict fast derivation, while Relevance theory goes together with slow derivation. Note that the respective experiments aim at measuring processing speed. Many studies have the result that deriving implicatures is costly, this being usually regarded as argument against a default view as proposed by Levinson (2000) and Chierchia (2004). However, the GCI/PCI-distinction may stay important for conceptual

reasons, and the strict opposition between the default view and context-driven approaches may turn out as problematical, because alternative views (e.g. **Interac-tionism**, see Breheny et al. 2006) lend themselves for testing, too (cf. also Katsos 2007, 2009; Zondervan, this volume).

3. Varieties of pragmatic enrichment

The scope of phenomena that fall under the heading of 'enrichment' is quite large, and, as far as we know, most phenomena have not been studied comprehensively. A general theory of enrichment is still to be developed. Levinson (2000: 170ff.) discusses disambiguation, interpretation of indexical and general expressions, ellipsis resolution and narrowing as cases of 'pragmatic intrusion into what is said'. Carston (2002: 21ff.), in her overview on underdeterminacy phenomena, also mentions ambiguities and indexicals, and furthermore adds missing constituents, unspecified scope, as well as under- and overdeterminacy of word meanings. Note that there is a strand of research that, under the heading of enriched composition, complement coercion, and cocomposition, also deals with phenomena like the ones discussed by Carston (cf. Pustejovsky 1995; Jackendoff 1997; Bezuidenhout 2009). However, this line of research refrains from any commitments to a pragmatic module and largely focuses on the syntax-semantics interface. For a comprehensive overview on psycholinguistic and neurolinguistic research dealing with enriched composition phenomena see Pylkkänen and McElree (2006).

A classical case of enrichment is of course 'conjunction buttressing' or **asymmetric coordination**. In cases like (2a) versus (2b) the impression of a temporal sequence (and, furthermore, a causal relation) may be either construed as an implicature (on the basis of the maxim of Modality), or traced back to a hidden constituent specifying time (Blakemore & Carston 2005; Hertwig, Benz & Krauss 2008).

- (2) a. Ann married and got pregnant.
 - b. Ann got pregnant and married.

Another much-disputed case of a seemingly missing constituent where temporal enrichment plays a role is the utterance (3a), where one might argue for a hidden temporal constituent, or, alternatively, derive this element as an 'explicature' or an implicature, as contexts like (3b) suggest.

- (3) a. I didn't have breakfast.
 - b. I didn't have sex.

Further cases where enrichment appears to take place are bridging and reference transfer. **Bridging** takes place in contexts like (4a), where the beer is to be understood

as part of the picnic. Several experimental studies have been carried out, among them Clark & Haviland (1977), Matsui (2000), and Burkhardt (2006), the latter taking P600 effects as evidence for enriched composition (cf. also Bornkessel-Schlesewsky & Schlesewsky (2008) and Drury & Steinhauer (2009) for linguistic interpretations of P600 effects).

- (4) a. Bob unpacked the picnic. The beer was warm.
 - b. The omelette left without paying.

Reference transfer concerns utterances like (4b), where *omelette* does not refer to the meal, but to the customer who ordered the meal (see Jackendoff 1977; Nunberg 2004; Schumacher, this volume).

In a broader perspective, **metaphor** and **irony** may also be viewed as enrichment phenomena. While the Gricean approach to these phenomena, despite being quite elementary, often serves as a starting point for experimental research, there are numerous studies intended to test details as well as general approaches (e.g. pretense versus echo theory of irony, cf. the collections by Gibbs & Colston 2007 and Gibbs 2008).

Pragmatic enrichment phenomena are certainly crucial for any attempt at exploring the semantics/pragmatics distinction. However, there are many more pragmatic phenomena that lend themselves to exploration, e.g. research on indirect speech acts (cf. Clark 1979; Shapiro & Murphy 1993) and the operation of maxims (cf. Engelhardt, Bailey & Ferreira 2006), research on presuppositions (cf. Sauerland 2009), anaphora resolution (cf. Holler & Irmen 2007), diverse aspects of information structure (cf. Burkhardt 2006, 2007; Zondervan, this volume), politeness phenomena, and, most importantly, the role of contextual knowledge in utterance interpretation (cf. Altmann & Steedman 1988; Sedivy et al. 1999; Meibauer & Schumacher 2010; Schmitz, this volume; Panizza & Chierchia, this volume). This is, however, only a sketch of crucial phenomena, and it goes without saying that there are many more. Hopefully, experimental research will provide many more data that could eventually lead to an empirically validated theory of pragmatic enrichment.

4. Developmental pragmatics

Children do not only acquire grammatical competence, but also pragmatic competence. At the age of 2, a child does not understand irony or metaphor (cf. Winner 1988; Creusere 2000), and she doesn't know much about speech acts like insulting or reproaching. Since the seminal work of Bates (1976), numerous studies on the acquisition of pragmatics have been carried out (cf. Hickmann 2000; Pan & Snow 1999). It is not by mere chance that experimental pragmatics gains much from experimental expertise in language acquisition research, since there is a longstanding tradition in psycholinguistic methods and standards of testing (see also Moscati, this volume; Müller et al., this volume; Paltiel-Gedalyovich, this volume; Rohlfing, this volume).

One recurrent finding is that children are more 'logical' than adults. Adults are more 'pragmatic', in that they observe wider aspects of the context and encyclopaedic knowledge. We will shortly consider the cases of asymmetric coordination and scalar implicature.

Noveck (2004: 310) reports on a study dealing with the question whether children are capable to observe the maxim of Orderliness, prescribing that in an **asymmetric coordination** p and q, p and q should be ordered according to the natural order of the narrated events. Noveck found that children were surprisingly tolerant against infringements of this principle. This could be interpreted as evidence for the fact that, for younger children, the facts described are more important than their particular relatedness, so that connective meaning is for them most important. For Noveck (2004: 310f.), such results hint to the superiority of Relevance theory, because it is enrichment of an underdetermined proposition that matters here, not a generalised conversational implicature, as Levinson (2000) would have it.

Scalar implicatures are another domain where children seem to act more logical than adults (and older children more adult-like than younger children). In one experiment dealing with scalar implicatures triggered by the French quantifier *certains* (scale: {tous, certains}), children and adult controls were asked whether they agreed with the (French translation of the) utterance *Some giraffes have long necks* (cf. Noveck 2001, Experiment III). 'Logical' children reacted with 'yes'. Possibly, they reasoned that even if all giraffes had long necks, it is at least true that some have long necks. Thus, 89 % out of 31 children aged 7–8, and 85% out of 30 children aged 10–11 agreed. In contrast, 'pragmatic' adults answered 'no'. For them, it would be underinformative to answer yes, because, as far as they know, all giraffes have long necks. From 15 adults, 41 % agreed.

In subsequent studies, contexts were more or less enriched in order to control for contextual knowledge. In Papafragou & Musolino (2003), where the Greek quantifier *meriki* (scale: (oli, meriki)) was tested, the focus was on felicity instead of truth. In their acting out-Experiment 1 using Truth Value Judgment Test Methodology, children were shown three toy horses which were about to jump over a toy fence. When all horses had jumped over the fence, a puppet commented on that event with, e.g. *Some horses jumped over the fence*, and children were asked whether the puppet 'answered well' (focusing on felicity). Adults rejected the puppet's statement in 92.5 % of the time, whereas 5-year-olds rejected the puppet's statement only in 12.5 % of the time.A narrative element, mimicking a certain aspect of a discourse situation, was also introduced into the methodology used by Bott & Noveck (2004) who otherwise used the material from Noveck (2001). In their Experiment II, basically a Truth Value Judgement Test, where sentences were presented on a screen, the stimuli were preceded by a declaration like "Mary says the following sentence is true/false". This declaration was intended to make a comparison between a 'logical' and a 'pragmatic' group of participants possible. Note, however, that in this case, a speech report becomes part of the context.

While the general impression from Noveck (2001) and Papafragou and Musolino (2003) was that children are more logical than adults indeed, Guasti et al. (2005:672) stress that children as young as seven years of age are able to derive (adult-like) implicatures "when the contexts meet all the cognitive and linguistic requirements for doing so". In their Experiment I, the authors replicated Noveck (2001) and found that children accepted "statements like Some giraffes have long necks much more often than adults do: 87% compared with 50%". (Again, it might be asked why so many adults were 'logical'.) The authors ventured that these results may not have to do with a simple lack in children to derive implicatures (Pragmatic Delay hypothesis), but with (unnatural) features of the experimental design (Pragmatic Limitation hypothesis). In further experiments the authors used stories presented in videos culminating in sentences like Some soldiers decide to ride a horse, and a puppet character commenting on them. Here, the rejection rate was 75% for the group of the seven-year-old children and, quite astonishingly, 83% for the adults. Thus, it is demonstrated that adults decide between agreement and disagreement similar to the children, and that there is a tension between opting for 'standard conversational norms' versus adhering to strict truth (cf. also Katsos 2009; Hendriks et al. 2009).

In sum, then, experimental designs cannot do without representing rich contexts, because participants usually react to such rich contexts and are at a loss when forced to act without them. If 'neutral' contexts are demanded, participants try to create contexts on their own (Guasti et al. 2005:684–85). And even if it is assumed that participants are able to construe a 'neutral' context, the use of certain language materials "can create their own context through a variety of presupposition triggers and information-structure triggers", as Breheny et al. (2006: 445) stress.

Implicatures are a classical pragmatic research topic, speech acts are another. Several studies focus on **felicity conditions** for promises. Experiments by Astington (1988) showed that many children between 5 and 9 years of age assume that promises are true statements related to past or future states. What is relevant from the adult's perspective (and the sincerity condition), namely that the future act must not only be (accidentally) realised, but (intentionally) caused by the speaker, is not very important for children in this age. In other words, the early concept of a promise resembles very much a prediction. The full distinction between promises and assertions is acquired between 11 and 13 years of age.

Other experiments focused on the preparatory condition. Bernicot & Laval (2004) wanted to know whether the fulfilment of the preparatory condition was relevant for children aged 3, 6 and 10 years. Interestingly, they tested also the explicit performative construction vis-á-vis a declarative sentence with reference to the future. Bernicot & Laval (2004:213) conclude that children aged 3 and 6 years did not manage the task as good as the 10 years old.

If the preparatory condition was fulfilled, then the form of the utterance (explicit performative versus future-related declarative sentence) was irrelevant. But if the preparatory condition was not fulfilled, the variable 'explicit performative' had a negative influence on the set of correct answers. In sum, then, the authors conclude that the preparatory condition (as studied by Astington 1988) is acquired earlier than the preparatory condition.

Although research into speech acts has a great psychological tradition (see Lee 2011 on lying), it goes without saying that there still is a long way to go until a comprehensive theory of speech act acquisition is reached. At the moment, research focuses on single illocutions or age groups, while social and emotional aspects are largely neglected. But experimental pragmatics is certainly a means to enlarge our knowledge of speech act acquisition.

5. Pragmatic impairment

If pragmatic competences are not innate, but acquired in the course of children's development, it is plausible that there might exist children who show difficulties with respect to an adequate pragmatic behaviour. Those children may be regarded as pragmatically impaired children. For example, a child has not grasped the felicity condition on promises, requiring that the speaker is obliged to do a future act. Then this child may be regarded as being insincere and not trustworthy, albeit he suffers from a pragmatic impairment. Pragmatic impairments nowadays are regarded as the proper object of clinical pragmatics (cf. Stemmer 1999; Perkins 2007; Cummings 2009). One particular question is whether children with abnormal speech behaviour are children with SLI (Specific Language Impairment, cf. Leonard 1998) or autistic children (Happé 1993; Bishop 2000). SLI children are those children who show language deficits, but are sane and possess a normal IQ. There are numerous attempts at developing a behavioural profile for pragmatically impaired children, but typically questions related to pragmatic competences are still lacking in questionnaires and language tests. According to one standard definition, pragmatically impaired children have the following properties or deficits: They want

to communicate, have difficulties with inferencing, give to much or to little information, do not understand metaphors and irony, are not able to narrate, have difficulties with turn taking and open, unstructured discourse situations (Leinonen 2000: 5f.).

Kurtz & Wilbur, this volume, measured the pragmatic abilities of SLI children on the basis of Bishop and Adam's (1989) categorial system. While it is plausible that there are in fact pragmatic impairments in language acquisition, it seems that – in comparison with the state of art in pragmatics – many pragmatic abilities checklists are still not fine-grained enough (cf. Dohmen 2009). Thus, research into pragmatic impairment certainly will profit from progress in experimental pragmatics.

6. Processing and the neuronal system

Of course, there are certain classical studies concerning the processing of pragmatics, namely studies dealing with **reaction times**. For example, in their study of scalar implicatures, Breheny et al. (2006) used short texts that made the background clear to the participants. Since they aimed at reading times of a trigger-containing text segment, the texts were presented on a video screen. In Experiment I, focusing on scalar implicatures with (the Greek correlates of) *or*, the materials contained short stories displaying an upper-bound context (activating the implicature 'not both') and lower-bound contexts (activating the conventional meaning 'both'). The overall result was that reading times were longer in upper-bound contexts, i.e. with the implicature.

Pragmatic processes certainly are products of the human brain. Thus, **neuropragmatics** is the study of neurological processes accompanying the production and comprehension of communicative sense. The parallel notion of **neurosemantics** obviously is established to a lesser degree. A classical question is whether specifically pragmatic processes correspond to specific neuronal systems, for instance that pragmatics is right-hemispherical while grammar is left-hemispherical. This simple picture has been abandoned: The majority opinion is that neuropragmatic functions arise through a complex interaction of cortical and subcortical neuronal systems, these systems being influenced by external and internal factors in turn (Stemmer 1999; Bara & Tirassa 2000). Nevertheless, it may be the case that semantics and pragmatics may be neurologically represented in a distinct way.

Kasher et al. (1999) looked at the ability to understand conversational implicatures with adults suffering from left- and right-hemispherical lesions. All in all, they found that the left versus right hemispheres contribute to the understanding of implicatures in a different way. In addition to clinical studies that concentrate on questions of location (cf. Cummings 2009), there are also neuropragmatic studies with speakers without a specific language impairment using functional magnetic resonance imaging (fMRI) (Bornkessel & Friederici 2007). Here the problem is how to bring together certain effects with what we know about typical measures form electrophysiological experiments (for instance the N400, indicating semantic processing), to check for larger text fragments, and to control for aspects of contextual or encyclopaedic knowledge (Bornkessel-Schlesewsky & Schlesewsky 2009). Moreover, standard experiments concentrate on comprehension and largely neglect production.

7. Variety of methodologies

Modern psycho- and neurolinguistic research uses a broad spectrum of different experimental methods: questionnaires, picture-selection tasks, truth value and speeded judgment tasks, self-paced reading and reaction time experiments, eye tracking, visual world design, event-related potentials (ERP) or fMRI (cf. Coulson 2004). Each method has, of course, specific (dis-)advantages and yields its own type of data. Although most of these online- and offline methods are frequently used in experimental pragmatics and semantics, there is still a lack of combining and comparing these methods to reach a more comprehensive experimental picture. Moreover, these methods may be complemented with corpus studies on semantic and pragmatic aspects of meaning and the interface between pragmatics and semantics (cf. Romero-Trillo 2008). This is another avenue that certainly will be walked along in future research on pragmatics/semantics.

8. The contributions to this volume

The articles in this volume take up many of the topics discussed in the previous sections and also add new topics. They discuss data from different languages and domains of pragmatic and semantic research using various experimental methods such as questionnaires, picture-selection tasks, truth value and semantic judgment tasks, eye tracking, and event-related potentials (ERP). Additionally, many papers discuss issues concerning experimental designs and provide methodological guide-lines for further research. Although the resulting picture is by no ways a coherent one, this volume clearly focuses on theoretical and experimental aspects of the interface between semantics and pragmatics and contributes findings and arguments that will foster future discussions.

Robert M. Kurtz & Ronnie B. Wilbur show in their contribution "The development of conversational competence in children with Specific Language Impairment" that SLI children have pragmatic difficulties, especially with regard

to the observation of the Gricean maxims of Relation and Quantity. The research method was rating of videotapes showing children with and without SLI in spontaneous conversations.

In his paper "The impact of literal meaning on what-is-said" Frank Liedtke contributes questionnaire evidence on the question which level of an utterance contributes to 'what is said'. The results of his study show that the literal meaning of an utterance is taken into account when native speaker judgments on 'what is said' are elicited.

Vincenzo **Moscati's** paper on "Discourse under control in ambiguous sentences" focuses on the acquisition of those sentences that contain interactions between modals and negation, thus challenging the child's capacities for disambiguation. In two experiments, children's preferences for scope assignment, e.g. in the case of Italian modal *potere* ('can') followed by clausal negation, were tested.

The contribution "Pragmatic children: How German children interpret sentences with and without *only*" by Anja Müller, Petra Schulz, and Barbara Höhle is an experimental study of the acquisition of the meaning of the German focus particle *nur* ('only'). The authors show that on the one hand 6-year-old children have difficulties with a target-like interpretation. On the other hand, children are able to take into account pictorial information as well as verbal information.

Leah **Paltiel-Gedalyovich** "Adult response uniformity distinguishes semantics from pragmatics: Implications for child language" reports on experiments investigating the interpretation of Hebrew *ve* ('and') and *aval* ('but'). While adults showed uniformity across their judgments, children were reluctant with regard to non-truth conditional meaning. Hence it is uniformity of judgment that matters with regard to a robust distinction between semantics and pragmatics.

In their contribution "Numerals and scalar implicatures" Daniele **Panizza** and Gennaro **Chierchia** discuss the results of a questionnaire and an eye tracking experiment both dealing with the relation between numerals and different types of contexts (upward entailing versus downward entailing contexts). The authors argue that the stronger interpretation of a numeral, which is supported by upward entailing contexts, results from a scalar implicature.

Katharina J. Rohlfing's paper "Meaning in the objects" argues that objects being present in an experimental setting influence children's linguistic and gestural behavior. Therefore, one must be cautious to confront children with novel situations in which they have no chance to build onto their non-linguistic experiences. Experimental data are presented that draw on spatial relations represented by prepositions.

Hans J. Schmitz' contribution on "Blocking modal enrichment (*tatsächlich*)" provides a detailed analysis of the German adverb *tatsächlich* ('in fact'), which is analysed as a pragmatic marker blocking enrichment. This particular item doesn't

contribute to the truth-conditional content, but triggers truth-conditional effects. Experimental evidence from paper-and-pencil tests shows that subjects are able to carry out modal enrichments and are sensitive to enrichment blocking.

Pragmatic enrichment is also discussed in Petra Schumacher's paper "*The hepatitis called*...: Electrophysiological evidence for enriched composition". She focuses on cases of reference transfer where a salient property of an entity is used to refer to this entity (i.e. the word *hepatitis* can be used to refer to a patient suffering from hepatitis). The ERP data discussed in this paper show that enrichment operations are connected to a late positivity.

The final contribution "The role of QUD and focus on the scalar implicature of *most*" by Arjen **Zondervan** reports on two experiments investigating the influence of focus on scalar implicatures. The experiments show that the Question under Discussion (QUD) affects the interpretation of the English scalar term *most*: When *most* is contained in the focused part of the sentence, more scalar implicatures will be derived. The same effect arises in the context of sentential answers to yes-no-questions.

Selected bibliography

- Altmann, G. & Steedman, M. 1988. Interaction with context during human sentence processing. *Cognition* 38: 419–439.
- Ariel, M. 2008. Pragmatics and Grammar. Cambridge: CUP.
- Astington, J.W. 1988. Children's understanding of the speech act of promising. *Journal of Child Language* 15: 157–173.
- Atlas, J.D. 2005. Logic, Meaning, and Conversation. Semantical Underdeterminacy, Implicature, and their Interface. Oxford: OUP.
- Bach, K. 1999. The myth of conventional implicature. Linguistics and Philosophy 22: 327–366.
- Bach, K. 2005. Context ex *machina*. In *Semantics versus Pragmatics*, Z.G. Szabó (ed.), 15–44. Oxford: Clarendon.
- Bach, K. 2006. The top ten misconceptions about implicature. In *Drawing the Boundaries of Meaning: Neo-Gricean Studies in Pragmatics and Semantics in Honor of Laurence R. Horn* [Studies in Language Companion Series 80], B. Birner and G. Ward (eds), 21–30. Amsterdam: John Benjamins.
- Bach, K. 2007. Regressions in pragmatics (and semantics). In *Pragmatics*, N. Burton-Roberts (ed.), 24–44. Basingstoke: Palgrave.
- Bara, B.G. & Tirassa, M. 2000. Neuropragmatics: Brain and communication. *Brain and Language* 71: 10–14.
- Bates, E. 1976. Language and Context: The Acquisition of Pragmatics. New York NY: Academic Press.
- Bernicot, J. & Laval, V. 2004. Speech acts in children: The example of promises. In *Experimental Pragmatics*, I.A. Noveck & D. Sperber (eds.), 207–227. Basingstoke: Palgrave.
- Bezuidenhout, A. 2009. Semantics-pragmatics-boundary. In *Concise Encyclopedia of Pragmatics*, 2nd edn, J.L. Mey (ed.), 913–920. Oxford: Elsevier.

Bezuidenhout, A. 2010. Experimental pragmatics. In *The Pragmatics Encyclopedia*, L. Cummings, (ed.), 148–154. London: Routledge.

Bianchi, C. (ed.). 2004. The Semantics/Pragmatics Distinction. Stanford CA: CSLI.

- Bishop, D.V.M. 2000. Pragmatic language impairment: A correlate of SLI, a distinct subgroup, or part of the autistic continuum? In Speech and Language Impairments in Children: Causes, Characteristics, Intervention and Outcome, D.V.M. Bishop and L.B. Leonard (eds.), 99–113. Hove: Psychology Press.
- Bishop, D.V.M. & Adams, C. 1989. Conversational characteristics of children with semanticpragmatic disorder. British Journal of Disorders of Communication 24: 211–263.
- Blakemore, D. 2001. *Relevance and Linguistic Meaning: The Semantics and Pragmatics of Discourse Markers*. Cambridge: CUP.
- Blakemore, D. & Carston, R. 2005. The pragmatics of sentential coordination with 'and'. *Lingua* 115: 569–589.
- Bloom, R.L. & Obler, L.K. 1998. Pragmatic breakdown in patients with left and right brain damage: Clinical implications. In *Pragmatics in Neurogenic Communication Disorders*, M. Paradis (ed.), 11–20. Oxford: Elsevier.
- Borg, E. 2004. Semantic Minimalism. Oxford: OUP.
- Bornkessel-Schlesewsky, I. & Friederici, A. 2007. Neuroimaging studies of sentence and discourse comprehension. In *The Oxford Handbook of Psycholinguistics*, M.G. Gaskell (ed.), 407–424. Oxford: OUP.
- Bornkessel-Schlesewsky, I. & Schlesewsky, M. 2008. An alternative perspective on 'Semantic P600' effects in language comprehension. *Brain Research Reviews* 59: 55–73.
- Bornkessel-Schlesewsky, I. & Schlesewsky, M. 2009. Processing Syntax and Morphology. A Neurocognitive Perspective. Oxford: OUP.
- Bott, L. & Noveck, I.A. 2004. Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language* 51: 437–457.
- Breheny, R., Katsos, N. & Williams, J. 2006. Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition* 100: 434–463.
- Burkhardt, P. 2006. Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language* 98: 159–168.
- Burkhardt, P. 2007. The P600 reflects cost of new information in discourse memory. Neuro-Report 18(17): 1851–1854.
- Cappelen, H. & Lepore, E. 2004. Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism. Oxford: Blackwell.
- Carston, R. 2002. *Thoughts and Utterance. The Pragmatics of Explicit Communication.* Oxford: Blackwell.
- Chierchia, G. 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In *Structures and Beyond*, A. Belletti (ed.), 39–103. Oxford: OUP.
- Chierchia, G., Guasti, M.T., Gualmini, A., Meroni, L., Crain, S. & Foppolo, F. 2004. Semantic and pragmatic competence in children's and adult's comprehension of *or*. In *Experimental Pragmatics*, I.A. Noveck & D. Sperber (eds.), 283–300. Basingstoke: Palgrave.
- Clark, H.H. & Haviland, S. 1977. Comprehension and the given-new-contract. In *Discourse Production and Comprehension*, R. Freedle (ed.), 1–40. Norwood NJ: Ablex.
- Clark, H.H. 1979. Responding to Indirect Speech Acts. Cognitive Psychology 11: 430-477.
- Clark, H.H. & Bangerter, A. 2004. Changing ideas about reference. In *Experimental Pragmatics*, I.A. Noveck & D. Sperber (eds.), 25–49. Basingstoke: Palgrave.

- Coulson, S. 2004. Electrophysiology and pragmatic language comprehension. In *Experimental Pragmatics*, I.A. Noveck & D. Sperber (eds), 187–206. Basingstoke: Palgrave.
- Creusere, M.A. 2000. A developmental test of theoretical perspectives on the understanding of verbal irony: Children's recognition of allusion and pragmatic insincerity. *Metaphor and Symbol* 15: 29–45.
- Cummings, L. 2009. Clinical Pragmatics. Cambridge: CUP.
- Dohmen, A., Dewart, H. & Summers, S. 2009. *Das pragmatische Profil. Analyse kommunikativer Fähigkeiten von Kindern.* Unter Mitarbeit von Athina Skordi. München: Urban & Fischer.
- Drury, J.E. & Steinhauer, K. 2009. Brain potentials for logical semantics/pragmatics. In *Semantics and Pragmatics. From Experiment to Theory.* U. Sauerland & K. Yatsushiro (eds), 186–215. Basingstoke: Palgrave.
- Engelhardt, P., Bailey, D. & Ferreira, F. 2006. Do speakers and listeners observe the Gricean maxim of quantity? *Journal of Memory and Language* 54: 554–573.
- Feeney, A., Scrafton, S., Duckworth, A. & Handey, S. 2004. The story of some: Everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology* 58: 121–132.
- Garrett, M. & Harnish, R. 2007. Experimental pragmatics: Testing for implicitures. *Pragmatics and Cognition* 15: 65–90.
- Gaskell, M.G. (ed.) 2007. The Oxford Handbook of Psycholinguistics. Oxford: OUP.
- Gibbs, R.W. 1997. Pragmatics in understanding What is said. Cognition 62: 51-74.
- Gibbs, R.W., Jr. 1999. Interpreting what speakers say and implicate. *Brain and Language* 68: 466-485.
- Gibbs, R.W., Jr. 2002. A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics* 34: 457–486.
- Gibbs, R.W. Jr. 2004. Psycholinguistic experiments and linguistic-pragmatics. In *Experimental Pragmatics*, I.A. Noveck & D. Sperber (eds), 50–71. Basingstoke: Palgrave.
- Gibbs, R.W., Jr. (ed.). 2008. The Cambridge Handbook of Metaphor and Thought. Cambridge: CUP.
- Gibbs, R.W., Jr. & Colston, H.L. (eds). 2007. *Irony in Language and Thought. A Cognitive Science Reader*. Mahwah NJ: Lawrence Erlbaum Associates.
- Glucksberg, S.S. 2004. On the automaticity of pragmatic processes: A modular proposal. In *Experimental Pragmatics*, I.A. Noveck & D. Sperber (eds), 72–93. Basingstoke: Palgrave Macmillan.
- Grice, P. 1989. Logic and conversation. In *Studies in the Way of Words*, P. Grice, 22–40. Cambridge MA: Harvard University Press.
- Grodner, D. & Sedivy, J.C. To appear. The effect of speaker-specific information on pragmatic inferences. In *The Processing and Acquisition of Reference*, N. Pearlmutter & E. Gibson (eds.). Cambridge MA: The MIT Press.
- Guasti, M.T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A. & Meroni, L. 2005. Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes* 20: 667–696.
- Happé, F.G.E. 1993. Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition* 48: 101–119.
- Hendriks, P. et al. 2009. A large-scale investigation of scalar implicature. In Semantics and Pragmatics. From Experiment to Theory, U. Sauerland & K. Yatsushiro (eds), 30–50. Basingstoke: Palgrave Macmillan.

- Hertwig, R., Benz, B. & Krauss, S. 2008. The conjunction fallacy and the many meanings of *and*. *Cognition* 108: 740–753.
- Hickmann, M. 2000. Pragmatische Entwicklung. In *Sprachentwicklung* [Enzyklopädie der Psychologie. Sprache, Bd. 3], H. Grimm (ed.), 193–227. Göttingen: Hogrefe.
- Holler, A. & Irmen, L. 2007. Empirically assessing the effects of the right frontier constraint. In *Anaphora. Analysis, Algorithms and Applications,* A. Branco (ed.), 15–27. Berlin: Springer.
- Horn, L.R. 2004. Implicature. In *The Handbook of Pragmatics*, L.R. Horn & G. Ward (eds), 3–28. Oxford: Blackwell.
- Horn, L.R. & Ward, G. (eds). 2004. The Handbook of Pragmatics. Oxford: Blackwell.
- Jackendoff, R. 1997. The Architecture of the Language Faculty. Cambridge MA: The MIT Press.
- Jaszczolt, K.M. 2005. Default Semantics. Foundations of a Compositional Theory of Acts of Communication. Oxford: OUP.
- Jaszczolt, K.M. To appear. Semantics and pragmatics: The boundary issue. In *Semantics: An International Handbook of Natural Language Meaning*, K. von Heusinger, P. Portner & C. Maienborn (eds). Berlin: Mouton de Gruyter.
- Kasher, A. et al. 1999. Effects of right- and left-hemisphere damage on understanding conversational implicatures. *Brain and Language* 68: 566–590.
- Katsos, N. 2007. The semantics/pragmatics distinction from an experimental perspective: The case of scalar implicature. *Synthese* (online publication 15/08/2007).
- Katsos, N. 2009. Evaluating under-informative utterances with context-dependent and contextindependent scales: Experimental and theoretical implications. In *Semantics and Pragmatics. From Experiment to Theory*, U. Sauerland & K. Yatsushiro (eds), 51–73. Basingstoke: Palgrave Macmillan.
- Katsos, N., Breheny, R. & Williams, J. 2005. Interaction of structural and contextual constraints during the on-line generation of scalar inferences. In *Proceedings of the* 27th *Annual Conference of the Cognitive Science Society*, B. Bara, L. Barsalou & M. Bucciarelli (eds), 1108–1113. Mahwah NJ: Lawrence Erlbaum Associates.
- Katsos, N. & Cummins, C. 2010. Pragmatics. From theory to experiment and back again. Language and Linguistics Compass 4–5: 282–295.
- Kutas, M. & Federmeier, K.D. 2000. Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences* 4: 463–470.
- Lee, K. 2011. Children and Lying: A Century of Scientific Research. Oxford: Blackwell.
- Leinonen, E. 2000. Children's Pragmatic Communication Difficulties. London: Whurr.
- Leonard, L.B. 1998. Children with Specific Language Impairment. Cambridge MA: The MIT Press.
- Levinson, S.C. 2000. *Presumptive Meanings. The Theory of Generalized Conversational Implicature*. Cambridge MA: The MIT Press.
- Matsui, T. 2000. Bridging and Relevance. Amsterdam: John Benjamins.
- McDonald, S., Togher, L. & Code, C. (eds). 1999. *Communication Disorders Following Traumatic Brain Injury*. Hove: Psychology Press.
- Meibauer, J. 2006. Implicature. In *Encyclopedia of Language and Linguistics*, 2nd edn, Vol. 5, K. Brown (ed.), 568–580. Oxford: Elsevier.
- Meibauer, J. To appear. Pragmatic evidence, context, and story design. To appear in Language Sciences (Special Issue Converging Data Sources in Cognitive Linguistics, M. Consten, A. Kertész and M. Schwarz-Friesel (eds)).
- Meibauer, J. & Schumacher, P. 2010. Comprehension despite lack of knowledge? ERP evidence on the relevance of expert knowledge in generating pragmatic inferences. Paper given to the Conference Beyond the Meaning of Words, Leipzig.

- Nicolle, S. & Clark, B. 1999. Experimental pragmatics and what is said: A response to Gibbs and Moise. *Cognition* 69: 337–354.
- Noveck, I.A. 2001. When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition* 78: 165–188.
- Noveck, I.A. 2004. Pragmatic inferences related to logical terms. In *Experimental Pragmatics*, I.A. Noveck & D. Sperber (eds.), 301–321. Basingstoke: Palgrave Macmillan.
- Noveck, I.A. & Posada, A. 2003. Characterising the time course of an implicature: An evoked potentials study. *Brain and Language* 85: 203–210.
- Noveck, I.A. & Sperber, D. (eds). 2004. *Experimental Pragmatics*. Basingstoke: Palgrave Macmillan.
- Noveck, I.A. & Sperber, D. 2007. The why and how of experimental pragmatics. In *Pragmatics*, N. Burton-Roberts (ed.), 184–212. Basingstoke: Palgrave Macmillan.
- Nunberg, G. 2004. The pragmatics of deferred interpretation. In *The Handbook of Pragmatics*, L. Horn & G. Ward (eds), 344–364. Oxford: Blackwell.
- Pan, B.A. & Snow, C.E. 1999. The development of conversational and discourse skills. In *The Development of Language*, M. Barrett (ed.), 229–250. Mahwah NJ: Lawrence Erlbaum Associates.
- Papafragou, A. & Musolino, J. 2003. Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition* 86: 253–282.
- Paradis, M. 1998a. The other side of language: Pragmatic competence. *Journal of Neurolinguistics* 11: 1–10.
- Paradis, M. (ed.). 1998b. Pragmatics in Neurogenic Communication Disorders. Special issue of Journal of Neurolinguistics 11.
- Perkins, M.R. 2007. Pragmatic Impairment. Cambridge: CUP.
- Potts, C. 2005. The Logic of Conventional Implicatures. Oxford: OUP.
- Pouscoulous, N., Noveck, I., Politzer, G. & Bastide, A. 2007. A developmental investigation of processing costs in implicature production. *Language Acquisition* 14: 347–376.
- Pustejovsky, J. 1995. The Generative Lexicon. Cambridge MA: The MIT Press.
- Pylkkänen, L. & McElree, B. 2006. The syntax-semantics interface: On-line composition of sentence meaning. *In Handbook of Psycholinguistics*, M.J. Traxler & M.A. Gernsbacher (eds), 539–580. Amsterdam: Elsevier.
- Recanati, F. 2004. Literal Meaning. Cambridge: CUP.
- Romero-Trillo, J. (ed.). 2008. *Pragmatics and Corpus Linguistics. A Mutualistic Entente*. Berlin: Mouton de Gruyter.
- Sauerland, U. 2009. Presupposition: From theory to experiment. In *Semantics and Pragmatics*. *From Experiment to Theory*, U. Sauerland & K. Yatsushiro (eds.), 219–227. Basingstoke: Palgrave Macmillan.
- Sauerland, U. & Yatsushiro, K. 2009. *Semantics and Pragmatics. From Experiment to Theory.* Basingstoke: Palgrave Macmillan.
- Schlesewsky, M. & Bornkessel, I. 2004. On incremental interpretation: Degrees of meaning accessed during sentence comprehension. *Lingua* 114: 1213–1234.
- Sedivy, J.C. 2007. Implicature during real time conversation: A view from language processing research. *Philosophy Compass* 2–3: 475–496.
- Sedivy, J.C., Tanenhaus, M.K., Chambers, C.G. & Carlson, G.N. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition* 71: 109–147.
- Shapiro, A.M. & Murphy, G.L. 1993. Can you answer a question for me? Processing indirect speech acts. *Journal of Memory and Language* 32: 211–229.

- Smith, C.S. 1980. Quantifiers and question answering in young children. *Journal of Experimental Child Psychology* 30: 191–205.
- Sperber, D. & Wilson, D. 1995. *Relevance: Communication and Cognition*. 2nd edn. Oxford: Blackwell.

Stemmer, B. 1999. Pragmatics: Theoretical and Clinical Issues. Special issue of Brain and Language 68.

- Stemmer, B. & Schönle, P.W. 2000. Neuropragmatics in the 21st century. *Brain and Language* 71: 233–236.
- Szabó, Z.G. (ed.). 2005. Semantics versus Pragmatics. Oxford: OUP.
- Van der Henst, J.-B. & Sperber, D. 2004. Testing the cognitive and communicative principles of relevance. In *Experimental Pragmatics*, I.A. Noveck & D. Sperber (eds.), 141–171. Basingstoke: Palgrave.
- Von Heusinger, K., Portner, P. & Maienborn, C. (eds). To appear. *Semantics. An International Handbook of Natural Language Meaning*. Berlin: Mouton de Gruyter.
- Wilson, D. & Sperber, D. 2004. Relevance theory. In *The Handbook of Pragmatics*, L.R. Horn & G. Ward (ed.), 607–632. Oxford: Blackwell.
- Winner, E. 1988. The Point of Words. Children's Understanding of Metaphor and Irony. Cambridge MA: Harvard University Press.

The development of conversational competence in children with Specific Language Impairment*

Robert M. Kurtz & Ronnie B. Wilbur Purdue University, West Lafayette

Two adult raters viewed videotapes of children with and without SLI in spontaneous conversations. Raters identified violations of conversational rules and assigned each to a category based on a system developed by Bishop and Adams (1989). Children with SLI produced significantly more violations than their typical peers. All but two of the violations were analyzed as falling under the Gricean maxims of relation and quantity, consistent with results of earlier studies of younger, typically-developing children. These findings confirm that children with SLI exhibit lower pragmatic competence than their typically developing peers, inviting further study comparing children with SLI with younger controls to investigate the contributions of language development and cognitive maturation.

1. Introduction

1.1 Children with Specific Language Impairment

Specific Language Impairment (SLI) is a developmental language disorder occurring in children who have none of the factors commonly associated with language disorders, including hearing loss, mental impairment, neurological disorders such as autism and epilepsy, or a history of frank neurological trauma such as a traumatic brain injury or stroke. Because children with SLI are cognitively similar

^{*} This research was supported by a U.S. National Institutes of Health Research Training Award in Communication Disorders; Ronnie B. Wilbur, Director. The Purdue University Department of Speech, Language, and Hearing Science provided travel funding for the presentation of this work at the 2008 Conference of the Deutsche Gesellschaft für Sprachwissenschaft. Laurence Leonard, Barbara Brown, and Patricia Deevy of the Purdue Child Language Research Laboratory provided valuable assistance in locating the archival data used in this study. Alexander Lipka of the Purdue University Statistics Department assisted in the analysis of the data from this study.

to other children, but deficient in their language skills, the study of their language output, and particularly their error patterns, can shed light on the structure of language while controlling for many possible confounding factors (Leonard 1998).

Children with SLI tend to begin talking at a later age than their typicallydeveloping peers; when they do begin to talk, their language development does not follow a typical course. Compared to their typically-developing peers, preschool aged (3–6 years) English-speaking children with SLI tend to exhibit limited vocabulary, difficulty acquiring new words, poor command of grammatical morphology, shorter utterances, and a limited range of syntactic structures. This atypical development pattern distinguishes SLI as a true disorder rather than a simple delay (Leonard 1998; Rice, Wexler & Cleave 1995).

The exact nature of grammatical errors produced by children with SLI depends on the typology of the language they speak. In English, which has some inflectional suffixes but is largely isolating, children with SLI often omit grammatical suffixes, such as 3rd person *-s* and past tense *-ed*, as well as plural *-s*. They also commonly omit auxiliary and copula forms of the verb *to be* and substitute object pronouns *him, her, them*, and *us* for subject pronouns *he, she, they*, and *we*. In Italian, which has a much more extensive inflectional system, children with SLI tend to substitute rather than omit inflectional suffixes, and to omit articles and clitics (Leonard 1998).

While SLI is frequently described in terms of deficits in morphosyntax, conversational skills are undoubtedly an important part of language development. As Hymes (1974) observed in an early argument for the inclusion of sociological data in the study of linguistics, it is of little use for a child to learn the rules necessary for producing all of the grammatical utterances in a language if such knowledge is not accompanied by an understanding of what verbal behaviors are appropriate and expected in a given context. In studies of social access behaviors (Brinton, Fujiki, Spencer & Robinson 1997; Craig & Washington 1993), negative reactions from peers with typically developing language skills focused not on language-impaired subjects' surface morphology and syntax, but rather on their socially and pragmatically odd behaviors.

1.2 Grice's maxims

In his landmark publication on conversational pragmatics, Grice (1975) outlined basic principles of behavior for participants in conversation, the primary requirement being that the participants cooperate. Even in the face of some popular and scholarly analogies of conversation as an adversarial process (e.g. Gergen 1999; Lakoff & Johnson 1980), Grice's characterization of conversation as a collaborative process has proven quite robust (see for example Attardo 1997). In elaboration of the rather general "cooperative principle," Grice specifies four "maxims" of a more specific nature:

Quantity:

- 1. Make your contribution as informative as required (for the current purposes of the exchange).
- 2. Do not make your contribution more informative than is required.

Quality:

- 1. Do not say what you believe to be false.
- 2. Do not say that for which you lack adequate evidence.

Relation: Be relevant.

Manner:

- 1. Avoid obscurity of expression.
- 2. Avoid ambiguity.
- 3. Be brief (avoid unnecessary prolixity).
- 4. Be orderly.

Utterances that fail to adhere to these maxims (whether they fail due to performance errors or lack of conversational skill on the part of the speaker) are perceived as pragmatically odd or inappropriate; they are likely to be met with puzzlement or perhaps even outright rejection by conversational partners, and may create an impediment to the progress of a conversation (Sala 2004). On the other hand, deliberate and obvious violation, or 'flouting', of the maxims can serve as the basis for humor (Attardo 1990, 1997; Brumark 2006; Sala 2004) or other culturally defined purposes such as modesty, politeness, and euphemism (Farghal 1995; Salih 2001; Zai-kauskas 2002). In such cases, the apparent 'violation', because it is intentional, is in fact not a violation at all, but a means of conveying a context-based meaning, or conversational implicature, other than the surface meaning expressed by the speaker. When this happens, it is the listener's responsibility to make the necessary inference in order to recognize the speaker's intended meaning.

2. Language-impaired children's conversational skills

2.1 General characteristics

Grice's maxims provide a useful framework for the examination of conversational competence in children. Considerable research has focused on the development of pragmatics in children with and without language disorders. Bishop & Adams (1989)

explored the question of whether inappropriateness is something that can be reliably identified in children's conversation. A study of language-impaired (LI, N = 24, ages 4–12) and younger normally-developing (ND-Y, N = 20, ages 4–6) children in conversation with adult partners resulted in pairwise interrater reliability of greater than 68%. Age-matched controls produced so few utterances judged inappropriate that they were excluded from the analysis (see Appendix for a list of utterance categories judged inappropriate).

The authors also found that children with semantic-pragmatic disorder (SPD, N = 14) displayed a profile that was distinct not only from that of the ND-Y group, but also children with other language impairments (OLI, N = 43). Children with SPD exhibit an error pattern in which formal aspects of language are relatively accurate, but adherence to conversational expectations is impaired. The authors concluded that children's command of conversational conventions was related to the type of language impairment with which they presented.

Conti-Ramsden & Friel-Patti (1984) compared the conversational skills of normally developing (ND, N = 14 age 1;7–2;9) and language impaired (LI, N = 14, age 3;6–5;3) children in interactions with their mothers. The conversational behaviors of both the mothers and the children in the LI group differed significantly from those of their counterparts in the ND group. While the difference in the number of turns taken did not differ significantly from group to group, the mothers in the LI group took the initiator role more frequently than mothers in the ND group, while children in the LI group more frequently took the role of responder. Furthermore, while children in both groups introduced more topics than their mothers, the ND group introduced significantly more topics than the LI group. Finally, the children in the ND group appeared more likely to interpret comments as requiring a response than LI children. Based on these significant differences in mother-child interaction patterns, the authors concluded that analysis of mother-child dialogue could form a valuable part of an assessment battery for children with language impairment.

While the study offers no clear answer as to the reasons for these differences, one may easily speculate that the mothers in the LI group were responding to the tendency of their children to initiate fewer exchanges, and to introduce fewer topics, by increasing their own behaviors in these areas to compensate and to maintain the flow of conversation.

Decreased adequacy and increased ambiguity in LI children's responses to mothers' comments versus obligators reflect lower levels of skill in responding to indirect speech acts, a task that requires command of Gricean principles. It is unclear, however, to what extent the data in this study can be generalized. The participants were ethnically homogeneous (white, middle-class, urban), and participation involved tape recording interactions in the home, suggesting a high interest on the part of the parents in their children's language development. This may have had an unknown effect on their interaction with their children.

Coulter (1998) asked what pragmatic features, if any, can distinguish children with semantic pragmatic disorder (SPD) from language-impaired children with other etiologies. This was a qualitative study, analyzing the conversational performance of three children with SPD. Descriptions of language behavior of children with SPD tend to overlap with behaviors commonly attributed to children with autism, Asperger syndrome and traumatic brain injury. In addition, it is difficult to identify whether semantic-pragmatic deficits can be attributed to language, pragmatic, or social-cognitive factors.

Conversation samples of three children with SPD were analyzed, yielding detailed profiles of children's use of pragmatic reference theory, Grice's theory of conversation, repairs, pauses, topic maintenance, and discourse markers. All children's profiles appeared similar except for topic maintenance, suggesting that children with SPD (who performed below their typically developing peers) may be identified using this area as a criterion. The author suggested that Gricean principles could be used for screening purposes.

2.2 Social skills of children with SLI

Three studies (Brinton, Fujiki, Spencer & Robinson 1997; Craig & Washington 1993; Fujiki, Brinton & Todd 1996) focused on social skills of children with SLI compared to their typically-developing peers. The first two examined children's behavior when accessing an ongoing conversational interaction. In both studies, successful access was defined as taking a verbal or non-verbal conversational turn that was noticed, and not rejected, by the participants in the ongoing interaction. Of the five SLI subjects in the study by Craig & Washington (1993), only two successfully accessed; these two had higher receptive scores than the children who did not access. All of the control children successfully accessed. Similarly, in the study by Brinton et al. (1997), the only children not to access successfully were those with SLI; the children with SLI also took on average longer to access; however, there was no significant difference in receptive skills between the successful and unsuccessful subjects with SLI in this study.

Fujiki, Brinton & Todd (1996) evaluated children's social skills as shown by a general measure of social skill, the number of peers with whom they interact, and their satisfaction with social relationships. Children with SLI displayed lower levels of social achievement on all three measures, leading the authors to conclude that language impairment appears to contribute to behavioral and social difficulties.

2.3 Adult interaction with language-impaired children

Much of the data in the literature has been gathered in conversational settings between children and adults, usually either one or both parents, or with experimenters. Conti-Ramsden & Friel-Patti (1984, see above) concluded that analysis of mother-child dialogue could be a valuable part of an assessment battery for children with language impairment, since mothers' child-directed speech to LI children is dissimilar to that directed to ND children.

Pellegrini, Brody & Stoneman (1987) examined the nature of children's violations of Grice's maxims in conversation with their parents, and parents' reactions to these violations. Children's utterances, if identified as violations, were coded according to the maxim they violated; for each violation, the parent's reaction was coded as *No reaction* (the parent continued on a related topic or changed topic), *Repetition* (the parent repeated the child's violation), *Clarification* (the parent questioned or rephrased the violation), or *Model/Correct* (the parent modeled a correct form of the child's violation). The children in their study showed emerging competence with the maxims of relation and quantity between ages two and three; two-year-olds violated these maxims more often than three-yearolds and four-year-olds; there was no significant difference between three- and four-year-olds.

Damico & Oller (1980) compared referral rates for language services when classroom teachers were trained to identify children based on pragmatic criteria to referral rates when teachers used syntactic/morphological criteria. Each group of teachers attended an in-service training, and was shown how to identify on the basis of morphosyntax (Group S) or pragmatics (Group P). Teachers were not told that the separate in-service trainings were different, and were naïve to the research questions of the study.

Group P referred more students for evaluation; in addition, this group showed a lower rate of both misses and false alarms, leading the authors to conclude that referral criteria based on informed judgments of pragmatic difficulty are more effective than criteria based on morphosyntactic factors.

3. Language development and early performance on the maxims

3.1 Findings of previous studies

Akhtar, Carpenter & Tomasello (1996) conducted two studies in which 24-monthold children played with three nameless objects. Following a time of play, the three objects were placed in a clear box with a fourth (novel) nameless object. An adult then named the novel object and displayed excitement while directing his/her gaze in the general direction of all four nameless objects, without revealing through gaze or gesture which of the four objects s/he was naming.

In the first study, adults and children played together with the first three objects but not with the fourth, so that the fourth object was novel to both; in the second, the adults and children played together with the first three objects and the children played with the fourth object alone in the absence of the adults, so that it was novel only to the adults. In both studies, the children associated the word with the object that was novel to the adult who was displaying excitement (i.e. the fourth object), presumably based on the understanding that adults are more likely to mention, and become excited about, items that are new to the discourse context, rather than items to which they have habituated. Although the purpose of the studies was to shed light on issues of word learning, these results suggest an awareness of Grice's maxims of relation and manner in children of this age, which may provide children with a 'bootstrapping' mechanism for lexical development, similar to that proposed for semantics and syntax (Landau & Gleitman 1985; Pinker 1984).

Similarly, Bishop & Adams (1989) do not specifically cite Grice; however, many of the behaviors listed in their profile of inappropriate utterance types correspond quite closely with Gricean principles. 'Too much information' and 'too little information' are clearly quantity issues, although some of the subcategories listed under them might better be thought to correspond with other maxims (e.g. 'unusual or socially inappropriate content or style' suggests the maxim of manner, and 'topic drift' evokes relation).

Coulter (1998) specifically includes Gricean conversation theory as an area of analysis, particularly violations of quality, quantity, and manner; poor understanding of implicature; and general cooperation. The author suggests using Gricean principles in screening for SPD.

Ferrier, Dunham & Dunham (2000) examined the development of Grice's maxim of quantity in children in their third year. Forty-two children were grouped by age into two groups: younger (2;3–2;4) and older (2;8–2;10). Experimenters manipulated a toy robot to converse with the children. The robot responded to children's declarative statements of two or more words with either a general query ("What?"), or a specific query (e.g. "Piggy's on the what?" when the child had said *Piggy on floor*). Younger children were more likely to produce full repetitions (*Piggy on floor*) than specifications (*Floor*) in response to general queries; they were equally likely to produce repetitions or specifications in response to specific queries. Older children usually responded to general queries with repetitions, but were more likely than younger children
to respond with specifications; they generally responded to specific queries with specifications.

The authors concluded that the younger children showed transitional understanding of the Gricean maxim of quantity; the older group showed a more mature understanding of the maxim. Both groups continued to use repetitions as a "default" strategy.

This use of repetitions as the default response may indicate that the principle of providing enough information is stronger than that of not providing too much. However, it is not clear whether poorer performance on the quantity maxim at this age is due to imperfect understanding of the maxim. An equally plausible explanation could be that repetition imposes a lower cognitive processing load on the child than the ellipsis required for a specific response. While a repetition simply involves reiterating a statement that the child has already formulated, a specific response must be contingent upon the adult utterance; to provide an appropriate specific response, the child must analyze the adult utterance and correctly identify the referent for *what* in "Piggy" rather than "Floor".

This hypothesis is not supported by the findings of Johnson, Miller, Curtiss & Tallal (1993). Their study examined the effect of questioning by adults on children's MLU in conversational setting, and whether the effect was the same for children with SLI as for children with typically developing language. Children with SLI were more likely than NL controls to produce utterances containing ellipsis in response to questions; not surprisingly, use of ellipsis correlated negatively with MLU: the mean ratio for MLU of non-elliptical to elliptical utterances was 2.1:1. In addition, examiners asked more questions of the SLI children than of the NL controls. The greater likelihood of children with SLI to use ellipsis suggests that its use is not excessively complex or difficult for their processing ability. Appropriate use of ellipsis is governed in part by Gricean principles; note that Bishop's & Adams's (1989) list of inappropriate conversational behaviors includes both inappropriate use of ellipsis and failure to use ellipsis where it is expected. Inappropriate ellipsis, or 'pseudo-ellipsis', occurred when the child wrongly presupposed knowledge of the 'elided' words on behalf of the listener, as in the following example (p. 250):

- A: so what did you do when you were sick?/
- *C:* I can't remember/ I did though when I was run over by a car/

A failure to use ellipsis where appropriate and expected typically involved a full response to a WH- question, as in the following example (p. 253):

- A: what's the doctor doing?/
- C: the doctor is looking at the boy/

The authors note that such responses, which may be perceived as 'learned' or stilted, may occur as a result of therapy activities in which the children in the LI group were required to use full subject-verb-object forms to describe pictured events; several LI children produced such forms during structured research activities but not during more informal exchanges.

Pellegrini, Brody & Stoneman (1987) examined the nature of 2-, 3-, and 4-year-old children's violations of Grice's maxims in conversation with their parents, as well as the parents' reactions to their children's violations. Audio recordings and transcripts of three 15-minute play sessions were observed with each child: one with the mother, one with the father, and one with both.

Two-year-old children committed frequent violations of the quantity and relation maxims. These decreased by age three, and there was no significant difference between 3- and 4-year-olds. Parents' reactions were coded as No reaction (the parent continued on a related topic or changed topic), Repetition (the parent repeated the child's violation), Clarification (the parent questioned or rephrased the violation), or Model/Correct (the parent modeled a correct form of the child's violation). Parents of 3- and 4-year-old children responded with No Reaction more often than parents of 2-year-olds; Clarifications occurred more often in dyads than in triads; Repetitions occurred more often in dyads than in triads; in dyads with 2-year-olds, fathers produced more *Repetitions* than mothers; and fathers in dyads produced *Modeling* responses more than mothers in dyads; both fathers and mothers produced Modeling more in dyads than in triads. The data for this study came from audio recordings, and therefore could not capture gestures and other nonverbal behaviors, which are often relevant in determining whether a conversational turn conforms to Grice's maxims. In addition, the parents' behavior was not standardized; depending on their view of the situation as a teaching or play context, their behavior may have been significantly different, even though toys and physical setting were similar. As in Ferrier et al. (2000), it is unclear whether violations (as defined by adult criteria) were due to imperfect command of the maxims or to language development.

3.2 Limitations of previous studies

No obvious operational definitions of the maxims or their violation have emerged from the studies directly or indirectly addressing Gricean principles. Several of the studies relied on raters judging utterances to be in compliance or in violation (Bishop & Adams 1989; Coulter 1998; Pellegrini, Brody & Stoneman 1987). The inter-rater reliability levels achieved in these studies suggest that this can be an effective way to assess Grice's maxims in children who produce a fair amount of spontaneous speech in conversational interaction. For children who are younger or who are language impaired, however, such an approach might present challenges if the children prove less talkative. Furthermore, it seems apparent from the findings in the literature that the performance of normally developing preschool children (i.e. age 3–5) has limited utility for examining the emergence of the maxims. For example, note that Pellegrini et al. (1987) found significant violations of only two of the four maxims (relation and quantity) in their youngest (2-year-old) subjects. Ferrier et al. (2000) created contexts in which ellipsis was either appropriate or inappropriate according to the maxim of quantity; their finding that competence in this area emerges in the third year of life is consistent with the spontaneous conversation studies.

Akhtar, Carpenter & Tomasello (1996), in focusing on word-finding, created a setting that elicited behavior consistent with understanding of the maxim of relation. The subjects of the study were similar in age (2 years) to those in Pellegrini et al. (1987), and the results were consistent with the findings of that study in that subjects showed emerging competence in exercising the relation maxim.

4. The present study

4.1 Research questions

From the evidence available, it appears that the maxims of quality and manner develop earlier than relation and quantity; however, it remains to be shown whether young children's difficulty with the latter two actually reflects lack of understanding of the maxims themselves, or whether these children simply lack the linguistic sophistication necessary to comply with them. Thus, it remains uncertain whether the maxims are present pre-linguistically, or whether they emerge as part of language development. Testing for their presence in pre-linguistic children would be valuable but challenging; another option, which we chose for this study, is to examine children with SLI, who are cognitively within normal limits, but whose language development lags in relation to that of their peers. We compared their performance to that of typically-developing age-mates, whose cognitive level and linguistic maturity were both age-appropriate.

We conducted a pilot study examining videotaped spontaneous conversations between children with SLI and adult examiners, compared with similar conversations involving typically-developing children of similar age. This involved the use of independent adult raters, following the example of Bishop & Adams (1989) and Coulter (1998). A qualitative analysis of utterances judged to be pragmatic violations was then conducted to determine whether the children with SLI produced a pattern that is distinct from the control subjects. We expected a greater number of violations in the SLI group than in the control group. Such a result would support the hypothesis that mastery of conversational rules is tied somehow to language development and cannot be attributed solely to cognitive maturity. If children's violations are due to cognitive immaturity, we can expect the SLI children's performance to show more similarity to that of the age-matched controls.

4.2 Participants

The videotapes selected for the study came from archival data used in a previous study of children with SLI and typically-developing children of the same age. All children were between 38 and 45 months of age (mean = 39.88, SD = 2.47). Children in the experimental group (n = 4) were between 38 and 45 months (mean = 40.5, SD = 3.11), and had been selected as experimental subjects for the previous study based on the following criteria:

- 1. scores within normal limits on the *Leiter International Performance Scale-Revised* (Roid & Miller 1997), a test of nonverbal cognitive ability;
- 2. responses at 20 dBHL in both ears during a hearing screen and no known history of hearing loss according to parent report;
- 3. no history of neural deficits or frank neurological trauma according to parent report;
- 4. scores at least one standard deviation below the mean (standard scores below 85 or percentile rank below 16) on at least one test of language ability.

The control group (n = 4) ranged in age from 38 to 42 months (mean = 39.25, SD = 1.89), and met the same qualification criteria as the experimental group, except that they scored within normal limits in both expressive and receptive language. Table 1 shows each subject's qualification information. Blank spaces in the table listing test scores for the control group indicate testing that had been conducted elsewhere (e.g. in school), the exact results of which were not available to us, other than that the scores were within normal limits. One of the children in the experimental group, EFLD, scored within normal limits on several language tests, and produced a higher mean length of utterance (MLU) than the others in the experimental group. This child had received extensive speech and language therapy and had demonstrated significant gains in her language ability; she qualified as an experimental subject based on her history of SLI, the significant gap (1.47 SD) between expressive and receptive scores on a number of measures: the Preschool Language Scales, 3rd Edition (Zimmerman, Steiner & Pond 1992), the Peabody Picture Vocabulary Test, 3rd Edition (Dunn & Dunn 1997) and the Structured Photographic Expressive Language Test (Werner & Kresheck 1983). As will be discussed later, her conversational skills in some ways shared similarities with the control group, and in others resembled her fellow experimental subjects.

SLI							
Subject	Age (Months)	PLS-3 AC	PLS-3 Exp	PPVT-III	SPELT (%ile)	Leiter	MLU- words
EFSB	38	79	78	79	1	98	2.45
EMAG	39	79	81	91	1	98	2.02
EMCJ	45	86	71	107	1	105	1.76
EFLD	40	112	90	74	1	126	3.18
Mean	40.50	89.00	80.00	87.75	1.00	106.75	2.35
SD	3.11	15.68	7.87	14.68	0.00	13.25	0.62
Control							
Subject	Age (Months)	Reynell-R	Reynell-E	PPVT-III	SPELT (%ile)	Leiter	MLU- words
CMCS	38	94	*	100	*	*	3.25
CMEW	42	93	*	*	42	*	4.13
CMJPG	38	123	123	130	60	131	4.31
CMTJV	39	118	118	113	*	*	4.73
Mean	39.25	107.00	120.50	114.33	51.00	131.00	4.11
SD	1.89	15.73	3.54	15.04	12.73		0.62

Table 1. Qualifying data for participants

*Exact scores not available due to testing at other facilities (e.g. school); abilities in these areas are within normal limits according to parent report.

Three graduate students from the master's programs in linguistics and speechlanguage pathology at Purdue University were employed as raters. Raters had familiarity with principles of conversational pragmatics through course work in their graduate programs. One rater withdrew from the study after training but before viewing the videotapes; to avoid the delays inherent in recruiting and training a replacement, the study proceeded with two raters, and the researchers mediated discrepancies between their responses.

4.3 Procedure

Eight videotaped, 8–10 minute segments of conversations between preschool children and adult examiners were viewed by adult raters. These video tapes were selected from archival data used for a previous study comparing grammatical development in children with SLI to that of children with typical language development. The interactions took place in the context of a qualifying evaluation session to determine eligibility for participation in the prior study based on standardized testing and analysis of spontaneous speech production. Recordings of children with SLI were selected randomly from those judged to be of adequate quality for the purposes of this study; recordings of the control group were similarly selected and had already been age-matched to the experimental group for the previous study. While the videotaped testing sessions lasted well over an hour, we selected a segment approximately ten minutes in length from each session to use in our study. Where possible, we selected interactions in which the focus was on eliciting spontaneous speech; however, in some of the sessions, this interaction lasted significantly less than ten minutes. In those cases we included interaction involving standardized testing where the child and the adult both produced spontaneous utterances with reasonable frequency to allow analysis of their interactions. Four of the segments involved children with SLI; the others portrayed children of similar age with typically developing language. The two adult raters made judgments of each child's adherence to or violation of specific rules of conversation. Raters were trained prior to viewing the videotapes to recognize the types of violations targeted by the study. Error judgments included (1) identifying a behavior as a violation; (2) assigning the violation to one of the categories based on Bishop and Adams (1989; see Appendix); and (3) describing in the rater's own words why the behavior constituted a violation. To prevent inattention due to fatigue, raters were allowed to complete no more than two, one-hour sessions on a given day, with no less than 30 minutes between sessions.

A nonverbal turn was defined as any situation in which the child held the floor, whether through his/her own initiative or as a result of an adult behavior interpreted as granting the floor (such as a question or a comment followed by a pause). Thus, a nonverbal turn may have contained either the occurrence of a gesture or other kinesthetic response to an utterance by the experimenter (e.g. a change in eye gaze), or it may have contained no overt communicative act on the part of the child.

A violation was defined as a verbal or nonverbal turn identified by both raters as fitting into any of the categories in Bishop & Adams (1989, see Appendix). A total of 30 violations were identified in this way, and on 21 of these, the raters agreed on the category to which the violation belonged; for the nine violations which the two raters assigned to different categories, the experimenters consulted the raters' written comments, the transcripts, and the video recordings, and assigned the violation to the category judged to be the better fit.

4.4 Results

The children in the control group produced 252 utterances (range: 47–79; mean: 63; SD: 16.83), a slightly higher number than the 237 produced by children in the experimental group (range: 31–100; mean: 59.25; SD: 31.24). Table 2 summarizes the results for each group. The experimental group had a higher number of nonverbal turns than the control group (130 *versus* 83) and a higher number of combined verbal and nonverbal turns (367 *versus* 335). Consistent with their lower MLU, and as is to be expected in comparisons involving children with SLI

Experimental gr	roup								
Child	Verbal turns	Nonverbal turns	Total turns	Utterance to turn ratio	Errors	Error to utter- ance ratio	Error to turn ratio	Adult Utterances	Total Utterances
EFSB	39	12	51	0.765	3	0.077	0.059	117	156
EMAG	100	26	126	0.794	11	0.110	0.087	242	342
EMCJ	67	51	118	0.568	6	0.090	0.051	326	393
EFLD	31	41	72	0.431	2	0.065	0.028	157	188
Total	237	130	367	0.646	22	0.093	0.060	842	1079
Mean	59.25	32.5	91.75		5.5			210.5	269.75
SD	31.24	17.1	36.11		4.04			92.98	115.52
Control group									
Child	Verbal turns	Nonverbal turns	Total turns	Utterance to turn ratio	Errors	Error to utterance ratio	Error to turn ratio	Adult Utterances	Total Utterances
CMCS	76	37	113	0.673	1	0.013	0.009	273	349
CMEW	47	20	67	0.701	2	0.043	0.030	137	184
CMJPG	50	17	67	0.746	2	0.040	0.030	156	206
CMTJV	79	9	88	0.898	3	0.038	0.034	122	201
Total	252	83	335	0.752	8	0.032	0.024	688	940
Mean	63	20.75	83.75		2			172	235
SD	16.83	11.79	21.87		0.82			68.76	76.58
	SLI	Control							
All violations	22	8							
All turns	267	225							
	307	333							
Proportion	0.060	0.024							

 Table 2. Results of transcript analysis and rater responses

and their typically developing peers, the experimental group produced shorter utterances and employed simpler syntax than the control group.

The experimental group produced a total of 22 pragmatic errors (range: 2–11; mean: 5.5; SD: 4.04), while the control group produced 8 (range: 1–3; mean: 2; SD: 0.82). Errors were analyzed as a proportion of the total number of turns for



Figure 1. Violations by group for each category, as a proportion of total conversational turns

	SLI	Control	
VES	0.027	0.009	
FCC	0.003	0.003	
QTY	0.022	0.006	
SIS	0.003	0.006	
TCV	0.000	0.000	
ОТН	0.005	0.000	
All categories	0.060	0.024	

Table 3. Violations by group for each category, as a proportion of total conversational turns

each group. Figure 1 and Table 3 summarize these results for each category. The proportion of errors to utterances for the experimental group was 0.060, compared to 0.024 for the control group. A one-tailed t-test confirmed the significance of this difference (p = .0056).

5. Discussion

5.1 Violation categories and the maxims

Both groups committed the highest number of violations in the category of Violations of Exchange Structure (VES). Table 4 shows the number of violations committed by each child in each category and subcategory. Ten of the violations by the children with SLI fell into the VES category, and of these, eight were coded

SLI	1 (VES)	2 (]	FCC)	<u>3 (Q</u>	TY)	4 (SIS)						
	Nil Response	Ignoring an initiation while remaining on tonic	Overliteral interpretation	Disregarding context set in prior turns	Too little information	Too much information	Topic drift	Unmarked topic shift	Stereotyped "learned" language	Inappropriate questioning	Socially inappropriate remarks	5 (TCV)	6 (OTH)	Total
EFSB	2				1									3
EMAG	2	2		1	3	1	1						1	11
EMCJ	3				2								1	6
EFLD	1				1									2
Subtotal	8	2	0	1	7	1	1	0	0	0	0	0	2	22
Total		10		1		8		1			1		2	
Control	1 (VES)	2 (]	FCC)	3 (Q	TY)	4 (SIS)						
	Nil Response	Ignoring an initiation while remaining on tobic	Overliteral interpretation	Disregarding context set in prior turns	Too little information	Too much information	Topic drift	Unmarked topic shift	Stereotyped "learned" language	Inappropriate questioning	Socially inappropriate remarks	5 (TCV)	6 (OTH)	Total
CMCS								1						1
CMEW	1	1												2
CMJPG	1		1											2
CMTJV					1	1		1						3
Subtotal	2	1	1	0	1	1	0	2	0	0	0	0	0	8
Total		3		1		2					2		0	

Table 4. Violations committed by individual children by subcategory

as nil responses. Three violations by children in the control group were coded as VES, with two of these in the nil response subcategory. Quantity violations (QTY) were the second largest category, accounting for eight violations by the experimental group and two by the control group. Seven of the eight quantity violations by children with SLI were in the subcategory of too little information, and only one involved too much information. Three of the remaining four categories

(FCC, SIS, and OTH) accounted for a total of seven violations, with neither group committing more than two violations in any one of these categories; neither group committed any truth condition violations (TCV).

The six violation categories used in this study, which were based on those used by Bishop & Adams (1989), do not correspond exactly with the four Gricean maxims. The QTY and TCV categories do match well with the maxims of quantity and quality; the VES and FCC categories both seem a likely fit for the maxim of relation, in that they have to do with producing contingent responses to verbal and nonverbal contextual cues. The SIS category contains five subcategories; two of these, *topic drift* and *unmarked topic shift*, would probably best fit within the maxim of relation, while the remaining three seem to have more to do with manner.

It is interesting to note that all SIS violations identified in this study belonged to the first two subcategories. Furthermore, if we do consider these two subcategories of SIS, along with the VES and FCC categories, as representing the maxim of relation, then our results are consistent with those of Pellegrini et al. (1987) in that 28 of the 30 violations in our study can be analyzed as falling under the maxims of relation (18) and quantity (10).

The two violations coded as OTH (*other violation*) were both committed by children in the experimental group. In one case, the child appeared to misunderstand the word *small* in the examiner's previous utterance:

- *E:* There's not much room on this table, [child's name]. This table is small.
- C: Smell.
- E: Smell?
- C: Yeah.

This quite possibly could be coded as an FCC (maxim of relation) violation, in that both the examiner's comment about lack of room and the syntactic structure in which the word *small* appears could have served as cues to the examiner's meaning; however both raters coded it as OTH. In the other instance, the child answered a question about an animal by using an animal sound rather than naming the animal. One might code this response as a use of socially inappropriate style (SIS), or one might even argue that it was not inappropriate at all, in that the child answered the question posed by the examiner, albeit using a nonverbal modality; however, as both raters coded this response as a violation, it was treated as such in the analysis.

Of the 30 violations identified by both raters, nine (30%) were assigned to different categories by the two raters. Five such examples occurred with respect to violations by the experimental group, and four were found in responses to the control group. In these cases, we consulted the raters' written comments, the transcripts, and the video recordings, and assigned the violation to the category judged

to be the better fit. In seven of the nine cases involving a discrepancy, one of the raters characterized the child's output as a QTY violation, while the other listed it as SIS (3), OTH (2), FCC (1), or VES (1). In five of these cases, the violation was ultimately designated as belonging to the QTY category; in one, it was assigned as FCC, and in one, VES. In the remaining two cases, one rater listed the violation as VES, while the other assigned it to either SIS or FCC. These violations were ultimately assigned as SIS and FCC, respectively. Violations involving discrepancies in the raters' characterizations were similar to the overall results, in that neither rater assigned them to violation categories corresponding to the maxims of quality (TCV) or manner (SIS c-e). In both cases where one rater used OTH, the rater wrote, "not sure what [the child] is talking about". It was determined from the transcript that this was due to the lack of information in the child's utterance, and did not represent confusion between quantity and quality, or between quantity and manner. Therefore, even in cases where the raters did not agree, it was clear that, consistent with the findings of Pellegrini et al. (1987) and with the rest of the data in this study, the children's violations involved the maxims of quantity and relation, but not quality or manner.

5.2 Conversational skills of a language impaired child with higher cognitive and language skills

One of the children with SLI, EFLD, produced an error to turn ratio of 0.028, significantly lower than the other children in the experimental group and similar to that of the control group. This child had received extensive speech and language therapy and had demonstrated significant gains in her language ability, scoring within normal limits on several language tests and producing a higher MLU than the others in the experimental group. She also scored significantly higher than her peers with SLI on the Leiter International Performance Scale-Revised, a test of nonverbal cognitive ability (Roid & Miller 1997). She qualified as an experimental subject based on her history of SLI, the gap between her receptive and expressive language skills, and her low scores on measures of receptive single-word vocabulary (PPVT-III) and grammatical morphology (SPELT). It would be premature to assume, based on our data, a causal relationship between EFLD's strengths in intelligence and expressive and receptive language and what appears to be a greater command of conversational maxims relative to the others in her group. Furthermore, it would be inaccurate to suggest that EFLD's pragmatic skills are similar to those of typically developing children. While her error to turn ratio better approximated that of the children in the control group, she produced the fewest verbal turns and the lowest utterance to turn ratio of any child in the study. Not surprisingly, then, her error to utterance ratio (0.065) was twice as high as that of the control group (mean = 0.032).

5.3 Quantifying and categorizing conversational violations by Children with SLI

Previous studies of social interaction skills of children with specific language impairment examined access behaviors and success rates (Brinton, Fujiki, Spencer & Robinson 1997; Craig & Washington 1993), and general measures of social skill and social achievement (Fujiki, Brinton & Todd 1996). Some experiments on children with SLI (Johnson, Miller, Curtiss & Tallal 1993) and with typically developing children (Akhtar, Carpenter, and Tomasello 1996; Ferrier, Dunham & Dunham 2000) were aimed at examining grammatical behaviors (i.e. ellipsis) in response to different types of questions presented by adult examiners.

This study differs from previous research in that we focused on quantifying and categorizing the violations committed by children with SLI using a framework similar to those used to analyze the conversational behaviors of children with other types of language disorders (Bishop & Adams 1989; Coulter 1998; Conti-Ramsden & Friel-Patti 1984) and with typically developing children (Pellegrini, Brody & Stoneman 1987). Our results suggest that the difficulty children with SLI experience in accessing social interactions (Brinton, Fujiki, Spencer & Robinson 1997; Craig & Washington 1993), and their poor performance on general measures of social skills and achievement (Fujiki, Brinton & Todd 1996), may be attributed to their difficulty in adhering to the Gricean maxims of quantity and relation.

The tendency of children with SLI to produce more elliptical responses to questions as compared to their typically developing peers (Johnson, Miller, Curtiss & Tallal 1993) predicts that they will commit more violations of the maxim of quantity; indeed our findings show that the children with SLI in our study did produce more quantity violations, and that all but one of these violations involved offering too little information rather than too much. The conversation analysis approach we used also revealed that children with SLI produced more nil responses than their peers in the control group. Consistent with the analysis of Bishop & Adams (1989), our data include nil response as constituting a violation of relation (Violation of Exchange Structure) rather than quantity. The inclusion of nil response as a violation is certainly appropriate in that conversations are full of contexts in which a response of some sort is obligatory. Including nil responses greatly influenced the results for both groups, as violations of this type accounted for eight of ten VES errors committed by children with SLI and two of three VES errors committed by children in the control group.

5.4 Future directions

The purpose of this study was to serve as an initial project to help inform the direction of future research in this area. Our findings thus far support the

hypotheses that (1) proficiency in the maxims of quality and manner develops earlier than in those of quantity and relation, and (2) children with SLI display a greater frequency of violations with regard to the maxims of quantity and relation than do their age-matched peers. These findings bring to mind the suggestions raised by earlier research on conversational behavior of children with language disorders. Conti-Ramsden & Friel-Patti (1984) suggested the analysis of mother-child dialogue be included as part of an assessment battery for language impairment, since the child-directed speech of the mothers in their study was dissimilar to that used by mothers of non-disordered children. Damico & Oller (1980) found that teachers' referrals for language services included fewer misses and false alarms when the teachers were trained to refer on the basis of pragmatics as compared to morphosyntax. Our results support the development of clinical diagnostic measures making use of conversational pragmatics to assist in the identification of children with SLI. The analysis we performed involved raters familiar with Gricean pragmatics, which seems more consistent with the type of analysis done by Conti-Ramsden & Friel-Patti (1984); however, it seems reasonable to expect that a checklist or rating system could be developed for use by teachers and caregivers as an initial identification process, as suggested by Damico & Oller (1980).

To develop such a measure, it is necessary to collect further data. In this study, we included a small number of subjects, four with SLI and four age-matched controls. This small sample size limits reliability and the development of norms. However, the significant difference in frequency of violations between the two groups does suggest that further research with a larger sample would be appropriate.

While our study confirms that children with SLI exhibit lower competence compared to their typically developing peers, we did not include younger, languagematched control subjects, and therefore cannot predict how children with SLI will compare to this population. The difficulty exhibited by the children with SLI relative to their typically developing age mates supports the position that conversational competence cannot be accounted for by cognitive maturity alone, and that linguistic development plays a significant role. Exactly what that role is, however, remains uncertain. One possibility is that development of conversational competence is by nature similar to the development of grammatical competence, and that the two therefore develop parallel to one another. The variation in age of mastery among the maxims would reflect the uneven development of linguistic skills such as morphosyntax (see Brown 1973). An alternate explanation is that conversational competence is primarily cognitive in nature, but that compliance with the maxims, particularly those of quantity and relation, requires a level of grammatical sophistication that preschool-aged children with SLI have not attained. A pattern of results showing violations similar in number and type between children with SLI and

younger control subjects with typically developing language skills would support the former explanation. Performance by children with SLI that shows greater conversational competence than younger, language-matched controls, but below that of their more language-proficient age mates, would support the latter. The inclusion of a second control group, comprised of younger typically developing children with language ability similar to that of the children with SLI but a lower level of cognitive maturity, could shed light on this question.

The use of archival data for our study had a number of beneficial effects, including speed and convenience of data collection and the guarantee that the children and adults on the video recordings were naïve to the research questions we were asking. A disadvantage of this approach is that we had no control over the quality of the recordings, the camera angles, the tasks chosen in which the children engaged, or the materials used in the sessions. Further research will therefore involve more control over the tasks, materials, and recording, and will not make use of archival data.

Our use of the categorization scheme developed by Bishop & Adams (1989) also presented both advantages and disadvantages. On the one hand, it enabled us to present the raters with consistent and detailed criteria for identifying violations, including examples of each type of violation. On the negative side, while certainly more detailed than Grice's (1975) four maxims, the Bishop & Adams (1989) categories are not without their areas of vagueness and ambiguity. The two raters in this study disagreed on categorization in nine of the 30 cases where both identified a violation, a 30 percent disagreement rate. Eight of these nine violations involved SIS, QTY, or both; this suggests that perhaps these two categories require more effective definition than the raters were provided.

The use of adult raters in itself also raises a number of methodological considerations, which should be considered here. First, adults will obviously tend to judge a conversation according to adult criteria, which may not correspond to what is normal and typical in child discourse. Rather than trying to train our raters to judge according to what *is* normal for preschoolers, we instructed them to judge according to adult standards. We recognized that this would likely result in violations being identified in the conversations of the control group (and indeed this was the case – the raters identified eight violations committed by the typically developing children), but we felt that this disadvantage would be offset by greater consistency in the raters' judgments.

We also anticipated that the adult raters' own criteria for judging an utterance as inappropriate would be influenced by the fact that they were being paid to identify violations. We predicted that this might lead the raters to over-identify violations in the conversations of both groups; to correct for this, we analyzed only those cases where both raters identified a violation. Finally, we began our analysis of the data with no *a priori* assumptions about which of the maxims, if any, would be violated the most. Given that none of the children in the study violated the maxims of quality or manner, future research will focus on the maxims of quantity and relation, which were violated by both groups, but with greater frequency by the children with SLI. It may be possible to implement tasks that specifically elicit application of these two maxims, resulting in a greater density of data than may be expected through spontaneous conversation. Such tasks may be used in addition to, or instead of, spontaneous conversation.

6. Summary

This study examined the ability of preschool children with specific language impairment (SLI), whose cognitive development was within normal limits, but whose language development lagged behind that of their peers. Two adult raters viewed videotapes of spontaneous, one-on-one conversations between four children with SLI and adult examiners, compared with similar conversations involving four typically-developing children of similar age. The raters identified occurrences of what they judged to be violations of conversational rules and assigned each violation to a category based on a system developed by Bishop & Adams (1989).

The children with SLI produced a significantly higher frequency of violations than their peers in the control group. All but two of the violations identified by the adult raters can be analyzed as falling under Grice's maxims of relation and quantity, consistent with results of earlier studies of younger, typically-developing children (Pellegrini et al. 1987). If children with SLI display a pattern of violations similar to that of younger, typically developing children of similar linguistic ability, it stands to reason that compliance with the maxims of relation and quantity are strongly linked to language development. We conclude that the development of conversational competence in children cannot be explained in terms of cognitive maturity alone; compliance with Grice's maxims appears to develop, at least in part, in connection with the development of grammatical competence.

References

- Akhtar, N., Carpenter, M. & Tomasello, M. 1996. The role of discourse novelty in early word learning. *Child Development* 67: 635–645.
- Attardo, S. 1990. The violation of Grice's maxims in jokes. Proceedings of the Berkeley Linguistics Society 16: 355–362.
- Attardo, S. 1997. Competition and cooperation: Beyond Gricean pragmatics. *Pragmatics and Cognition* 5: 21–50.

- Bishop, D. & Adams, C. 1989. Conversational characteristics of children with semanticpragmatic disorder, II: What features lead to a judgment of inappropriacy? *The British Journal of Disorders of Communication* 24: 241–263.
- Brinton, B., Fujiki, M., Spencer, J. & Robinson, L. 1997. The ability of children with specific language impairment to access and participate in an ongoing interaction. *Journal of Speech, Language, and Hearing Research* 40: 1011–1025.
- Brown, R. 1973. A First Language, the Early Stages. Cambridge MA: Harvard University Press.
- Brumark, A. 2006. Non-observance of Gricean maxims in family dinner table conversation. *Journal of Pragmatics* 38: 1206–1238.
- Conti-Ramsden, G. & Friel-Patti, S. 1984. Mother-child dialogues: A comparison of normal and language impaired children. *Journal of Communication Disorders* 17: 19–35.
- Coulter, L. 1998. Semantic pragmatic disorder with application of selected pragmatic concepts. International Journal of Language and Communication Disorders 33 (supplement): 435–438.
- Craig, H. & Washington, J. 1993. Access behaviors of children with specific language impairment. *Journal of Speech and Hearing Research* 36: 322–337.
- Damico, J. & Oller, J. 1980. Pragmatic versus morphological/syntactic criteria for language referrals. Language, Speech, and Hearing Services in the Schools 11: 85–94.
- Dunn, L. & Dunn, L. 1997. *Peabody Picture Vocabulary Test*, 3rd edn. Circle Pines MN: American Guidance Service.
- Farghal, M. 1995. Euphemism in Arabic: A Gricean interpretation. Anthropological Linguistics 37: 366–378.
- Ferrier, S., Dunham, P. & Dunham, F. 2000. The confused robot: Two-year-olds' responses to breakdowns in conversation. *Social Development* 9: 337–347.
- Fujiki, M., Brinton, B. & Todd, C. 1996. Social skills of children with specific language impairment. *Language, Speech, and Hearing Services in Schools* 27: 195–202.
- Gergen, K. 1999. An Invitation to Social Construction. London: Sage.
- Grice, H. 1975. Logic and conversation. Syntax and Semantics 3: 41-58.
- Hymes, D. 1974. *Foundations in Sociolinguistics: An Ethnolinguistic Approach*. Philadelphia PA: University of Pennsylvania Press.
- Johnson, J., Miller, J., Curtiss, S. & Tallal, P. 1993. Conversations with children who are language impaired: Asking questions. *Journal of Speech and Hearing Research* 36: 973–978.
- Lakoff, G. & Johnson, M. 1980. Conceptual metaphor in everyday language. *Journal of Philosophy* 77: 453–456.
- Landau, B. & Gleitman, L. 1985. *Language and Experience: Evidence from the Blind Child*. Cambridge MA: Harvard University Press.
- Leonard, L. 1998. Children with Specific Language Impairment. Cambridge MA: The MIT Press.
- Pellegrini, A., Brody, G. & Stoneman, Z. 1987. Children's conversational competence with their parents. *Discourse Processes* 10: 93–106.
- Pinker, S. 1984. *Language Learnability and Language Development*. Cambridge MA: Harvard University Press.
- Rice, M., Wexler, K. & Cleave, P. 1995. Specific language impairment as a period of extended optional infinitive. *Journal of Speech and Hearing Research* 38: 850–863.
- Roid, G. & Miller, L. 1997. Leiter International Performance Scale-Revised. Wood Dale IL: Stoelting.
- Sala, M. 2004. Cooperation, bona-fide mode of communication, and the violation of Grice's maxims. *Linguistica e Filologia* 19: 75–96.

- Salih, M. 2001. A pragmatic analysis of certain interactional units in Jordanian Arabic. *Journal* of Arabic Linguistics 39: 90–96.
- Werner, E. & Kresheck, J. 1983. *Structured Photographic Expressive Language Test*. DeKalb IL: Janelle Publications.
- Zaikauskas, E. 2002. Strategies, social roles, and types of speech act participants in Lithuanian. *Acta Linguistica Lithuanica* 46: 145–172.
- Zimmerman, I., Steiner, V. & Pond, R. 1992. Preschool Language Scales, 3rd edn. San Antonio TX: The Psychological Corporation.

Appendix

2.

Categories of violations, based on Bishop & Adams (1989).

- 1. Violation of exchange structure (VES)
 - a. Nil response
 - b. Ignoring an initiation while remaining on topic
 - Failure to use context in comprehension (FCC)
 - a. Overliteral interpretation
 - b. Disregarding context set in prior turns
- 3. Quantity violation (QTY)
 - a. Too little information
 - i. Inappropriate presupposition ('pseudo-ellipsis'):
 - ii. Unestablished referent
 - iii. Logical step omitted
 - b. Too much information
 - i. Unnecessary assertion or denial
 - ii. Excessive elaboration
 - iii. Unnecessary reiteration
 - vi. Ellipsis expected but not used
- 4. Unusual or socially inappropriate content/style (SIS)
 - a. Topic drift
 - b. Unmarked topic shift
 - c. Stereotyped "learned" language
 - d. Inappropriate questioning
 - e. Socially inappropriate remarks
- 5. Truth Condition Violation (TCV): deliberate attempt to mislead or deceive.
- 6. Other violation (OTH)

The impact of literal meaning on what-is-said

Frank Liedtke University of Leipzig

Whithin the Gricean tradition, the question which stratum of an utterance one would intuitively judge as what-is-said (and what not) has been debated intensely. Several experimental approaches have been developed which are reported here. In addition, a study with examples in German has been performed. One result of the study is that the literal meaning of the utterance has a strong impact upon the tendency to judge something as what-is-said, thus it does not seem to be equal whether the suggested level of what-is-said is close to its literal meaning – the results of the studies are stronger in this case.

1. Contextualism: The state of the debate

Following the traditional, Gricean account of utterance-meaning (s. Grice 1989a,b), there are two aspects of an utterance which are relevant for grasping what the speaker has communicated: the conventional meaning of the word(s)/sentences uttered (what is said) and the conversational implicature(s) (what is implicated). Within this account however, *saying* is not restricted to conventional semantic aspects alone, but contains pragmatic ingredients too. These components enable the addressees to specify the reference, the resolution of indexicals and the disambiguation of ambiguous expressions, and together they constitute the pragmatic *saturation* of what-is-said.

Thus, in an utterance like

(1) I am an artist

we have the conventional meaning of the words (1) (reference to the speaker), *am* (copula), *an* (indefinite determiner) *artist* (noun, person creating art or music ...) and their syntactic linkage. Besides this, one has to fix the reference of the indexical expression (1) and to narrow down the vague expression *artist* to the sense of, e.g. *musician* in order to arrive at a full proposition. In a third step, one has to work out the conversational implicature, for example if the utterance is an answer to the question in the following dialogue:

(2) A: Do you like the music of Richard Clayderman? – B: I am an artist.

In this case by exploiting the maxim of relevance, the conversational implicature is: +> No. This reading of the uttered sentence however does not correspond to anything that actually has been said. There are no elements which are indicating this reading. So far, this is the Gricean picture of what-is-said.

Since the eighties of the last century, the problems and deficiencies of this account have been under debate (cf. Searle 1979a; Sperber & Wilson 1986). Not only reference assignment and disambiguation, which are triggered "bottom up" via the use of e.g. indexical expressions, but also further information is held to be undispensible for determining a representation of what-is-said by an utterance. Typically this information is added "top down" to what has been overtly expressed. In spite of its seemingly non-literal character, it is treated as having been communicated or *said* by the speaker. This second type of pragmatic intrusion is called *free enrichment* by Récanati (2004) or pre-semantic pragmatics by Levinson (2000).

Consider the following example:

(3) John turned the switch and the motor started. (cf. Levinson 2000: 38)

The fact that the starting of the motor happened after John's turning the switch, that the former is causally responsible for the latter etc. is treated by some authors as a communicative stratum which is part of or very close to what-is-said. The position which consists in claiming that not only saturation, but also free enrichment plays a constitutive role in building up an instance of what-is-said, is called the contextualist position, and it is held roughly e.g. by F. Récanati, who labels this a primary pragmatic process. Others like Levinson introduce a separate layer of meaning, called utterance-type-meaning, which is different from the conventional meaning of the sentence on the one side and particularized conversational implicatures on the other. It is the level of generalized conversational implicatures, "a level of systematic pragmatic inference based ... on general expectations about how language is normally used." (Levinson 2000: 22) Another group of pragmatists assumes a third layer too, located between conventional sentence meaning on the one hand and particularized conversational implicatures on the other, but they don't share Levinson's assumption concerning its nature. They too claim that the process of free enrichment is inferential in nature, but that it is not an implicature. Thus, D. Sperber and D. Wilson introduce the notion of explicature, which concerns the development of the logical form of an utterance, i.e. all enrichments of semantic representations necessary to achieve informational content to count as conversational contribution (Sperber & Wilson 1986). K. Bach coined the term impliciture in order to account for completions and expansions of the semantic content of an utterance (s. Bach 1994). In Bach's approach, what-is-said is restricted to the semantic representation of the sentence plus reference resolution, without

even reaching the level of a full proposition. For a synoptic representation of these different accounts up to 2000 see the scheme in Levinson (2000: 195).

In contrast to these inferential approaches, which implement inferences into the architecture of the standardized, but not conventionalized layer of utterancemeaning, F. Récanati holds that these processes are associative in nature, not inferential, and that they arise locally, incrementally in the course of the production of an utterance (cf. Récanati 2004). He is the most ambitious thinker in this area, insofar as he includes many things into what-is-said which other authors would refrain from categorizing on this level. Beneath sentence meaning, what-is-said includes so called primary pragmatic processes, which are - as we have seen composed of two parts: saturation on the one hand, which is triggered by some expression in the utterance and which is therefore a mandatory process in proposition-building, and further optional processes such as free enrichment, which are not triggered and therefore non-mandatory for proposition-building. According to Récanati, enrichment is a sort of process which constitutes what-is-said in a straightforward manner: nothing has been said without enrichments, at least in most cases. Just to have an impression of the realm comprising enrichments some examples may be given. First there are processes Levinson called conjunction buttressing, like the one in (3), which Récanati rubrifies under the species of what-is-said, contrary to most other accounts including Levinson who files this case under generalized conversational implicatures, thus classifying it as opposed to what-is-said. Récanati doesn't allow any inferences here, although he concedes that there is a non-minimal departure from sentence-meaning in this and similar cases. As addressees of utterances like (3), we arrive directly at an interpretation, "as a result of the interaction of the t-literal meaning [= type-literal meaning, F.L.] of the words (and constructions), salient features of the speech situation, expectations created by the discourse, schemata stored in memory and evoked by the words, and so on." (Récanati 2004:73).

Moreover, there are cases that are traditionally counted as indirectness or irony, which Récanati equally subsumes under the heading of primary pragmatic processes, i.e. what-is-said. So

(4) John is a fine friend

uttered in a situation in which the opposite is obviously the case creates a primary reading such as: John isn't honest *as part of what-is-said*. (cf. Récanati 2004:77) Récanati relies on the fact that the act of asserting (4) is staged or simulated rather than actually performed. Within the layer of primary pragmatic processes one has to discern two subparts or -layers, that of the surface speech act which the speaker pretends to perform, and the ironical act of staging the performance of that act. This layering characterizes staged communicative acts.

The strategy of including irony and other forms of verbal behaviour like sarcasm, teasing, overstatement and understatement into the area of what-is-said is radical, and it is shared by no other authors except Récanati. In one way or another, this reflects our linguistic intuition, that it is "the message" of the utterance which counts as being said, which is obviously the ironical and not the literal reading of (4). Similar considerations may be the reason behind Searle's choice to call the indirect speech act the primary one, the literal speech act the secondary one (cf. Searle 1979b). Nevertheless, there might be some doubts about those cases, because it is not the notion of an illocution, but the notion of what-is-said that is under discussion, and this is more or less the propositional/locutionary layer of the speech act.

After all, the idea of delimiting the notion of what-is-said at all may be doubted fundamentally. So, in order to define the middle ground between sentence meaning and particularized conversational implicatures, Levinson pleads for another strategy. He holds that all terminological efforts to delimit what-is-said against what-is-unsaid are fruitless because they create more problems than they solve. Our pretheoretic notion of what-is-said is mainly dependent on the mode of talk. Levinson refers to cross-examinations in court, where everything which the speaker has not said literally is routinely queried. He claims that there is no consistent way of cutting up the semiotic pie such that what-is-said excludes what is implicated. Rather one has to define "the types of content by the processes that yield them and the important semantical properties they have (e.g. default presumption, defeasability under distinct conditions)" (Levinson 2000: 198). In one word, one has to distinguish monotonic from nonmonotonic reasoning.

For reasons of space, I cannot further elaborate the arguments of this controversy. But it should be clear that a consistent theory of different layers of the significance of an utterance has to respect the different nature of cognitive processes which govern semantic interpretation on the one hand and pragmatic reasoning on the other. Nevertheless, the protagonists (and antagonists) of the debate seem to operate on different levels. Looking for intuitions concerning the limits of what-is-said and elaborating conceptual tools for the description of the different types of reasoning processes do not seem to conflict necessarily, rather they might "cooperate" in search for an adequate account of layer-specific aspects of meaning. In addition to that, pragmatic intuitions about exemplary utterances might be backed by an experimental account that might strenghthen the evidence of certain claims about the limits of what-is-said (or the bare legitimity of questions about these limits).

Thus, in a first shot, one can change the style of argumentation and look for intuitions of language-users who are faced with sentence-tokens of a certain type. The idea behind this is that normal language-users are – and should be – able

to distinguish between what has been explicitly communicated ("said") and what has not, because a decision about this is immediately relevant for the communicative status of the utterance in question. In short, what-is-said counts as something what speakers are obliged to, it is a normative aspect of meaning. What is implicated is something that has been calculated on the side of the addressee, not something that speakers can be held responsible for alone (cf. Kriempardis 2002). This means that speaker-obligations are minor to an extent in which the part of the "said" is smaller - and the speaker's responsibility is stronger to an extent in which the weight of the "said" is stronger. If this is the case, meaning-intuitions have an important impact on the way communicative enterprises are going on, on the extent to which they succeed or don't succeed. One condition for successful communication is certainly that speaker and addressee are in concord about the status of the utterances performed: is it something which is intended to be said or something which is intended to be implicated? Much controvery and bargaining is going on in everyday conversation concerning this point, and a good intuition is critical not only for consensual, but also for conflictive language-use.

2. Experimental approaches

2.1 Noveck and Sperber

Following Noveck & Sperber (2007), a further reason in favour of an empirical approach is a methodological one. As they point out, semantic intuitions "are not just *about* semantic facts; they are semantic facts themselves." (Noveck & Sperber 2007:185) The intuition that (a) "John knows that it is raining" entails (b) "It is raining" is, so they argue, not about some semantic property that this sentence has regardless of its accessibility to speakers' and addressees. The meaning of (a) consists in the fact that it is intuitively understood as entailing (b) (ibd.). Compared with semantic intuitions, we could say that pragmatic intuitions are those "that addressees have about the intended meaning of an utterance addressed to them." (ibd.) They are able to assign irony to an utterance like (c) "Carl is a fine friend" if it is uttered towards them in a specific setting in which the contrary of (c) is obviously true.

At this point, Noveck & Sperber (2007) argue as follows: The typical scenario of a pragmatic example doesn't consist of an utterance directed towards its reader, but it contains different persons, who are communicating with each other. The author of these settings is not part of the cast, and the reader of the example introduced into an article isn't either. For Noveck and Sperber pragmatic intuitions "are not about how an utterance is interpreted, but about how an utterance would be interpreted if it were produced in a specific situation. ... These intuitions ... are not themselves pragmatic facts and they may well be in error." (Noveck & Sperber 2007: 186).

Of course, Noveck and Sperber point to a specific property of pragmatic settings, because they figure more as an invented scene in which S and A are interacting, and in the course of the construction of such a scene idiosyncratic opinions of the creator of it may interfere with common sense intuitions. Moreover, the elements of the context which constitute the interactional setting of an utterance-token are not part of the pragmatic competence of the speaker and hence may not be subject to his/her intuitions. But the relation of utterance-types to types of contexts forms indeed an essential part of the speaker's pragmatic competence and it is consequently subject to his/her intuitions. Knowledge of the kind of relation between utterance and context is comparable to the knowledge of semantic conventions governing linguistic expressions. These too, by the way, contain relations between types of states-of-affairs and types of expressions, requiring more often than not contextual assumptions.

On the other hand, a great deal of properties which are ascribed to semantic intuitions by Noveck and Sperber are related to pragmatic intuitions as well. The figures are not acting autonomously, but are determined by the decisions of the writer, who is involved in them. He makes them speak, and he would not admit anything contrary to his intuitions. Hence the semantic and the pragmatic case are not fundamentally different from each other, the latter being a case of fictional utterance. Of course, pragmatic intuitions should be checked empirically, because – as was mentioned – they might be idiosyncratic up to a point, *a fortiori* in the case of utterances with a sparse context. But this is a risk to which all sorts of linguistic intuitions are exposed, pragmatic, semantic and even syntactic ones. Consequently, one may consider pragmatic intuitions as pragmatic facts in a similar way as one may consider semantic intuitions as semantic facts.

2.2 Gibbs and Moise

Prior to Noveck & Sperber's studies, since the late nineties, several approaches to experimental pragmatics have been developed, e.g. by Gibbs & Moise (1997)¹ and Gibbs (2004). Following Gibbs & Moise (1997), enriched pragmatic information "similar to that used in inferring conversational implicatures may very well come into play as part of how people determine what speakers say, or what is said." (Gibbs & Moise 1997:54) Against Grice's implicature hypothesis, as they call it, claiming that only some aspects of our understanding of what a speaker says are

^{1.} See Nicolle & Clark (1999) and Gibbs (1999) for discussion. See also their contributions in Noveck & Sperber (2005).

influenced by pragmatics, they vote for the independence hypothesis advanced by R. Carston (1988) – who holds that implicatures have to be functionally independent of what-is-said (in the sense that they must not entail and must not be entailed by what-is-said). On the other hand, they argue for the availability hypothesis by Récanati (1989) – who holds that both what-is-said and conversational implicatures are consciously available to the speaker, in contrast to sentence meaning. If this kind of availability is given, then a speaker-hearer should consciously recognize the difference between his understanding of what-is-said and what-is-implicated (s. Gibbs & Moise 1997:55). In performing four experiments, Gibbs and Moise examined how ordinary people determine what speakers say as opposed to what they implicate. In doing so, they focused on "indicative utterances that Grice referred to as generalized conversational implicatures" (Gibbs & Moise 1997:56).

A first study was devoted to the question whether the participants of the experiment (thirty undergraduate students) included pragmatic enrichments into their notion of what-is-said or whether they relied upon the bare sentencemeaning of the utterance. Five groups of sentences had been presented to them, each accompanied by two paraphrases, one referring to what is traditionally seen as sentence meaning, the other referring to what is called a generalized conversational implicature.

Being faced with the utterance.

(5) Jane has three children

participants had to choose between the paraphrases

- a. Jane has at least three children, but may have more.
- b. Jane has exactly three children, but no more than three.

(s. Gibbs & Moise 1997: Appendix B, 70)

These paraphrases represent the minimal interpretation (a) vs. the enriched interpretation (b) respectively. In this and other cases the majority of the participants (means 84%) chose the enriched interpretation of utterance (5) as what was said, which was expected by the authors. The remaining three experiments served to sharpen the distinction between the minimal interpretation and the enriched one on the one hand, the distinction between the enriched interpretation of what-is-said and the conversational implicature on the other. The purpose of the second experiment was to investigate whether it is possible to train people to recognize the distinction between the minimal and the enriched interpretation of what-is-said, and then to measure the amount to which they would chose the minimal paraphrase as something what-is-said. But even in this case, the participants voted for the enriched paraphrase as candidate for what-is-said (means 80%). The third experiment tested whether the participants were able to distinguish between enriched interpretations of what-is-said and implicatures. Thus the same set of sentences was given to them at the end of short stories which provided a small context. The (particularized conversational) implicature in question was

(5) c. Jane is already married.

(c) was given to the participants as one alternative paraphrase together with the enriched version (b). Even in this setting, which provided a certain drift towards an implicature-reading of the cited utterance, the participants chose the enriched interpretation as that one which had been said, and not the implicature-driven interpretation (mean 86%).

The result shows that the participants did not choose the richest interpretation of an utterance available – this being the implicature, but that they chose the level of enriched interpretation as what-is-said even in a context strongly suggesting an implicature-based reading of the whole utterance.

In the fourth experiment Gibbs & Moise tested whether one could construe contexts in which minimal readings of uttered sentences can gain the status of what-is-said. In this case the items contained e.g. cardinals which had to be read as "at least x and possibly more", which corresponds to the minimal interpretation of their use. In this case (only) the participants voted for the minimal reading as what-is-said and not for the enriched one (means 90%). This result shows "that in *some* cases people understand what speakers say as conveying minimal, and not enriched, pragmatic meaning." (Gibbs & Moise 1997:65).

As a general finding, Gibbs and Moise hold that (contrary to the Gricean view) many aspects of enriched meaning are not to be rubrified under conversational implicatures but figure as a part of the speaker's and hearer's understanding of what-is-said. Thus they see the results of their experiments as supporting the claim that pragmatic inferences strongly influence our understanding of what-is-said. With these findings they laid the foundation-stone for experimental approaches in pragmatics and confirmed part of the results of Sperber and Wilson's approach to relevance-theory, especially their claim that inferential processes play a central role in processing explicit utterance-meaning (Sperber & Wilson 1986).

2.3 Nicolle and Clark

However, the results of the experiments have been challenged by Nicolle & Clark (1999), who raise some principled objections against the method and the material of the tests. First they deny that sentences containing cardinals, quantifiers and expressions denoting a time-distance do express a minimal proposition at all. As Carston (1988) puts it, uttered sentences like (5) need to be pragmatically enriched before they may express a proposition at all. From this in turn follows that both readings ("at least three children" *and* "exactly three children") have to be conceived as the result of pragmatic enrichment.² The same holds for quantifiers like *everyone*, which require pragmatic enrichment in order to determine the domain to which they apply, and for time-distance expressions like *some time*, who may have been lexicalized in the sense of "considerably". Thus the examples of the first experiment of Gibbs & Moise (1997) would not contain paraphrases of minimal propositions and consequently, the alternative "minimal vs. enriched" would be void in this case.

The second objection aims at the analysis of inalienable possession-sentences, which in most cases do not give rise to the enriched interpretation being indicated. Indeed it is not plausible that utterances like "He was walking a dog yesterday" or "She was weeding in a garden" should express inalienable possession such as "He broke a finger yesterday." This might explain the results of the tests by Gibbs & Moise (1997), which amount only to 57% for enriched choices in the first and 59% in the second experiment. The authors themselves admit in a footnote that nonpossession in the dog and garden cases could be the reason for these unfamiliar results (s. Gibbs & Moise 1997:59). Indeed the treatment of this type of possession-sentences in the context of Gibbs and Moise's argument is surprising, bearing in mind that the Gricean analysis of the use of determiners like "a" pointed to the contrary, i.e. the rise of a non-possession reading via a generalized conversational implicature (Grice 1989 a,b). Of course one may argue that in the given examples a possession-reading is expectable because as a rule one walks his own dog and weeds in his own garden. But this is by no means self-evident, considering that non-possessed dogs may be walked and non-possessed gardens may be weeded.

The third and main objection of Nicolle and Clark is that it is not clear whether the participants chose between the enriched reading and the conversational implicature (as was stated in experiment 3, s. Gibbs & Moise 1997:62), or whether they chose the reading which was the most plausible depending on the number of contextual effects in the sense of relevance-theory (s. Nicolle & Clark 1999: 345). Within their own experimental approach, they formulate the hypothesis that the participants will choose an implicature as preferred reading if it achieves the same set of cognitive effects as the original utterance, given the same set of contextual assumptions – which is opposed to the hypothesis of Gibbs & Moise (1997), claiming that people do not select an implicature in preference of an enriched explicature as something what-is-said.

^{2.} With this argument, she anticipates the distinction between two processes within the domain of what-is-said which has been drawn by Récanati (2004), i.e. saturation (necessary for propositionality) and free enrichment.

In their experiments, Nicolle & Clark demonstrate that the implicature-based interpretation for what-is-said may well be chosen in the case of a suitable context. The first experiment corresponds to the third experiment of Gibbs & Moise, but additionally ten further stimuli have been presented in such a way that implicaturebased paraphrases were strongly suggested (Nicolle & Clark 1999: 346). Beneath the question what was said or what the speaker's words meant, participants had been asked what the speaker wanted to communicate (Nicolle & Clark 1999:346). Under all three conditions, people chose the implicature as the best paraphrase (mean 79%). Even under conditions in which people were given paraphrases beginning with ""S said that ..." on all levels (minimal, enriched, implicature plus one false paraphrase) - this was the design of the second study - they voted for implicatures in four of six sentence types (except the false paraphrase which has been rejected). In the third experiment, the participants were instructed to distinguish between the technical usages of the phrases "what-is-said" and "what-is-communicated" (in the sense of "corresponding to the meaning of the words" and "going beyond the meaning of the words"). After this instruction, the participants were able to assign the paraphrases in line with their instructions, i.e. they identified enriched meanings as what-is-said (mean 87%) and implicatures as what-is-communicated (s. Nicolle & Clark 1999: 349). This indeed shows that the participants were able to distinguish between the levels of analysis of the utterances in question, but that the choices made in everyday-contexts vastly ignored the limits of saying and implicating. The consequences Nicolle & Clark are drawing are critical for the design of experimental settings, especially concerning the choice of example utterances and the relation between the predicted and the factual results of the studies.

Thus the authors suggest that the significant difference between their own studies and those of Gibbs & Moise results at least in part from the degree of determination of the implicature-based paraphrases. If the utterances presented to the participants were connected with a range of possible implicatures, then enrichments were chosen as preferred interpretations of what-is-said (e.g. (5b) "Jane has exactly three children ..."). The set of possible implicatures ("Jane is married", "You should like children", ...) is indeterminate and is therefore not suited for being treated as what-is-said. If, as in the examples of Nicolle & Clark (1999), the implicature is relatively determinate, then this is the first choice for what-is-said. This is plausible, because in a setting in which the question arises whether there are enough footballs for two playing teams, the answer

(6) Billy's got two footballs

is more relevant with the interpretation

c. There are enough footballs to play two matches (i.e. the implicature) as with

- b. Billy has exactly two footballs (enrichment)
- or as
- a. Billy has two footballs and possibly more (sentence meaning).

(s. Nicolle & Clark 1997:351)

The authors follow from this that subjects do not focus their attention upon the meaning of the word *said* in the context of the experimental studies, but that they "try to work out the overall communicative intention behind the utterance" (Nicolle & Clark 1997:351). This might explain the difference in the outcomes of the studies of Gibbs & Moise and Nicolle & Clark.

2.4 The status of implicatures

The discussion between Gibbs & Moise (1997) and Nicolle & Clark (1999) - which is followed by Gibbs (1999) - points to important aspects of experimental approaches to pragmatics in general and the difference between layers of utterance-meaning in particular. A general conclusion one may draw from the results is that the context in which the utterances are given is crucial for possible assessments of the participants about what-is-said. Generalizing the debate reported so far, one may say that the reading which has the highest degree of obligation for the speaker is the one which has been chosen as what-is-said. If it becomes apparent (in example (6)) that one of the footballs Billy has is not suited for playing, the speaker will be charged for having mislead his audience. This is an argument for (c) as the preferred reading and not for, e.g. (a). If, in example (5), Jane is not married, this may not lead to the same consequences for the speaker, since there may be other reasons which may turn out to be relevant for the shy guy who is interested in dating with her. This may be an argument for (b) as the preferred reading in this case. Despite the difficulties of operationalizing this criterion, it seems to be the one which guides the decisions of the interlocutors when they are calculating what has been said with an utterance - in contrast to what has merely been communicated or implicated.

Despite this, one should bear in mind that implicatures as they are defined by Grice do have an essential character of indeterminacy, and that's why they are chosen in the cases which are illustrated in the Gricean examples. If they are built in the domain of what-is-*said* by an utterance, this indeterminacy is no longer preserved. This means that Nicolle and Clark are advocating another notion of implicature, different from that defined by H.P. Grice (s. 1989a,b).

In his reply to Nicolle & Clark, Gibbs argues that the decisive criterion of the authors, that of contextual effects, has not been defined independently, so that their approach is simply unfalsifiable (Gibbs 1999: 357). The hypothesis concerning the identity of contextual effects would explain the performance of the participants, no matter what they actually did. Thus, as Gibbs holds, the relevance-theoretic

approach as it has been fostered by Nicolle & Clark is not suited for experimental studies like those performed by Gibbs & Moise (1997).

In addition, there is another conjecture which can be made, concerning the notion of intended contextual effect. If this notion is being defined in terms of conversational implicatures, then there is a strong tie between the nature of intended contextual effects and the interpretation of utterances with respect to what-is-said. The reason why in some cases preferred readings are implicatures is not that they have (a certain amount of) contextual effects. Rather, the reason why they figure as intended contextual effects is that they are implicatures. That is, at least in part the results of the study are a confirmation of the decision token in advance to *define* readings resulting from contextual reasoning as implicatures. This does not destroy the value of the experiments as a whole, but it is nevertheless a factor which influences some of their results.

It is worth noting that Nicolle & Clark too are based on the paradigm of relevance theory. In another respect, Gibbs seems to have reacted on the critique of Nicolle & Clark (1999) in that he has taken into account the context delivered to the participants in more recent studies. In (2004), Gibbs reports of a series of experiments aiming at recovering the speed with which people understand expressions. In one case, speakers' communicative intentions had to be associated with what-is-said, in another the communicative intentions had to be recovered by conversational implicatures. In this case, the same utterance was presented with a narrative background which suggested only a pragmatically enriched meaning, and with another narrative background which suggested a reading as conversational implicature. The critical utterance was.

(7) I drive a sports utility vehicle.

In one case (suggesting what-is-said), (7) was the response to a question posed by a friend of the utterer, who wanted to by a new car and was not sure which one to buy. In the other case, (7) was a kind of assertion that the vehicle is a good one to drive facing the danger of an upcoming storm in the area in which a trip is planned. The intended implicature in this case is that sports utility vehicles are suitable for making such a trip. The result of the study was that participants needed significantly less time to comprehend utterances in which "what speakers mean is identical to what they pragmatically say than to understand messages in which what speakers say underdetermines what they mean" (Gibbs 2004:66) – that is in the case of implicatural reasoning.

3. The method of the study

In order to catch the pragmatic intuitions of language-users about the limits of what-is-said, an experimental design has been chosen by presenting German

sentence tokens to native speakers of German with the aim of testing their intuitions. Thus, some of the standard examples in the literature have been presented to 42 undergraduate students which had not been trained in pragmatics before. They have been asked to judge whether the given paraphrases were part of what is said or whether it would be part of what is intimated, or whether it cannot be accepted at all as a paraphrase of the given utterance. The division between what-is-said and what-is-implicated is one of the best established in the pragmatic literature, but the terminology is not very common in everyday language use, in particular the notion of an implicature. For this reason an instruction was given to the participants to the effect that sometimes speakers/ hearers distinguish between what they explicitly say on the one hand and what they intimate (*was sie andeuten*) or give to understand (*was sie zu verstehen geben*) on the other, and that their task consisted in assessing the given paraphrases as to their communicative status, i.e. as something what is said or intimated. Of course there are other candidates for the "translation" of 'implicature' into normal language use, but the option which had been chosen does not exclude any important reading of the given utterances. In most cases of the given examples we would say that the message communicated by an implicature is one to which the predicate in question (das Angedeutete) is applicable. But of course language specific expressions for an inference based meaning component may and should be tested additionally. This however cannot be achieved within the realm of this study.

In the study performed, two paraphrases had been offered to the participants and they had been asked to assign one of the communicative statuses to them. This study contained examples with temporal or local relations on the one hand and the use of quantifiers in connection with scalar implicatures on the other. In the latter case it seemed to be important to test e.g. the "at least three" or the "exactly three"-version of the use of the number in question – bearing in mind Carston's (1988) claim that there is no minimal meaning for quantifiers. The idea behind this was that these two readings could be weighed against each other with regard to the tendency of the participants to classify them as said or implicated.

The experiment contained the following 8 items. They were of the type:

(Möchtest du etwas essen?) Nein danke, ich habe schon gefrühstückt. (Do you want something to eat?) No thanks, I've had breakfast.

- P1 a: Nein danke, ich habe heute morgen schon gefrühstückt. *enriched proposition.*
 - No thanks, I've had breakfast this morning.
- P1 b: Nein danke, ich habe früher schon einmal gefrühstückt. *minimal sentence meaning.* No thanks, I've had breakfast earlier in my life. (s. Appendix).

The percentages were as follows (the absolute number is given in brackets):

	Said	Intimated	Non-acceptance
P1 a	52,4 (22)	45,4 (19)	2,2 (1)
P1 b	33,3 (14)	38,1 (16)	28,6 (12)
P2 a	18,6 (8)	39,5 (17)	41,9 (18)
P2 b	67,4 (29)	30,2 (13)	2,4 (1)
P3 a	9,5 (4)	26,2 (11)	64,3 (27)
Р3 b	47,6 (20)	50,0 (21)	2,4 (1)
P4 a	14,3 (6)	35,7 (15)	50,0 (21)
P4 b	62,0 (26)	38,0 (16)	0
P5 a	4,8 (2)	23,8 (10)	71,4 (30)
P5 b	23,8 (10)	33,3 (14)	42,9 (18)
P6 a	19,0 (8)	16,7 (7)	64,3 (25)
P6 b	43,0 (17)	52,3 (22)	4,7 (2)
P7 a	9,6 (4)	40,4 (17)	50,0 (21)
P7 b	71,4 (30)	21,4 (9)	7,2 (3)
P8 a	19,2 (8)	16,6 (7)	64,2 (27)
P8 b	42,9 (18)	52,7 (22)	4,4 (2)

Table 1. Results of the experiment

4. Discussion of the results

I will comment on four of the 8 given examples in order to clarify the idea behind the approach. Of course it may be doubted whether the intuitions of everyday language users are precise enough to capture a clear cut distinction between the layers of the said and of the intimated – but they must be sufficiently clear in order to separate the levels of utterance-meaning in the above sense. The ability to identify what was said by an interlocutor is part of our basic pragmatic competence, on the basis of which we organize our talk exchanges. So we react differently upon a meaning component that has been insinuated by an utterance, opposed to a meaning component that has been meant by saying something. Not to possess this ability equals to not being able to communicate successfully. The results of the studies presented here – and the outcomes of the tests performed by the authors presented in the last paragraph – are rather encouraging in favour of the idea that knowledge of the difference between the said and the insinuated is part of the pragmatic competence of language users.

Consider the breakfast example 1 (s. Appendix). This classical example is discussed *inter alia* by Récanati (2004:8), who states that from a minimalist point of view the proposition that the speaker has had breakfast before time t*is expressed (= I've had breakfast in my life). This claim results from the idea that what-is-said consists of nothing more than of the sentence-meaning plus indexical resolution and reference assignment. Intuitively the linguist would say that this is not what the speaker meant, but that the meaning is more specific, namely that S had breakfast on that very day (the day which includes t*). This is what is more or less confirmed by the test. Facing the paraphrases P1a and P1b, the readers were asked to decide whether the paraphrases were about something which had been *said* by uttering B or something which had been *intimated* by the utterance of B. A majority voted for P1a as having been said by B (52,4%), but a rather strong minority voted for P1b as having been said (33,3%). On the other hand, 28,6% did not accept P1b as a paraphrase at all.

Contrastive to this example, the next item 2 (see Appendix) focuses a different action type: sailing. In this case, the distribution of answers is more clear cut: 18,6% classified P2a as something what has been *said*, 39,5% as *intimated*, 41,9% did not accept this as a possible paraphrase at all; concerning P2b, 67,4% accepted it as having been said, 30,2% as having been insinuated, 2,4% did not accept *b* at all. Thus, in this case, the lifelong reading was accepted by a majority, which may be explained by their lexical semantic knowledge concerning the verb 'sailing' in addition to the context, which provokes a reading of 'sailing' which focuses on the ability to sail and not on a single action. This in turn is not possible with the breakfast example.

Example 6 ("party") includes the use of quantifiers. In this case, 19% accepted version P6a as a paraphrase of what is *said*, 16,7% as *intimated*; 64,3% did not accept this reading at all. 43,0% accepted P6b as *said*, 52,3% as *intimated*, and 4,7% did not accept the reading. In the case of paraphrase b, which was accepted by most of the participants, we do not have any significant result. This might be an outcome of Carston's observation that quantifiers lack a minimal reading, i.e. the interpretation hinges severely upon the utterance-context.

Example 8 is about metaphorical language use ("swallow"). The majority of the answers concerning paraphrase P8a was non-acceptance (64,2%), the majority of answers concerning P8b classified it as having been *said* (42,8%).

What we can see with these four examples is that there are intuitions on the side of language-users regarding the extent to which one would classify something *uttered* as something *said*. Thus, an account which relies on these intuitions is not condemned to voidness, as Levinson claims. Rather, there is an opportunity for testing intuitions by means of elicited data. How precise they are, is open to debate. Sometimes, however, such intuitions enable speakers to distinguish between meaning-levels which clearly belong to the realm of *what-is-said* and those which do not in a clear cut way.

Let us start with Example 8 ("swallow"). The sentence which had been presented contains a conceptual conflict leading to a metaphorical interpretation. Since the activity of swallowing can only be performed by animate beings, we have two possible strategies in order to appease this conflict – interpret "bank" metonymically (in the sense of "its employees") or interpret "swallow" metaphorically (in the sense of "withdraw"). In the case in question the decision is clear; the metonymical interpretation of "bank" is ruled out, the metaphorical interpretation of "swallow" is the only one which makes sense. This decision isn't always so easy. As Récanati points out, there are cases in which some sort of "trade-off" is going on concerning the question which constituent has to be interpreted literally or non-literally. Thus with.

(8) The city is asleep.

Récanati notes "[...] if 'the city' applies literally to a city, 'asleep' will be taken nonliterally; conversely, if 'asleep' is literal, 'the city' will not be." (Récanati 2004: 34) In each case, the truth conditions are distinct: in one case "the speaker means that the inhabitants of the city are sleeping, in the other she means that the city itself is quiet and shows little activity" (35). In Example 8, the case is clear, however there is a naturalistic retention of the metaphor in that the card really disappeared with no return – so a kind of swallowing has been going on. If we decide to follow this path, "the bank" has to be interpreted metonymically in another respect – the machine standing for the whole bank. This interpretation however is less attractive because it is not the physical process which is crucial here but the institutional fact of confiscating the card.

If we suppose that "swallow" is interpreted metaphorically in order to save the coherence of the sentence, the crucial question is whether this is a reason for interpreting the whole utterance as something which has not been *said*. 42,8% treat paraphrase P8b as being about what has been *said*, 52,3% classify this as an *insinuated* reading, opposed to what-is-said. Thus, looking at the majority, we may question Récanati's position that metaphorical readings of constituents of uttered sentences do not lead to a non-said-reading, but that they have to be classified as what has been *said*, in spite of their metaphoricity. Now let us have a look on the other tests.

If we compare Examples 1 and 2, the result of the latter seems to be clearer than that of the former. The questions of the scenarios 1 and 2 have tendencies which are opposed to each other: The question in 1 has been understood by the majority as aiming at a "this-morning"-reading, the question in 2 focusses a context obviously provoking a "life-span-reading" (earlier in my life ...). Now in the case of 1, we have 52,4% voting for P1a as *said* ("this-morning"), but also 33,3% voting for P1b as *said* ("in my life"). The result of the test is somewhat mixed, paraphrase b is not completely ruled out as a candidate of what has been *said*. On the other hand, in the case of 2, the result is more clear-cut. P2a has

18,6% said as against 67,4% for P2b, and it has 41,9% non-acceptance as against 2,4% non-acceptance for P2b. Why is this the case – i.e. why are the results more clear cut in the case of a preferred life-span-reading than in the case of a nonce reading? Obviously an answer preferring a life-span-reading is more accessible than an answer which is directed towards an everyday-reading, and speaker-intuitions concerning life-span-readings are stronger than intuitions in favour of an everyday-reading of utterances. The reason for this seems to be that the sentence-meaning is in concord with the life-span-reading, such that it contributes easier to the final reading of the utterance than in case of the everyday-reading. From this in turn follows that sentence meaning does in fact play a strong role in determining the total signification of the utterance. If we accept this as a tentative outcome of the tests, we arrive at a position which is more differentiated than the idea that the minimalistic construction of (a part of) what-is-said determined by sentence-meaning is not accessible - this in fact being open to Récanati's critique. Sentence-meaning has an influence upon the tendency of addressees to judge paraphrases as something what-is-said - they are stricter in their judgement if the paraphrase has a close relation to sentence meaning. This interpretation reflects the important role of sentence meaning without making statements about inaccessible counterintuitive readings of utterances in everyday conversation.

The result of 6 is remarkable in that the alternative for the readers is either non-acceptance or intimated, the said being in the focus of 19% of the participants. 64,3% did not accept P6a at all, and 52,3% classified P6b as intimated. This points to the fact that the non-enriched reading of the quantifier 'all' is barely considered as playing any role in the calculation of the overall meaning of the sentence. In contrast, the qualified reading 'all invited people' is accepted as the basis for interpretation, but it has to be noted that the qualification as having been *said* is not a central category for the participants in this test. This is caused by the blatant absurdity of paraphrase P6a which relies on the unqualified application of the quantifier. If we want to avoid the unaccessible-meaning trap, we have to assume two readings of "all": one reading relevant for propositional logic and one reading relevant for natural language use, which is indexical and which has to be "completed" by contextual information.

5. Summary

The debate about the definition of what-is-said and its delimitation against implicatures has shown that intuitions of language users encompass the ability to ascribe a level of utterance meaning which has a set of obligations for the utterer. As a rule, this is not the level of implicatural meaning, but as has been shown by Nicolle & Clark (1999), this may be in some cases the crucial level. In transferring the studies performed with English examples into German, the results of the preceding studies could be confirmed in the main. That is a certain argument that the results of the experimental approaches may be crosslinguistically confirmed. A remarkable point is that in comparing the results of 1 and 2, a certain drift towards the literal meaning may be seen in that the answers in the latter case have been more firm, more sure than in the former.

Of course, questions about the category non-acceptance may arise. The interpretation of these answers is not determinate, meaning either that the utteranceinterpretation is not acceptable, or that the suggested dichotomy itself is not acceptable. Thus in a further study one has to investigate whether the participants of a study do accept the discrimination between the said and the implicated at all or in the suggested terminology. Some experiments performed by Gibbs & Moise and Nicolle & Clark are dealing with this question, but further inquiry has to be undertaken in order to clarify the acceptance of suggested distinctions in the experimental setting.

References

- Bach, K. 1994. Conversational impliciture. Mind and Language 9: 124-162.
- Carston, R. 1988. Implicature, explicature, and truth-theoretic semantics. In *Mental Repre*sentations. The Interface Between Language and Reality, R.M. Kempson (ed), 155–181. Cambridge: CUP.
- Grice, H.P. 1989a. Logic and conversation. In *Studies in the Way of Words*, P. Grice, 22–40. Cambridge MA: Harvard University Press.
- Grice, H.P. 1989b. Further notes on logic and conversation. In *Studies in the Way of Words*, P. Grice, 41–57. Cambridge MA: Harvard University Press.
- Gibbs, R.W. 1999. Speakers' intuitions and pragmatic theory. Cognition 69: 355-359.
- Gibbs, R.W. 2004. Psycholinguistic experiments and linguistic-pragmatics. In *Experimental Pragmatics*, I. Noveck & D. Sperber (eds), 50–71. Basingstoke: Palgrave.

Gibbs, R.W. & Moise, J.F. 1997. Pragmatics in understanding what is said. Cognition 62: 51-74.

- Nicolle, S. & Clark, B. 1999. Experimental pragmatics and what is said: A response to Gibbs and Moise. *Cognition* 69: 337–354.
- Kriempardis, E. 2002. Die Klassifikation von Illokutionen anhand interpersonal relevanter Kriterien. *Sprache und Sprachen* 27(28): 68–81.
- Levinson, S. 2000. Presumptive Meanings. Cambridge MA: The MIT Press.
- Noveck, I. & Sperber, D. (eds). 2005. Experimental Pragmatics. Basingstoke: Palgrave.
- Noveck, I. & Sperber, D. 2007. The why and how of experimental pragmatics: The case of 'scalar inferences'. In *Pragmatics*, N. Burton-Roberts (ed.), 184–212. Basingstoke: Palgrave.
- Récanati, F. 1989. The pragmatics of what is said. Mind and Language 4: 295-329.
- Récanati, F. 2004. Literal Meaning. Cambridge: CUP.
- Searle, J.R. 1979a. Literal meaning. In Expression and Meaning, Studies in the Theory of Speech Acts, J.R. Searle, 117–136. Cambridge: CUP.

Searle, John R. 1979b. Indirect speech acts. In Expression and Meaning, Studies in the Theory of Speech Acts, J.R. Searle, 30–57. Cambridge: CUP.

Sperber, D. & Wilson, D. 1986. Relevance. Cognition and Communication. Oxford: Blackwell.

Appendix

- (Möchtest du etwas essen?) Nein danke, ich habe schon gefrühstückt. (Do you want something to eat?) No thanks, I've had breakfast.
 - P1 a: Nein danke, ich habe heute morgen schon gefrühstückt. No thanks, I've had breakfast this morning.
 - P1 b: Nein danke, ich habe früher schon einmal gefrühstückt. No thanks, I've had breakfast earlier in my life.
- 2. (Musst du einen Segelkurs besuchen?) Nein, ich bin schon gesegelt. (Do you have to take sailing lessons?) No, I have sailed.
 - P2 a: Nein, ich bin heute morgen schon gesegelt. No, I have sailed this morning.
 - P2 b: Nein, ich bin früher schon einmal gesegelt. No, I have sailing experience.
- 3. (Ein Kind hat sich verletzt, die Mutter sagt:) Du wirst nicht sterben. (*A child has harmed him/herself. His/her Mother said*) You will not die.
 - P3 a: Du bist unsterblich. *You are immortal.*
 - P3 b: Du wirst von der kleinen Wunde nicht sterben. *You will not die because of this little wound.*

4. Es regnet.

It is raining.

- P4 a: Es regnet an irgendeinem Ort der Welt. It is raining at some place in the world.
- P4 b: Es regnet hier. It is raining here.
- 5. Der Tisch ist voll mit Büchern. *The table is covered with books.*
 - P5 a: Der in der Welt einzig existierende Tisch ist voll mit Büchern. *The only existing table in the world is covered with books.*
 - P5 b: Der Tisch in der Bibliothek ist voll mit Büchern. *The table in the library is covered with books.*
- 6. Alle kamen zur Party.

Everybody went to the party.

- P6 a: Alle Menschen der Welt kamen zur Party. All existing human beings went to the party.
- P6 b: Alle Eingeladenen kamen zur Party. All invited people went to the party.
- 7. Helmut hat drei Kinder. *Helmut has three children.*
 - P7 a: Helmut hat mindestens drei Kinder. Helmut has at least three children.
 - P7 b: Helmut hat genau drei Kinder. Helmut has exactly three children.
- 8. Die Bank hat meine Kreditkarte geschluckt. *The bank swallowed my EC-card.*
 - P8 a: Die Bank hat meine EC-Karte gegessen. *The bank ate my EC-card.*
 - P8 b: Die Bank hat meine EC-Karte eingezogen. *The bank confiscated my EC-Card.*

Discourse under control in ambiguous sentences

Vincenzo Moscati University of Siena

Sentences with more than one logic operator may be ambiguous between different interpretations and a recent research question is whether children have access to all the possible readings available to adults. Early studies focused on the interpretation of nominal quantifiers and negation, suggest that children may have only a subset of the possible meanings. In this paper, we extend the inquiry to the interactions between modals and negation. We report on two experiments. In the first, we tested children's interpretation of the Italian modal *potere* followed by clausal negation, a construction which results unambiguous in the adult language and not compatible with a reading expressing impossibility. In a second experiment, we manipulated the context in order to evaluate the effects of the Question Answer Requirement (Husley et al. 2004) on children's scope assignment.

1. Introduction

A recent topic in language acquisition is children's comprehension of sentences containing multiple scope-bearing elements. The main problem posed by these sentences is that the relative scope of each operator might be reversed, creating ambiguity between the two possible interpretations of a single sentence.

Sometimes languages have means to reduce ambiguity, as in negative sentences, where elements such as quantifiers present a specification relative to polarity. However, this information also needs to be acquired and, even if there is evidence that children are sensitive to polarity from early on (Thornton 1994), several studies have shown that under certain circumstances, children interpret multiple-operator sentences differently from adults. Initial findings by Musolino (1998) and Musolino et al. (2000) suggested that in an early developmental stage, children misinterpret sentences (1), assigning to the quantifier *some* narrow scope, as in the non-adult reading (1b).

- (1) the detective didn't find some guys
 - a. There are some guys that the detective didn't find
 - b. It is not the case that some guys were found

This observation lead to the proposal that polarity may be overruled by other principles active on early grammar. However, such observation of an *isomorphic* LF-PF mapping has been revealed to be too strong and more recent studies (Gualmini 2004; Husley et al. 2004; Krämer 2000) challenged the conclusion that surface relations determine scope assignment. According to Husley et al. (2004), structural relations are not the only way to look at the problematic meanings and the link with the discourse has also to be taken into account. According to this view, Isomorphism is a by-product of the violation of certain discursive requirements which have not been considered in the experimental setting.

Additional evidence against the hypothesis of an isomorphic structural mapping between LF and PF comes from the study of modality. Extending the domain of inquiry to modal verbs, Moscati and Gualmini (2008) looked at children's interpretations of sentences as (2) and (3) below:

- (2) The red ball cannot be with the yellow ball
 - a. it is not possible that the red ball is together with the yellow ball
 - b. *it is possible that the red ball is not together with the yellow ball
- (3) L' Indiano non deve cavalcare l' ippopotamo the Indian not must ride the hippopotamus
 a. it is necessary that the Indian does not ride the hippopotamus
 - b. it is not necessary that the Indian rides the hippopotamus

The inverse scope readings for sentences (2) and (3) are given in (2a) and (3a). Such readings are fully target consistent and they are perfectly licit in the adult language. In the case of (2), the inverse-scope reading (2a) is the only one allowed in adult English while sentence (3) -with plain intonation- is instead ambiguous in Italian. According to the prediction of Isomorphism, readings (2a) and (3a) should be problematic. Contrary to this expectation, Moscati and Gualmini (2008) found that children do not have problems with these interpretations.

Taken in isolation, these results suggested that children already have an adultlike competence with respect to the interpretation of modals in negative sentences. However, the picture is more complex and in sentences similar to (4), from Moscati and Gualmini (2008b), children incorrectly select the meaning (4b), which is not allowed in adult English:

- (4) you need not feed the zebra
 - a. it is not necessary to feed the zebra
 - b. *it is necessary not to feed the zebra

Although this sentence is unambiguous and its only interpretation is (4a), children show a preference for reading (4b). This case, where the forbidden reading is selected, is reminiscent of the deviant interpretations with nominal quantifiers initially reported in Musolino et al. (2000). Following Husley et al. (2004), we cannot ignore the possibility that children choose this particular interpretation by virtue of some pragmatic factor and that the deviant reading (4b) will vanish under a contextual maneuver able to better fit interpretation (4a) within the overall discourse. The problem is how to capture discourse adherence of negative sentences. Many ideas have been proposed to capture the link between the sentence and its previous context (Rooth 1992; von Stechow 1991), however, in the rest of the paper we will adopt the Question-Answer-Requirement (QAR) proposed by Husley et al. (2004), which seems to be a sufficiently explicit model to make empirical predictions. We will discussed it in the next section.

In Section 3 new data from Italian are presented, showing the existence of children's misinterpretations similar to those seen for sentence (3). This case will be used in a successive experiment (Section 3) to test the predictions of the QAR.

2. Satisfaction of the Question-Answer-Requirement

The general view that extra-linguistic context and subject expectations may favor particular interpretations is relatively uncontroversial. Early studies by Wason (1965) show that a negative sentence requires an additional processing with respect to a positive declarative. However, Wason also showed that the processing load required by negation consistently diminishes if an appropriate visual context is provided. This supports the idea that the logic processor is sensitive to the background and that discrepancies between a linguistic stimulus and the overall context might affect the computation of the meaning.

What we can learn from this is that, every time we want to test children's comprehension of negative sentences, we should be aware of the extra work required by their cognitive system. In particular, a given proposition should be made salient by the context in order to facilitate the processing of its negation. This experimental requirement has been expressed also by Crain & Thornton (1998) through the need to satisfy the *condition of plausible dissent*: if we want to test the comprehension of a negative proposition $\neg p$, its processing will be easier if children have already considered *p*.

The position expressed in Crain & Thornton (1998) and in Wason (1965) privileges a view of the background as a priming factor. However, in cases where

negation might originate ambiguous interpretations as in (4), it is unclear how the context might favor one interpretation over the other. It seems that we need a way to evaluate competitive interpretations along some other dimension.

Since Grice (1975), it is a fairly uncontroversial assumption that an utterance must be maximally informative. This consideration is particularly relevant in the case of ambiguous sentences and we might expect that the more informative – the more contextually relevant – interpretation should be preferred. This intuition can be captured if we find a way to assess how the possible interpretations of an ambiguous sentence fit the context.

A recent proposal in this direction has been put forward by Husley al. (2004). This proposal is based on the assumption, common to many theories of communication, that every assertion has to be informative with regard to a salient question. If a sentence is perceived as out-of-topic or redundant, this sentence does not constitute a direct answer to the Question Under Discussion (QUD). The felicity of a sentence can then be evaluated through the Question-Answer-Requirement by its capacity to satisfy the salient QUD.

Going back to logically ambiguous sentences as (1), we may evaluate what is the role of the QAR in guiding children's interpretation. Musolino (1998) claimed that children, in an early developmental stage, are unable to access the adult meaning of (1). However, more recent studies have shown that this conclusion is too strong. In one experiment reported in Gualmini (2004), children were divided into two experimental groups and each group heard a sentence with negation and the indefinite *some*. The story context was the same for the two groups and it was about a firefighter playing hide-and-seek with some dwarves: at the end of the story the firefighter found only two out of the four dwarves. Each group heard either sentence (5) or (6):

- (5) the firefighter didn't find some dwarves
- (6) the firefighter didn't miss some dwarves

The test sentences only differed in the choice of the lexical verb and they were truth conditionally equivalent since two dwarves were found and two were missed. However, Gualmini found that while children accepted sentence (5) around 90% of the time, they only accepted (6) at chance level. The difference in the acceptance rate for (5) and (6) can be accounted for if we consider how well the two sentences answer the QUD primed by the story. Children know that the goal of hide-and-seek is to find everybody. Therefore, the question has to be the following:

(7) Did the fire-fighter find all the dwarves?

Now consider how the different interpretations of (5) and (6) answer the QUD in (7). Let us look first at (5):

QUD: Did the firefighter find all the dwarves?

the firefighter didn't find some dwarves

- a. There are some dwarves that were not found \rightarrow NO (relevant answer)
- b. *It is not the case that some dwarves were found →NO (relevant answer)

Here both interpretations are good (negative) answers to the QUD and the discourse does not pose any special pressure on the deviant non-isomorphic interpretation (5b). Children then accept the adult reading (5a) which is true. Consider now the case of (6):

QUD: Did the firefighter find all the dwarves?

the firefighter didn't miss some dwarves

- a. There are some dwarves that were not missed \rightarrow ? (uninformative answer)
- b. *It is not the case that some dwarves were missed →YES (relevant answer)

The adult reading (6a) only says that some of the dwarves were not missed: here we do not know if all of them were actually found. This reading is not relevant and it violates the QAR. In contrast, the non-adult reading (6b) is more informative, since it says that no dwarf was missed and it constitutes a good (affirmative) answer to the QUD in (7): here the context forces the reading (6b), even if this is not allowed in adult English.

It seems that children fail to correctly interpret (1) only in special cases, namely when the adult interpretation does not satisfy the QAR. What remains to be explained is the difference between children and adults where the requirement of the QAR are not fulfilled. The answer, according to Husley et al. (2004), is that children are more sensitive than adults to the contextual pressure. While adults might accommodate the sentence to the QAR by exploring different QUDs, children fail to backtrack their initial assumptions and they are committed to the first/salient QUD.

The QAR makes predictions that are not limited to the case of nominal quantifiers. In the rest of this paper, we will try to determine if children are also sensitive to the QAR in negative modal sentences.

3. Non-adult interpretations with modals

In principle, any sentence with more than one logic element might be ambiguous between two interpretations and modal verbs, when combined with sentential negation, are no exception. If we look at languages whose modal system presents underdetermined forms with respect to the scope of negation, we can find cases of real ambiguity,¹ as in sentence (3), repeated here as (8):

- (8) L' Indiano non deve cavalcare l' ippopotamo
 - the Indian not must ride the hippopotamus
 - a. it is necessary that the Indian doesn't ride the hippopotamus
 - b. it is not necessary that the Indian rides the hippopotamus

In (8) the modal *dovere* 'must' takes either wide or narrow scope with respect to negation and interpretations (8a–b) are both possible.

In one experiment, reported in Moscati & Gualmini (2008a), the predictions of isomorphism were tested against sentences as (8). If children lack covert logic operations, able to reverse the surface scope of the two operators, we expect that reading (8a) should be hard to access. This prediction was not confirmed and children were able to access the meaning (8a) in the 93% of the cases, showing no problem with such interpretation.

In another experiment, Moscati & Gualmini (2008b) also tested the comprehension of unambiguous sentences as (9).

- (9) you need not feed the zebra
 - a. *it is necessary not to feed the zebra
 - b. it is not necessary to feed the zebra

finding that English children accepted the adult reading (9b) only in the 30% of the cases, preferring to interpret this sentence with wide scope of the modal over negation (9a). This interpretation, equivalent to the one selected by Italian children in (8a), is forbidden in adult English. How can we explain this result? A first possibility is that readings in (8–9a) have a special status in early grammar and that they are favored over the alternatives (8–9b). This idea is conform to the Semantic Subset Principle proposed in Crain et al. (1994) and it is grounded on the fact that those interpretations are stronger in that they have more restrictive truth-conditions if compared with (8–9b). In fact, (8–9a) asymmetrically entail readings (8–9b):

(8-9a) necessary > not \rightarrow (8-9b) not > necessary

However, it is also possible that the preference found for the non-adult reading (9a) is due to some language specific properties relative to construction (9) in English. For this reason, before formulate hypotheses about the semantic/pragmatic

^{1.} Intonation can make one of the two interpretations more prominent. However, with a plain intonation, both interpretations are equally possible.

properties of negative modal sentences, it is necessary to determine to what extent the errors found in child language are amenable to be explained trough a languagespecific account. A way to check this, is to look at a meaning equivalent to (9a) in a different language. If analogous problems are found, the language-specific explanation becomes less plausible.

3.1 Experiment I: Children's interpretation of *potere* and negation in Italian

The combination of a modal verb with negation might be unambiguous also in Italian, for reasons which presumably are not related with polarity.² In sentences where negation appears between the modal and the non-finite lexical verb, it can only receive narrow scope with respect to the modal, as in (10):

- (10) Il contadino può non dare le carote all' elefante the farmer can not give the carrots to-the elephant
 - a. *it is not possible that the farmer gives carrots to the elephant
 - b. it is possible that the farmer doesn't give carrots to the elephant

this sentence only allows reading (10b), expressing the permission not to give carrots to the elephant. Now remember, from what has been said in the previous section, that English children find difficult to access the weaker permissive reading in sentences as (9). A plausible hypothesis is that also Italian children have problems in the understanding of (10b), given that this meaning is logically equivalent to the one in (9a):

(11) $not > necessary \equiv possible > not$

If Italian children also misinterpret sentence (10), we have an indication that deviant interpretations are related to a particular meaning, the one in (11).

To verify this prediction, we tested 20 monolingual Italian speaking children (Age: 3;9 – 5;7. Mean 4;5) from two kindergartens in the Siena area using a Truth Value Judgment Task (Crain & Thornton, 1998).

Each child heard 4 sentences similar to the one in (10), after a short story as the following:

There is a farmer with some turnips and some carrots and he has one tiger and two elephants to feed. The farmer first tries to feed the tiger, but it says that it doesn't like carrots. Thus the farmer decides to give a turnip to the tiger. Then he tries to feed the

^{2.} The particular interpretation might be related to the properties of the "restructuring rule" in Italian (Rizzi 1982). Assuming that the logic scope of negation is clause-bounded and that lack of Restructuring will preserve the biclausal structure of the sentence, negation cannot take wide scope over the higher modal.

elephants. They say that they don't have any preference. The farmer decides to give a turnip to the first elephant and a carrot to the second one.

Given this story context, narrated with the help of props and toys, sentence (10) results True under the interpretation (10b), since there is one elephant without carrots. Conversely, since the other elephant received a carrot, it also follow that the reading in (10a) is False.

Two kinds of controls were also interspersed within the trials. In the first one (Type A), children were asked to judge a sentence after the following story:

There is a king who has some magic objects: two swords and a harp. Each sword gives the ability to run fast to whom is able to hold it, while the harp gives the ability to fly to whom is able to play it. Semola goes to the king and he tries first to play the harp. He has success and he is able to fly. Then he tries with the swords. He tries the first sword, but it is too heavy and he fails. He tries then with the second one, and this time he manages to lift it and he can run fast.

At the end of the story, children heard sentence (12).

(12) Semola può sollevare la spada. 'Semola can lift the sword.'

This control is included to test whether children tend to reject sentences with the modal *potere* and with a definite determiner in presence of two swords. Definite determiners were chosen also in the target condition in order to avoid the insertion of another scope-bearing element as an indefinite. Since this choice may be a source of confusion, we use the definite determiner in sentence (12) to check if rejections could be related to this factor.

In the second control (Type B), we gave children sentences with a negative operator linearly preceding the modal 'potere'. Again, we report below one example:

> There is a Pilot who wants to go out and he is torn between taking a motorbike or an F1 car. He decides to take the motorbike, but when he tries to climb on it, he falls down since it is too high. He tries again, but he falls another time. He changes his mind and takes the car.

At this point, the puppet uttered sentence (13):

- (13) Il pilota non può guidare la moto the pilot not can drive the motorbike
 - a. It is not possible that the pilot drives the motorbike.
 - b. *It is possible that the pilot doesn't drive the motorbike.

this sentence is true under both interpretations (13a–b). This second control has been included to test whether children reject sentences with negation and the modal for which the reading expressing *impossibility* was available in the adult language.

Let us now look at the results for this first experiment. In Table 1, we report the results relative to the target and the controls. Notice that the acceptance for the target sentences is 34.7%, considerably lower than controls of both types.

	Type 4	Type B	Target
Acceptance	33/40 (82.5%)	26/40 (65%)	25/72 (34.7%)

Table 1. Children's acceptance to controls and Target sentences

By looking at the data reported in Table 1, we also found that children accept Type B sentences only 65% of the times. Even if acceptance of Type B controls is sensibly higher than the one found in target sentence, it seems that negation itself may be a source of confusion. However, looking at the individual data, we found that this low acceptance is due to the performance of few children, which presented scarce attention to the task and have a strong tendency to answer 'wrong'. Notice that this is also the answer related to the experimental hypothesis and we may want to exclude the answers of those children which present traces of a negative bias. If we consider only the data relative to children correctly answering to at least one control for each kind (Table 2), these 15 children show no particular problems with controls of Type B, although the acceptance for the target sentence is about the same than the one reported in Table 1.

Table 2. Children with at least 50% of correct answers to each control

Condition	Affirmative False	Negative True	Target
Acceptance	28/30(93.3%)	26/30(86.6%)	19/55 (34.5%)

The conclusion is that also children that correctly process negation and modal in sentences expressing impossibility still have problems with the target sentences. This seems to confirm that the low acceptance found with sentence (10) is due to the particular interactions between negation and modality. Notice that, as an additional control, every time that children judged a sentence as False, we asked for explanation. Children always gave an explanation consistent with the deviant reading (10a), except in 5 cases, where the explanation was not based on logic scope (for example: *because elephants are strange*). Those answers are not included in Table 2. Moreover, we also test a control group (N=10) with adult subjects. They always accepted the target sentences.

Results from this experiment support the idea that the meaning in (10b) is harder to access, suggesting that the difficulties generated by this particular interpretation are not limited to English. A language specific explanation does not seem appropriate to account for our data and a semantic/pragmatic-based account seems to be more adequate. At this point, the next step is to determine whether it is the logic meaning of sentence (9) and (10) in itself to be harder. Alternatively, difficulties may be due to the efforts required to cast the particular meaning in the context created by the discourse. In the next sub-section, we will test the predictions of the QAR.

3.2 Experiment II: Evaluation of the Question-Answer-Requirement

The results of Experiment I show that children find problematic, also in Italian, to process modals in negative sentences when the meaning does not express impossibility but instead a weaker permissive reading. Isomorphism cannot be observed in this case, since children prefer the non-isomorphic interpretation even when this is not allowed by the adult grammar.

Sentence (9) and (10) share few syntactic properties and they differ with respect to the force and to the logic scope of the modal. A syntactic explanation in general does not seem to be adequate to account for children's deviant interpretation. Instead, the fact that the interaction between negation and the modal results in an equivalent logic representation (see 11) suggests that the problem has to be semantic, or pragmatic.

In fact, the meaning might not be problematic *per se* but might be problematic given the story context: we already saw that if the story favors a certain QUD, children tend to choose a salient interpretation even if not permitted by the adult grammar. For nominal quantifiers, children accept the adult interpretation of *some* only when this interpretation does not violate the QAR. Instead, whenever the sentence is not entirely appropriate in the discourse, children may accommodate the felicity conditions choosing the alternative interpretation.

This consideration may be relevant also in the case at issue here. In fact, we cannot exclude that children fail to access the permissive reading in (9) and (10) because it is not consistent with the QUD made salient by the story. The goal of this second experiment is to assess if the QAR biases children interpretation also in the case of modals.

A problem for the QAR account is that it is not easy to assess what is the actual question selected by the subject. A way to overcome this problem is to state explicitly the QUD, a solution which has been adopted in the experimental design.

In this experiment, we used the same 4 target sentences and the same 4 stories as in the previous one, but this time the target sentence was presented in response to a precise question. In order to evaluate the role of the QAR, we use two different kinds of questions: in the fist condition, the QUD made felicitous both interpretations, while in the second condition, only the deviant reading satisfied the QAR.

As in the case of nominal quantifiers, we expect that if both readings are appropriate (Condition 1), the adult reading will be selected, since this is the interpretation that minimizes violations. Conversely, when only the deviant reading satisfies the QAR (Condition 2), we expect a competition between this discursive requirement and the syntactic constraints. If the child decides to avoid violations of the QAR, he will select the pragmatically appropriate interpretation even if this violates the syntax.

The only difference between the first and the second experiment is that before the target sentence, now the puppet heard a question. Below an exchange is reported to illustrate the first experimental condition:

QUD1- First Experimenter:

(14) il contadino <u>deve</u> dare le carote all' elefante? 'Must the farmer give carrots to the elephant?'

ANSWER- Second Experimenter, manipulating the puppet:

- (15) Il contadino può non dare le carote all'elefante
 - a. *it is not possible that the farmer gives carrots to the elephant →NO (relevant answer)
 - b. it is possible that the farmer doesn't give carrots to the elephant \rightarrow NO (relevant answer)

In this condition, both interpretations represent a good (negative) answer to the QUD. Either the prohibitive reading (15a) or the permissive reading (15b) entail that it is not necessary to give carrots to the elephant: the QAR is satisfied by any interpretation of (15). Here, we expect that if children have the adult interpretation (15b) they will choose this reading. Since (15b) is true in the story context, children should accept the target sentence.

In the second condition, we varied the QUD in a way that only the deviant interpretation (15a) satisfies the QAR.

QUD2- First Experimenter:

(16) il contadino <u>può</u> dare la carote all' elefante?
'Can the farmer give carrots to the elephant?'

ANSWER- Second Experimenter, manipulating the puppet:

- (17) Il contadino può non dare le carote all'elefante
 - a. *it is not possible that the farmer gives carrots to the elephant \rightarrow NO (relevant answer)
 - b. it is possible that the farmer doesn't give carrots to the elephant \rightarrow ? (uninformative answer)

In (17) the modal *potere* 'can' was used instead of *dovere* 'must' in the QUD and here the only informative answer is (17a), expressing the prohibition to give carrots to the elephant. The other reading, the adult (17b), is uninformative since it only states the possibility not to give carrots to the elephant. The context created by the QUD2 adds discursive pressure, which will favor the non-adult

interpretation (17a). Since this interpretation is false given the story, children are expected to reject the target sentence. We then have a clear prediction: children will accept the sentence in the first condition but will reject it in the second condition.

The experimental procedure is that each child heard the same 4 stories as in the previous experiment, but this time he heard at the end of the story the QUD1 or the QUD2, randomized in a way that subjects heard two questions of each kind. The participants were 16 monolingual Italian-speaking children aged between 4;0 and 5;5 (mean 4;9).³

Looking at the results (Table 3), the acceptance rate for the two QUDs is nearly the same and it is even higher for QUD2. The predictions of the QAR are not confirmed and instead, we found a slightly higher acceptance in condition 2.

	QUD 1	QUD 2		
Acceptance	8/26 (30.7%)	10/25 (40%)		

Table 3. Acceptance for the target sentence in response to QUD1 and QUD2

Children in general show no sensitivity to variations of the QUD and the interpretation does not seem to be affected by the QAR. In the absence of contrary evidence, we can assume that the adult logic representation for the target sentences is harder to acquire and we have no evidence that at this developmental stage this reading is already available.

4. General discussion

Combining together the results of the two experiments illustrated in the previous sections, the conclusion is that in presence of logical ambiguity, children prefer the reading expressing *impossibility*. Even when this interpretation is not allowed in the adult grammar.

Moreover, the fact that this pattern is not confined to Italian but can be found also in English, indicates that an explanation which relies only on syntactic variations between the two languages cannot capture the substantial homogeneity of the results when the meaning is taken into account. A more general mechanism is required and a candidate might be the immature isomorphic mapping between LF/PF in child grammar. However, the deviant reading preferred by children in

^{3.} Two children gave unintelligible responses in both conditions. Those trials are excluded from the count.

the case of sentence as (10) is the *non-isomorphic* one and we are in front, if anything, of an observation of anti-isomorphism. A better way to try to capture the data presented in Experiment I and in Moscati & Gualmini (2008b) is to verify the felicity conditions of the target sentence given the discursive context. In this way, we may control if the experimental setting contains elements which bias children interpretation toward the wrong interpretation. However, when both interpretations of sentence (10) where made salient with respect to an explicit Question Under Discussion, the wrong interpretation was still selected in the majority of the cases.

Keeping in mind that it is possible that the QUD made salient by the story may not be easily manipulated and that it is also possible that subjects stop paying attention to the changes in the modal used in the QUD,⁴ for the moment it is safer to assume that the discourse is not a key factor, at least if weighed in accordance with the QAR.

A way to account for the deviant readings found for sentence (9) and (10) may be in the same spirit of the Subset Principle (Berwick 1985; Crain et al. 1994), an idea already present in Musolino et al. (2000).

A way to look at the pattern is to place the possible interpretations along a scale created by the strength of each interpretation's truth conditions. Consider the possible readings for (10), repeated here as (18):

- (18) Il contadino può non dare le carote all'elefante
 - a. *not > possible
 - b. possible > not

As already notices in Section 3, the interpretation (18a) is stronger than (18b) since (18a) entails (18b) while the inverse entailment does not hold:

- (19) not > possible \rightarrow possible > not
- (20) possible not $-\to$ not > possible

The conclusion that can be draw is that if a sentence permits two logical interpretations, a bias toward the stronger interpretation exists in early grammar.

However, the present study only considered one way of looking at the contextual background. Even if no role associated with the discourse has been detected here in the case of modals, further investigation is needed to exclude the role of contextual pressure on the selection of deviant interpretations.

^{4.} This problem can easily be overcome using different groups of subjects for each condition.

5. Summary

We tested children's interpretation of logically ambiguous sentences with the Italian modal 'potere' and sentential negation, in order to check i. whether the deviation found with negative modal sentences (Moscati & Gualmini 2008b) is confined only to English and ii. whether the QAR proposed in Husley et al. (2004) might account for the non-adult readings.

Our first experiment confirms that also Italian children have problems with the correct interpretations of negative modal sentences expressing the possibility not to do a given action. In the second experiment, we tested if an effect related to variations in the Question Under Discussion was detectable. We found no difference between the two conditions and the adult reading was not selected to an higher rate when it was an appropriate answer to the QUD.

References

Berwick, R. 1985. The Acquisition of Syntactic Knowledge. Cambridge MA: The MIT Press.

- Crain, S, Conway, L. & Ni, W. 1994. Learning, parsing, and modularity. In *Perspectives on Sentence Processing*, C. Clifton, L. Frazier & K. Rayner (eds), 443–467. Hillsdale NJ: Lawrence Erlbaum Associates.
- Crain, S. & Thornton, R. 1998. Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics. Cambridge MA: The MIT Press.
- Grice, P. 1975. Logic and conversation. In *Syntax and Semantics*, P. Cole & J. Morgan (eds), 41–58. New York NY: Academic Press.
- Gualmini, A. 2004. Some knowledge children don't lack. Linguistics 42: 957-982.
- Hintikka, J. 1962. Knowledge and Belief. Ithaca, NY: Cornell University Press.
- Horn, L.R. 1989. A Natural History of Negation. Chicago IL: University of Chicago Press.
- Husley, S., Hacquard, V., Fox, D. & Gualmini, A. 2004. The question-answer requirement and scope assignment. In *Plato's Problem: Papers on Language Acquisition*, A. Csirmaz, A. Gualmini & A. Nevins, 71–90. Cambridge MA: MITWPL.

Krämer, I. 2000. Interpreting Indefinites. Wageningen: Ponsen & Looijen.

- Ladusaw, W. 1979. Polarity Sensitivity as Inherent Scope Relations. Ph.D. dissertation, University of Texas at Austin.
- Lewis, D. 1993. Counterfactuals. Oxford: Basil Blackwell.
- May, R. 1985. Logical Form: Its Structure and Derivation. Cambridge MA: The MIT Press.
- Moscati, V. 2008. Negative chains and LF negation raising, In *Proceedings of XXXIII Incontro di Grammatica Generativa*, A. Bisetto & F.E. Barbieri (eds), 106–120. Bologna: University of Bologna.
- Moscati, V. 2006. The Scope of Negation. Ph.D. dissertation, Università degli Studi di Siena.
- Moscati, V. 2010. Negation Raising: Logical Form and Linguistic Variation. Newcastle: Cambridge Scholars Publishing.
- Moscati, V. & Gualmini, A. 2008a. More facts that isomorphism cannot explain. In *Proceedings* of SALT XVII, T. Friedman & M. Gibson (eds), 202–209. Ithaca NY: Cornell University.

- Moscati, V. & Gualmini, A. 2008b. Negation and modality in child language. Paper presented at the Incontro di Grammatica Generativa, Padova.
- Musolino, J. 1998. Universal Grammar and the Acquisition of Semantic Knowledge: An Experimental Investigation into the Acquisition of Quantifier-Negation Interaction in English. Ph.D. dissertation, University of Maryland at College Park.
- Musolino, J., Crain, S. & Thornton, T. 2000. Navigating negative quantificational space. *Linguistics* 38: 1–32.
- Palmer, R.F. 1986. Mood and Modality. Cambridge: CUP.
- Rizzi, L. 1982. Issues in Italian Syntax. Dordrecht: Foris.
- Rooth, M. 1992. A theory of focus interpretation. Natural Language Semantics 1: 75-116.
- von Stechow, Arnim. 1991. Focusing and backgrounding operators. In *Discourse Particles*, W. Abraham (ed.), 37–84. Amsterdam: John Benjamins.
- Szabolcsi, A. 2004. Positive polarity-negative polarity. *Natural Language and Linguistic Theory* 22(2): 409–452.
- Thornton, R. 1994. Children's negative questions: A production/comprehension asymmetry. *Proceedings of ESCOL*: 306–317.
- Wason, P. 1965. The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior* 4: 7–11.

Pragmatic children

How German children interpret sentences with and without the focus particle *only**

Anja Müller, Petra Schulz & Barbara Höhle Goethe University Frankfurt/Goethe University Frankfurt/ Potsdam University

Our study investigated the abilities of 6-year-old German-speaking children to interpret sentences with and without the focus particle *nur* (only). We report two experiments: In Experiment 1 the study by Paterson et al. (2003) on English was replicated in German. We found that German-speaking children do not interpret *only*-sentences target-like. This supports Paterson et al. that children ignore information that is not verbally given. The second experiment investigated children's pragmatic ability to judge underinformative sentences. The results indicate that children take into account information that is showed on a picture, but not verbally introduced. We argue that children's performance in Experiment 1 is not caused by an insufficient pragmatic knowledge but rather to its methodological set up.

1. Introduction

Previous comprehension studies reported that up to school age children interpret sentences containing the focus particle *only* not in an adult-like fashion (for English: Crain, Ni & Conway 1994; Philip & Lynch 1999; Gualmini, Maciukaite & Crain 2003; Paterson, Liversedge, Rowland & Filik 2003; for Dutch: Bergsma 2002;

^{*} Portions of this study were presented at the 30th Child Language Seminar (CLS) 2007 at the University of Reading (Müller et al. 2007a) and at the 32nd Boston University Conference on Language Development (BUCLD) 2007 (Müller et al. 2007b).

This study was supported by the German Research Foundation (DFG) as part of the SFB 632 "Information structure: the linguistic means for structuring utterances, sentences and texts" with a grant to Barbara Höhle and Jürgen Weissenborn. We are grateful to Angela Grimm for discussions of the topic. The paper also benefited from the comments of the anonymous reviewers. Thanks to Paul Abbott for lending his native eye to the English. Special thanks go to Dr. Kevin Paterson for providing the materials for the current studies.

Szendrői 2004; for Portuguese: Costa & Szendrői 2006). Crain et al. (1994) and Gualmini et al. (2003) argued that children prefer a default interpretation because they misconstrue the scope domain of the focus particle. In contrast, Paterson et al. (2003) propose that children fail to process the intended set of alternatives because this information is not verbally introduced in the actual discourse. Consequently, according to Paterson et al. children assign the same meaning to sentences with and without *only*.

In the present study we investigated how 6-year-old German-speaking children interpret sentences with nur (only). As to date there is no study on how German children interpret focus particle sentences with nur, we replicated the study by Paterson et al. (2003), which was carried out in English. The results we found are in line with Paterson et al.'s assumption that children do not consistently instantiate the set of alternatives in their actual discourse model, as required upon encountering a focus particle in a sentence. In the second experiment we investigated how children interpret sentences without a focus particle like nur. Our results show that in fact children are able to take into account information that is visually present in the context, even though it is not verbally introduced in the discourse. Hence, it seems that the children's difficulties in Experiment 1 and in the Paterson et al. study are not caused by difficulties to take into account information that is not verbally given but important for the sentences meaning. Rather we suggest that the observed performance is due to an infelicitous use of the focus particle in the experimental task. We assume that children solved the task by creating a discourse model which is adequate with respect to the given test situation, i.e. the discourse model accepts an interpretation of the sentence without taking into account the meaning and the function of the focus particle.

This paper is organized as follows: Focus particles are described in more detail in Section 2. Previous acquisition studies on the comprehension of focus particles are reported in Section 3. The study by Paterson et al. (2003) is presented in more detail, since our Experiment 1 is a close replication of that study. In Section 4 of the paper we present our experiments. We conclude with a discussion of our results in the light of recent research on focus particles.

2. Focus particles and children's acquisition task

According to Rooth (1992), focus particles are a special set of quantifying expressions that behave like semantic operators. They take scope over the part of a sentence they c-command in the parse tree (Jacobs 1983; König 1991). Within their scope domain focus particles take a specific constituent as their argument. This constituent is called *related constituent* (rc) (Reis & Rosengren 1997; Dimroth

2004). Typically, the related constituent corresponds to the focus of the sentence and is prosodically highlighted by a pitch accent (Altmann 1976; Jacobs 1983; König 1991; Dimroth & Klein 1996).¹

In order to interpret a sentence with a focus particle the child has to master a demanding task which requires syntactic, semantic as well as pragmatic knowledge about the felicitous use of focus particles. Consider examples (1) and (2) with the focus particle *nur*:²

- A: Mögen Popeye und Olivia Spinat? 'Do Popeye and Oliva like spinach?'
 - B: Nein. Nur [POPEYE]rc mag Spinat. 'No. Only [POPEYE]rc likes spinach.'
- (2) A: Mag Popeye Spinat und Möhren?'Does Popeye like spinach and carrots?'
 - B: Nein. Popeye mag nur [SPINAT]rc. 'No. Popeye only likes [SPINACH]rc.'

As shown in (1) and (2) focus particles can occur in different sentence positions. In (1) the focus particle *nur* occurs in utterance initial position. The following subject-NP *Popeye* is the related constituent and thus the focus of the sentences. In contrast, in (2) *nur* appears in postverbal position. In this case the objekt-NP *spinach* is the related constituent of the focus particle. As both examples show, in German typically the focus particle precedes its related constituent. Additionally, the focus particle typically occurs adjacent to its related constituent. In general, in German focus particles are syntactically more restricted than in English (König 1991). The restriction for the position of the focus particle and its related constituent is part of the general constraint, called *Maximale Fokusnähe* (Jacobs 1983). This constraint requires that a focus particle selects the next constituent as its related constituent. Therefore, example (3a) in German is marked, whereas the English equivalent (3b) is felicitous.

^{1.} Pitch accent is marked by capital letters.

^{2.} Note that there are other meaning variants of *nur* (e.g. Altmann 1976; Lerner & Zimmermann 1981). For instance, the particle *nur* can also be used as a modal particle, e.g. *Kommen sie nur herein* ('Just come in please') and as particle with a scalar reading, e.g. *Ich bin nur ein Lehrer* ('I am just a teacher'). Hence, the child has to learn to distinguish between the use of *nur* as a focus particle and as a modal particle, respectively. In the present paper we focus on children's interpretation of the focus particle *nur*.

- (3) a. ? Popeye hat nur Olivia [BLUmen]rc geschenkt. Popeye sent only Olivia flowers.
 'Popeye sent Olivia only flowers.'
 - b. Popeye only sent Olivia some [FLOWers]rc.

Furthermore, the previous examples (1) to (3) show that by selecting a related constituent the focus particle nur affects the meaning of the carrier sentence. The underlying structure of sentence (1) is *Popeye likes spinach*. The proposition of that sentence is [LIKE(Popeye, Spinach)]. This proposition is the presupposition of the only-sentence (1). In other words, Only Popeye likes spinach presupposes that 'Popeye likes spinach' is true. Nur belongs to the group of restrictive focus particles, i.e. nur establishes an exclusive contrast between its related constituent and a so-called set of alternatives that is typically verbally introduced in the actual discourse. The entities of that set signal that there are referents which represent alternatives to the focused element (cf. Rooth 1992). The set of alternatives is required for the sentence interpretation because only by taking into account this set the hearer is able to interpret the focus particle sentence. In example (1) the use of nur signals that in the given discourse 'Popeye' is the only person who likes spinach. In the given discourse the second individual 'Olivia' represents the set of alternatives for which the property of the related constituent is understood as being false. The reading of example (2) in the present discourse is that wrt both types of food in question Popeye likes only spinach. Hence, the set of alternatives consists of the NP 'carrots'.

As shown in the examples above a felicitous use of a focus particle requires that the set of alternatives is present in the actual discourse. Typically, the set of alternatives is either verbally introduced during the preceding conversation or is already part of the common background knowledge of speaker and hearer (cf. Jackendoff 1972). In order to interpret a focus particle sentence the child has to be able to identify the set of alternatives in the discourse. As she has to take into account this information for the sentence interpretation she has to integrate this information into the actual discourse model, which provides the basis for the sentence interpretation. Thus, the identification of the set of alternatives is required for a correct interpretation of a focus particle sentence. This suggests that the preceding verbal discourse, providing the set of alternatives, plays an important role in focus particle comprehension.

In sum, to interpret sentences with a focus particle the child has to master the following tasks: She has to identify the sentence position of the focus particle in the sentence and the related constituent. Then, she must evaluate the set of alternatives from the preceding verbal discourse and must integrate this information into her current discourse model. Furthermore, the child has to establish a contrast between the focus particle and the set of alternatives and to take into account this contrast when interpreting the sentence.

3. Previous studies on children's comprehension of sentences with only

Several acquisition studies have investigated how children acquire focus particles. Although an early and target-like production of focus particles has been documented (Penner, Tracy & Weissenborn 2000; Nederstigt 2003; Hulk 2003; Höhle, Berger, Müller, Schmitz & Weissenborn 2009), several comprehension studies have reported a non adult-like interpretation of focus particle sentences up to school age.³

The study by Crain, Ni & Conway (1994) was one of the first to investigate children's ability to interpret sentences with the focus particle *only*. Using a sentencepicture-matching task, Crain et al. found that 3- to 6-year-old English-speaking children interpreted sentences with the focus particle in pre-subject position (4a) as having the meaning of sentences with the focus particle in preverbal position (4b).

- (4) a. Only the cat is holding a flag.
 - b. The cat is only holding a flag.

In this study participants were presented with a picture depicting a cat holding a flag, a duck holding a flag and a balloon, and a frog holding a balloon. Crain et al. reported that the majority of the children judged the sentence (4a) as a true description of the picture, thus assigning the meaning of (4b) to the sentence. To account for that error pattern, they suggested that children had difficulty correctly restricting the scope of the focus particle *only*. As a consequence, children selected as a default the direct object as the related constituent, regardless of the surface position of *only*. A study by Philip & Lynch (1999) seems to support these findings. They found that 3- to 5-year-old English-speaking children accepted the sentence *Only the dog is holding a starfish* as a true description of a picture showing a dog and a cat both holding a starfish, corresponding to the interpretation of *only* in pre-object position.

A further study was conducted by Gualmini, Maciukaite & Crain (2003), which also seems to support the account of Crain et al. (1994). In line with Crain et al.'s results Gualmini et al. reported a default interpretation of sentences with *only*. Within a truth-value-judgement task 4- to 5-year-old English-speaking children were asked to judge sentences like (5a) and (5b).

- (5) a. The farmer only sold a [BANANA]rc to Snow White.
 - b. The farmer only sold a banana to [SNOW WHITE]rc.

^{3.} In the following we concentrate on studies which tested unambiguous sentences with *only*. For research which investigated children's comprehension of ambiguous sentences with *only* see Crain et al. (1994) and Paterson, Liversedge, White, Filik & Jaz (2006).

In contrast to Crain et al. (1994) the focus particle sentence was preceded by a short story narrated by the experimenter. After the story the experimenter prompted a puppet with *What happened in the story?* The child's task was then to judge whether the puppet's answer was felicitous given the story. The results showed that children interpreted sentences like (5a), in which the direct object was prosodically highlighted, as having the meaning of (5b), i.e. children identified the indirect object instead of the direct object as the related constituent. Gualmini et al. concluded that children are insensitive to prosodic information as a reliable cue for interpreting sentences containing *only*. Instead, children resort to a default interpretation by assigning *only* to the indirect object. These findings were replicated using the same experimental design by Szendrői (2004) for Dutch and Costa & Szendrői (2006) for European Portuguese. Hence, it seems that children are not able to use prosodic information to unambiguously identify the related constituent of a focus particle.⁴

Paterson et al. (2003) pointed to another possible cause of the problems that children can have when interpreting sentences with *only*. They suggested that due to an insufficient representation of the set of alternatives children neglect the meaning of the focus particle and thus interpret sentences with focus particles in the same way as sentences without focus particles. To test this hypothesis Paterson et al. presented children with sets of six pictures and corresponding test sentences. Each picture set comprised six drawings (see Figure 1). The corresponding set of test sentences represented three different experimental conditions (see Example 6 a-c).

- (6) a. *The fireman is holding a hose.* Sentence without *only*:
 - b. Only the fireman is holding a hose. Sentence with pre-subject only:
 - c. *The fireman is only holding a hose.* Sentence with pre-verbal *only*:

Using a picture-selection task, the experimenter presented the whole picture set to the participant and read one of the test sentences (6a) to (6c) aloud. The participant was asked to point to all pictures that matched the given sentence. Paterson et al. grouped the different response patterns into three main categories, reflecting the correct interpretation for each of the three experimental conditions. For

- (i) a. Toby hat AUCH eine Puppe. 'Toby possesses ALSO a doll.' (like e.g. Anna)
 - b. Toby hat auch eine PUPPE. 'Toby possesses also a DOLL.' (in addition to e.g. a ball)

^{4.} Findings by Höhle et al. (2009) cast doubt on the assumption that children generally ignore prosodic information when identifying the related constituent of a focus particle. In an eye-tracking study with three- and four-year-olds, they tested sentences like (ia) and (ib) with the focus particle *auch* (also) and found that children reacted differently to the two accent patterns.



Figure 1. Picture set used by Paterson et al. (2003:272)

the condition without *only* (6a), pictures A, C, D, and E constitute the correct response, because in all four pictures the proposition of the test sentence, in this case [HOLD(fireman, hose)] was fulfilled. For the test sentence with pre-subject *only* (6b) pictures A and D should be chosen because in both pictures the fireman is performing the described action and nobody else is. When presented with preverbal *only* (6c) the participants should point to pictures A and C because in both pictures the fireman is holding a hose and there is nothing else that he is holding. Paterson et al. hypothesized that if children do not process the information triggered by the focus particle, they should exhibit the same response pattern for sentences with and without *only*. However, if children misconstrue the scope of *only* as predicted by Crain et al. (1994), then children should point to pictures A and C for both conditions (6b) and (6c).

Paterson et al. tested 6- to 7-year-old, 8- to 10-year-old and 11- to 12-year-old English-speaking children as well as a control group of adults. All participants showed a better performance for sentences without *only* than with *only*. The group of adults gave more correct responses than the other age groups on all types of test sentences. For further analyses Paterson et al. concentrated on the group of 6-year-olds. When presented with test sentences without *only* these children pointed to the expected response pattern (pictures A, C, D, E) in 84% of all cases. For pre-subject *only* sentences Paterson et al. found 26% correct responses (pictures A, D) and for pre-verbal *only* sentences 34% correct responses (pictures A, C). A more detailed analysis of the unexpected responses in these two *only*-conditions revealed that a relatively high proportion of children's responses (45% in the pre-subject *only* condition; 49% in the pre-verbal *only* condition) could be ascribed to ignoring the focus particle when interpreting the sentence. In these cases children selected the pictures A, C, D, E. Only a minority of responses was compatible with Crain et al.'s assumption of problems with scope restrictions (16% in the pre-subject-only condition).

These findings were taken by Paterson et al.'s to support their account, i.e. that children tend to ignore the information given by the focus particle when interpreting an *only*-sentence. More precisely, children are not able to consistently instantiate the set of alternatives in the actual discourse model when this is only triggered by the presence of the focus particle in the sentence. Paterson et al. argued that this inability is caused by the still insufficiently developed pragmatic knowledge of the children.

In summary, previous research on the comprehension of sentences containing the focus particle *only* indicates that children up to age 6 do not interpret these sentences in adult-like fashion. Gualmini et al. (2003) argued that children are not able to use prosodic focus information to identify the related constituent of *only*. Crain et al. (1994) claimed that children misconstrue the scope of the focus particle. Both accounts suppose a default interpretation for associating *only* with the last constituent in the sentence. However, Paterson et al. (2003) argued that previous studies on *only* did not clearly show that children's difficulties are due to problems with the identification of the correctly related constituent of *only*. Instead, they suggested that children fail to instantiate the set of alternatives in the actual discourse model and thus tend to interpret the focus particle sentences as sentences without *a* focus particle.

One way to obtain further evidence that may be helpful in deciding between these different accounts is a crosslinguistic comparison. If children up to a specific age show problems in integrating the set of alternatives into their sentence interpretation then across languages no variation in the interpreting sentences with focus particles is expected. Thus, if Paterson et al.'s account is correct, we should obtain the same results for German-speaking children. On the other hand, if children have problems with the identification of the scope of the particles then differences across languages related to structural variation between the languages should occur. In German there are stronger positional restrictions for the focus particle *only* as compared to English. *Only* typically precedes its related constituent directly which might help the child to identify the related constituent of the particle. Therefore we replicated the Paterson et al. (2003) study with German learners.

4. The study

4.1 Experiment 1

Experiment 1 replicated the Paterson et al. (2003) study in German. As Paterson et al. focused on the results of the 6-year old English-speaking children, we tested German-speaking children of the same age.

4.1.1 Participants

Thirty 6-year-old German-speaking children (15 girls and 15 boys) participated in this experiment (mean age 6;8 years; range 6;1 – 7;2 years). The children were recruited from several schools in Potsdam, Brandenburg. All children are growing up in monolingual German-speaking homes. In addition, 30 adults were tested as a control group.

4.1.2 *Materials and procedure*

We used the original test pictures from Paterson et al. (2003) as visual stimuli and translated their sentence material into German. In some cases adequate translation required the use of a verb with a separable particle. These items were excluded. Nine out of the twelve original test sets were then included in the replication (see Appendix A). Each of the nine test sentence types was used in the three experimental conditions: in a sentence without *nur* (7a), in a sentence with *nur* in pre-subject position (7b) and in a sentence with *nur* in pre-object position (7c).

(7)	a.	Der Feuerwehrmann hält einen Schlauch. 'The fireman is holding a hose.'	Without <i>nur</i>
	b.	Nur der Feuerwehrmann hält einen Schlauch. 'Only the fireman is holding a hose.'	Pre-subject nur
	c.	Der Feuerwehrmann hält nur einen Schlauch. 'The fireman is holding only a hose.'	Pre-object nur

As shown in the examples the German sentences in *pre-object* condition (7c) differ systematically from the English sentences as the focus particle in German always directly precedes its related constituent (cf. 6c). The sentence material was divided into three lists such that each list contained nine sentences, three of each condition. None of the test sentence types appeared more than once in a list. Each participant was only tested with one list, thus no participant was presented more than once with each test sentence type and the corresponding picture set. The participants were randomly assigned to one of the three test lists. The test sentences were orally presented to the participants. The experimenter read aloud the test sentences in which the focussed constituent of the sentences was prosodically highlighted. In the case of the test sentences without *nur* the experimenter presented the sentence with normal intonation, i.e. with an accent on the final object-NP.

The children were tested after school in their day care center. The child and the experimenter sat in a separate room from the other children in the day care center. At the beginning the experimenter presented the first picture set to the child and asked to name the persons depicted at the pictures. After that the experimenter presented the test sentence and asked the child to point to the pictures which match the test sentence. When the child finished the experimenter showed the next picture set and the experimental procedure was repeated. An experimental session lasted for about ten to 15 minutes. The participants of the adult control group were also tested individually.

4.1.3 Results

As in Paterson et al. (2003) the responses of each participant were assigned to one of the four following categories: *no scope*, *subject scope*, *object scope*, and *other*. If the participant pointed to the pictures A, C, D, E, this response was coded as *no scope* because this response was the expected pattern for sentences without the focus particle. If only the pictures A and D were selected, this response counted as *subject scope*. Selections of only the pictures A and C were assigned to the category *object scope*. The remaining responses were coded as *other*.

As Table 1 illustrates adults gave the expected responses in the majority of the cases in all three conditions. In contrast, children's major response type corresponded to the *no scope* interpretation.⁵ A 2 (age group) × 3 (sentence condition) ANOVA based on the number of correct responses revealed a main effect for the factor age group (F(1,58) = 311,128; p < .001). Furthermore, there was a main effect for the factor sentence condition (F(1,58) = 89,963; p < .001) and also an interaction between both factors (F(1,58) = 65,629; p < .001). Separate one-way ANOVAs for both age groups revealed an effect for the factor sentence condition for the factor sentence condition for the children (F(1,29) = 133,218; p < .001) but not for the adults (F(1,29) = 1,482; p = .236) A paired t-test showed that children made significantly more errors with sentences with pre-subject *nur* and pre-object *nur*, respectively, than with sentences without *nur* ($t_{(29)} = 13,646; p < .001; t(29) = 12,892; p < .001$). There was no significant difference between the performance for sentences with *nur* in pre-subject position and in pre-object position ($t_{(29)} = .405; p = .689$). A comparison of children's and

^{5. 13} of the 30 children always selected pictures A, C, D, and E when interpreting sentences with *nur* in either condition. Only one child gave the expected responses in all three conditions, with the exception of one sentence in the pre-object *nur* condition. The remaining children showed various responses for the *nur*-sentences.

sentence		no scope (pointing A, C, D, E)	subject scope (pointing A & D)	object scope (pointing A & C)	other
Without nur	children	94%	0%	1%	5%
	adults	100%	0%	0%	0%
pre-subject	children	70%	13%	6%	11%
<i>nur</i>	adults	2%	96%	0%	2%
pre-object	children	71%	5%	13%	11%
<i>nur</i>	adults	1%	1%	98%	0%

Table 1. Children's and adults' responses for the sentences with *nur* in pre-subject and pre-object position and without *nur*. (Grey coloured fields indicate the expected/ correct response.)

adults performance revealed that adults gave significantly more expected responses in the pre-subject and pre-object condition than children (unpaired t-test: t(58) =12,618; p<.001; (t(58) = 14,523; p = .001). For sentences without *nur* no significant difference was found (t(58) = 1,248; p = .217).

In a further analysis we compared the frequency of responses in which children unambiguously ignored the set of alternatives (*no scope errors*) and in which they unambiguously misconstrue the scope of *nur* (*scope analysis errors*). A paired t-test revealed that children produced significantly more *no scope errors* than *scope analysis errors* (t(29) = 7,753; p<.001).

In summary, both groups of participants interpreted the test sentences without the focus particle *nur* as expected. However, unlike the adults the children frequently judged the pictures A, C, D, E as a felicitous depiction for *nur*-sentences which was actually the expected response for sentences without *nur*.

4.1.4 Discussion

Our first experiment replicated the findings of Paterson et al. (2003). Whereas adults selected the expected pictures in all three sentence conditions, 6-year-old children most frequently selected the picture set that corresponds to the reading of the sentence without a focus particle, and they did so even when presented with sentences including the focus particle. In summary, our German data are compatible with the account by Paterson et al.

As mentioned above, Paterson et al. assume that children have difficulties instantiating information in their current discourse model that is not verbally introduced. Therefore, the set of alternatives which is implicated by the focused constituent, i.e. the related constituent, but not verbally given is ignored in children's interpretation. According to Paterson et al. this non adult-like performance is rooted in the children's insufficient pragmatic knowledge.

Apart from Paterson et al.'s account, two alternative accounts could explain children's performance as well: First, assume that children in general exhibit difficulties instantiating information that is not verbally introduced in the discourse, regardless of whether they interpret a sentence with or without a focus particle. Then it would follow that preschool children also fail to compute the pragmatic contribution of nur to the sentence interpretation. Second, wrt children's acquisition task (Section 2) the interpretation of a focus particle sentence requires the existence of a set of alternatives. Either the set of alternatives is presented in the preceding verbal discourse or it is already part of the common background knowledge of speaker and hearer. That means, that the ability to identify the set of alternatives in a visual presented picture belongs not to the abilities which are required for the mastery of focus particle sentences. Therefore, it is possible that children calculate focus particles only if the set of alternatives is explicitly introduced in the preceding discourse. In that case, children interpret focus particle sentences adult-like when the set of alternatives is explicitly mentioned in the discourse, but not in the absence of such a context as in the Paterson et al. study.

The first account will be investigated in Experiment 2, discussed in the following, for the second account (cf. Müller, in prep.).

4.2 Experiment 2

To test whether children in general have difficulties in instantiating information that is not verbally introduced in their discourse model we chose sentences without focus particles as test material for Experiment 2. These were taken from Experiment 1. We presented a subset of the pictures from Experiment 1 that varied with respect to their informational complexity. Furthermore the pictures represent more information than is mentioned in the sentence. The sentences are underinformative with respect to the information contained in the pictures shown. The children's task was to evaluate whether a sentence matched a picture. If children are not sensitive to the degree of informativeness of a given sentence in relation to a given picture, i.e. they are insensitive to information that has not been explicitly mentioned, we expect no differences in the children's judgements for the different pictures. In contrast, sensitivity to informativeness should result in the following response pattern: The higher the informational complexity of the picture the lower should be the acceptance rate.

As test materials we used the picture types A, C, D, E (see Fig. 2) and the test sentences without *nur* from Experiment 1. Although all these pictures are a logically true description of the test sentences without the focus particle, they differ with respect to the amount of additional information depicted. For example, in Figure 2 all four pictures depict a fireman holding a hose. However, all pictures

contain additional information, which is not mentioned in the test sentence: A policeman (all pictures), a hose held by the policeman (pictures C and E) and a ladder held by the fireman (pictures D and E). This difference we refer to as (visual) informational complexity. On this complexity scale, Picture A is informationally the least complex picture and picture E is the most complex one, with pictures C and D being in between. With respect to the maxim of quantity (Grice 1975) the sentence The fireman is holding a hose is most informative wrt picture A, and least informative wrt picture E. Besides the mentioned event, Picture E displays another character (the policeman) who performs the same action as the fireman. Moreover, in picture E the fireman performs a second activity, i.e. holding a ladder. Thus, there are at least three pieces of information depicted in picture E that are not verbally expressed in the test sentence. Therefore, with respect to the maxim of quantity the sentence The fireman is holding a hose is underinformative with respect to picture E. In picture A, the police man is also present, but no other activities are portrayed. As a result, the degree of underinformativeness of the sentence is dependent on the complexity of the picture. If underinformativeness effects children's judgements along a scale of visual informational complexity, we would expect the highest acceptance rate for picture A and the lowest acceptance rate for picture E.

4.2.1 Participants

Thirty-two 6-year-old German-speaking children (16 girls and 16 boys) participated in our second experiment (mean age 6;9 years; range 6;6 – 7;0 years). As in our first experiment all children are growing up in monolingual German families and were recruited from several schools in Potsdam. In addition, 30 adults were tested as a control group.



Figure 2. Example of test pictures in Experiment 2

4.2.2 *Material and procedure*

A felicity judgement task was developed. In this task children were asked to judge whether a given sentence matches a given picture, which was presented simultaneously. The picture types A, C, D, E (see Figure 2) and the sentences without *nur* from the stimulus sets used in Experiment 1 (8) served as material for this experiment

(cf. Appendix for a complete list of the test items). To balance the sentence-picture pairs with respect to the four picture types we added three test sentences and the corresponding four picture types. In all, we used twelve test sentences and 48 test pictures (twelve for each picture type).

(8) Der Feuerwehrmann hält einen Schlauch.'The fireman is holding a hose.'

Four item lists were created. Each list consisted of twelve test sentences and twelve test pictures that were distributed across the four different lists so that no participant had to judge one sentence with two picture types of the same set. List 1 contained picture type A of the fireman example, list 2 contained picture type C, list 3 included picture type D and list 4 included picture type E. In addition, four supplementary control picture-sentence pairs were added to each list in which a sentence's proposition was not depicted on the picture. For example, for the sentence *The woman is walking a chicken* the picture showed a women walking a dog and a cat and a man who is walking by himself. These items provided clear cases of a mismatch between picture and sentence and thus should evoke no-responses. All in all each list consisted in 16 sentences and 16 test pictures.

Children were tested after school in their day care center. After introducing a puppet the experimenter showed the first picture to the child and read the test sentence aloud. Afterwards the puppet asked the child whether the sentence matched the picture or not. After the child answered the question the experimenter presented the next picture and the experimental procedure was repeated. At the beginning of the experiment two practice items were presented: Once prompting a yes-response and the other a no-response. This way, children should become aware that a no-response could be a correct response. The whole test session lasted for about ten to 15 minutes and was audio-taped. The adults of the control group were also tested individually.

4.2.3 Results

Table 2 presents the percentages of sentences that were accepted (yes-responses) as felicitous matches to the picture types A, C, D, and E.

The number of yes-responses in both children and adults decreases from the least informative picture type A to the most informative picture type E. A 2 (age group) × 4 (picture type) ANOVA revealed a significant main effect for the factor age group (F(1,62) = 9,630; p < .001) and a significant main effect for the factor picture type (F(1,62) = 6,072; p < .05) but no interaction between the two factors (F(1,62) = .564; p = .455). A paired t-test revealed that children accepted picture type A significantly more often than picture type C (t(31) = 3,371; p < .01), picture type D (t(31) = 3,127; p < .05), and picture type E

	Picture type A	Picture type C	Picture type D	Picture type E
Children	88%	76%	65%	58%
Adults	100%	93%	81%	79%

Table 2. Percentages of yes-responses for children and adults

(t(31) = 4,244; p <.001). There were also significantly more yes-responses for the picture type C than for type E (t(31) = 2,672; p < .05). No difference was found between picture types C–D (t(31) = 1,259; p = .218) and D–E (t(31) = 1,677; p = .104). For the group of adults the paired-t-test revealed a significant difference between picture types A–D (t(31) = 2,982; p < .05) and A–E (t(31) = 3,374; p < .05) but not between picture types A–C (t(31) = 1,923; p = .064). Furthermore the t-test showed a significant difference between picture types C–E (t(31) = 3,371; p < .05). No difference was found between picture types D–E (t(31) = 1,249; p = .221).

In summary, both groups rated the picture types A, C, D, and E differently according to their assumed complexity of additional information, and children rejected all picture types much more often than adults.

4.2.4 Discussion

The results of our second experiment suggest that children as well as adults are affected by the informational complexity of a picture when they have to decide whether a sentence matches a picture. The sentence-picture pairs were judged better matches when the degree of underinformativeness of the sentence wrt the picture is small. Overall the amount of no-responses given by children is higher than by adults. This suggests that children might be even more affected by the degree of underinformativeness of the sentence than adults. Obviously children expect that sentences are maximally informative about the depicted events. For the same assumption see also Paterson et al. (2006).

Most importantly, these results clearly indicate that 6-year-old children do not generally ignore information that is only visually present. Our results indicate that this kind of information is relevant in children's evaluation of how well a sentence describes a given situation. Thus, they seem to adhere to the maxim of quantity (Grice 1975). In other words, children's judgment of a sentence-picture match is not purely based on the match between the sentence's proposition and the picture, because in this case all four types of pictures would have been judged alike. The fact that children's reaction to the pictures was significantly different suggests that they are sensitive to the (pragmatically flavoured) felicity of a sentence wrt a picture. This finding casts doubt on the assumption that children's responses to sentences with focus particles found in Experiment 1 resulted from children's ignorance of information that is not explicitly mentioned. We will come back to this in the general discussion.

Interestingly, the rejection pattern of children and adults were not the same. For adults, the four picture types were split in two groups: picture types A and C as one group, and picture types D and E as the second group. Type A/C was accepted significantly more often than type D/E. In contrast, the children showed a more gradual response pattern. A closer examination of the adult responses shows that they pattern according to the presence or absence of a second object attributed to the subject mentioned in the sentence (the fireman in our example). Presence or absence of an object attributed to the second protagonist, in contrast, does not seem to influence the acceptance. This suggests that the concept of underinformativeness is different for children and adults. For children every person or object present in a picture but not mentioned in the verbal discourse increases the underinformativeness of the sentence in that case. For adults underinformativeness seems to relate only to the subject of the sentence. Further evidence for this difference is the fact that children rejected picture type A, which only shows an additional protagonist, as not matching the sentence in 12% of the cases, while all adults accepted that picture. The behavior of the adults might be explained within Rooth's (1992) account of Alternative Semantics. A central claim of this proposal is that focus in general indicates the presence of alternatives that are relevant for the interpretation of a linguistic expression. The sentences presented in this experiment were prosodically unmarked, i.e. the nuclear stress was on the utterance's final element, the object of the sentence (cf. Cinque 1993). This prosodic marking supports a reading of the sentence in which the object is the focused constituent of the sentence that may turn the hearer's attention to the set of alternatives that consisted of the second object held by the protagonist (the ladder in our example). This assumption could be tested by changing the focus of the sentences to the subject, for instance by prosodic highlighting. If our hypothesis is correct we would expect a change in the adults' response pattern but not in the children's.

5. General discussion

The aim of the present study was to determine how German-speaking children interpret sentences with and without the focus particle *nur*. The first experiment replicated the findings of Paterson et al. (2003) for German-speaking children. Children seem to assign the same reading to sentences with and without the focus particle *nur*. The error pattern observed did not provide any evidence that misconstrued scope restrictions are the basis for children's difficulties in the interpretation of sentences with the focus particle *nur*. In contrast, the similarity

of the response patterns for sentences with and without *nur* is compatible with the Paterson et al.'s assumption that children do not implement the set of alternatives in their actual discourse model when interpreting sentences with *nur* in isolation.

Experiment 2 tested whether children are unable to integrate information into the discourse model that is not explicitly given in the verbal context and has to be inferred from the visual context. The results of this experiment indicate that children do not ignore information provided only visually. In addition, we found that visual information influenced children's decision about how well a sentence matched a given picture. In sum, the results of the second experiment suggest that children's failure to integrate the information signalled by the focus particle into their current discourse model as shown in our first experiment is not due to a general problem in implementing information that is not verbally introduced in the discourse.

The comparison of children's responses in the two experiments suggests that children responded differently when asked to judge the felicity of the sentence without *nur* given the four picture types. Whereas in Experiment 1 these four pictures were judged to match the sentences without *nur* in almost all of the cases, there were much fewer match responses for the same pictures and the same sentences in Experiment 2 (see Table 3).

We suggest that this discrepancy is rooted in the differences in the procedures that were used in the two experiments. In Experiment 1 we used a picture-selection

		Picture			
		Α	С	D	E
Sentences without <i>nur</i>	Experiment 1	99%	98%	97%	98%
	Experiment 2	88%	76%	65%	58%

Table 3. Comparison of the match responses for pictures A, C, D, and E in Experiment 1 and 2

task: The child saw six pictures at once and was asked to "point to the pictures that matched the sentence". Hence, the formulation of the task or more precisely the use of plural noun 'pictures' may have signalled to the child that she could or had to point to more than one picture in order to solve the task correctly. In contrast, in Experiment 2 we used a felicity judgement task: The child saw only one picture and was asked to judge whether the given sentence matched the picture or not.

A central question of our study is why children seem to ignore the meaning of *nur* when interpreting sentences with this focus particle. In the following, based on the results for Experiment 2, we propose an alternative account to Paterson et al. (2003) that relates children's performance in our first experiment (as well as in the

Paterson et al. study) to methodological aspects of that study. The task in Experiment 1 called on the participants to draw pragmatic inferences. Importantly, the *nur*-sentences were not embedded in a verbal context motivating the use of the focus particle. Instead, the information licensing the use of the focus particle in the test sentence was contained in the visual stimulus. Thus, children had to infer the set of alternatives from the picture. The focus particle in the sentences may not have served as a cue for the children to search for a set of alternatives within this test situation, as was obviously the case for the adults. The test situation, i.e. the presentation of the *nur*-sentence together with a set of six pictures, obviously did not suffice to make the children understand why the focus particle is used. Consequently, the children were not able to establish the discourse model, which was intended by the experimental set up, solely on the basis of the visual information. Based on this assumption, we are led to conclude that the investigation into children's ability to interpret sentences with a focus particle was confounded by the requirement to (re)construct an appropriate discourse model.

Our assumption is in line with the discussion on the probable impact of the requirements of a given task on children's performance when interpreting scalar terms (cf. Noveck & Sperber 2004). In this field the discussion concentrates on the extent to which experimental settings provide an adequate verbal context for the use of a scalar term in a given utterance (cf. Noveck & Sperber 2004; Papafragou & Musolino 2003). Papafragou & Musolino (2003) and Papafragou & Tantalou (2004) compared children's comprehension of scalar implicatures when the sentences were presented with and without an appropriate verbal context. They found that 5-year-old children performed better when the scalar term was embedded and motivated by a verbal context.

In the case of focus particles, a felicitous use moreover depends on the specific kind of verbal context. Gualmini et al. (2003), for example, embedded the focus particle test sentences in a short story, which was read aloud and acted out by the experimenter in front of the child. As mentioned before, children showed difficulties when interpreting sentences like (9a) where the related constituent of *only* is the direct object. In the Gualmini et al. study the question *What happend in the story*? directly preceded the *only*-sentence. However, Papafragou & Tantalou (2004) pointed out that this type of question is too general for licensing the use of quantifying expressions. Likewise, we suggest that this question does not licence a felicitous use of a focus particle in a sentence like (9a). The question *What happened in the story*? is felicitous if the questioner has no specific knowledge about the story. The answer to this question is characterized by wide focus with elements on the right periphery carrying the nuclear stress (Cinque 1993). Thus, (9a) with an accented direct object does not provide an adequate answer to the question.

- (9) What happened?
 - a. ? The farmer only sold a [BANANA]rc to Snow White.
 - b. The farmer only sold a banana to [SNOW WHITE]rc.⁶

From this point of view children's tendency to interpret sentence (9a) as having the meaning of sentence (9b) as found by Gualmini et al. (2003) might be due to the fact that the children were aware of this mismatch between the question and the prosodic pattern of the test sentence.

To conclude, our study aimed at a better understanding of how children interpret sentences with (and without) focus particles. In Experiment 1 we replicated the study by Paterson et al. (2003) and found that the German-speaking children behaved like their English-speaking peers. The 6-year-olds did not seem to integrate the set of alternatives into their current discourse model when the set of alternatives was not introduced in the verbal context. As the same design yielded the same pattern of results, crosslinguistic differences between English and German in the syntactic restrictions for focus particles did not seem to play a role in interpreting the focus particle sentences. Therefore, we predict that instantiating a set of alternatives based upon accommodating implicit information to be a challenge at age 6 across languages.

Second, our findings in Experiment 2 suggest that children's inability to integrate this information in the discourse model in Experiment 1 is not caused by the fact that this information was not verbally introduced in the discourse. Children did not ignore the information visually present in the pictures, without being mentioned in the verbal discourse. Hence, we claim that children's difficulties observed in Experiment 1 may be caused by an insufficient activation of the set of alternatives on the child's part. Possibly children's performance would increase if the focus particle sentence is embedded in a verbal context, which introduces the set of alternatives and thus established it in the discourse. This is tested in ongoing work (Müller, in prep.).

References

- Altmann, H. 1976. Die Gradpartikeln im Deutschen: Untersuchungen zu ihrer Syntax, Semantik und Pragmatik. Tübingen: Niemeyer.
- Bergsma, W. 2002. Children's interpretations of Dutch sentences with the focus particle alleen ('only'). In The Process of Language Acquisition. Proceedings of the 1999. GALA Conference, I. Lasser (ed.), 263–280. Frankfurt: Peter Lang.

^{6.} Note that sentence (9b) seems to be out too as a felicitous answer because of the contrastive stress on the indirect object.
Cinque, G. 1993. A null theory of phrase and compound stress. Linguistic Inquiry 24(2): 239–297.

- Costa, J. & Szendrői, K. 2006. Acquisition of focus marking in European Portuguese Evidence for a unified approach to focus. In *The Acquisition of Syntax in Romance Languages*, V. Torres & L. Escobar (eds), 319–329. Amsterdam: John Benjamins.
- Crain, S., Ni, W. & Conway, L. 1994. Learning, parsing and modularity. In *Perspectives on Sentence Processing*, C. Clifton, L. Frazier & K. Rayner (eds), 443–467. Hillsdale NJ: Lawrence Erlbaum Associates.
- Dimroth, C. 2004. Fokuspartikeln und Informationsgliederung im Deutschen [Studien zur Deutschen Grammatik 69]. Tübingen: Stauffenburg.
- Dimroth, C. & Klein, W. 1996. Fokuspartikeln in Lernervarietäten. Ein Analyserahmen und einige Beispiele. Zeitschrift für Literaturwissenschaft und Linguistik 104: 73–114.
- Grice, P. 1975. Logic and conversation. In *Syntax and Semantics*, Vol. 3., P. Cole & J. Morgan (eds), 41–58. New York NY: Academic Press.
- Gualmini, A., Maciukaite, S. & Crain, S. 2003. Children's insensitivity to contrastive stress in sentences with ONLY. In *Proceedings of PLC 25*, S. Arunachalan, E. Kaiser & A. Williams (eds), 87–110. Philadelphia PA: University of Pennsylvania.
- Hulk, A. 2003. Merging scope particles: Word order variation and the acquisition of *aussi* and *ook* in a bilingual context. In *Information Structure and the Dynamics of Language Acquisition*, C. Dimroth & M. Starren (eds), 211–234. Amsterdam: John Benjamins.
- Höhle, B., Berger, F., Müller, A., Schmitz, M. & Weissenborn, J. 2009. Focus particles in children's language: Production and comprehension of auch 'also' in German learners from 1 year to 4 years of age. Language Acquisition 16(1): 36–66.
- Jackendoff, R.S. 1972. Semantic Interpretation in Generative Grammar. Cambridge MA: The MIT Press.
- Jacobs, J. 1983. Fokus und Skalen. Zur Syntax und Semantik von Gradpartikeln im Deutschen. Tübingen: Niemeyer.
- König, E. 1991. The Meaning of Focus Particles A Comparative Perspective. London: Routledge.
- Lerner, J.-Y. & Zimmermann, T. 1981. Mehrdimensionale Semantik: Die Präsuppositionen und die Kontextabhängigkeit von *nur* [Working Paper 50 of the SFB 99]. Konstanz: University Konstanz.
- Müller, A., Schulz, P. & Höhle, B. 2007a. What Sentences without only can Tell us about Children's interpretation of only: Evidence from German. Poster presented at the 30th Child Language Seminar (CLS), Reading.
- Müller, A., Schulz, P. & Höhle, B. 2007b. Do Children Interpret only Sentences? Evidence from German. Poster presented at the 32nd Boston University Conference on Language Development, Boston.
- Müller, A. In preparation. Wie interpretieren Kinder *nur?* Ph.D. dissertation, University Potsdam.
- Nederstigt, U. 2003. Auch *and* noch *in Child and Adult German* [Studies in Language Acquisition 23]. Berlin: Mouton de Gruyter.
- Noveck, I. & Sperber, D. 2004. Experimental Pragmatics. London: Palgrave.
- Papafragou, A. & Musolino, J. 2003. Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition* 86: 253–282.
- Papafragou, A. & Tantalou, N. 2004. Children's computation of implicatures. *Language Acquisition* 12: 71–82.
- Paterson, K., Liversedge, S., Rowland, C. & Filik, R. 2003. Children's comprehension of sentences with focus particles. *Cognition* 89: 263–294.

- Paterson, K., Liversedge, S., White, D., Filik, R. & Jaz, K. 2006. Children's interpretation of ambiguous focus sentences with 'only'. *Language Acquisition* 13(3): 253–284.
- Penner, Z., Tracy, R. & Weissenborn, J. 2000. Where scrambling begins: Triggering object scrambling at the early stage in German and Bernese Swiss German. In *The Acquisition of Scrambling* and Cliticization, S.M. Powers & C. Hamann (eds), 127–164. Dordrecht: Kluwer.
- Philip, W. & Lynch, E. 1999. Felicity, relevance, and acquisition of the grammar of *every* and *only*. In *Proceedings of* 24th *Annual Boston University Conference on Language Development*, C. Howell, S.A. Fish & T. Keith-Lucas (eds), 583–596. Somerville MA: Cascadilla Press.
- Reis, M. & Rosengren, I. 1997. A modular approach to the grammar of additive particles: The case of German *auch. Journal of Semantics* 14(3): 237–309.
- Rooth, M. 1992. A theory of focus interpretation. Natural Language Semantics 1(1): 75-116.
- Szendrői, K. 2004. Acquisition evidence for an interface theory of focus. In Proceedings of Generative Approaches to Language Acquisition 2003, J. van Kampen & S. Baauw (eds), 457–468. Utrecht: LOT.

Appendix A

(Nur) das Mädchen streichelt (nur) ein Pferd. '(Only) the girl is (only) stroking a horse.' (Nur) der Mann trägt (nur) eine Tasche. '(Only) the man is (only) carrying a briefcase.' (Nur) der Junge angelt (nur) einen Fisch. '(Only) the boy is (only) catching a fish.' (Nur) der Hund jagt (nur) eine Katze. '(Only) the dog is (only) chasing a cat.' (Nur) der Postbote bringt (nur) einen Brief. '(Only) the postman is (only) delivering a letter.' (Nur) die Frau schiebt (nur) einen Kinderwagen. '(Only) the woman is (only) pushing a pram.' (Nur) der Junge hält (nur) einen Drachen. '(Only) the boy is (only) holding a kite.' (Nur) der Junge spielt (nur) mit einem Ball. '(Only) the boy is (only) playing with a ball.' (Nur) der Feuerwehrmann hält (nur) einen Schlauch. '(Only) the fireman is (only) holding a hose.'

Appendix B

Test sentences

Das Mädchen streichelt ein Pferd. 'The girl is stroking a horse.' Der Mann trägt eine Tasche. 'The man is carrying a briefcase.' Der Junge angelt einen Fisch. 'The boy is catching a fish.' Der Hund jagt eine Katze. 'The dog is chasing a cat.' Der Postbote bringt einen Brief. 'The postman is delivering a letter.' Die Frau schiebt einen Kinderwagen. 'The woman is pushing a pram.' Der Junge hält einen Drachen. 'The boy is holding a kite.' Der Junge spielt mit einem Ball. 'The boy is playing with a ball.' Der Feuerwehrmann hält einen Schlauch. 'The fireman is holding a hose.' Der Junge hält einen Hammer. 'The boy is holding a hammer.' Der Mann trägt einen Korb. 'The man is carrying a basket.' Die Krankenschwester trägt einen Eimer. 'The nurse is carrying a bucket.'

Control sentences

Der Weihnachtsmann bringt einen Roller. 'Santa Claus is bringing a scooter.' Die Katze spielt mit einer Puppe. 'The cat is playing with a doll.' Der Polizist stoppt einen Fahrradfahrer. 'The policeman is stopping a cyclist.' Die Frau spaziert mit einem Huhn. 'The woman is walking with a chicken.'

Adult response uniformity distinguishes semantics from pragmatics

Implications for child language*

Leah R. Paltiel-Gedalyovich Ben-Gurion University of the Negev

This paper reports data from 17 adult and 141 child Hebrew-speakers from experiments testing knowledge of the semantics and pragmatics of coordination, using a variant of the Truth-Value Judgment Task (Crain & Thornton 1998). Adults showed uniformity in judgments of semantic meaning, the truth-conditions of conjunction (ve/'and', aval/'but') and disjunction (o/'or') and the non truth-conditional contrast associated with *aval/*'but'. By contrast, judgments of the pragmatic meanings, the scalar quantity implicatures associated with the use of disjunction, and the pseudo-scalar quantity implicature associated with the use of aval/'but' implicatures, varied. Children from the age of 5 showed uniform adultlike knowledge of semantic, truth-conditional meaning, while the non-truth-conditional semantic and pragmatic meanings were not demonstrated even at the age of 9;6. I argue that it is uniformity which distinguishes semantics from pragmatics for adults. For children, it has been argued that earlier acquisition distinguishes semantics from pragmatics (e.g. Hyams 1996). I argue that the distinction between semantic and pragmatic meanings in age of acquisition is a reflection of the relational complexity of these meanings, for instance as measured by Halford, Wilson & Philllips' (1998) relational complexity metric, and is not related to their classification as semantic or pragmatic meaning.

^{*} The study presented here was carried out in preparation of my doctoral thesis (2003) completed under the supervision of Dr. Jeannette Schaeffer and Prof. Nomi Shir. This current work was supported by grant no. 236-06-07 from the National Institute for Psychobiology in Israel to Dr. Jeannette Schaeffer. I would like to thank the Zlotowski Center for Neuroscience, the Department of Foreign Literatures and Linguistics and the Faculty of Social Sciences and Humanities at Ben-Gurion University for their generous support in funding my participation at the Workshop on Experimental Semantics and Pragmatics (Bamberg, February, 2008) where the ideas in this paper were originally presented. Thanks are also due to the participants of this workshop whose comments contributed to the refining of my ideas. Finally, I would like to thank an anonymous reviewer for careful and thought-provoking comments.

1. Introduction

This paper discusses the proposal that *response uniformity* is a characteristic distinguishing semantic from pragmatic meaning in psycholinguistic experiments. Furthermore, the uniformity of adult responses, or lack thereof, is a crucial factor in interpreting child acquisition data for which adult behavior is the reference point. Only when uniform adult responses are obtained can a clear target for acquisition be identified. Adult and child experimental data on Hebrew coordination are used to support this proposal.

By *response uniformity* I refer to 100% or near 100% consistency of responses on experimental tasks, both within-individuals and between-individuals. This proposed experimental criterion is predicted by a theoretical distinction between (compositional) semantics and pragmatics. Semantic meaning has been characterized as arbitrary, context independent, constant meaning (e.g. Grice 1975; Horn 1991; Levinson 2000). Thus by definition, given the speakers'/hearers' familiarity with the standard arbitrary meaning of an expression, this semantic meaning does not vary with context, and should not be dependent on the individual knowledge or personality of the speaker or hearer. This leads to the premise that responses to semantic psycholinguistic tasks should not vary between-individuals. Furthermore, the constant nature of semantic meaning renders it resistant to contextual influences such as those that vary with experimental item or experimental task. This leads to the second premise, specifically that responses to semantic psycholinguistic tasks should not vary within individuals.

Pragmatic meaning, on the other hand, has been characterized as non-arbitrary, context dependent, variable meaning (e.g. Grice 1975; Horn 1991; Levinson 2000). This characterization leads to premises directly opposed to those suggested for semantics. As this meaning is arbitrary and in the case of pragmatic implicatures, is calculated for each usage of an expression, variability of calculation is likely. Hence, it follows that pragmatic responses in psycholinguistic experiments will show between-individual variability. The context dependence of pragmatic meaning is directly translated into sensitivity to specific experimental item effects and task effects. This leads to the proposed premise that responses to pragmatic psycholinguistic experiments will show variability within-individuals.¹

A possible exception to this proposal is the case of Generalized Conversational Implicatures (hereafter, GCIs, e.g. Levinson 2000). These implicatures are considered to be the default interpretation of an utterance and as such should

^{1.} It could be argued that within individual variability is a result of task difficulty or 'guessing'. This point will be discussed in detail below.

be calculated uniformly. However, these implicatures remain context dependent in that they are cancelled in particular contexts. Non-uniformity for these implicatures can then be seen across contexts, i.e. there is a predicted lack of uniformity between the responses to the same utterance in a default versus a particularized context.

The data I bring to support these premises come from the (compositional) semantics and pragmatics of (Hebrew) coordination.² Specifically, I investigated adult and children's knowledge of the semantic (truth-conditional and non-truth-conditional) meaning of *ve*/and, *o*/or and *aval*/but, and the pragmatic implicature which leads to the preference of the exclusive (over the inclusive) interpretation of disjunction. Note that I argue that the non-truth-conditional contrastive meaning of *aval*/but is semantic; this is elaborated on below.

The paper is organized as follows: in Section 2 I briefly review the compositional semantics of conjunction and disjunction and related implicatures, in Section 3 I present the experimental hypotheses and predictions, in Section 4 I describe the experiments carried out, Section 5 reports the results, Section 6 discusses the adult results in the context of the premises brought in this introduction and the implications for interpreting the child data, Section 7 discusses these results in the context of cross-linguistic semantic and pragmatic research, with an emphasis on adult data, Section 8 discusses the relationship between theoretical and experimental distinctions between semantic and pragmatics, and Section 9 summarizes and concludes the paper.

2. (Some of) the semantic and pragmatic meanings of coordination

2.1 Truth-conditional semantic meaning

The (compositional) semantics of coordination can be roughly divided into truthconditional and non-truth-conditional meaning. Note that a distinction is made here between lexical and compositional semantic meaning. Lexical meaning is restricted to the meaning associated with a word without reference to how this meaning interacts with other word-meanings in a phrase or sentence. By contrast, compositional semantic meaning deals with the rules of how word meanings interact in a phrase or sentence. Often, formal logical methods are used to describe compositional semantic meaning. Thus, in accordance with traditional analyses, truth-conditional meaning of the coordinators parallels the Boolean

^{2.} The experiments were carried out in Hebrew. However, I expect the same response patterns for these phenomena to be observed cross-linguistically.

definitions of conjunction and disjunction (e.g. Keenan & Faltz 1985). As such the truth-conditions of disjunction include the truth-conditions of conjunction and conjunction entails disjunction, but not vice versa. This circumstance results in the scale \langle conjunction, disjunction \rangle or as can be translated into natural language, in our case $\langle ve/and, o/or \rangle$. This relationship is illustrated in (1).

- (1) Example of truth-conditional meaning of conjunction and disjunction
 - a. Conjunction

Utterance: Cookie Monster eats cookies and cake. Truth-conditions: It is true that Cookie Monster eats cookies and it is also true that Cookie Monster eats cake.

- b. Disjunction: Cookie Monster eats cookies or cake. Truth-conditions:
 - i. Cookie Monster eats cookies (and doesn't eat cake).
 - ii. Cookie Monster eats cakes (and doesn't eat cookies).
 - iii. Cookie Monster eats cookies and also eats cake.

2.2 Pragmatic meaning

The semantic scale provides the necessary conditions for a scalar quantity implicature to be calculated so that the use of disjunction leads the hearer to calculate that the use of conjunction is prohibited as false (at least as far as the speakers knows). This implicature, exemplified in (2), is the pragmatic meaning investigated in the current study.

(2) Example of scalar implicature of disjunction

Utterance: Cookie Monster eats cookies or cake. Implicature calculated by hearer: To the best of the speaker's knowledge, Cookie

Monster eats cookies (and not cake) or Cookie Monster eats cake (but not cookies) but he does not eat both.

Calculation: If the speaker knows that Cookie Monster eats both cake and cookies it would be more felicitous (based on the maxim of quantity) to use 'and' and not 'or'.

Note that the pragmatic meaning of coordinators is not limited to this implicature. Other pragmatic inferences include non-scalar quantity implicatures (such as temporal sequence), relevance implicatures and clausal implicatures.

2.3 Non-truth-conditional arbitrary meaning

A further part of the meaning of coordinators is the contrast associated with *aval*/but. This meaning has been alternatively considered a conventional implicature (e.g. Grice 1975) or presupposition (e.g. Winter & Rimon 1994) or even an 'impliciture' (Bach 1994). This type of phenomena is often placed on the border between semantics and pragmatics. I argue that based on the distinguishing theoretical characteristics given above, as an arbitrary, context independent, constant meaning this contrast should be seen as semantic. However, this meaning is not truth-conditional, i.e. the truthful use of the term *aval*/but is dependent on this meaning holding in the world or a model of the world, only to the extent that *aval*/ but means the truth-conditions of conjunction. In other words, the truthful use of *aval*/but requires that the coordinated conjuncts both be true. In opposition to this, the contrastive meaning of *aval*/but, that part of its meaning that contrasts the conjuncts, does not affect its truth-conditions. In other words, even if the conjuncts are not contrasted, the coordination will be considered true. In this latter case the conjunction may be rejected, but not on the grounds of its being false. The suggested criterion of experimental response uniformity can be used to settle the classification of the contrast of *aval*/but. The contrastive meaning is illustrated in (3).

 (3) Illustration of the contrast of *aval*/but
 Utterance: haish hixlik aval lo nafal 'the man slipped but didn't fall'
 Conventional implicature: There is a contrast between slipping and not falling.

I return to the issue of the semantic versus pragmatic classification of this type of meaning in the discussion of the experimental results below.

By taking the theory that uniformity distinguishes semantic meaning from pragmatic meaning experimentally, and applying this theory to (Hebrew) coordination, hypotheses and predictions can be formulated. These are presented in the following section.

3. Hypotheses and predictions

3.1 Hypotheses and predictions for adults

The underlying idea I suggest is that the theoretical difference between semantics and pragmatics translates into an experimental difference, that is, adult responses to semantic experiments will be uniform while adult responses to pragmatic experiments will be non-uniform. This is expressed in (4).

 (4) Hypothesis 1: Semantic tasks yield uniform responses within and between individuals. Hypothesis 2: Pragmatic tasks yield non-uniform responses within and between individuals. In addition, I propose the theoretically motivated hypothesis that the contrast of *aval*/but is semantic since it is arbitrary, non-cancellable, context independent meaning. This hypothesis appears in

(5) Hypothesis 3:

The contrastive meaning associated with *aval*/but is semantic.

Predictions regarding the semantics and pragmatics of coordination follow from these hypotheses. Hypothesis 1 (and 3) leads to Predictions 1 and 2, presented in (6), while Hypothesis 2 leads to Prediction 3, presented in (7).

- Predictions for adult behavior on the semantics of coordination
 Prediction 1: Adults will uniformly reject false coordinations (conjunctions and disjunctions).
 - Prediction 2: Adults will uniformly reject true non-contrastive coordinations using *aval*/but.
- (7) Predictions for adult behavior on the pragmatics of coordinationPrediction 3: Adults will reject infelicitous coordinations non-uniformly.

3.2 Hypotheses and predictions for children

When carrying out child acquisition studies, adult behavior serves as the standard. To this end, the children's responses are classified as 'adultlike' or 'nonadultlike' rather than 'correct' or 'incorrect'. In the case where adults give clear and uniform responses, adultlike ("correct") behavior is defined as adultlike uniformity. However, in the case where adults give non-uniform responses, how will adultlike 'correct' behavior be defined?.

It has been previously suggested (e.g. Hyams 1996) that age of acquisition distinguishes semantics from pragmatics in child language. In this model, language is considered to be comprised of three main components: lexicon, grammar and pragmatics. The grammatical component includes morphology, syntax, compositional semantics and phonology. Compositional semantic meaning, as part of grammar, is predicted to develop early (if it is not innate, as some versions of the theory propose). This is particularly the case when we consider those parts of compositional semantic meaning which appear to be (virtually) universal, such as conjunctive and disjunctive meaning (see Chierchia, Crain, Guasti & Thornton 1998). On the other hand pragmatic meaning and lexical meaning, external to the grammar are predicted to develop later, exactly when depending on general cognitive processes. Hence, there may be found an interaction between lexical and pragmatic development in early acquisition, where early words are restricted to specific contexts (e.g. Clark 1993; Bloom 2000). Of course, if the adult pragmatic behavior is inconsistent, this second prediction becomes very difficult to falsify. In other words, when children are inconsistent in their responses to pragmatic tasks, is this a reflection of non-adultlike knowledge or of adultlike non-uniformity?

I have previously suggested an alternative hypothesis, whereby child knowledge of semantics and pragmatics are not distinguished by age/order of acquisition, rather, both semantic and pragmatic knowledge develop in accordance with a general cognitive ability to process complex relations (e.g. Paltiel-Gedalyovich 2003).³ The truth-condition of conjunction and disjunction are analyzed as ternary.⁴ Ternary level relations are predicted to be mastered by 5 years (Halford et al. 1998), thus the predicted age of acquisition for the truth-conditions of the coordinators is 5. This is summarized in (8).

(8) Hypothesis 4:

The truth-conditional meanings of conjunction and disjunction are ternary level relations.

Prediction 4: 5-year-old children who are able to process ternary level relations will reject false coordinations uniformly.

^{3.} A full explanation of this analysis is beyond the scope of this paper. The interested reader is referred to Halford et al. (1998) for a full description of the relational complexity metric and to Paltiel-Gedalyovich (2003) and Paltiel-Gedalyovich & Schaeffer (2004) for the application of the metric to coordination. For the purposes of the current discussion, suffice to say that in applying this metric to all types of linguistic knowledge, lexical, compositional semantic and pragmatic meaning are all considered as relations which can be analyzed for complexity level. This particular analysis is not crucial to the argument that compositional semantic, as opposed to pragmatic, responses are uniform. The metric does serve to explain developmental difficult of various phenomena and thus clarify when children's responses are non-uniform for developmental reasons and when because they have adultlike variability in pragmatic tasks.

^{4.} The analysis of the complexity of conjunction/disjunction is as follows:

⁻ The truth value of the coordinated sentence derives from the truth-values of each of the conjuncts/disjuncts.

The relationship between the truth values can be seen as an ordered set. For conjunction this set is (T,T,T) where the first two values represent the truth-values of the conjuncts and the third the value of the coordinated sentence. This is a binary function having ternary level difficulty.

For disjunction, a superset of ordered sets can be envisaged $\{\langle T,F,T \rangle, \langle F,T,T \rangle, \langle T,T,T \rangle\}$ where again, each possibility is a binary function having ternary level difficulty. This analysis of conjunction differs from that proposed by Halford et al. (1998).

The application of the relational complexity metric to the contrast of *aval*/but results in a quaternary level relation.⁵ Quaternary level relations are predicted to be mastered by 11 years (Halford et al. 1998). Thus, children will not achieve the uniform adultlike rejection of non-contrastive coordinations using *aval*/but until the age of 11 years. This is summarized in (9).

(9) Hypothesis 5:

The contrast of *aval* 'but' is quaternary(Paltiel-Gedalyovich, 2003)Prediction 5:11 year-old children who are able to process quaternary level
relations will reject true non-contrastive coordinations using
aval/but uniformly.

Regarding non-uniform adult pragmatic responses, I could adopt the adult prediction as the target, resulting in the hypothesis and prediction suggested in (10).

 (10) Hypothesis 6:
 Child responses in experiments testing pragmatic knowledge will be non-uniform.
 Prediction 6: Children will reject infelicitous coordinations inconsistently.

Using the relational complexity metric provides a possible framework for interpreting child data in the face of inconsistent adult data. By providing an independent analysis of the difficulty of different meaning relations, the metric can help distinguish children's non-uniformity resulting from adultlike variability for pragmatics from non-uniformity resulting from lack of adultlike pragmatic knowledge. I return to this central issue below.

- Existential quantification over r (unary function)
- Negation of r (unary operation)
- Material implication $(p \rightarrow r)$ (binary operation)
- Possibility operator $(p \rightarrow \neg r)$ (unary function)
- Material implication $(q \rightarrow r)$ (binary operation)

^{5.} A full description of the application of the metric is beyond the scope of this paper, the interested reader is referred to Paltiel-Gedalyovich (2003). Briefly, the analysis makes use of Winter's & Rimon's (1994: 377) formalization of the contrast: "*Contrast relation of aval*/but -A proposition **r** *establishes contrast* between two (ordered) propositions **p** and **q** iff \Diamond (**p** $\rightarrow \neg$ **r**) \land (**q** \rightarrow **r**) is true. This relation is denoted by Θ r(**p**,**q**)." is as follows:

Processing of the contrast relation of aval/but.

Conjunction of result of second and third steps (binary operation of unary function and binary operation resulting in quaternary level relation)

4. The experiment

The experimental conditions described here were part of a larger experiment into adult and child knowledge of the semantics and pragmatics of coordination reported in Paltiel-Gedalyovich (2003). Only those conditions bearing on the present argument are presented.

4.1 The task and materials

The task was a Truth/Felicity-Value Judgment task adapted from Crain & Thornton (1998). Note that in the case of a semantic judgment the task involves judgment of truth, while in the case of the pragmatic judgment, the task involves judgment of felicity. In the case of the non-truth-conditional contrast of *aval*/but what is being judged is not clear cut. On one hand, use of *aval*/but when there is no contrast between the conjuncts does not render the coordinated sentence false, on the other hand, it is not clear that the speaker who utters such a coordination is being infelicitous. For current purposes I group these judgments with the truth-judgments. Note that I assume that participants are unaware as to whether they are judging truth or felicity.

Participants were individually presented with color pictures.⁶ These were first described by the experimenter, in the presence of a puppet.⁷ The puppet then described the picture using a coordinated sentence. The participant then judged the puppet's description for truth/felicity. Participants were also asked to 'teach' the puppet by providing an improved description. These corrections were used to verify the motivation for participants' judgments but were not used for quantitative analysis.

There were 10 target items and 5 control items per condition.

4.1.1 Condition 1

Condition 1 investigated knowledge of the truth-conditions of *ve*/and. The 10 experimental items were false coordinations describing color pictures, the 5 control items were true descriptions using coordinations. The predicted responses are

^{6.} The pictures were prepared using the Sesame ARTWorkshop (1995) software program.

^{7.} The puppet was used to facilitate responses from the young children who participated in the same experiment. The use of the puppet with older children and adults was maintained in the interest of retaining the same procedures for all participants. The use of this device was explained to the older participants as necessary since the same experiment was to be carried out on younger children. These older participants accepted the 'childish' ploy and the accompanying explanation.

uniform rejection of the experimental items and uniform acceptance of the control items. A sample experimental item is presented in (11).

(11) Condition 1 – experimental item Picture stimulus:⁶ Puppet: arik oxel tapua

4.1.2 Condition 2

Condition 2 investigated knowledge of the truth-conditions of *o*/or. The 10 experimental items were false disjunctions describing color pictures, the 5 controls were true descriptions using conjunctions. True conjunctions rather than true disjunctions were used to avoid the complication of rejection on pragmatic grounds. The predicted responses are uniform rejection of the experimental items and uniform acceptance of the control items. A sample experimental item is presented in (12).

'Ernie eats (an) apple and (a) banana.'



4.1.3 Condition 3

Condition 3 investigated knowledge of the contrast associated with *aval*/but. The 10 experimental items were true non-contrastive conjunctions describing color pictures, using *aval*/but, the 5 controls were true contrastive conjunctions. The predicted responses for adults are uniform rejection of the experimental items and uniform acceptance of the control items. This prediction follows from the hypothesis that this contrast is semantic and therefore should yield uniform responses. A sample experimental item is presented in (13).

^{8.} It has been suggested to me that this picture invites a variety of descriptions, aside from the coordination which the puppet presents for judgment. Although other descriptions would certainly be true of this picture, rejection of the puppet's description when occurring was explained by negation of the coordinaton, as evidenced by the corrections offered to the puppet. In an expressive task it would certainly be necessary to refine the picture stimuli.

(13) Condition 3 – experimental item Picture stimulus:^{6,8}

> Puppet: ugifletset ohev ugiyot aval hu oxel harbe 'Cookiemonster loves cookies but he eats a lot.'

4.1.4 Condition 4

Condition 4 investigated knowledge of the scalar implicature associated with disjunction. The 10 experimental items were true disjunctions describing color pictures, where both disjuncts are true, the 5 controls were false descriptions using disjunctions. The predicted responses for adults are non-uniform rejection of the experimental items and uniform rejection of the control items. A sample experimental item is presented in (14).

 (14) Condition 4 – experimental item Picture stimulus:⁶
 Puppet: tsiporet loveshet meil o tseif

'Bigbird wears a coat or a scarf.'

4.1.5 Condition 5

Condition 5 investigated knowledge of the pragmatic non-scalar implicature associated with contrastive conjunction. The 10 experimental items were true contrastive conjunctions describing color pictures, using *ve*/and, the 5 controls were false conjunctions. The predicted responses for adults are non-uniform rejection of the experimental items and uniform rejection of the control items. A sample experimental item is presented in (15).

(15) Condition 5 – experimental item Picture stimulus:⁶



Puppet: arik ohev glida vezorek ota la pax 'Ernie loves ice cream and throws it in the bin.'

4.1.6 Condition 6

Condition 6 investigated the possibility of cancellation of the scalar implicature associated with disjunction. The 10 experimental items were true disjunctions

describing color pictures, where both disjuncts are true, in a special context of doubt. The 5 controls were true disjunctions with one true disjuncts, under the same special context of doubt.⁹ The predicted responses for adults are non-uniform acceptance of the experimental items and non-uniform acceptance of the control items. A sample experimental item is presented in (16).

(16) Condition 6 – experimental item
 Picture stimulus:⁶



Puppet [hiding]: wai shaxaxti, eix ani ezkor bli lehistakel ... ulay, ulay ani zoxeret, ulay arik mesaxek ba xevel o bakadur?

'Oh, I forgot, how can I remember without looking ... Maybe, maybe I remember, maybe Ernie's playing with a rope or with a ball?'

4.2 Participants

Participants were 17 monolingual Hebrew speaking adults and 141 monolingual typically developing children. The children were divided into seven age groups. These groups were 3-year-olds (n = 17), 4-year-olds (n = 31), 5-year-olds (n = 11), 6-year-olds (n = 6), 7 year-olds (n = 20), 8-year-olds (n = 21) and 9-year-olds (n = 19). There were roughly equal numbers of girls and boys in each age group.

5. Results and discussion

Results for adults and children are presented separately for semantic and pragmatic meaning. Results for all groups for all conditions are summarized in Table 1.

Note that in Table 1 the results are presented as percent acceptance of stimuli, where the prediction for conditions 1, 2, and 3 is that stimuli will be rejected

^{9.} This context is phrased in a downward entailing environment to facilitate cancellation of the scalar implicature. In retrospect the controls were not appropriate and should probably have been presented as experimental items in a separate condition. However, as the experimental items of conditions 3, 4, and 5 and their controls were randomly presented, this is not considered to have bearing on the experimental results, especially not with regard to the specific question of response uniformity under discussion.

This condition becomes especially important in light of the uniformity of responses actually collected for the scalar implicature in condition 4. This condition adds the context that provides the variability of responses.

Age Group	Conditions					
	1 truth conditions of <i>ve</i> /and	2 truth conditions of <i>o</i> /or	3 contrast of <i>aval</i> /but	4 calculation of the GCI of disjunction	5 calculation of the PCI of contrastive conjunction	6 cancellation of the GCI of disjunction
3	16.7	17.8	92.2	89.0	85.0	93.0
4	11.7	20.0	96.2	87.2	80.5	86.8
5	3.0	7.0	97.0	87.5	97.5	90.0
6	1.4	2.9	97.1	85.0	88.3	86.7
7	2.5	2.5	99.5	81.0	96.5	82.0
8	0.0	0.5	85.5	42.5	98.0	41.5
9	0.5	2.0	64.5	48.5	79.5	57.0
Adults	1.3	1.3	6.3	2.9	50.0	25.9

Table 1. Summary of results for all groups and conditions as percent stimuli accepted

if participants have knowledge of the compositional semantic knowledge. For conditions 3 and 4 the prediction is that stimuli will be rejected if the implicatures are calculated, and for condition 6, the prediction is that the stimuli will be accepted if the implicature is cancelled. The results in the text are presented as percentage accepted or rejected dependent on the condition and varying for clarity of the argument.

5.1 The adult results

The adult results are the test for response uniformity as an experimental criterion for distinguishing semantics and pragmatics.

5.1.1 Semantic meaning

As predicted, adults responded uniformly in conditions 1 and 2 testing truthconditional semantic meaning of conjunction and disjunction, respectively. They rejected false coordinations 98.7% of the time.

Also as predicted, adults responded uniformly in condition 3, testing nontruth conditional meaning of the contrastive coordinator *aval*/but, rejecting noncontrastive coordinations using *aval*/but 93.7% of the time. The uniformity of the responses supports the claim that this contrast is semantic.

5.1.2 Pragmatic meaning

Recall that, in general, performance on tasks testing pragmatic meaning is predicted to be non-uniform.

Adults gave unpredictedly (for pragmatics) uniform responses in condition 4, rejecting true disjuncts when both disjuncts are true 97.1% of the time. However, recall that the predicted exception to the predicted non-uniformity is the case of GCIs. As predicted in condition 6, where the special context was provided, adults gave non-uniform responses, rejecting the true disjuncts with both disjuncts true only 74.1% of the time. Note that this lack of uniformity does not reflect an arbitrary guessing pattern. The adults can be seen as two groups. The first group, comprised of the majority of the adults (12 or 70.6%), did not cancel the GCI, despite the specialized context and calculated the GCI at least 80% of the time. The second group, comprising only two of the adult participants (11.8%), consistently cancelled the GCI in the specialized context.¹⁰ On the one hand, this result highlights the generalized nature of this implicature and the strong contextual changes required to cancel it. On the other hand, this result differs significantly (p < 0.05) from the results for the default context, indicating that given a special context, the cancellability - or variability - of the implicature can be demonstrated. Of course the small number of participants in the second group restricts conclusions that can be drawn here and further research would strengthen this argument. Similar within (adult) group non-uniformity has been found by other researchers (see below).

For the non-generalized implicature tested in condition 5, as predicted, adults rejected the use of a non-contrastive coordinator, *ve*/and in a contrastive context only 50% of the time. Again, the adult participants can be seen as two distinct groups. However, in this case, the groups were of the same size. The first group was comprised of seven adults (41.1%) who calculated the PCI at least 80% of the time. The second group was comprised of seven adults who accepted the descriptions of the stimulus pictures at least 80% of the time, thus indicating that they did not calculate the PCI.¹¹

5.2 The child results

5.2.1 Semantic meaning

For semantic meaning, the adult results provide a clear target. Thus, the child results may be reported not only in terms of percentages, but also in terms of identity/ difference as compared to the adult results.

^{10.} The remaining adult participants did not perform consistently, cancelling the GCI in the specialized context less than 80%, but more often than 20% of the time.

^{11.} Here, too, the remaining adult participants did not perform consistently, calculating the PCI less than 80%, but more often than 20% of the time.

As predicted, by 5 years uniform rejection of false conjunctions and disjunctions was found (93.0%–97.0%, respectively), i.e. by 5 years children do not behave significantly different from adults (F (1,127) = 6.12, MS error = 1.98, p < 0.05, for conjunction, and F (7,127) = 5.779, MS error = 1.9235, p < 0.05 for disjunction).

The contrast of *aval*/but proved more difficult, as predicted, and even children as old as 9;6 years were not adultlike and rejected non-contrastive true conjunctions using *aval* 'but' only 36.5% of the time ((F(1,127) = 57.767, MS error = 5.221, p < 0.05).

5.2.2 Pragmatic meaning

As predicted, calculation of the generalized scalar quantity implicature was nonuniform for children. Even the oldest group, the 9 year olds, rejected true disjunctions having both disjuncts true only inconsistently (40.5% rejection, F(7,115) =13.07, MS error = 11.623, p < 0.05). In the case of a specialized context, all children calculated the implicature inconsistently (41.5%–93.0% acceptance). Note that although the results for the adults were also non-uniform, there was a significant effect for age with all age groups calculating the implicature significantly less often than the adults (F(7,115) = 8.1999, MS error = 10.496, p < 0.05).

Similarly, calculation of the particularized quantity implicature was non-uniform in all child age groups with rejection of contrastive true conjunctions using *ve*/and only very rare even for the 9;6 year olds (10.5% rejection). Although the adult calculation was non-uniform (50.5%), the children's behavior differed significantly form the adults (F(7,115) = 9.364, MS error = 4.461, p < 0.05).

5.3 Summary of results

As predicted, adults produced uniform responses for conditions addressing (semantic) truth-conditional meaning and the contrast of *aval*/but, thus supporting the classification of this contrast as semantic, within the context of the current hypothesis. Also as predicted, adults provided non-uniform responses for pragmatic meaning. For the particularized implicature this is seen in non-uniform calculation within a single condition; for the generalized implicature, this is seen in a comparison of the difference between consistent responses in a default context and variable responses in a specialized context.

The uniform adult results for semantics provide a clear target for child language such that the child data can be seen as adultlike or non-adultlike. Knowledge of truth-conditional meaning was demonstrated consistently from the age of 5 years, while knowledge of the contrast of *aval*/but was not demonstrated even by the oldest children tested. Pragmatic implicatures were calculated only inconsistently by all age groups, but given the non-uniformity of the adult results, this result is problematic for interpretation.

6. Summary and general discussion

6.1 Adult response (non)uniformity for semantics and pragmatics

Recall that I hypothesized that adults' responses in experiments testing semantic meaning would be uniform, both within and between participants. When translated into specific predictions regarding (Hebrew) coordination, this anticipates that adults will uniformly reject false conjunctions and disjunctions, since truth is a semantic meaning. This prediction was upheld, supporting the semantic uniform response hypothesis.

The uniformity of the responses, as well as being experimentally demonstrable, is theoretically consistent with a characterization of semantic meaning. Characteristics of semantic meaning have included arbitrariness/non-calculability, context independence, and non-defeasibility (as brought in the introduction above). Since the contrast of *aval*/but is consistent with this characterization, the prediction that follows is that experiments into knowledge of this contrastive meaning would yield uniform responses from adults. This prediction, too, was supported.

The opposite characterization of pragmatics, non-arbitrariness/calculability, context dependence and defeasibility (see introduction) is consistent with the characteristics of quantity implicatures. This leads to the prediction that adult responses to tasks testing this type of meaning should be non-uniform. This prediction too was borne out. For the case of the particularized implicature, the results were clearly non-uniform. For the case of the generalized implicature, uniform, semantic-like responses were found for the default context, apparently counter-evidence to the hypothesis of pragmatic non-uniformity. However, this uniformity can be explained by the generalized nature of the implicature. Yet, once all contexts are considered, i.e. both default and particularized contexts, non-uniform responses are obtained, again lending support to the hypothesis.

The primary hypothesis investigated in this paper is that semantic meaning by definition is interpreted consistently and therefore yields uniform responses when investigated experimentally. By contrast pragmatic meaning is by definition interpreted inconsistently and therefore yields non-uniform responses when investigated experimentally. The data brought from Hebrew coordination supports this hypothesis. I now turn to the implications of this result for interpreting child language data.

6.2 Interpreting child responses to tests of semantic and pragmatic meaning

The premise of investigations into child language acquisition is that child language can be compared to an adult target and based on this comparison conclusions can be made about how closely a child's language behavior (and therefore grammar) resembles the behavior (and grammar) of adult speakers of the language. When clear patterns of response are obtained, such as (near) 100% responses of a particular type to a particular type of test or alternatively (near) 0% responses of a particular type, we have a clear model of the adult target, a defined adult behavior and a basis for hypothesizing the adult grammar. On the other hand when adult response patterns are inconsistent or non-uniform, it is difficult to derive the adult grammar and there is therefore no clear reference point for interpreting the children's results.¹²

The data brought here suggest an inherent problem in the interpretation of child data from experiments into pragmatic behavior. If we take the case of the particularized implicature, we cannot tell if the inconsistency of child responses is the result of immaturity, i.e. a difficulty in calculating the target implicature, or if the inconsistency simply mirrors the adult non-uniformity. It has been argued that adult inconsistency is the result of difficulty (e.g. Pouscoulous, Noveck, Politzer & Bastide 2007). If this is the case, then clearly there is a point in comparing the child data to the adult data and considering how well the children do as compared with the adults in the processing of this difficult meaning. We would need to identify the source of the difficulty and explain why this difficulty is present in the calculation of PCIs, for instance, but not in the calculation of GCIs. Recall that for GCIs, the adult results are consistent and inconsistency is found across contexts. It would be difficult to argue that the context itself provides the difficulty. It could be argued that cancellation of a GCI is more difficult than its calculation, and hence the discrepancy between the consistency of responses on the calculation of the GCI and its cancellation. However, the question remains as to why the calculation of the

^{12.} As has been suggested to me by an anonymous reviewer, a (near) 100% criterion for adult behavior to be considered uniform may be unrealistic, given that some variability exists in all (linguistic) behavior. This level was upheld by the experimental results obtained. Given that in those tasks which were (theoretically) clearly semantic, adult responses approached 100%, this was taken to be the criterion for uniformity. Admittedly, less consistent responses may still be considered uniform. Note, that for the pragmatic tasks, adults who accepted or rejected stimulus items 80% of the time were considered to be consistently calculating/cancelling an implicature. It is worthy to note however, that this considerably lower level of consistency was found for pragmatic meaning, not for semantic meaning. It remains to be seen in further investigations of adult semantic knowledge, if the near 100% criterion is realistic or not for an experimental situation.

GCI is consistent while the calculation of the PCI is not. As I have previously argued (Paltiel-Gedalyovich 2003), based on a relational complexity metric suggested by Halford et al. (1998), the difficulty of the calculation of the GCI of disjunction and the PCI of contrastive conjunction is similar. Thus, there is no analysis at present attributing the difference in the consistency of the calculation of these different implicatures to their difficulty. However, the GCI/PCI distinction can explain the difference in the consistency of responses for adults.

A further point to note is that even the youngest children tested in the current experiments, the three year old group, demonstrated considerable knowledge of scalar implicatures (over 80% calculation). Yet, they did not reach adult levels of uniformity. Similarly, other researchers have found non-chance level calculation of scalar implicatures by young children (e.g. Papafragou & Musolino 2003; Pouscoulous et al. 2007). I do not argue that young children are not capable of calculating these implicatures, but rather that they do not reach adult levels of uniformity because the complexity of the relations involved in the scalar implicatures creates too great a burden on processing to allow consistent calculation.

It has been suggested, that a critical point to be considered, in a discussion such as this on children's acquisition of pragmatic behaviors, is the influence of adult pragmatic behavior, as observed by the children, on the children's own developing behavioral patterns.¹³ Regarding the current issues, do children observe adults (in)consistency in pragmatic behavior and model their own behavior accordingly? Although this point is certainly of interest, it is beyond the scope of the current paper. The thrust of the current argument is the need to take into account adult variability, and the source of such variability, when evaluating the maturity of children's performance on pragmatic tasks.

In summary, as yet, there does not appear to be a definitive answer to the question posed regarding the correct source of non-uniformity in children's responses on (pragmatic) tasks where adults also perform non-uniformly. In my view, the answer can be found in an independent evaluation of possible sources of inconsistent behavior. In cases where a theoretically based analysis of the behavior suggests difficulty, and adult inconsistency appears to reflect that difficulty, this analysis constitutes justification of attributing non-uniform responses to this difficulty. One such analysis is that based on the relational complexity metric (Halford et al. 1998). In these cases, I would predict a learning curve to be seen in which there is an increase to adult levels of consistency (even if these levels never reach uniformity), and less than adult level consistency will be interpreted as immaturity. In fact, this is what

^{13.} This point was made by an anonymous reviewer.

is seen in the calculation of the GCI in the experiments reported above. A similar learning curve was found by Pouscoulous et al. (2007).

On the other hand, in those cases where there appears to be no theoretical analysis that suggests difficulty, the non-uniformity of the adult responses will be considered to be a result of the pragmatic nature of the task. In these cases I would not explain children's non-uniformity as a result of immaturity, but rather as a result of adult-like pragmatic variability. Teasing out the mature versus immature non-uniform behavior in children can only occur by comparison with performance on other tasks, analyzed of comparable complexity. In the experiments reported here, the children's non-uniform, non-adultlike performance on the GCI lends support for their performance on the cancellation of the GCI being the result of immaturity and not adultlike non-uniformity. Furthermore, there does not appear to be a clear learning curve which suggests increased mastery with age, with children moving closer to adultlike levels with increased age.

A closer look at the cross-linguistic data in the following section emphasizes the non-uniformity of calculation of implicature phenomena across languages. Note that this non-uniformity is expressed both in that adults show non-uniform behavior and in the lack of uniformity in patterns of behavior in different languages.

7. Response uniformity cross-linguistically

The idea that response uniformity distinguishes semantics from pragmatics experimentally has implications for interpreting experimental differences found crosslinguistically for apparently identical, or at least very similar, phenomena.

In the case of semantics, very little data is available for adults. In addition to the uniformity (97.1%) of the judgments of the semantic truth-conditions found in the present study for Hebrew conjunction, evidence is reported of adult consistent judgments of the truth of quantifiers in French (95.5%–98.1%, Noveck & Posada 2003). Far more data is available from pragmatic experiments. Non-uniform calculation of implicatures for disjunction has been found in this study, when considering the discrepancy between default and particularized contexts (91.7% and 74–1%, respectively). Similarly, Chierchia, Crain, Guasti, Gualmini & Meroni (2001) found that calculation of this implicature ranged from 98% to 4.5% in varied contexts.¹⁴ Likewise, the investigations of the calculation of a scalar implicature associated with existential quantifiers (all/every, some), have yielded

^{14.} The varied contexts reflect default contexts in which the implicature is largely calculated and other contexts which promote cancellation of the implicature.

varied results cross-linguistically. For French, results have varied from 4% to 93%, with many results in-between (59% Noveck 2001; 4%–92% Noveck & Posada 2003; 53%–93%, with negation 47%; Pouscoulous et al. 2007). For Italian, results vary from 53% to 83% (Guasti et al. 2005). Finally, for Greek, 50% calculation was found (Papafragou & Musolino 2003).

The response uniformity of the children's responses is subject to two factors. The first is the same factor that affects the adult responses, namely the nature of the meaning, semantic versus pragmatic. The expectation is that semantic responses will be uniform and pragmatic responses will be non-uniform. The second factor affected the uniformity of the children's responses is the developmental difficulty of the meanings involved.

First I consider cross-linguistic child data for semantic knowledge, particularly, truth-conditional knowledge. Again, this data comes primarily from studies aiming to investigate pragmatic knowledge, such that data on knowledge of semantic meaning is secondary. In general, when children are found to be "more logical" than adults, when they fail to enrich their semantic interpretations with pragmatic meaning, they demonstrate semantic knowledge, similar to the case where adults cancel an implicature. So in an implicature erasing context as in the first experiment described in Chierchia et al. (2001), where adults' responses demonstrating knowledge of semantic meaning were relatively uniform (96.5%), children demonstrated this knowledge more than 90% of the time. In other words, children, like adults, show uniform responses on tasks requiring knowledge of (compositional) semantic meaning.

Further considering the child data from the study by Chierchia et al. (2001) cited above, in the case where adults showed only 4.5% calculation of the implicature (a downward-entailing, implicature erasing context), the children performed similarly, appearing to calculate the implicature less than 10% of the time.¹⁵ In the default context where the adults calculated the implicature 98% of the time, the children could be divided into two groups. One group calculated the implicature 90% of the time, while the second group calculated the implicature 90% of the time, while the variability in the calculation of the implicature is explained with the varied contexts; pragmatic meaning lacks uniformity because it varies with context. For the children, explaining the non-uniformity is more problematic. The non-uniformity cannot be explained in the same way as the adult variability because the child data is variable within a single context. The adults are consistent for pragmatic meaning within context, in this study, the children

^{15.} No statistical analysis is provided to show whether or not these apparently similar results are in fact not significantly different.

vary. The fact that the children are divided into two groups suggests that those children, who calculate the implicature 90% of the time, are showing adult consistency within context, and varying between contexts, as evidenced by calculation less than 10% of the time in the second context. I argue that those children who show low calculation of the implicature (less than 10% and 7%), regardless of context, are demonstrating immature pragmatic knowledge. Chierchia et al. (2001) explain the difficulty as a problem with reference sets, adopting Reinhart's (1999) Reference Set Hypothesis. Thus, in a further experiment where these investigators provided the children with the choices, or in other words, provided the reference set, the children calculated the implicature with adultlike consistency (90% calculation). Hence, this is one suggestion for teasing out the difference between adult-like variability and developmental difficulty in interpreting pragmatic meaning.

Other examples of inconsistent adult performance can be found in experiments in French, reported by Noveck (2001). In Noveck's Experiment 1, adults calculated and implicature precluding 'must' when 'might is used, 35% of the time. The adult participants showed internal consistency with roughly one third failing to calculate the implicature and two-thirds calculating the implicature. Furthermore, in his Experiment 3, adults are reported to calculate an implicature precluding 'all' when 'some' is used 41% of the time. His sample sentence is 'Some giraffes have long necks.' The implicature involved is considered to be based on the scale (all, some) where the use of the weaker member of the scale implicates that the use of the stronger member of the scale would be false.¹⁶ In each of these cases the adult data is nonuniform on a pragmatic task. Noveck suggests that this is a reflection of the adults equivocation between two possible interpretations - the literal (semantic) interpretation and the non-literal (pragmatically, implicature enriched) interpretation. The group results comprises three adults goups, those who consistently calculate and implicature, those who consistently do not calculate the implicature and those who are inconsistent (the last constituting the smallest group). Noveck brings (among

(i) a. 'All unicorns eat sausages.'

b. 'Some unicorns eat sausages.'

The sentence in (ia) does not entail the sentence in (ib). Given that the set of unicorns is empty, (ia) is true in the world as we know it, while (ib) is false, since sentence (ib) requires the set of unicorns to have at least one member to be true. It has been argued that in natural language such cases should be ignored. I would propose that such cases should not be ignored since the inference that a sentence using 'all' the quantified set is non-empty is a pragmatic inference and not a semantic one. However, a discussion of this point is far beyond the topic of the present paper.

^{16. &#}x27;All' and 'some' only form a scale, (all, some) if we ignore cases such as below

others) an account in the context of Relevance theory. As I understand it, within this theory, the choice of the literal or non-literal interpretation is ultimately dependent on whether or not the hearer's search for relevance is satisfied in the context; if the adult hearer is satisfied with the literal interpretation, and then there is no need to calculate the implicature. The non-uniform results would then reflect differences in the individual adults' requirements for relevance within the same context, similar to the results obtained in my study for the PCI of contrastive conjunction.

When comparing the children's results in these experiments, Noveck argues that they are not as pragmatically aware as the adults, although they appear to have the necessary cognitive abilities to perform in an 'adultlike' manner. When the task is difficult, children are more likely than adults to opt for an easier solution, However, to my view, a decision to classify less-pragmatic behavior as 'immature' is consistent with a problematic judgment that a significant percentage of typical adult speakers show 'immature' behavior. In a case like this, where the children appear 'adultlike' on tasks of similar difficulty, a developmental argument does not appear legitimate. It could well be, however, that personal experience and personality affects adult pragmatic behavior and that certain types of personal styles are more prevalent in children than others. Importantly, the role of context in pragmatic behavior, where context includes hearer attributes such as personality and general experience, remains central.

Furthermore, an argument that computation of scalar implicatures is difficult for adults does not correlate with the adult behavior in the current experiments, where adults calculated these implicatures nearly 100% of the time. Levels of computation of the scalar implicatures were similar to levels of truth-value interpretation of the coordinators. This constitutes counter-evidence to the argument that non-uniformity results from difficulty.¹⁷

In the light of Noveck's experiment, Guasti et al. (2005) investigated factors affecting the variability of adult and child variable calculation of implicatures in Italian. The data brought in this later study support a similar picture to that of Noveck's earlier study. In contrast to some previous accounts, and to the view taken in the current paper, Guasti et al. adopt Chierchia's (2004) argument that implicatures are computed locally, as part of the semantic interpretation. If implicatures are in fact part of the semantic interpretation, then of course the experiments used in the current paper cannot distinguish semantics from pragmatics

^{17.} Note however, that computation of specific scalar implicatures may require different levels of effort in different languages. Thus, the explanation may hold for the languages studied, although not for the Hebrew. This further supports the lack of uniformity of pragmatic phenomena across languages.

as all of the phenomena investigated would be considered semantic. Thus, this later study examines the developmental difficulty of implicatures as a semantic phenomenon. I return to this point below.

Some previous attempts to account for the inconsistency of pragmatic responses in children (and adults) have included those relating different result rates to differences in task methodologies, while others have attributed the inconsistency to processing difficulty. Regarding methodological issues, some researchers have noted difficulty in eliciting consistent responses from adults for judgments of pragmatic meaning (e.g. Hacohen p.c.; Zondervan, this volume). Furthermore, task differences have been shown to greatly influence children's performance (Pouscoulous et al. 2007). Pouscoulous et al. (2007) found that children's (and adults') performance varied dependent on the number of distracter items and the type of task (metalinguistic judgment versus action). In this study, Noveck's initial study, investigating implicatures drawn form the use of 'some' as opposed to 'all', was repeated and similar results were found with 9 year olds calculating the implicature 9% of the time, while adults calculated the implicature 48% of the time. Again the adult rates are far from uniform. Furthermore, when negation was added to the sentences, the calculation of the implicature rates was reduced for both adults and children (37% and 30%, respectively). In this case, the rates for adults and children were similar. The lack of uniformity is explained as reflecting a processing difficulty. This is further investigated in the same study in three further experiments.

However, if the non-uniformity can be attributed to methodological differences, why is the effect of different methodologies so much greater for pragmatic than for semantic meaning? Although little semantic data is available to date, the little that is available appears to support the claim that semantic results are uniform across methodologies, and across languages. The claim that the nonuniformity is the result of processing difficulties has been made to account for data by Pouscoulous et al. (2007). This claim, in the spirit of Reinhart (2006), claims that there is an inherent difficulty in processing of competing options (including scales). The difficulty with this claim is that it does not explain why processing of certain types of sets (such as the for instance the scale $\langle and, or \rangle$) should be easier (as seen by the uniformity of adult responses) than processing of an implicature based on another type of set (such as a contrast set {but, and}). Thus, such an explanation does not account for the results in the present study.

To summarize, children, and to a lesser extent adults are often inconsistent in their calculation of scalar implicatures. Often the explanation given for this inconsistency is the relative difficulty of the calculation of the scalar implicatures. However, detailed explanations of why scalar (particularly as opposed to non-scalar) implicatures are so complex have not been given. I now turn to the relationship between theoretical and experimental distinctions between semantics and pragmatics.

8. Theoretical and experimental criteria for distinguishing semantics and pragmatics

I have presented here an argument for response uniformity as a distinguishing criterion in experimentation into semantics and pragmatics. The criterion relates to adult performance with consequent implications for the interpretation of child data a primary consideration. The suggestion that such a distinction exists does not propose that the semantics/pragmatics distinction should be made experimentally only. The relative contributions of experimental and theoretical arguments in the case can be seen in light of a general perspective on the relative contributions of theory and experimentation in the study of language and language development.

The hypotheses proposed in this paper, and the resulting predictions regarding the semantic and pragmatic meaning of coordinators, are firstly the result of a theoretical definition of semantics and pragmatics. This definition, in a Gricean tradition distinguishes 'what is said', i.e. literal, arbitrary, context independent, conventional meaning, currently, 'semantic meaning', from 'what is meant', i.e. non-literal, non-arbitrary, context dependent, non-conventional meaning, currently, 'pragmatic meaning'. Returning to the specific meanings investigated here, this definition results in the following classification: since any meaning which is arbitrary, conventional and context independent is semantic, both truth-conditional meaning such as the truth conditions of the coordinators, and non-truth-conditional meaning, such as the conventional implicature of *aval*/but are semantic. On the other hand, since both the PCI discussed here, implicating non-contrast when a non-contrastive coordinator (e.g. *ve*/and is chosen over a contrastive coordinator, e.g. *aval*/but), and the GCI of disjunction, are non-arbitrary, context dependent and non-conventional, these meanings are pragmatic.

The classification of the semantic meanings may be transparent, however, regarding the conventional implicature of *aval*/but, let me point out that a serious challenge to this view that this semantic meaning would be, for instance, a suggested context in which *aval*/but does not indicate contrast, or an explanation for how the contrastive meaning is derived.¹⁸

The classification of the pragmatic meanings can be explained in the following way. The GCI and PCI are both calculable (i.e. non-arbitrary, non-conventional)

^{18.} The use of *aval*/but as a discourse marker is not considered to be a counterexample. A discussion of this use of *aval*/but is beyond the scope of this chapter.

since adherence to the maxim of quantity, give as much information as you are able to, explains the need to use the more informative member of a set, *ve*/and rather than *o*/or, and *aval*/but rather than *ve*/and. Furthermore, the calculation of these implicatures are context dependent as we can see from the ease in which contexts can be described in which the GCI is not calculated (e.g. doubt, prediction, conditionals) or the PCI is not calculated (e.g. cases where there is nothing in the context to suggest that a question of contrast or lack thereof arises).

This then, is the theoretical basis of the semantic/pragmatic distinction. This theoretical distinction leads directly to an experimental prediction. Primarily based on the idea of context (in)dependence, in behavioural investigations, semantic meaning is predicted to be uniformly present while pragmatic meaning is predicted to be only non-uniformly present.

As has already been mentioned (in Section 8), some of the more recent experimental research accepts Chierchia's (2004) analysis whereby scalar implicatures are hypothesized to be calculated locally, as part of the semantics. Note that in the 'local implicature' framework, the classification of an implicature a semantic is based purely on when it takes place in the language process. Whereby previous accounts held that semantic meanings were proposed first and then refined by pragmatic implicatures, in this view the literal and enriched meanings are proposed and refined in the semantics, before context comes into play. I find this reasoning problematic primarily because the distinction between semantic and pragmatic meanings appears to be made on the basis of 'where' the refining takes place, as opposed to on the basis of some intrinsic characteristic of semantic versus pragmatic meaning.

It appears that context dependency remains the most robust theoretical distinction between semantic and pragmatic meaning. With regard to the phenomena investigated here, this context dependency theoretically distinguishes the conventional meanings (truth conditions and the conventional implicature) from the pragmatic meanings (GCI and PCI). This hypothesized distinction is reflected experimentally as a difference between uniform and non-uniform behaviour. Uniform behaviour is found for non-context dependent meanings, since the meanings do not change in the context of different tasks or with different extra-linguistic contexts, such as individual style differences. Non-uniform behaviour is found for context-dependent meanings. In case of the GCI, this non-uniform behaviour is clearly related to a change in experimental (linguistic) context.¹⁹ In the case of the

^{19.} Note that the characterization of an implicature as a GCI may not be maintained with the translation of terms, consider the relatively low levels of calculation of the disjunctive implicature in studies in French (Noveck & Posada 2003) and Italian (Guasti et al. 2005).

PCI, the non-uniform behaviour appears to occur within a stable context, seemingly reflecting internal differences in the adult participants.

Experimentally, there are two important implications of this argument. The first is that adult behaviour on pragmatic tasks may be non-uniform, simply because the tasks test pragmatic meaning, and pragmatic interpretations are variable. Second, when studying the development of pragmatic meaning in children, non-uniform performance by adults confound data analysis. In these cases, an independent theoretical argument needs to be made for the complexity of the phenomenon involved in order to argue that less than adultlike consistency is the result of immaturity or processing difficulty. If no such argument is made, we will be forced to accept the children's less consistent demonstration of knowledge of the pragmatic meaning as reflecting adultlike pragmatic variability.

9. Conclusions

In this paper I have brought evidence from Hebrew coordination to support the hypothesis that response uniformity distinguishes semantics from pragmatics experimentally. The phenomenon of GCIs proves an exception to this rule. However, if cross-context data is considered, allowing for contexts facilitating cancellation, the notion of pragmatic non-uniformity is upheld. The implications of the adult (non)uniformity for the interpretation of child acquisition data are important as for pragmatic experiments we often do not have a clear target for comparison. Finally, response uniformity distinguishes semantics from pragmatics cross-linguistically. Note however that little data is available on semantic meaning, so that there is more support at the moment for the inconsistency of pragmatic meaning than for the consistency of semantic meaning cross-linguistically. Further research extending the number of semantic and pragmatic phenomena investigated cross-linguistically will clarify the usefulness of response uniformity as an experimental criterion for distinguishing semantics and pragmatics.

References

Bach, K. 1994. Conversational impliciture. Mind & Language 9: 124-162.

- Bloom, P. 2000. How Children Learn the Meanings of Words. Cambridge MA: The MIT Press.
- Chierchia, G. 2004. Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. In *Structures and Beyond*, A. Belletti (ed.), 39–102.Oxford: OUP.
- Chierchia, G., Crain, S., Maria Teresa Guasti, M.T. & Thornton, R. 1998. Some and or: A study in the emergence of logical words. In Proceedings of the Boston University Conference on Language Development 22, A. Greenhill, M. Hughes, H. Littlefield & H. Walsh (eds), 97–108. Somerville MA: Cascadilla Press.

- Chierchia, G., Crain, S., Guasti, M.T., Gualmini, A. & Meroni, L. 2001. The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In *BUCLD 25 Proceedings*, A.H.J. Do et al. (eds), 157–168. Somerville MA: Cascadilla Press.
- Clark, E. 1993. The Lexicon in Acquisition. Cambridge: CUP.
- Crain, S. & Thornton, R. 1998. Investigations in Universal Grammar: A Guide to Experiments in the Acquisition of Syntax and Semantics. Cambridge MA: The MIT Press.
- Grice, H.P. 1975. Logic and conversation. In *Syntax and Semantics*, Vol.3, P. Cole & J.L. Morgan (eds), 41–58. New York NY: Academic Press.
- Guasti, M.T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A. & Meroni, L. 2005. Why children sometimes but not always compute scalar implicatures. *Language and Cognitive Processes* 20: 667–676.
- Halford, G.S., Wilson, W.H. & Phillips, S. 1998. Processing capacity defined by relational complexity: Implications for comparative, developmental and cognitive psychology. *Behavioral* and Brain Sciences 21 (6): 803–831.
- Horn, L.R. 1991. Given as new. Journal of Pragmatics 15: 313-336.
- Hyams, Nina 1996. The underspecification of functional categories in early grammar. In *Generative Perspectives in Language Acquisition* [Language Acquisition and Language Disorders 14],
 H. Clahsen (ed.), 91–128. Amsterdam: John Benjamins.
- Keenan, E.L. & Faltz, L.M. 1985. *Boolean Semantics for Natural Language* [Synthese Language Library 3]. Dordrecht: Reidel.
- Levinson, S. 2000. Presumptive Meanings. Cambridge MA: The MIT Press.
- Noveck, I.A. 2001. When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition* 78: 165–188.
- Noveck, I.A. & Posada, A. 2003. Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language* 85: 203–210.
- Paltiel-Gedalyovich, L.R. 2003. *Towards an Explanation of First Language Acquisition of Hebrew Coordination*. Ph.D. dissertation, Ben-Gurion University of the Negev.
- Paltiel-Gedalyovich, L.R. & Schaeffer, J. 2004. On children's knowledge of scalar implicatures. In Proceedings of GALA 2003, Vol. 2, J. van Kampen & S. Baauw (eds), 341–352. Utrecht: LOT.
- Papafragou, A. & Musolino, J. 2003. Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition* 86: 253–282.
- Pouscoulous, N., Noveck, I.A., Politzer, G. & Bastide, A. 2007. A developmental investigation of processing costs in implicature production. *Language Acquisition* 14: 347–375.
- Reinhart, T. 1999. The processing cost of reference-set computation: Guess patterns in acquisition. *OTS Working Papers*.
- Reinhart, T. 2006. Interface Strategies: Optimal and Costly Computations. Cambridge MA: The MIT Press.
- Sesame ARTWorkshop 1995. Children's TelevisionWorkshop. Creative Wonders ABC Electronic Arts.
- Winter, Y. & Rimon, M. 1994. Contrast and implication in natural language. *Journal of Semantics* 11: 365–406.

Numerals and scalar implicatures

Daniele Panizza & Gennaro Chierchia University of Trento/Harvard University

In this paper we explore the relation between the meaning of number denoting determiners ('numerals') and the polarity of the context in which they occur. We claim that when numerals are embedded in positive (i.e. Upward Entailing) contexts they are given an upper bounded ('exactly') reading more often than when they are embedded in minimally different Downward Entailing contexts. Since this corresponds to the behavior of scalar triggers, we suggest that the stronger interpretation of a numeral is really due to a Scalar Implicature. We review the outcome of two experiments, a questionnaire and a reading task where eye movements were recorded, which tested this claim empirically. Finally, we discuss these findings in light of the current approaches on the semantics of numerals, analyzing in detail the relation between the meaning of a numeral and scalar strenghtening.

1. Introduction

Sentences quantified by numerals may have (at least) two different meanings: an 'exact' (or upper bounded) meaning and an 'at least' (or lower bounded) one. They are exemplified by (1a) and (1b) respectively:

- (1) a. If I get in the Summer competition, I'll buy four golf clubs
 - b. If I buy four golf clubs, I'll never use my loose old clubs again

We feel that (1a)'s most salient reading is an upper bounded one (..., I'll buy four golf clubs and no more), while (1b) is more naturally interpreted as lower bounded (If I buy four golf clubs or more, ...). This paper is devoted to arguing that this 'feeling' is both general (across speakers) and grounded in grammar and processing.

More specifically, although the interpretation of numerals is influenced by extra-linguistic factors (e.g. the utterance context, speaker and hearer's intentions, etc.) in the present paper we argue that there are also purely structural factors (i.e. the entailment pattern of the construction in which the numeral is embedded, that is, the polarity of the context) that systematically affect the distribution of

lower bounded vs. upper bounded readings of numerals. In (1a) the numeral 'four' occurs in the consequent of a conditional. This position licenses entailments from subsets to supersets, and is therefore an 'Upward Entailing' (UE) environment. That is, (2a) logically entails (2b) (but not vice versa):

- (2) a. If I go to the pub, I'll drink a stout beer.
 - b. If I go to the pub, I'll drink a beer.

In (2), specifically, the entailing inference goes from the 'set of *stout beer* drinkers' to the 'set of *beer* drinkers', with the latter a superset of the former. The antecedents of conditionals, on the other hand, licence the opposite entailment pattern and hence constitute 'Downward Entailing' (DE) contexts. Sentence (3a) logically entails sentence (3b) (but not viceversa):

- (3) a. If I drink a beer, I'll get a headache.
 - b. If I drink a stout beer, I'll get a headache.

The UE/DE contrast has been shown to be relevant to many grammatical phenomena (e.g. the distribution of negative polarity items), and is also relevant, we claim, to the distribution of the 'exact' vs. 'at least' construals of numerals. Here is, specifically, the theoretical claim we put forward in the present paper:

(4) Claim:

The upper bounded interpretation of numerals occurs preferentially in UE contexts with respect to minimally different DE contexts.

The lower bounded interpretation of numerals occurs preferentially in DE contexts with respect to minimally different UE contexts.

Two UE and DE contexts are minimally different in case they share the same lexical material with the exception of the replacement of a DE functor for an UE one, or vice versa.

This claim does not say anything about the absolute proportion of upper/lower bounded interpretations intended by the speaker in any context. It is well known, for instance, that contextual (i.e. knowledge or discourse context based) factors may affect the interpretation of a scalar item (cf. Breheny, Katsos & Williams 2006). Our thesis posits that, whatever the base rate of interpreting a numeral with the upper or lower bounded reading is, structural grammatical factors, like the entailing property of the proposition, systematically affect readers' interpretation as stated in (4). Clearly, native speakers' intuitions of the type often used in current linguistic theorizing are insufficient to establish (4), as the frequency of the upper/lower bounded interpretation of numerals may vary across speakers, or even within a single one. For this reason in the present paper we will discuss an experimental study (cf. Panizza, Chierchia & Clifton 2009) that we conducted to specifically test the claim pointed out in (4). As for processing concerns, (4) does not say anything particular about the cost of deriving an upper/lower bounded reading in either a UE or a DE context. However, many studies in psycholinguistics attest that a dispreferred interpretation of a word or a sentence imposes a processing load with respect to the preferred one (cf. Rayner 1998, for a review of research using the eyetracking methodology we will use). We return to the question of processing cost after a discussion of the interpretive preferences for numerals.

The structure of the paper is as follows. We first discuss, in Section 2 and 3, some competing approaches on the meaning of numerals. Then, in Sections 4 and 5, we present two experiments investigating the offline and online behaviour, respectively, of subjects while they are presented with sentences containing numerals. Finally, in Section 6 we discuss how our findings affect the theoretical debate reviewed in Sections 2 and 3.

2. An overview of theoretical approaches to scalar implicatures

Claims similar to (4) have been put forth in the literature, especially in the context of work on Scalar Implicatures (SIs; Grice 1967). Consider the variation in interpretation of items like *some*, displayed in (5a,b):

- (5) a. The professor saw some of his students and he'll go out for dinner with them.
 - b. If the professor saw some of his students, he'll go out for dinner with them.
 - c. The quantifier scale: (some, many, most, every)

Sentences (5a) and (5b) differ minimally in form and meaning. Yet, in sentence (5a) *some* seems to mean 'some though not all' (upper bounded interpretation), while clearly that isn't how *some* is interpreted in (5b) (lower bounded interpretation). According to Grice and the Neogricean literature (e.g. Levinson 2000; Horn 1972, 2007), this phenomenon is discussed in terms of SIs. Sentences involving *some* are often considered against the background of alternatives constituted by other quantifiers that might be relevant, such as those in (5c). Use of *some* prompts the hearer to assume that the stronger alternatives do not hold (on the basis of simple assumptions on how conversation proceeds – the Gricean maxims). The same phenomenon is argued to affect the interpretations of other scalar items like connectives (*or, and*), modals (*may, must*), gradable adjectives (like *warm, hot*), etc. Within the Neogricean tradition, the 'some and possibly all' meaning of *some* is taken to be basic, and the denial that a stronger interpretation of *some* holds (which results in the 'some but not all' reading) is analyzed as a scalar implicature (SI).

An alternative development of Grice's insights on these matters, namely Relevance Theory (Sperber & Wilson 1985, 1995), maintains that interpretive alternations such as those in (5a,b) arise in a different manner. The basic meaning of *some* is lower bounded ('at least some'), and inferences that specify such a meaning towards the upper bounded construals come about only if it is required to maximize relevance. 'Relevance' is a property of the situation and the stimulus (a sentence, in this case). To arrive at the intended meaning, the hearer puts his or her whole cognitive domain (perceptions, world knowledge, intentions, emotions etc.) in relation with the sentence. If the hearer's expected relevance is satisfied with *some* being interpreted with its basic (i.e. lower bounded) meaning, additional effort will not be invested in drawing an inference towards the upper bounded meaning. This approach seems designed to predict the interpretations of cases like (6) vs. (7).

- (6) Mary, who is a big eater, will surely eat some of that cake.
- (7) Mary, who is scared of dentists, will surely have to spend some of her time tomorrow at the dentist.

Some in (6) might have greater chances to be interpreted in a lower bounded way than in (7) given the context and the nature of the task. We will discuss the implications of this approach with respect to the meaning of numerals in the next pages.

More recently, Chierchia (2004) and Chierchia, Fox & Spector (2009) have argued for the view that scalar implicatures are really part of core grammar. According to them, scalar implicatures arise through mechanisms analogous to those that drive association with focus (cf. Rooth 1985). Building more closely on the insights of the Neogriceans, Chierchia et al. (2009) propose to represent strengthened (upper-bounded, scalar implicature) meanings grammatically by adding a silent operator O, with an effect similar to that of focus sensitive operators like *only*. Consider (8).

- (8) a. John graded some of the homework.
 - b. John graded only some of the homework.

Sentence (8a) may or may not receive an upper bounded interpretation; sentence (8b) must be interpreted in an upper bounded way. The word *only* can be said to overtly 'exhaustify' the proposition, making it incompatible with any stronger interpretation of the sentence (see Chierchia et al. 2009, for extended discussion).

Exhaustification can be covert as well as overt. Consider (9):

- (9) a. i. Speaker A: So did you see the students?ii. Speaker B: Well, I saw MARY
 - b. O [I saw Mary] where O = only

Speaker B's utterance, in which *Mary* is focused, is clearly interpreted as equivalent to 'I saw only Mary'. This might be understood by assuming that there is a covert counterpart of only that is used in interpreting (9a.ii), as schematically indicated in the Logical Form (9b). Clearly, the kind of covert exhaustification that must be at play in cases like (9) might well be responsible for the upper bounded interpretation of sentences like (8a). According to this view, the lower bounded interpretation would correspond to the (unexhaustified) Logical Form in (10a); the upper bounded one would correspond to (10b).

a. some homework_x[John graded x]
b. O[some homework_x[John graded x]]

The Logical Form in (10b) would be interpreted just like (8b), giving rise to the upper bounded construal.

One prediction that sets aside the Grammatical View from Neogriceans or relevant theorists concerns the possibility of embedded implicatures. On the Grammatical View, where implicatures arise through a covert counterpart of *only*, one expects there to be embedded implicatures, for such an operator will be able to occur in embedded positions (just like its overt counterpart). On the other hand, in approaches where implicatures arise through a 'global', post-compositional process, involving speakers' intentions and the like, it is hard to see how implicatures could ever enrich the meaning of embedded constituents (thereby affecting the compositional part of semantics). This problem (acknowledged in, e.g. Levinson 2000) along with the different predictions of the two families of approaches is discussed at length in Chierchia et al. (2009).

3. Numerals and scalar implicatures

Much of the analysis of SIs has involved quantifiers such as *some*. Whether the analysis applies to numerals, which are clearly scalar terms, is controversial. For example, Horn (1992, 2004) argues that quantifiers and numerals behave differently under negation. Moreover, it has also been argued that numerals can also get an 'at most' reading, which is not available with other scalar quantifiers and could not be derived as a SI. Carston (1998) and Musolino (2004) present examples of the following sort in favor of the existence of an 'at most' reading.

(11) If you miss two shots, you will still win

We believe this claim to be factually wrong. The appearance of an 'at most' construal for the numeral in (11) has, we think, a different account. Sentences like
(11) have an 'exactly' interpretation (something we independently know to be possible), *plus* a presupposition (in the case at hand, triggered by 'still') that can be paraphrased as.

(12) If you miss no or exactly one shot, you will win.

Sentences (11) and (12) together are equivalent to 'at most'. Be that as it may, the main point we want to make in this paper is orthogonal to the (non) existence of 'at most' readings for numerals, and concerns the distribution of the 'at least' and the 'exactly' reading.

Huang & Snedeker (2009) conducted an experimental study that speaks to this issue. In this study eye movements of subjects were recorded while they attended to a visual scenario depicted on a monitor and listened to some sentences (visual world paradigm). These sentences contained numerals like *two* and *three* and scalar quantifiers like *some* and *all*. Subjects had to pick the character, present in the scenario, who was described by the sentence they heard. The authors found a delay in looking toward the referent of a phrase containing *some* (e.g. "point to the boy with some of the socks", where the scenario dysplayed a boy with two socks, a boy with three soccer balls, a girl with two socks and another girl with no objects) compared to unambiguous phrases containing *all* and phrases containing the numerals *two* or *three* (where *three* was unambiguous but *two* could arguably have a 'two or more' meaning). They argued that resolving *some* as 'some but not all' incurred into higher processing costs compared to unambiguous sentences and sentences with possibly ambiguous numerals.

One goal of our study was to determine whether the generalization in (4) holds for numerals as much as it seems to hold for quantifiers and other scalar items. For our purposes, we may briefly classify the positions on the semantics of numerals in four main families, schematically laid out in (13):

- (13) a. Lexicalist approaches. Numerals are lexically ambiguous between two construals. (Horn 1992)
 - b. Numbers are underspecified. Pragmatic enrichment is driven by relevance.

(Relevance Theory: Sperber & Wilson 1985, 1995; Carston 1998)

- c. Numbers are exact. The 'at least' reading comes through by a pragmatic operation of existential closure. (Breheny 2008)
- Scalar Implicatures. The upper bounded construal of numerals is a SI derived on the basis of their lower bounded construal.
 Approaches of this third type divide further into two streams:
 - i. Purely pragmatically based (Neogricean; cf. Kadmon 2001)
 - ii. Grammatically based
 - (Chierchia 2004; Chierchia, Fox & Spector, 2009)

Let us now comment briefly on each proposal in turn. Concerning position (13a), typical lexical ambiguities are not sensitive to the entailing properties of the context in which a lexical item is inserted. For example, the interpretation of the words like 'bank' or 'run' will typically remain constant across (14a) or (14b) even though they occur in a UE or a DE context, respectively.

- (14) a. i. If it isn't too crowded, you'll like that bank
 - ii. If you like that bank, you'll go there often
 - b. i. If you pray, the car will run
 - ii. If the car will run, we are in luck[run = functioning/vs. go fast, or partake in a competition]

Notice that this holds even if the relation among different meanings of a word do display subsets/superset relations, and hence entailment would be potentially relevant (if a car runs in a competition, it has to be running – in the sense of functioning – but not viceversa). So if (4) is right, lexicalist approaches would be faced with the task of explaining why the lexical ambiguity of numbers turns out to be sensitive to entailment patterns while other lexically ambiguous words are not (even when they potentially could).

In so far as underspecification analyses in (13b) are concerned, the point of contention is the recognition of a purely structural factor in the emergence of the lower bounded vs. upper bounded contrast in numeral interpretation. If (4) above is right, then the entailment characteristic of two minimally different local environments of the numeral would affect our interpretation, regardless of any relevant contextual factor like world knowledge, discourse context, speaker's intentions, etc. We do not see how the UE/DE context could be claimed to affect relevance (apart from building into relevance a condition equivalent to (4)).

Breheny (2008) may be viewed as a representative of theories of type (13c). It might be useful to sketch here a simplified approach loosely inspired to his ideas (as Breheny's own proposal cannot be summarized within the bounds of the present work). One might take (15b) as the basic interpretation of (15a):

- (15) a. John loves two cats
 - b. $|\operatorname{cat}_{D} \cap \{x: \operatorname{John} \operatorname{loves} x\}| = 2$

Let us explain (15b) in more detail. Formula (15b) says that the intersection of the set of cats in a domain D with the set of things that John loves has cardinality two. Such a formula is true if and only if John loves exactly two cats in D. It therefore appears to be an accurate rendering of the upper bounded construal of sentence (15a). D in (15b) is a variable over the domain of discourse. This is meant to represent the fact that we may utter (15b) having a specific domain in mind (e.g. the cats that live in John's neighbourhood). Such a domain is usually contextually specified. However, conceivably, some contexts, for a variety of reasons, may drive a process of existential closure of such domain variable, which would result in something like:

```
(16) \exists D \mid cat_D \cap \{x: John loves x\} \mid = 2
```

This formula says that there is a way of picking a domain D which would make (15a) true. For this to be so, it has to be the case that for some domain D, John loves exactly two cats in D (which is of course compatible with John loving more than two cats). Formula (16) has, in other words, the truth conditions characteristic of the lower bounded reading. This illustrates one way of adopting as basic the 'exact' reading of numerals and deriving through a semantic/pragmatic process the 'at least' one.

Our claim (4) could be compatible with theories of the type (13c), *if* one is willing to add to them the condition that an operation like (16) is performed preferentially in DE contexts than UE ones. This may be difficult to reconcile with the view that (16) is a purely *pragmatic* inference, for it is not clear why (or how) a pragmatic tendency should be sensitive to whether numerals are embedded in UE vs. DE contexts. If, on the other hand, the existential closure of domain is an option provided by grammar (much like different scope options), it would be quite natural to maintain that such an operation might be subject to a processing constraint that links its preferential occurrence to contexts in which this operation leads to stronger (i.e. more informative) statements.¹

Finally, recall that the main assumption of SI-approaches such as (13d) is that the lower bounded interpretation constitutes in some sense the core meaning of sentences with numerals, with the upper bounded one derived as an implicature.² Generalization in (4) is a priori compatible with such an assumption. However, it also carries some implications that seem to favour the grammatical approach (13d.ii) over the purely pragmatic one (13d.i). Consider the natural explicit paraphrases of the upper vs. lower bounded readings of sentences in (17) and (18), with the upper bounded ('exactly') meanings obtained by adding an operator (O) either in the consequent (17b) or in the antecedent (18b) of a conditional clause.

^{1.} However, on such an approach one would have to explore whether existential closure of domains is possible with other quantifiers, and if not, why not.

^{2.} There is an important caveat to be made in this connection. Saying that the basic meaning of a *sentence* containing a numeral is the lower bounded interpretation doesn't entail that the *lexical* meaning of the numeral itself is the lower bounded one. It is perfectly conceivable that the lexical meaning of a numeral is the exact/upper bounded one and that the lower bounded interpretation arises as part of the compositional semantics of sentences (see e.g. Landman 2003, for an approach along these lines).

- (17) a. If I get in the summer competition, I'll buy (at least) four golf clubsb. If I get in the summer competition, O[I'll buy four golf clubs]
 - = If I get in the summer competition, I'll buy exactly four golf clubs
- (18) a. If I buy (at least) four golf clubs, I'll never use my loose old clubs again
 b. If O[I buy four golf clubs], I'll never use my loose old clubs again
 - = If I buy exactly four golf clubs, I'll never use my loose old clubs again

On a SI approach, (17a) and (18a) would correspond to the core ('at least') meaning, while (17b) and (18b) would arise as implicatures (when warranted). In other words (17a) would be the core meaning of (1a) and (17b) would be the implicature enriched reading of (1a), and similarly for (18a,b) with respect to (1b). A possible account of claim (4), consistent with the SI approach, is to appeal to a principle of the following sort:

(19) Optimize Informativeness: Preferably, embed an implicature in contexts where it leads to a stronger statement than its alternative without the implicature.

The nature of Optimize Informativeness should be fairly clear: sentence (17b) (i.e. the upper bounded reading) entails sentence (17a) (i.e. the lower bounded reading). Namely, (17b) is logically stronger (and hence more informative) than (17a). On the other hand, sentence (18a) (the lower bounded reading) logically entails (18b) (the upper bounded reading). Namely (18a) is logically stronger (and hence more informative) than (18b). In other words, embedding an implicature in an UE context leads to strengthening (with respect to the statement without the implicature), while embedding an implicature in a DE context leads to weakening. This is a general property of DE environments: they reverse the entailment patterns of UE environments. So the point is that if we embed an implicature, we prefer to do it when this leads to strengthening with respect to the sentence without the implicature, as exemplified in (20) (where the arrow stands for the entailment, i.e. strength, relation).

(20) UE: 'at least four' \leftarrow 'exactly four' DE: 'at least four' \rightarrow 'exactly four'

While (19) is a natural, if sophisticated, principle, it is not obviously compatible with the standard Neogricean views, if for no other reason, that such an approach cannot motivate smoothly the very existence of embedded implicatures.

In conclusions, both approaches of the type in (13c) and (13d) are consistent with our proposed generalization (4), if the process that drives the derived interpretation of numerals is subject to a principle sensitive to logical strength such as (19).

It might be worth underscoring that the generalization in (4) says that (17b) as interpretation of (1a) is preferred to (18b) as interpretation of (1b). This specifically

means that (17b) should occur more often as an interpretation of (1a) than (18b) as an interpretation of (1b). What (4) says *nothing* about is whether the addition of an implicature to a core meaning is per se a costly process. We address this issue in discussing Experiment 2, which investigated the on-line processing costs of sentences with numerals.

Summing up, our goal is to investigate whether claim (4) is true or not. If it is, this will have rich consequences for our understanding of how numerals are interpreted, and a host of related issues at the interface between grammar, pragmatics and processing. It would be moreover quite striking to find out that we spontaneously submit to a rather abstract logical regularity like the one (4) relies on.

4. Experiment 1: An off line semantic judgment test

In this experiment, we explicitly asked 48 undergraduate students to indicate their interpretation of the numeral determiner. They had to choose between the stronger *exactly* meaning and the weaker *at least* one by checking the appropriate box, after reading a sentence of the kinds displayed in (21) and (22).

- (21) a. Giovanni ha due macchine in garage <u>e</u> parcheggia una motocicletta nel cortile esterno.
 John has two cars in the garage <u>and</u> he parks a motorcycle in the courtyard.
 b. <u>Se</u> Giovanni ha due macchine in garage, parcheggia una motocicletta
 - Se Giovanni ha due macchine in garage, parcheggia una motocicletta nel cortile esterno.
 <u>If</u> John has two cars in the garage he parks a motorcycle in the courtyard.
- (22) a. Nel mio quartiere ogni ragazza ha due fratelli più grandi e desidera una sorellina di età inferiore.
 In my neighborhood every girl has two older brothers <u>and</u> she wishes a younger sister.
 - Nel mio quartiere ogni ragazza <u>che</u> ha due fratelli più grandi desidera una sorellina di età inferiore.
 In my neighborhood every girl <u>who</u> has two older brothers wishes a younger sister.

As can be observed, the sentences in (21) and (22) differ minimally. Items (21a) and (21b) correspond to the *conditional type* and those in (22a) and (22b) are an example of *quantified type* items. In (21a) and (22a), the numeral is embedded in a UE environment (the second conjunct of a coordinated structure, and the scope of a universally quantified Noun Phrase, respectively); they therefore constitute the UE condition. In (21b) and (22b), on the other hand, the numeral is embedded in

a DE environment (the antecedent of a conditional and a relative clause adjoined to a universally quantified NP, respectively) and thus they constitute the DE condition. Let us note that the first argument of the universal quantifier shares the same semantic property (i.e. DEness) with the antecedent of conditional clauses, which we already discussed in (3). That is, sentence (23a) entails sentence (23b), with the latter considering a subset of the former (i.e. 'the set of guys who drink a *stout beer*' is included in 'the set of guys who drink a *beer*').

- (23) a. Every guy who drinks a beer will get an headache
 - b. Every guy who drinks a stout beer will get an headache

Our predictions are as follows. If numeral strengthening occurs more readily in UE context, people should select the 'exactly' interpretation significantly more often in phrases like (21a) and (22a) than in phrases like (21b) and (22b). Further, if this result is caused by the semantic property affecting numerals interpretation, rather than a specific grammatical construction, we should observe the same trend for both *conditional* and *quantified* sentences.

Participants were asked to make a choice between two alternatives by ticking the one preferred and turn over the page without altering their previous choice. The pivotal question was always posed in the following way: "we are talking about... exactly two cars/at least two cars". Participants were asked to carry out the task without lingering too much and to answer freely and naturally. It's worth underscoring once more that the material was almost the same across the experimental conditions. *UE* and *DE* items varied only from two words: the presence of "se" or "che" ("if" and "who" in English) in the latter condition versus the presence of "e" (the conjunction "and") in the former one.

The data we are focusing on is the percentage of strengthened choices, i.e. the proportion of "exactly N" answers over the totality of answers. The mean strengthened choices percentage (see Tab. 1) for the *conditional type* items was 78% in the *UE condition* vs. 49% in the *DE* condition. The mean for the *quantified type* items was 55% in the in the *UE condition* vs. 27% in the *DE* condition. The *polarity* factor (i.e. UE vs. DE) turned out to be significant, as well as the *type* of item (conditional vs. quantified), whereas no significant interaction was found.

Polarity	Senteno	се Туре	
	Conditional	Quantified	
UE	78%	55%	
DE	49%	27%	

Table 1. Means of "exactly N" choices

These results clearly show that the linguistic context, more specifically the polarity of the context embedding the numeral determiner, affects the participants' interpretation choices. Subjects selected significantly more often a strengthened, upper bounded reading in UE contexts with respect to DE ones. This confirms what was found for other scalar terms, like disjunction (Noveck 2001; Noveck, Chevaux, Guelminger, Sylvestre & Chierchia 2002; Chierchia, Crain, Guasti, Gualmini & Meroni 2001) and suggests that the strengthening of numerals is an analogous phenomenon. However, even though a strikingly similar pattern was found in the conditional vs. the quantified types, the impact of these two types was also a significant factor, as we can see by looking at the interpretation percentage difference between the UE and the DE condition (78% - 49% = 29% vs. 55% - 49% sc)27% = 28%). Accordingly, it must be acknowledged that there are contextual (e.g. the plausibility of the sentence) and structural (e.g. the grammatical construction of the sentence) factors other than polarity that can influence the participants' off-line interpretation of numerals. The influence of polarity is always combined with such factors, which is why the differential behaviour within the same items is important and telling. The crucial point, however, is that the *polarity* and the *type* of items affect participants' choice independently. That is, the entailment property of the context containing the numeral has the same influence on the task regardless of whether the numeral is in a *conditional* or *quantified* type sentence.

5. Experiment 2: An online processing experiment

In Experiment 2, we measured what happens in real time when readers are presented with numerals embedded in UE vs. DE contexts by recording their eye movements. We explored two possible effects. The first is the conventional expectation that if readers commit themselves to one interpretation of a numeral when they read the clause containing the numeral, then reading of the following clause will be disrupted if the interpretation of the numeral is inconsistent with it. The second possible effect is suggested by the proposal advanced in our discussion of the SI hypothesis, that the 'at least' interpretation is the core interpretation and the 'exactly' interpretation is created as a scalar implicature, which may take processing resources. The basic design is to have numerals in UE or DE contexts followed by continuations that do or do not force the upper bounded ('exactly') reading. Reading time on the phrase containing the numeral could reflect any processing cost of constructing an implicature, as explained below. Reading time on the continuation could reflect the cost of revising the initial interpretation of the numeral, and thus provide information about what the initial interpretation is in different contexts.

We prepared a set of two-clause discourses, each containing the first clause of one of the 24 items used in the offline experiment, followed by one of three second clauses. The first clause sets up either an upward or a downward entailing context for the numeral, just as in Experiment 1. The second clause could be one of three types: (a) neutral (the same sentences used in Experiment 1), without mention of the entity that was quantified in the first clause; (b) biased towards an upper bounded construal of the numeral in the first clause; or (c) a negative version of the biased continuation. Because of the role of negation, this third continuation canceled the upper bounded reading of the numeral in the first clause, making it functionally equivalent to the neutral reading.

- (24) First line
 - a. UE

Giovanni ha **due** macchine in garage <u>e</u> John has **two** cars in the garage <u>and</u>

- DE <u>Se</u> Giovanni ha due macchine in garage <u>If</u> John has two cars in the garage
- (25) Second line
 - a. *neutral continuation* parcheggia una **motocicletta** nel cortile esterno. he parks a **motorcycle** in the courtyard.
 - b. positive continuation parcheggia una terza macchina nel cortile esterno. he parks a third car in the courtyard.
 - negative continuation
 non parcheggia una terza macchina nel cortile esterno.
 he doesn't park a third car in the courtyard.
- (26) First line
 - a. UE

Nel mio quartiere ogni ragazza ha **due** fratelli più grandi <u>e</u> In my neighboorhood every girl has **two** older brothers <u>and</u>

b. UE

Nel mio quartiere ogni ragazza <u>che</u> ha **due** fratelli più grandi In my neighboorhood every girl <u>who</u> has **two** older brothers

(27) Second line

- *neutral continuation* desidera una **sorellina** di età inferiore. wishes a younger sister.
- b. *positive continuation* desidera un **terzo fratello** di età inferiore. wishes a younger **third brother**.

c. *negative continuation* non desidera un **terzo fratello** di età inferiore. doesn't wish a younger **third brother**.

The expected results may be divided into two categories. The first involves the firstpass reading times of the first line. The aim of the first-pass first line analysis is to look for any reading difference between the UE vs. DE contexts, regarding specifically the numeral region. If upper bounded ('exactly') readings are preferentially computed in UE contexts, and if they are enrichments of basic lower bounded readings, one might expect slower reading times for the numeral in the UE than in the DE context. The second category of predictions regards the first-pass indices computed on the second line after the ordinal (third car) is read, and the second-pass indices in all the regions. The key prediction is that only the positive continuation with the ordinal forces the upper bounded reading of the numeral in the first sentence. The positive continuation is incompatible with the lower bounded reading of the numeral whereas the neutral and the negative continuations are compatible with such reading. To see this consider the typical positive continuation of the DE conditional sentences, an example of which is repeated here in (28) for convenience.

(28) If John has two cars in the garage, he will park a third car in the courtyard.

If *two* in (28) is not (yet) upper bounded at the level of the antecedent of the conditional, its truth conditional import may be spelled out as follows:

(29) a. In any situation in which John has two or more cars in the garage, he will park a third car in the courtyard.

This cannot be true. For take any situation *s* in which John has three cars in the garage; under normal assumptions on ordinals (i.e. assuming that the ordering of cars matches the order of presentation in discourse), a third car will already be in the garage and hence cannot be parked elsewhere. Technically, we have a presupposition clash. The ordinal numeral *third* presupposes that its referent is the third in the most salient ordering available in the context. If John has three cars in the garage, such a presupposition could never be met. Hence this sentence is incoherent (and the same holds, mutatis mutandis of all other examples of this form). On the other hand, if *two* in (28) is upper bounded, the result is coherent, as the following paraphrase makes clear:

(29) b. In any situation in which John has exactly two cars in the garage, he parks a third in the courtyard.

So the continuation in (28) does force an upper bounded reading to be embedded in the antecedent of the conditional. Consider next what happens in the *DE* condition when we have the *neutral continuation*.

(30) If John has two cars in the garage, he will park a motorcycle in the courtyard.

Clearly, the continuation in (30) does not conflict with the *at least* interpretation of the numeral, as we may see by considering the following explicit paraphrase:

(31) In any situation in which John has two or more cars in the garage, he will park a motorcycle in the courtyard.

Trivially, the consequent of (30) can be true in any situation *s* in which John parks two or more cars in the garage. Therefore if the numeral gets an *at least* interpretation it will not need to be strengthened since the continuation in (30), unlike the one in (28), does not lead to a contradiction.

Let us now finally turn to a DE sentence followed by a *negative* continuation.

(32) If John has two cars in the garage, he won't park a third car in the courtyard.

In spite of differing minimally from (28), just by the presence of negation, this sentence is not contradictory under the lower bounded construal of the numeral. The following explicit paraphrase may make this claim clear:

(33) In any situation in which John has two or more cars in the garage, he doesn't park a third car in the courtyard.

54 native Italian speakers (between the age of 19 and 29 years old) took part to this experiment. They were told to read the sentences appearing on a screen silently and naturally, in order to answer simple comprehension questions which would follow randomly after reading some sentences. While they read each sentence their eye movements were recorded by an eyetracker.

The 24 experimental items were interspersed with 80 filler sentences and 44 simple comprehension questions. The two experimental factors (*polarity* and *type of continuation*) gave rise to six different conditions (24a+25a, 24a+25b, 24a+25c, 24b+25a, 24b+25b, 24b+25c are examples of the conditions for *conditional items*; 26a+27a, 26a+27b, 26a+27c, 26b+27a, 26b+27b, 26b+27c are examples of the conditions for *quantified items*). Notice that 12 items displayed a conditional clause in the *DE* conditions while the other 12 items contained a universal quantifier restriction in the same conditions, exactly like the first experiment. Thus, items can be divided into several regions of interest, and they were designed so that the regions of major interest (namely the first line numeral and the second line ordinal) were composed by only one word and were the same in every experimental condition (except for the *neutral* continuations where the ordinal was absent). In each item the first clause was ended by a line break, and the second clause (including the conjunction in the UE conditions and the negation in the negative conditions) appeared on the second line.

5.1 Results

The first-pass reading indices reveal whether participants spent more time in reading a specific region without having accessed to the following verbal material (i.e. the second clause of the sentence). In the first line analysis (24a and 26a vs. 24b and 26b) only the *polarity* factor was considered as the reading of the continuation did not affect the first-pass reading times of first clause nor did the verbal material in the first clause vary with respect to the type of continuation factor. The main point of interest, here, is to check whether the polarity of the context influenced the reading time of the numeral region (the word "two", in boldface in 24 and 26). We found a significant main effect of the *polarity* on this region in the conditioned regression-path duration. This index equals the gaze duration time (the mean of the sum of all fixation times starting with the reader's first fixation inside the region until the reader's gaze leaves the region either to the right or to the left) plus the time spent re-reading just the preceding word, which was the verb1 region. According to this index participants spent 14ms more in reading the numeral in the UE conditions than in the DE ones, with no significant impact of the type of construction (conditionals vs. quantified sentences) that we employed in this experiment (i.e. the factor type of item did not interact with the polarity). To ensure that this effect was not caused by a difference in the probability of skipping the *numeral* region we checked that the skipping rate did not vary significantly across DE and UE trials (40.8% and 37.6% of probability of skipping the numeral, respectively).

As for the first-pass reading times on the second line (25a-c and 27a-c), the core results involve the interaction between *polarity* and the *type of continuation* factors. In the *regression-path duration* (the mean of the sum of all fixation times starting with the reader's first fixation inside the region until the reader's gaze leaves the region to the right) computed on the last region (the last words of the sentence, e.g. "in the courtyard") we found significant interactions between *polarity* and the *type of continuation* factors. The interaction for the *neutral* vs. *negative continuation* comparison, in contrast, was not significant. This results show that participants spent more time in re-reading part of the sentence in the DE condition than the UE one, when they encountered the *positive continuation*. This motivates us to look carefully at the second-pass indices to investigate where the regressions coming from the second line were directed to.

The second-pass indices we will focus our attention on are *second-pass fixation number* (the count of every second pass fixation made on a word) and *secondpass fixation probability* (the probability that the reader made at least a second pass fixation on that word).³ In the second-pass indices we found that the numeral region, in the first line (24 and 26), displayed the very same pattern of the second line first-pass indices. That is, in the positive continuation, in all the second pass indices readers made more regressions towards the numeral in the DE conditions than in the UE ones, as displayed in Tab. 2 (second-pass fixation probability: DE: 26% vs. UE 33%; second-pass fixation number: DE: 0.34 vs. UE: 0.46). The neutral and negative continuation conditions, instead, both displayed the same pattern. Here participants behaved in the opposite way with respect to the positive continuation. That is, they made more regressive eye movements towards the numeral in the UE conditions than in the DE ones (neutral continuation. Second-pass fixation probability: DE: 31% vs. UE 23%; second-pass fixation number: DE: 0.4 vs. UE: 0.3; negative continuation. Second-pass fixation probability: DE: 33% vs. UE 25%; second-pass fixation number: DE: 0.46 vs. UE: 0.34). This pattern of results gave rise to significant interactions between the polarity and type of continuation factors in the postive vs. neutral continuation and positive vs. negative continuation comparison, but no interaction between those factors in the neutral vs. negative continuation comparison.

Index	Continuation						
	Polarity	Positive	Neutral	Negative			
Second Pass Fixation	UE	26%	31%	33%			
Probability	DE	33%	23%	25%			
Second Pass Fixation	UE	0.34	0.4	0.46			
Number	DE	0.46	0.3	0.34			
Conditioned Sec. Pass	UE	11%	17%	17%			
Fixation Probability	DE	16%	13%	9%			
Conditioned Sec. Pass	UE	0.14	0.18	0.22			
Fixation Number	DE	0.2	0.17	0.12			

Table 2.	second-pa	ass indices	for the	numeral	region
	over p	NOO III GILCOU	101 0110		10,1011

5.2 Discussion

The first important finding we have to underscore is that the phrases for which the first experiment participants preferred an upper bounded reading for the numeral in an UE context, exhibit an early processing penalty on the *numeral* region in the

^{3.} For a more exstensive analysis of the results of this experiment, see Panizza, Chierchia & Clifton (2009).

second experiment. This effect cannot merely be explained as a general influence of a specific grammatical construction since we tested two different environments (conditional and quantifiers) and the result remained stable across both constructions. It seems that the polarity of the context is a factor systematically exploited by a reader: if the embedding context is UE, the upper bounded meaning (on some approaches the SI) is computed (or at least considered) on line, as soon as possible. If the local embedding context is DE, the upper bounded meaning is not considered and if needed, a SI is computed only later while the reader fixates on other regions.

Consider next the effect of the second clause. Our materials were designed to force the *exact* numeral interpretation within the DE context of the *positive continuation*, so that the ordinal numeral was supposed to act as a trigger of a reanalysis if the *exactly* interpretation had not been computed. On the other hand, in the two control sentences this was predicted not to happen. The first control sentence, i.e. the *neutral continuation*, lacked the ordinal numeral, which triggers the need to recalculate the meaning of the numeral in the DE condition. The second, the *negative continuation*, differed by one word from our test sentences (namely, negation), and yet, for semantic reasons, it was not expected to force a reinterpretation of the numeral in the DE condition numeral in the second clause. Hence, only the *positive continuation* was expected to force an interpretation of the numeral in the direction of the upper bounded reading. In other words, we expected an interaction between the *polarity* of the first clause and the *type of continuation*.

Our findings are as follows. We found the expected interaction effects in the second-pass indices, resulting in significant interactions between *polarity* and *type of continuation*. The same interaction, along the same direction, was found in the *regression-path duration* computed on the last region, which includes all the regressions made by the reader after reading the whole sentence for the first time. According to these measures, the pattern surfacing from the difference between the *UE* and the *DE* condition of the *negative continuation* is strikingly similar to that coming from the difference between the same conditions of the *neutral continuation*, whereas the pattern in the *positive continuation* is diametrically opposed. Participants made more frequent second-pass fixations when the numeral was embedded under a DE context, in the *positive continuation*, while they behaved in the opposite way in both the *neutral* and the *negative continuations*.

Now, if we compare the first-pass results to those coming from the secondpass we see how in the *positive continuation*, and only there, the participants' reading pattern of the *numeral* in the first line was reversed. During the first-pass it was more difficult to read in the *UE* condition, while according to the secondpass it received more and more often fixations in the *DE* one. In contrast, the readers' behaviour was uniform, across the first and the second-pass, in both the *neutral* and the *negative* continuation. In the latter cases, the numeral always received more first- and second-pass processing in the *UE* condition. This global picture perfectly fits our main claim and its implications for processing. Moreover, to the extent that SIs triggered by other scalar items (like *some* or *or*) are subject to the same generalization, our results provide evidence in favour of the view that the alternation between the lower bounded vs. upper bounded construals of numerals may well be a scalar implicature. In an upward entailing environment a scalar operator is mostly strengthened locally whereas in a downward entailing environment it is typically strengthened only globally. Therefore, in the latter case, additional reading time on the numeral is observed only after the reading of the sentence.

Finally, both early and second-pass effects on the *numeral* region (as well as on the others in the *positive* vs. *negative continuation* comparison) were unaffected by the phrasal structures selected to create a downward entailing environment (*type of item* factor). This shows that the readers' behaviour with respect to the *numeral* region was influenced by the semantic diversity of the two environments (DE vs. UE), in interaction with the type of continuation, rather than other contingent factors like the specific words or the syntactic construction adopted in building in the sentences.

6. Conclusions

The results of these experiments show that structural factors (entailment properties of the local context) affect the interpretation of numerals. Furthermore they bring evidence that the lower bounded interpretation of numerals occurs preferentially in DE contexts (with respect to minimally different UE ones), while the upper bounded one occurs preferentially in UE contexts (with respect to minimally different DE ones).

We have addressed this task by investigating two types of functors, *every* and *if* (both DE in their restriction) and contrasting them with minimally different UE contexts (*and*, in the case of *if*; while in the case of *every*, we have simply displaced the numeral from the restriction to the scope). The choice of material (and the various controls we have run) makes it implausible that our results may stem from idiosyncratic features of the selected items or of the context. An off-line question-naire confirmed that readers interpret numerals in conformity with (4) most of the time. An on-line experiment based on the recording of eye movements seems to reveal a systematic processing penalty associated with contexts in which one forces readings that go against the generalization (4).

These findings have rich consequences. For one thing, they support something like 'Optimize Informativeness', i.e. the idea that the readings of numerals are unconsciously chosen in such a way as to avoid interpretations that lead to weakening with respect to available alternatives (unless forced to). This suggests that the parser somehow checks entailments in selecting a reading. Of great interest, for the future research, might be to investigate how such principle interacts with the contextual factors (i.e. the saliency of an entailing scale in a given context). A plausible hypothesis, stemming from this work, buys on the idea that the kind of computations performed while drawing a scalar implicature are encapsulated from extra-linguistics factors. This is not tantamount to saying that extra-linguistics factors do not affect the final outcome of scalar processes, but rather that the scalar computation itself is performed by a cognitive system relatively blind to contextual information. Where the context may be playing a crucial role is in determining whether the scalar alternatives are active or not. Although this 'blindness to context' hypothesis remains still rather speculative, the lack of interaction between the effects of polarity and type of constructions that we used in the experimental items could be taken as going in this direction.

A second important point that emerges is the following. Other scalar terms (e.g. quantifiers like *some*, connectives like *or*, etc.) have been argued to be subject to a similar constraint in their interpretation. This suggests that the variation in meaning of numerals is probably due to one and the same mechanism, presumably a scalar implicature, no matter how much numerals may otherwise differ in meaning and processing from, e.g. quantifiers like *some* (cf. on this Huang & Snedeker 2009). More experimental work is called in to test whether other scalar dimensions are affected at the same way by the experimental manipulation of the polarity of the embedding context.

The points made above are quite general. There are many other elements emerging from the present study that we think are relevant to ongoing theoretical debates on the nature of implicatures. One worth mentioning is the following. With numerals it is easy to see (and, as a matter of fact, quite uncontroversial) that both lower and upper bounded readings can occur in embedded positions. If the distribution of these readings follows a pattern similar to that of other scalar terms, and hence may be due to the same general mechanism responsible for SIs, we would have further confirmation of the existence of embedded implicatures, a frequently disputed claim (defended by Chierchia et al. 2009). Be that as it may, we think the present work provides strong evidence in favour of the fact that the polarity of the context where the numeral is embedded affects both the interpretation and the processing of numerals.

References

- Breheny, R., Katsos, N. & Williams, J. 2006. Are scalar implicatures generated by default? Cognition 100(3): 434–463.
- Breheny, R. 2008. A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of Semantics* 25(2): 93–139.
- Carston, R. 1998. Informativeness, relevance and scalar implicature. In *Relevance Theory: Applications and Implications* [Pragmatics & Beyond New Series 37], R. Carston & S. Uchida (eds), 179–236. Amsterdam: John Benjamins.
- Chierchia, G. 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In *Structures and Beyond*, A. Belletti (ed.), 39–103. Oxford: OUP.
- Chierchia, G., Crain, S., Guasti, M.T., Gualmini, A. & Meroni, L. 2001. The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In *Proceedings of the* 25th *Annual Boston University Conference on Language Development*, A.H.-J. Do, L. Dominguez & A. Johansen (eds), 157–168. Somerville MA: Cascadilla Press.
- Chierchia, G., Fox, D. & Spector, B. 2009. The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In *Handbook of Semantics*, P. Portner, C. Maienborn & K. von Heusinger (eds). Berlin: Mouton de Gruyter.
- Chierchia, G., Frazier, L. & Clifton, C. Jr. In press. When basic meanings are (not) enough: Processing scalar implicatures in adult comprehension. In *Issues in Pragmatics* [University of Massachusetts Occasional Publications in Linguistics 39], M. Biezma, C. Davis, and J. Harris, (eds).
- Grice, H.P. 1967. Logic and conversation. In Syntax and Semantics, Vol. 3: Speech Acts, P. Cole & J.L. Morgan (eds), 41–58. New York NY: Academic Press. (Reprinted in H.P. Grice 1989, 22–40).
- Grice, P. 1989. Studies in the Way of Words. Cambridge MA: Harvard University Press.
- Horn, L. 1972. On the Semantic Properties of the Logical Operators in English. Ph.D. dissertation, UCLA.
- Horn, L. 1992. The said and the unsaid. Proceedings of SALT 2, 163-191.
- Horn, L. 2004. Implicature. In *The Handbook of Pragmatics*, L.R. Horn & G. Ward (eds), 3–28. Malden, MA: Blackwell.
- Horn, L. 2007. Issues in the investigation of implicature. To appear in *Volume in Honor of Grice*, K. Petrus (ed.).
- Huang, Y. & Snedeker, J. 2009. On-line interpretation of scalar quantifiers: Insight into the semantic-pragmatics interface. *Cognitive Psychology* 58(3): 376–415.
- Kadmon, N. 2001. Formal Pragmatics. Cambridge MA: Blackwell.
- Landman, F. 2003. Predicate-argument mismatches and the adjectival theory of indefinites. In *The Syntax and Semantics of Noun Phrases* [Linguistik Aktuell/Linguistics Today 55], M. Coene & Y. D'hulst (eds), 211–237. Amsterdam: John Benjamins.
- Levinson, S. 2000. Presumptive Meanings. Cambridge MA: The MIT Press.
- Musolino, J. 2004. The semantics and acquisition of number words: Integrating linguistic and developmental perspectives. *Cognition* 93(1): 1–41.
- Noveck, I.A. 2001. When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition* 78(2): 165–188.
- Noveck, I.A., Chierchia, G., Chevaux, F., Guelminger, R. & Sylvestre, E. 2002. Linguisticpragmatic factors in interpreting disjunctions. *Thinking and Reasoning* 8(4): 297–326.

- Panizza, D., Chierchia, G. & Clifton, C. 2009. On the role of entailment patterns and scalar implicatures in the processing of numerals. *Journal of Memory and Language*.
- Rayner, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124: 372–422.
- Rooth, M. 1985. Association with Focus. Ph.D. dissertation, University of Massachusetts at Amherst.
- Sperber, D. & Wilson, D. 1985/1995. *Relevance: Communication and Cognition*. Oxford: Basil Blackwell.

Meaning in the objects*

Katharina J. Rohlfing University of Bielefeld

In this paper, I am arguing that objects being present in the external situation ground the linguistic meaning. Furthermore, I will show that the nature of objects can change not only linguistic but also gestural behavior. Instead of simply excluding materialistic factors, I therefore suggest a careful inclusion of object knowledge into experimental conditions. I also argue that we have to calculate the risk of eliminating important components for children's reasoning, when we adapt this method to studies on children's language development. Children are good learners because they are biased towards certain solutions (Dabrowska 2005). For this reason, it seems to be problematic to create novel or abstract situations in which children cannot draw from their nonlinguistic experiences.

1. Introduction

Traditionally, a distinction has been made between a symbol on the one side and an object that this symbol refers to on the other side. Within a semiotic triangle, the interpreter makes the effort to connect the symbol with the object. In this view, a symbol is seen as an autonomous *signifier* as it can be flexibly applied to changing situations. An object, in contrast, is viewed as a part of this changing situation. It is the *signified* within the reference process. Some psycholinguistic evidence, however, provides doubts on the strict distinction. Accordingly, not only deictic utterances draw their meanings from the environment. The reference process in general does not seem to be a matter of mental states only. For the distinction between pragmatics and semantics, it is of question here whether the environment (encompassing objects and people setting up an object's normative values)

^{*} The pilot data reported here was partially collected during my postdoctoral work with Soonja Choi at San Diego State University: I am very grateful for this collaboration. The final version benefited immensely from the comments of an anonymous reviewer and Jörg Meibauer. The research reported in this paper was partially supported by the Dilthey Fellowship (VolkswagenStiftung) and by the European Community under the Innovation and Communication Technologies programme of the 7th Framework for ITALK-project (ICT-214668).

contributes to the meaning or is (just) a part of it. It is important to note that it is not the goal here to claim that meaning cannot emerge in the absence of materiality. Instead, it is the goal to "foreground [...] the semiotic aspect of materiality, and the material aspects of meaning" (Sinha & Rodriguez 2008: 365). As I will show below, there are findings suggesting that a symbol not only can refer to an external situation but can create objects in the sense that it determines their perception (Feist & Gentner 1998). Here, the signifier can influence the signified. The converse influence of an object onto the symbol becomes apparent especially in light of studies on language development. The signified (the object) influences the signifier (the linguistic symbol) as an object can display semiotic character and affords how symbols are understood in a specific situation.

1.1 A symbol influences object perception

A tight coupling between the meaning of a word and the perception of the referee can be seen in the semantics of spatial terms. It is not sufficient to only show the influence of a symbol on the perception of an object to exemplify that object properties correspond to the semantics of spatial terms (as it has been shown by e.g. Hottenroth 1993). It has rather to be shown that "one and the same real situation may be associated with different mental representations, and hence it may be assigned different semantic values" (Zelinsky-Wibbelt 1993: 3). A bias for the perception of objects caused by a linguistic symbol was reported in a study with adults by Feist & Gentner (1998). The design of this study on understanding spatial prepositions considered the constellation between the Figure (trajector object) and the Ground (landmark object). In order to investigate the influence of conceptual/functional information, Feist & Gentner (1998) used different nouns to refer to the Ground. In one condition, the inanimate Ground was called a *dish*, in the second a *plate*, and in the third a *bowl*. The participants' task was to circle IN or ON on their answer sheets, which contained sentences of the form The Figure is IN/ON the Ground. The results of this study suggest that the use of spatial prepositions in English is influenced by materialistic factors. Participants circled much more often IN when they were told that the Ground is a bowl in contrast to the condition, in which they were told that the Ground is a plate.

Another study reported by Malt, Sloman & Gennari (2003) studied participants' categorization of motion events and show that language can determine the perception (more specifically the perceptual categorization) of events in conditions in which linguistic judgments are requested. In this study, the subjects were coming from different linguistic background. Spanish and English. The goal was to investigate whether speakers of English encode manner of motion (e.g. sneak, stroll) while Spanish speakers encode the path in the motion (e.g. [enter], salir [exit]). The participants were shown films. The target film showed a motion event (e.g. somebody walking into a room) and two alternatives showing variations in either the manner (e.g. somebody was sneaking towards a door) or the path (e.g. somebody walking out of the room). After viewing the films, participants were asked to do a recognition memory task and a similarity judgment task. In the recognition task, participants have to decide as quickly as they can whether they have seen a particular clip already. In the similarity judgment task, the participants decided which clips were most similar to the target. In a naming first condition, participants described the films while watching them.

In the recognition memory task, no effect of language was found. In the similarity task, however, an effect of language was found. But this effect applied only to conditions in which the similarity task took place after verbal encoding. This suggests that when encoding of actions was non-linguistic, speakers of both languages display a similar performance. However, when participants named the films first, they were more likely to see similarities according to their linguistic categories in their language. Malt and her colleagues (2003) conclude that language seems to affect the perception under specific conditions in which linguistic judgment are requested.

These studies provide evidence against a dichotomy between a word as a signifier on the one hand and the object as signified on the other hand. Instead the results suggest that linguistic performance is grounded in materiality and depends on the provided context, the task and perceived scene (Sinha & Rodriguez 2008; Zelinsky-Wibbelt 1993; Langacker 1987). From this perspective, semantics and pragmatics are inseparable (Langacker 1987). The distinction between pragmatics and semantics might be a solution for specific situations, where a linguistic decision is required (as in Malt et al. 2003), but not a default or, as will be shown below, the starting point during learning.

1.2 Semiotic character of objects

Gibson & Pick (2000) suggested that the world is relevant to an organism when it affords action. An object, thus, will be relevant when an organism can engage with its properties in behavior (Sinha 2005). An affordance of an object can be its graspability. However, when we look at the objects surrounding an infant, we can be sure that they are largely constructed for specific social purposes. A button on a TV affords the action of pushing it because it has been constructed for this purpose to switch the TV on or off. Even natural objects such as stones and sticks are often presented to the child as tools in terms of their properties that are relevant for achieving a specific goal (Rakoczy, Tomasello, & Striano 2005). Although children may discover the properties through their own exploration, it is necessary to see affordances in light of social materialist sense (Sinha 2005).

Csibra & Gergely (2006) argue that unlike infants of most other animals, human infants are fascinated by objects and enjoy manipulating them. Even though chimpanzees use objects as goal-directed tools, and can even modulate the object's properties to achieve the goal, once the goal is reached, they tend to discard it. In contrast, human infants are educated to handle objects and to view them as having permanent functions. Thus, as a result of this social learning, infants will expect objects to have functions. This linkage of object perception and object manipulation is a neurophysiological finding and applies to canonical neurons within the mirror neuron system (Rizzolatti et al. 2001). In the mirror neuron system, the same neurons are activated by both, when a goal-directed action is observed as well as executed. While mirror neurons fire by execution and observation of a goal-directed action, canonical neurons (Rizzolatti & Arbib 1998; Grèzes et al. 2003) are activated when an object is in focus of attention; they discharge by both, when the object is manipulated and when it is solely perceived as if the activity with this object would be foreseen.

The knowledge of the object manipulation is endogenous and includes prior experiences and expectations. The child will bring it to a situation. Here, recent neuroimaging studies show that motor knowledge is a part of it. More specifically, Pierno, Becchio, Wall, Smith, Turella & Castiello (2006) have shown that when the presence of an object is signaled through gaze, a similar neural response in an observer is elicited to an observation of a reach-to-grasp action on the same object. This finding suggests that referring to an object we can relay on a manipulative experience the persons have with the object.

How object knowledge in the form of affordances influences even linguistic behavior has been shown in Clark's (1973) investigations of the spatial language development process. When instructing two years old children to relate an object to another, she observed a bias. Accordingly, if the object presented included a surface, further objects were assigned to the trajector role and placed ON the initial landmark. If the object presented was a container, the infants put the other object IN this landmark. This behavior dominated language understanding to such an extent that infants typically placed something ON a table, even if they have been requested to put the it UNDER the table. Clark (1973), therefore, proposed that infants' responses to linguistic instructions are guided by the *physical properties* of objects. This bias towards perceptual features of objects establishes a basis for the "first linguistic hypotheses" (Clark 1973: 180) in the form of non-linguistic strategies. Even though in Clark's approach, it is difficult to understand how children go beyond the stage, where they relay on non-linguistic strategies to a stage where they rely on full semantic knowledge of the words, the findings strongly suggest that perceptual knowledge is involved in the process of building up linguistic capabilities. This argument has been repeated in other developmental research (e.g. Freeman, Lloyd & Sinha 1980; Rohlfing 2001), in which it has been shown that linguistic performance of children is grounded in materiality and external situations. Based on this research, the materiality of objects that influences perception and even linguistic behavior can be operationalized as (1) familiarity and (2) canonicality.

Familiarity. There are several studies suggesting that familiarity of objects affects perception, conceptualization and linguistic behavior. As to the perceptional processes, Flom & Pick (2003) tested 60 children and showed that the effect of the familiarity of objects influencing infants' perception of parental pointing behavior. When mothers point to novel objects in contrast to familiar objects, 18-month-olds follow their point more reliably. Thus, the novelty of objects seems to affect children's attention.

An effect of familiarity on conceptualization in infants has been shown in studies by Casasola & Cohen (2002) with 15 infants. In this study, six-month-olds were first habituated to some examples of a spatial relationship, either containment (one object going into the other), tight-fit (one object fitting into or onto the other) or support (one object supported by the other). Following habituation phase, the participating infants were tested with: (a) one of the familiar events seen during habituation, (b) with familiar objects in a novel relationship, (c) novel objects in the familiar habituation relationship and (d) novel objects in a novel relationship. The results show that infants who were habituated to the containment relationship discriminated reliably between different relationships regardless of the nature of objects. However, infants habituated to support and tight-fit responded only to changes in the objects and not the relationship. This suggests that infants earn to categorize a relationship between familiar objects prior to novel objects (Casasola & Cohen 2002).

The proposition that the effect of familiarity is involved in learning processes has been tested by Rohlfing (2006) for word learning. In this study, following a pretest about children's understanding of *under*, 42 children at the age of 20 to 26 months were trained to learn this preposition. In the test phase, three conditions were designed in which children were instructed to put an object under the other: (a) a familiar condition, in which the same objects as in training were used (b) a transfer condition, in which new toy objects were used and (c) a neutral condition, in which abstract objects in form of a sphere located in a wooden construction without functional properties were presented. Two year olds children showed very good learning effects with familiar objects, while only moderate learning effect could be observed in novel objects. No learning effect was observed in the neutral condition. This suggests that children need to know about what to do with objects in order to build on this information. This point holding that object knowledge is involved in object perception is less remarked in the language learning theories. Yet, it is a known phenomenon in theories about perceptual development and even linguistic research in adults. In a study with adults, for example, Coventry and his colleagues (2001) tested the understanding of prepositions like *over*, *under*, *above* and *below* by asking the participants to evaluate the appropriateness of a spatial description. They found that the participants' judgment is influenced by object knowledge. Only in cases when there are no objects properties of relevance, seem the participants to rely on more abstract and objective evaluation method based on geometric features.

Canonicality. Another form of object knowledge is canonicality. Similarly to the familiarity effect sketched above, it affects handling of a single object and the way the object is held for specific purposes (Freeman, Lloyd & Sinha 1980). A pot, for example, must be oriented in an appropriate way to fulfill its function as a container. Young infants need to learn this proper orientation. They start often with associating two objects together like a pen with a paper, which they simply put closely together and find out next that the pen needs to be held at a specific angle with its ink top toward the paper in order to write. Seeing a particular orientation (e.g. umbrella in an upright position), a child is already provided with some information about the role of the object (Rohlfing et al. 2003). This personal experience with objects adds a social perspective to the notion of affordances (Sinha 2005). Thus, even though people eventually come to an agreement on objective properties of objects (by the power of collective intentionality as is suggested in Rakoczy et al. 2005) so that they are accessible to everybody in a similar way, the crucial point here is that what objects afford is not strictly objective. Objects are not perceived by all humans in the same way. Instead, what objects afford (1) depends from the intrapersonal experience that a person have gained (Rohlfing et al. 2003) and (2) is culturally mediated (Freeman, Lloyd & Sinha 1980; Sinha 1982; Jensen de López 2006). When a child perceives an object, the culture-dependent variable is already playing its role, because the object is presented in a particular way within a particular social group. Nelson (1996:96) emphasizes: "Event knowledge is social knowledge and social knowledge is event knowledge". This argument opposes the idea that the basis for perceiving a situation is constituted by objects as given entities.

The effect of canonicality was experimentally investigated by Freeman, Lloyd & Sinha (1980). The authors observed 38 infants at the age of 12 to 15 months in a search task, in which two cups on a tray were presented. The child was supposed to search for a toy hidden in one of the cups, which was moved round behind the distractor cup in order to transpose them. The cups were either both upright or both inverted. The authors found that when cups were used in their customary orientation, the children's performance was better. This finding suggests

that children have a concept of an object in its "customary orientation" (Freeman, Lloyd & Sinha 1980: 259).

Canonicality affects also the relationship between two objects. For example, a canonical relationship is established when two objects are put together for the most common function (Nelson & Ware 2002) like a lid and a pot, in which case the lid goes ON the pot. These conventions are derived from cultural values and determine how to handle objects/artifact (Sinha 1983). Handling and manipulating of objects is inseparable from object knowledge (Sinha calls it "background knowledge", 1983: 269). For the operationalization of object knowledge in form of canonicality, it is important to differentiate between the canonical function of an object and canonical relationships of objects. A canonical function is linked to the role of the object. A table has the canonical function of supporting things; when a child puts a toy horse on a table, the canonical function of the table is fulfilled. It is not the case for a canonical relationship. For canonical relationship, both objects should be involved in a conventional way like, for example, a pot on a table.

How the background knowledge is communicated to children was of issue in a study by Choi & Rohlfing (2008). In this cross-linguistic study, 16 English and 16 Korean speaking mothers were asked to instruct their two years old children to do either a canonical or noncanonical relationship between two objects. For the analysis, the discourse of the mothers was transcribed and analyzed according to whether it follows *bring-in* or *follow-in* strategies. Using the *follow-in* strategy (Tomasello & Farrar 1986) mothers engage in joint attention and follow up on the child's action. Accordingly, mothers say something like "no, put the horse under, not up"). Applying the *bring-in* strategy (Rohlfing & Choi 2004) mothers say, for example, "let's have some tea!" and bring in background knowledge about tea drinking or evoke familiar situations that relate to the requested situation. The analysis revealed a striking difference in the use of *bring-in* and *follow-in* strategies as a function of canonicality. In both Korean and English, there was significantly more *bring-ins* than *follow-ins* for the canonical relationship, and the reverse was true for the non-canonical relationship.

In sum, the studies presented above show that objects function not only as signified but also as signifiers as they are semiotic resources. If this conclusion is true, the influence of materialistic dimension should also be seen for gestural behavior, because it often accompanies speech serving communicative function (Goldin-Meadow 1999). However, so far, little is known about how gestures are impacted by the nature of objects.

Even though gestural behavior is known as a crucial contributing factor to language performance (e.g. Kita & Özyürek 2003), language understanding (Goldin-Meadow, 1999; Goldin-Meadow & Sandhofer 1999) and language development (Bates, Thal, Whitesell, Fenson & Oakes 1989; Iverson & Thal 1998), relatively little is known about whether and how parental input, and in particular the semiotic information about objects, is conveyed to children's nonverbally. In a study with 17 preschool (mean age was 51 months) and 15 kindergarten (mean age 64 months) children, McNeil, Alibali & Evans (2000) showed that the effect of gestures on children's comprehension of spoken language is dynamic and manifold. Gestures guide children "toward the semantic content of the spoken message" (McNeil, Alibali & Evans 2000: 133). They either reinforce the verbal message as they convey the same semantic content or add to the verbal message providing additional semantic information (Goldin-Meadow 1999). Whether reinforced or supplemented, the additional information in gesture has been investigated to support word learning; not only does the use of gesture facilitate fast mapping processes but it also helps the child retain the relevant semantic information (Capone & McGregor 2006).

In talking to young children, especially the deictic gesture was encountered in the literature. Iverson and her colleagues (1999) investigated the synchronization of verbal semantics and pointing of Italian mothers when they were engaged in an interaction with objects varying in their degree of familiarity. They found that mostly reinforcing information is provided in deictic gesture towards children at the age of 16 and 20 months. Özçalişkan and Goldin-Meadow (2005) reported the same finding for 22 month olds. It seems that maternal nonverbal behavior does not change as a function of children's age. However, Iverson and her colleagues (1999) found that the production of pointing correlated positively with children's vocabulary size suggesting that children whose mothers pointed more were advanced in building their vocabulary. This correlation was significant in 16 month olds but no more significant in 20 month olds.

For our pilot study, it is hypothesized that in the given task, mothers instructing for a spatial relation may become aware about their child's knowledge of objects as they get immediate feedback from the child about their understanding of the instruction. The aim was to explore whether and how gestures will change as a matter of objects' canonicality.

2. Method

2.1 Subjects

17 American English-speaking and 17 German-speaking mother-child pairs participated in this study. Among the English-learners, 8 were boys and 9 girls. The age of the children varied from 20 to 26 months (22 months and 11 days on average). Participants were selected from an existing subject pool of the Child

Language Laboratory at San Diego State University. All were being raised in a monolingual American English-speaking environment in and around San Diego. Among the German-learners, 7 were boys and 10 girls. The age of children varied from 20 to 26 months (22 months and 27 days on average). The age of children in our samples spanned over 6 months, because starting at around 20 months of age, children are able to engage in such task-oriented dyads. After the age of 26 months, children's performance in understanding the prepositions ON and UNDER becomes closer to ceiling (Rohlfing 2005). Participants were selected from a subject pool of interested parents answering a call for study participation posed in a local newspaper in Bielefeld (Northern Germany). All were being raised in a monolingual German-speaking environment in and around Bielefeld.

2.2 Stimuli

The sets of objects contained toys that represented their real counterparts (e.g. a toy table stands for a real table). Such miniatures were reported to be treated as real objects by children at the age of 19 to 30 months (DeLoache, Uttal & Rosengren 2004). The sets differed with respect to: (1) the geometry of the spatial relation (horizontal ON, vertical ON, and UNDER) (2) the functionality of the spatial relation (canonical versus noncanonical relation). A canonical relation relates to the most common function between two particular objects. For a tunnel and a train, the canonical relation is IN (usually, the train goes in or through the tunnel). In contrast to the canonical relation, a noncanonical relation was defined as a relation that is possible and plausible with the objects involved but does not relate to their customary function, like a train ON a tunnel.

The relations ON and UNDER were chosen because of the different level of proficiency children demonstrate in understanding them. While the preposition ON is reported to be understood very early, the understanding of UNDER is relatively poor at the age of 1;8 to 2;2 (e.g. Sinha et al. 1999; Rohlfing 2005).

Figure 1 shows that the stimuli and the relations chosen for the study were varied with regard to the target preposition as well as the canonicality of the relationships. What might appear counterintuitive is the selected stimulus for the noncanonical UNDER-relation, i.e. 'the horse under a bridge', because adults tend to assign UNDER to be the most common function of a bridge. Rohlfing (2001) reports, however, that when no preposition is given, children put a horse on a bridge. This suggests that children consider at least this kind of a bridge to be more like stairs with the canonical function of going over or being on. Thus, the relation requiring a horse to be under the bridge was chosen as a noncanonical one.

As can be viewed from Figure 1, the major limitation of the design resides in the fact that only 1 item per condition was used, which restrains the outcome of



Figure 1. Objects in different relationships

statistical analyses. Ideally, more items should be involved. However, when children are two years old, there are not many relations that are reliably canonical, especially for the UNDER relation. It remains a challenge to find spatial relations that (a) offers not only one valid relation but also at least two alternatives – our sets offer the UNDER relation in addition to the canonical 'pot ON table' and the ON to the canonical 'boy UNDER the umbrella'; the experience was that children perceive both alternatives, and some of them put the boy ON the umbrella (b) should be reliably canonical to all participating children.

2.3 Procedure

A session in this study lasted about half an hour. At the beginning, the experimenter engaged the child in free play at a small table, while the mother was filling out a language survey (see appendix) about the child's understanding and production of 49 spatial terms. More specifically, the 49 terms that have been asked for in the language survey referred to actions (like open, put, hide), relations between objects (e.g. in, on, under, to), nouns (such as front, inside, top) and other, deictic terms (here, where). These words were chosen because of their semantic relevance for this experiment. For each child, the reported terms were summed up for the production and understanding separately, on which basis the percentage of the in the language survey 49 asked terms was then calculated. The perceptual scores for production and understanding were later submitted to further analyses. Kickert (2008) has shown that the German version of this language survey correlated strongly with the values of ELFRA-2 (Grimm & Doil 2000), which is the German version of the MacArthur-Bates Communicative Development Inventories, CDI. The productive vocabulary, for instance, in the language survey presented here correlated very significantly with the productive vocabulary in ELFRA-2 (r = 0.88, p < 0.001).



Figure 2. Left side: a session from American English sample; right side: a session from German sample

After a few minutes, all toys and books were removed from the table and all three participants (the child, the mother and the experimenter) sat at the table (see Figure 2). Telling the child that she was going to show some new toys and they would all play a game, the experimenter presented the first pair of objects. Next, she showed a photo to the mother, depicting a relation, which are all shown in Figure 1. From that point on, the data were transcribed from the video-recordings. The mother was told that the relation on the photo was the goal of the task, and she was instructed to feel free to use verbal and nonverbal behavior to get her child to perform the task but not actually perform the relationship on the objects directly. If a child was not successful in understanding after several attempts at instructing on the part of the mother, the experimenter continued with new toys. The order of the photos presented to the mothers was randomized.

2.4 Category system for nonverbal performance

Table 1 gives an overview of the coding system for nonverbal behavior (see Table 1) and its synchronization with the verbal utterances.

The mothers' gestures were assigned to the type of deictic gesture when they extended their pointing finger to an object, a location or the path of motion. Gestural behavior was coded as iconic gesture, when it depicted the physical characteristics of an objects or action, for example an action of building blocks. Iconic gestures were coded as reinforcing when they depicted either the object or the action that was uttered at the same time. In addition to the common gesture taxonomy,

Gesture type	Characterisation	Example in picture
deictic	 Especially pointing, i.e. gesture performed with the index finger in reference to objects, locations or actions. <i>Reinforcing pointing</i>: labeling what is being pointed at, e.g. pointing at the location and saying "under" (as on the picture). <i>Supplementing pointing</i>: pointing clarifies a deictic term, i.e. without a point it is not clear to what the terms "this" (when pointing at an trajector object) or "there" (when pointing at a location or landmark object) refer 	
iconic	Gesture depicting the physical characteristics of an object or an action, e.g. an action of putting a train (right hand) on the tunnel (left hand on the picture)	
manipulative	 Manipulation on objects in order to highlight their: Orientation: mostly used with the landmark object (as on the picture), which was taken and set up on a table so its orientation was prepared for the target relationship like on the picture. Role: mostly used with the trajector object, which was taken and given to the child, by which action its active role as an trajector object was highlighted (see Figure 2 in both examples) 	

 Table 1. Coding system for nonverbal behavior

gestural behavior that was typical for the tasks given in the study was identified as manipulative gesture (cf. Fritsch 2003). As already described in the Introduction, this type of gesture includes manipulation of objects highlighting their orientation or their relational role. These gestures can be seen as different from "enactive names" that were suggested as object gestures (Bates et al. 1989: 1005). One reason is that in the study, the manipulative gestures are performed by mothers in contrast to children in the study by Bates and her colleagues (1989). Another reason is a much more complex function that goes beyond the function of recognition and association suggested in Bates et al. (1989). This type of gesture is comparable to a pre-stage of what Zukow-Goldring (2006) calls embodying gestures. She describes that this gesture takes place when a caregiver puts an infant through the motions of some activity; for example when a child tries to peel an orange, a caregiver might help and put control over the child's hand using her or his hand. In contrast to the study presented in Zukow-Goldring (2006), in the present study, mothers were asked to instruct their child to perform a spatial action (e.g. put the boy under the umbrella) but not to perform the action themselves. Therefore, it could only be observed how the caregivers set up the objects (e.g. a boy and an umbrella), so the child can do it by her-/himself. This was the case when a mother, for example, took an umbrella and put it upright, so its proper orientation was already set up for the child. Another case of manipulative gestures occurred when the mother handed one object (e.g. the train) to her child and instructed to put this particular object on another one (the tunnel). In this example, she set up the relational role of the train as an active object by giving it to the hand of her child and suggesting an activity, and therefore, the trajector-role.

Each type of gesture was coded for its referent, because it was of relevance for our analysis, whether the gesture relates to a trajector-, a landmark-object (both often accompanied by a noun or the article), a location (often accompanying a locative word) or an action. Manipulative gestures were coded as reinforcing when the referent (object or action) was also mentioned verbally. For example, the case when a mother set up the orientation of the umbrella and said "Stell' mal hin! [put it that way]", was coded as reinforcing gestures.

With regard to the correspondence between verbal and nonverbal behavior, all gestures were described as either reinforcing or supplementing the verbal information. Here, not the strict temporal synchrony but rather the equivalence to the continuing utterance was crucial. Implementing the correspondence categories to the spatial task in our study was straightforward concerning the trajector object. In the case a mother, for example, pointed at or manipulated a horse and said "this horsey", this gesture was coded as reinforcing, because the mother labeled what she pointed to; when a mother pointed to a horse and said "this", it was coded as supplementing gesture, because the reference of the deictic term is clarified by the point. Similarly, concerning the landmark object, one could argue that reinforced pointing occurs when a speaker point to a location and says, for example, "under the bridge", so the targeted location and the landmark are verbalized. However, some challenges arose in the process of coding deictic gestures when for some landmark objects (especially the tunnel and the fence), it was difficult to differentiate whether mothers pointed to the landmark object or to the requested location. Therefore, these two referential categories were put together. Accordingly, a deictic gesture was coded as reinforced pointing as soon as one of the two referents (location or the landmark object) was mentioned in the accompanied speech. For example if a mother said "hook it to this!" and pointed to a location, even though the landmark was not explicitly mentioned in this utterance, it was counted as reinforced pointing, because the location was specified by the term "to". If a mother said "hook it right here" and pointed to a location, it was counted as supplementing pointing, because the utterance contains deictic terms referring to the location.

Finally, children's performance was scored as successful in the case when a child performed the spatial relation the mother requested. The child's performance was scored as not successful in the case when a child performed other spatial configuration instead of the requested relation – this could be often observed in the noncanonical conditions, when children kept performing the canonical relationship.

The reliability was tested on 25% of the American English and 25% of the German data randomly coded by an independent coder. The Cohen's kappa scores for agreement between coders were 0.74 for identifying gestures, 0.87 for identifying gesture types, 0.77 for classifying gestures (whether supplementing, reinforcing, or in form of a saccades) according to the correspondence between gesture and speech and 0.93 for coding the referent of the gestures.

3. Results and Discussion

3.1 Cross-linguistic findings on age, lexicon and performance

The initial analyses focused on the differences between the North American and German groups (summarized in Table 2) and encompass the Mann-Whitney U-test for independent samples on records that were obtained from the language survey filled out by the mothers before the experiment.

There were no statistically significant differences between the German and the North American samples for age and the data on their spatial vocabulary (see Table 2 for means in both samples). A further dependent variable was children's performance in the task, analyzed by giving scores for the outcome of the mother-child dialogue. When a child successfully solved the task, a score of '1' was given; when a child performed another relation or when a child did not perform at all, a '0' was given. Since the design consisted of 4 tasks, a maximum score of 4 was possible. The Mann-Whitney U-test analyses for independent

	North American sample		German Sample		Mann-Whitney U test for independent samples	
	M	SD	M	SD	Z (N=17 in each group)	
Age in months	22.7	2.0	22.5	1.9	-0.19, p = 0.86	
Production of spatial words in%	30.6	19.3	35.6	20.6	-0.73, p = 0.43	
Understanding of spatial words in%	63.3	13.9	68.1	19.6	-1.12, p = 0.27	
Performance in # of successful tasks	3.0	0.87	3.2	0.73	-0.60, p = 0.59	

Table 2. Comparison of data on age, lexicon and performance in North American versusGerman sample

samples revealed no differences between the German and North American sample on children's performance. The result is, therefore, that with regard to the age, reported lexicon of the children and their performance in the tasks, the two samples are comparable.

Concerning the overall performance, out of the four tasks (Table, Umbrella, Tunnel, Bridge), children performed 3.1 tasks on average (2 were the minimum and 4 the maximum). Because of the comparability were the North American and German samples together into one group to obtain a greater statistical power for correlational analyses of children performance with their age and lexicon. Children's performance correlated positively with their age (r = 0.50, df = 32, p < 0.01) and with reported production of spatial words (r = 0.49, df = 32, p < 0.01) suggesting that older children and/or children with a more advanced spatial vocabulary did better on solving the tasks. The production of spatial prepositions as reported by the mothers was, therefore, included as a covariate into further analyses. There were no significant findings on the age of children correlating with their spatial lexicon (r = 0.24, df = 32, p = 0.17 for understanding spatial terms, r = 0.29, df = 32, p = 0.1 for production of spatial terms). It suggests that children's growth of spatial lexicon is not only a matter of their age.

3.2 Gestural type in canonical vs. noncanonical settings

In order to analyze different types of maternal gestural behavior, the Mann-Whitney U-test first analyses independent samples on different gestural types in dependence on canonicality of the settings were conducted. The analyses revealed no differences between the German and North American sample on mothers' gestural performance. Because of this comparability, the North American and German samples were put together into one group to obtain a greater statistical power.



Figure 3. The mean number of gestures produced in North American and German samples

As introduced above, the maternal gestures were classified into three types: deictic, iconic and manipulative. The distribution of the maternal gesture types in both samples is presented in Figure 3.

An ANCOVA of the rate of maternal gestures with 3 gestural types (a withinsubject variable, language as a between-subject variable and the production of spatial terms as reported by the mothers being a covariate) revealed a main effect of type F(2, 62) = 17.83, p < 0.001, Eta² = 0.36 with more deictic gestures that were observable than any other type. This supports findings from previous studies about the majority of gestures produced by mothers being deictic (Iverson et al. 1999; Özçalişkan & Goldin-Meadow 2005). Furthermore, an interaction between the type of maternal gestures and the covariate, namely children's production of spatial terms as reported in the language surveys F(2, 62) = 3.57, p < 0.05, $Eta^2 = 0.10$ was found. Accordingly, there was a significant negative correlation between the deictic (r = -0.37, df = 32, p < 0.05) and iconic (r = -0.35, df = 32, p < 0.05), but not the manipulative (r = -0.08, df = 32, p = 0.65), gestures with children's productive vocabulary. The obtained interaction suggests that mothers of children with a reported less advanced spatial lexicon produced more deictic and iconic gestures. Even though the data do not provide longitudinal insights, this finding seems to be amenable to the explanation in terms of an adaptation process. Mothers seem to be sensitive to the developmental level of the interlocutor (Iverson et al. 1999; Namy & Nolan

2004) that is determined by children's communicative capabilities (as reported by mothers in language survey) rather than their age.

As to the different types of gestures, it might be questioned whether the category of manipulative gesture is a gesture at all and if so whether it should be regarded along the same lines as deictic and iconic gestures. Another possibility is to consider the manipulative gesture as glue between attentional device and communicative gestures (Salas Poblete & Rohlfing 2008). The similarity between the manipulative gestures and other communicative gestures was assessed by means of correlation coefficients. The resulting positive correlation with deictic (r = 0.41, df = 32, p < 0.05) but not iconic gestures (r = 0.01, df = 32, p = 0.95) suggests that manipulative gestures seem to be a part of mothers' gestural repertoire. Related to this, in some cases, it could be observed that immediately after a mother prepared the orientation of the objects, she pointed to it (see Figure 4).



'Can it go also up here on top? The train?'



Especially in the task train/tunnel, it could be observed that mothers prepared the tunnel orienting it with the opening away from the child, so it could be rather perceived like a block. It appears that this kind of manipulation can draw child's attention to the object. At the same time, it can highlight (the horizontal surface in the case of the tunnel) and also limit (the cavity of the tunnel) specific object's characteristics. This way, the child's attention is guided for the purpose of the task. I therefore want to argue that instead of being only an attentional device, manipulative gestures might rather express a planned motor action with the attentional function included highlighting specific characteristics of an object (Salas Poblete & Rohlfing 2008).

The main focus of analysis was to explore whether the canonicality of the spatial relationship can have an influence on the nonverbal behavior. As can be seen in Table 3, in noncanonical settings mothers gestured more.

	canonical relationship (Table, Umbrella)	noncanonical relationship (Tunnel, Bridge)
deictic gesture	110	146
manipulative gesture	30	58
iconic gesture	6	8
pointing saccades	24	17

Table 3. Gestural types in dependence on the canonicality of spatial relationships

As to the most expressed gestural type, the deictic gestures, it was analyzed whether pointing was reinforcing or supplementing the verbal expression (see Table 4). Then the relationship between mothers' gestural behavior and reported vocabulary competence was studied.

Table 4. Pointing gestures in dependence on the canonicality of spatial relationship taken together

	canonical relationship (Table, Umbrella)	noncanonical relationship (Tunnel, Bridge)
number of all points	110	146
reinforcing pointing	88	120
supplementing pointing	22	26
pointing saccades	24	17

The data revealed that reinforced pointing correlating negatively with children's reported productive lexicon in the noncanonical (r = -0.44, df = 32, p < 0.01) but only marginally significant for canonical condition (r = -0.32, df = 32, p = 0.06). Consistent with these findings, children's production of spatial prepositions also correlated negatively with the increase of reinforced pointing in the noncanonical (r = -0.40, df = 32, p < 0.05) but not canonical condition (r = -0.20, df = 32, p = 0.24). This supports the interpretation that in noncanonical tasks, the proficiency in spatial prepositions seems to play an important role and mothers reinforced their verbal message by pointing to children who produced few spatial prepositions. Thus, the correlational findings indicate that noncanonical relationships trigger the reinforcing function of deictic gestures.

In the canonical condition, in contrast, a specific pointing behavior in mothers' input was noticed. Here, the examination was motivated by observations reported in Rohlfing (2005). Accordingly, especially in canonical tasks (like the Umbrella

task), a pointing behavior that can be described as "pointing saccades" (Rohlfing, 2005) emerges, i.e. mothers indicated the trajector object first, and then their pointing finger moved immediately from there to the landmark object (or trajector's would-be-location) as it is captured in Figure 5.



Figure 5. A pointing saccade: A mother says "Put the dolly under the umbrella" and her gestures accompanies the words "dolly" when she points to the little doll and "under" when her point moves to the umbrella (duration almost one sec)

Based on these findings, all pointing behaviors were singled out that followed this saccade pattern and conducted a correlational analysis (see Figure 6).



Figure 6. Mean number of pointing saccades in canonical vs. noncanonical settings

The analysis revealed a negative correlational effect of the frequency of pointing saccades with children's age (r = -0.36, df = 32, p < 0.05). A regression analysis with children's age and their reported production of spatial terms entered in that order revealed that only children's age significantly accounted for 13% of variance (F(1,33) = 4.9, p < 0.05) in the production of gesture saccades in a canonical condition but not in noncanonical condition. The following ANCOVA on the
number of pointing saccades with canonicality as a within-subject variable and reported production of spatial terms and age as covariates revealed a statistically marginal main effect of canonicality F(1, 30) = 3.58, p = 0.068, Eta² = 0.11 (see Figure 6) and also statistical trend for an interaction between canonicality and age, F(1, 30) = 3.08, p = 0.09, Eta² = 0.09. Together these findings indicate that when instructing for a canonical relationship, mothers of younger children tended to point in saccades, i.e. from the trajector object to the landmark object. Thus, gestures in canonical conditions bear the potential to convey information about the canonicality of this relationship.

Although statistically weak, the findings can thus provide some indications for the hypothesis that gestures can guide children to the social convention of the spatial relationship (canonical or noncanonical). However, more research contrasting different canonical and noncanonical relationships is needed to fully answer the question whether the pointing saccades indicate a canonical relationship. For the time being, it can be proposed that while the reinforcing function of pointing (as found by Iverson et al. 1999) is a method integrated in the verbal system, so it corresponds to and reinforces the verbal semantic information, the pointing saccades may be a method used predominantly to children who are less advanced in lexicon. Pointing saccades, similar as iconic gestures (described above), may convey early verbally a semantic message about the social function ('togetherness') of the relationship between two objects.

3.3 Discussion

Concerning the gestural type, the presented findings confirmed the results found in Italian and North American mothers of toddlers by Iverson and her colleagues (1999) and extended them to the domain of space. The data presented above reveal that overall, mothers produced more deictic gestures than any other type. Further, pointing gestures were used more to children with a less advanced spatial lexicon. Thus, children's linguistic capabilities seem to have an impact on maternal nonverbal behavior. In addition, support for the semantic correspondence of verbal and nonverbal behavior was found. According to Iverson et al. (1999), mothers use gestures in order to reinforce their verbal message to children lexically less advanced. It should be noted that even though the values reported in the language survey developed for the study correlate with the values of ELFRA-2 (Kickert 2008; Grimm & Doil 2000), it is not a standardized measurement. It is possible that the checked words reflect the subjectively perceived dialogical comfort or an overall talkativeness of the child rather than the child's spatial vocabulary. However, both remain related to linguistic capabilities of the interlocutor. A further exploration of the function of manipulative gestures revealed that their numbers correlate with deictic gestures. I took this finding as an indication that

this type of gesture is related to the communicative types. Based on these observations, I would like to argue that this type can not only guide the interlocutor's attention, but also bears lots of motor impulses, which may facilitate attention and reference. Similar to this idea, Booth and her colleague (2008) compared whether manipulative gestures such as moving an object to a side can function as a referential gesture toddlers aged 28 to 31 months. They tested 80 children and found that in comparison to eye-gaze alone, manual forms such as pointing, touching or manipulating in concert with gazing facilitate word learning. Furthermore, gestural cues that involved contact between speaker and referent (such as in touching and manipulation) were superior to those that did not in facilitating comprehension of new words. In the authors' view, manual forms of gesture might have an advantage because they integrate a greater number of body parts into lengthier and more elaborate actions, which are more visible. In addition, they terminate in closer proximity to the location of the target, and therefore require less extrapolation from speaker to referent, which might facilitate the reference process.

Consistent with McNeil and her colleagues (2000) who found that gestural behavior can change as a function of the complexity of the spoken message, – a complex message required lexical and syntactic rather than context-based comprehension strategies – in the study presented above, it was also found that gestural behavior can change depending on the task demands. The canonicality of the spatial configuration between the objects determined the task. A canonical relationship can trigger more pointing trajectories conveying information about the relation ('togetherness') of objects, which conveys a social rule. It was also found that this behavior occurred especially when the younger children in our sample were addressed suggesting that it is a form of early verbal communication. In the noncanonical tasks, in contrast, mothers' pointing fulfills a linguistic function, because the gesture corresponds with the verbal message. It could be concluded that canonicality is an important contextual factor influencing nonverbal behavior in a particular situation (Rohlfing et al. 2003).

To summarize, in support of previous studies and in extension to the domain of space, the results also suggest that a simple causal relationship between children's age and the production of maternal gestures does not exist. Instead, it was found that mothers' nonverbal behavior is impacted by children's productive lexicon as reported by the mothers and the canonicality of the spatial relationship. The latter results suggest that the perception of objects, for the purpose of referring to them, cannot be separated from the social character of the objects, because artifacts "have an intrinsic meaning given by their canonical function of use value" (Sinha & Rodriguez 2008: 371). In this sense, semantics and pragmatics are inseparable (Zelinsky-Wibbelt 1993).

Clearly, in learning theories such as the Emergentist Coalision Model (Hollich et al. 2000; Golinkoff & Hirsh-Pasek 2006) or usage-based theory (Tomasello 2003), the notion of social cues should be extended to characteristics of objects that the children have experience with. Child's attention to object characteristics seems to be educated already in early verbal stages, when caregivers present objects in a particular manner and demonstrate what to do with them (Jensen de López 2006; Zukow-Goldring 2006; Rakoczy et al. 2005). From this experience, children gain background knowledge about these objects, which will accompany them in their task-oriented performances. Objects should therefore be considered as belonging to the "semiosphere" (Lotman 1984). Sinha & Rodriguez (2008) assume that before children know what the word "chair" refers to, they know about the particular function that a chair has. Thus, such action knowledge about "socially constructed and normatively regulated affordances" (ibid: 368) proceeds symbolic knowledge. These social facts create a semiosphere, i.e. an environment, in which an object or an event bears a meaning. A similar conception can be found in Strohner's (1995) eco-system metaphor stating that not the mind alone constitutes the cognition but the mind in the interaction with its environment. In contrast to known approaches towards semiotics, Lotman (1984) puts the environment in the middle of thinking. In this approach, a symbol does not exist from the beginning but is actually a product of the analysis. For the language acquisition, this perspective means that a child is surrounded by a meaning-giving environment, in which she or he can rely on different sources of meaning. Canonicality or familiarity of objects can be such a source.

Is the concept of a semiosphere containing different sources of meaning valid only for the developmental processes? As I showed above, mothers also make use of these social facts not only in their verbal but also nonverbal behavior. It is possible that such a behavior can be also elicited towards other partners with a knowledge deficiency.

3.4 Conclusions

In this paper, two operationalizations of the materiality were suggested: (i) the familiarity and (ii) the canonicality of objects. With reference to previous literature, it was shown that the influence of the nature of objects is crucial for verbal behavior. The results of the pilot study presented here further suggest that materiality also impacts nonverbal behavior as different gestural pattern could be observed in dependence of the canonicality of the relationships between objects. However, as only limited number of objects and only two relationships (ON and UNDER) were tested, these findings serve rather as further hypotheses that should be tested under experimental conditions with careful variations in the nature of objects, further relationships and more objects.

Based on the findings from verbal behavior in the previous literature and the indications from the pilot study presented here, it can be concluded that ecologically valid objects will restrict the answers and people (children and adults) are guided by them in their linguistic behavior. In experimental pragmatics, a careful variation of the nature of objects – i.e. their typical or familiar appearance, canonical functions, other use values – is therefore necessary.

References

- Bates, E., Thal, D., Whitesell, K., Fenson, L. & Oakes, L. 1989. Integrating language and gesture in infancy. *Developmental Psychology* 25: 1004–1019.
- Booth, A.E., McGregor, K.K. & Rohlfing, K.J. 2008. Socio-pragmatics and attention: Contributions to gesturally guided word learning in toddlers. *Journal of Language Learning and Development* 4: 179–202.
- Capone, N.C. & McGregor, K.K. 2006. The effect of semantic information on toddlers' word retrieval. *Journal of Speech, Language, and Hearing Research* 48: 1468–1480.
- Choi, S. & Rohlfing, K. 2010. Discourse and lexical patterns in mothers' speech during spatial tasks: What role do spatial words play? In *Japanese/Korean Linguistics* 17, S. Iwaski, H. Hoji, P.M. Clancy & S.-O. Sahn (eds), 117–133. Stanford: CSLI Publications.
- Clark, E.V. 1973. Non-linguistic strategies and the acquisition of word meanings. *Cognition* 3: 161–182.
- Coventry, K.R., Prat-Sala, M. & Richards, L. 2001. The interplay between geometry and function in the comprehension of *over, under, above* and *below. Journal of Memory and Language* 44: 376–398.
- Csibra, G. & Gergely, G. 2006. Social learning and social cognition: The case for pedagogy. In *Attention and Performance XXI*, M.H. Johnson & Y. Munakata (eds), 249–274. Oxford: OUP.
- Dąbrowska, E. 2005. Language, Mind and Brain. Some Psychological and Neurological Constraints on Theories of Grammar. Washington D.C.: Georgetown University Press.
- DeLoache, J.S., Uttal, D.H. & Rosengren, K.S. 2004. Scale errors offer evidence for a perceptionaction dissociation early in life. *Science* 304: 1027–1029.
- Feist, M.I. & Gentner, D. 1998. On plates, bowls, and dishes: Factors in the use of English in and on. In Proceedings of the Twentieth Annual Conference of the Cognitive Science Society, M.A. Gernsbacher & S.J. Derry (eds), 345–349. Mahwah NJ: Lawarence Erlbaum Associates.
- Flom, R. & Pick, A.D. 2003. Verbal encouragement and joint attention in 18-month-old infants. *Infant Behavior and Development* 26: 121–134.
- Freeman, N.H., Lloyd, S. & Sinha, C.G. 1980. Infant search tasks reveal early concepts of containment and canonical usage of objects. *Cognition* 8: 243–263.
- Fritsch, J. 2003. Vision-Based Recognition of Gestures with Context. Ph.D. dissertation, Universität Bielefeld.
- Gibson, E.J. & Pick, A.D. 2000. An Ecological Approach to Perceptual Learning and Development. Oxford: OUP.
- Grimm, H. & Doil, H. 2000. *Elternfragebogen für die Früherkennung von Risikokindern (ELFRA-2)*. Göttingen: Hogrefe.
- Goldin-Meadow, S. 1999. The role of gesture in communication and thinking. *Trends in Cognitive Science* 3: 419–429.

- Goldin-Meadow, S. & Sandhofer, C.M. 1999. Gesture conveys substantive information about a child's thoughts to ordinary listeners. *Developmental Science* 2: 67–74.
- Golinkoff, R.M. & Hirsh-Pasek, K. 2006. Baby wordsmith. *Current Directions in Psychological Science* 15: 30–33.
- Grèzes, J., Armony, J.L., Rowe, J. & Passingham, R.E. 2003. Activations related to 'mirror' and 'canonical' neurons in the human brain: An fMRI study. *Neuroimage* 18: 928–937.
- Hollich, G., Hirsh-Pasek, K. & Golinkoff, R. 2000. Breaking the language barrier: An emergentist coalition model of word learning. *Monographs of the Society for Research in Child Development* 65 [3, Serial No. 262].
- Hottenroth, P.-M. 1993. Prepositions and object concepts: A contribution to cognitive semantics.
 In *The Semantics of Prepositions: From Mental Processing to Natural Language Processing*, C. Zelinsky-Wibbelt (ed.), 179–219. Berlin: Mouton de Gruyter.
- Iverson, J.M. & Thal, D.J. 1998. Communicative transitions. There's more to the hand than meets the eye. In *Transition in Prelinguistic Communication*, A.M. Wetherby, S.F. Warren & J. Reichle (eds), 59–86. Baltimore MD: Paul Brookes.
- Iverson, J.M., Capirci, O., Longobardi, E. & Caselli, C.M. 1999. Gesturing in mother-child interactions. Cognitive Development 14: 57–75.
- Jensen de López, K. 2006. Culture, language and canonicality: Differences in the use of containers between Zapotec (Mexican Indigenous) and Danish children. In. *Doing Things with Things*, A. Costall & O. Dreier (eds),87–112. London: Ashgate.
- Kickert, K. 2008. UNTERstützen des Präpositionserwerbs bei 2-jährigen Kindern eine Interventionsstudie. Diplomarbeit, Abteilung für Psychologie, Universität Bielefeld.
- Kita, S. & Özyürek, A. 2003. What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language* 48: 16–32.
- Langacker, R.W. 1987. *Foundations of Cognitive Grammar*, Vol 1: *Theoretical Prerequisites*. Stanford CA: Stanford University Press.
- Lotman, Y.M. 1984. O semiosfere. Sign Systems Studies (Trudy po znakovym sistemam) 17: 5-23.
- Malt, B.C., Sloman, S.A. & Gennari, S.P. 2003. Speaking versus thinking about objects and actions. In *Language in Mind. Advances in the Study of Language and Thought*, D. Gentner & S. Goldin-Meadow (eds), 81–112. Cambridge MA: The MIT Press.
- Namy, L.L. & Nalan, S.A. 2004. Characterizing changes in parents labelling and gesturing and their relation to early communicative development. *Journal of Child Language* 31: 821–835.
- McNeil, N.M., Alibali, M.W. & Evans, J.L. 2000. The role of gesture in children's comprehension of spoken language: Now they need it, now they don't. *Journal of Nonverbal Behavior* 24: 131–150.
- Nelson, K. 1996. Language in Cognitive Development. Emergence of the Mediated Mind. Cambridge: CUP.
- Nelson, K. & Ware, A. 2002. The reemergence of function. In *Representation, Memory and Development. Essays in Honor of Jean Mandler*, N.L. Stein, P.J. Bauer & M. Rabinowitz (eds), 161–184. Mahwah NJ: Lawrence Erlbaum Associates.
- Özşalişkan, Ş. & Goldin-Meadow, S. 2005. Do parents lead their children by the hand? *Journal* of Child Language 32: 481–505.
- Pierno, A.C., Becchio, C., Wall, M.B., Smith, A.T., Turella, L. & Castiello, U. 2006. When gaze turns into grasp. *Journal of Cognitive Neurocience* 18: 2130–2137.
- Rakoczy, H., Tomasello, M. & Striano, T. 2005. On tools and toys: How children learn to act on and pretend with 'virgin objects'. *Developmental Science* 8: 57–73.

- Rizzolatti, G. & Arbib, M.A. 1998. Langauge within our grasp. *Trends in Neuroscience* 21: 188–194.
- Rizzolatti, G., Fogassi, L. & Gallese, V. 2001. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neurocience* 2: 661–670.
- Rohlfing, K.J. 2001. No preposition required. The role of prepositions for the understanding of spatial relations in language acquisition. In *Applied Cognitive Linguistics* I: *Theory and Language Acquisition*, M. Pütz, S. Niemeier & R. Dirven (eds), 229–247. Berlin: Mouton de Gruyter.
- Rohlfing, K.J. 2005. Learning prepositions. *Perspectives on Language Learning and Education* 12: 13–17.
- Rohlfing, K.J. 2006. Facilitating the acquisition of UNDER by means of IN and ON A training study in Polish. *Journal of Child Language* 33: 51–69.
- Rohlfing, K.J. & Choi, S. 2004. Getting you to understand. How mothers instruct their children to put two things together. Paper presented at the IADA (International Association for Dialogue Analysis) Conference, Chicago, 30. March–3. April.
- Rohlfing, K.J., Rehm, M. & Goecke, K.U. 2003. Situatedness: The interplay between context(s) and situation. *Journal of Cognition and Culture* 3: 132–157.
- Salas Poblete, J. & Rohlfing, K.J. 2008. Manipulative gestures or actions? Paper presented at the 12. Internationale Kongress der Deutschen Gesellschaft für Semantik, Stuttgart 9.– 12. October.
- Sinha, C. 1982. Representational development and the structure of action. In Social Cognition: Studies of the Development of Understanding, G. Butterworth & P. Light (eds), 137–162. Brighton: Harvester.
- Sinha, C. 1983. Background knowledge, presupposition and canonicality. In *Concept Development and the Development of Word Meaning*, T. Seiler & W. Wannenmacher (eds), 269–296. Berlin: Springer.
- Sinha, C. 2005. Blending out of the background: Play, props and staging in the material world. *Journal of Pragmatics* 37: 1537–1554.
- Sinha, C. & Rodriguez, C. 2008. Language and the signifying object: From convention to imagination. In *The Shared Mind: Perspectives on Intersubjectivity*, J. Zlatev, T. Racine, C. Sinha & E. Itkonen (eds), 357–378. Amsterdam: John Benjamins.
- Sinha, C., Thorseng, L., Hayashi, M. & Plunkett, K. 1999. Spatial language acquisition in Danish, English and Japanese. In *Language and Thought in Development. Cross-linguistic Studies*, P. Bcoeder & J. Murre (eds). 95–125. Tübingen: Narr.
- Strohner, H. 1995. Kognitive Systeme. Eine Einführung in die Kognitionswissenschaft. Opladen: Westdeutscher Verlag.
- Tomasello, M. 2003. *Constructing Language. A Usage-Based Theory of Language Acquisition.* Cambridge MA: Harvard University Press.
- Tomasello, M. & Farrar, M.J. 1986. Joint attention and early language. *Child Development* 57: 1454–1463.
- Zukow-Goldring, P. 2006. Assisted imitation: Affordances, effectivities, and the mirror system in early language development. In *From Action to Language*, M.A. Arbib (ed), 469–500. Cambridge: CUP.
- Zelinsky-Wibbelt, C. 1993. Introduction. In *The Semantics of Prepositions: From Mental Processing to Natural Language Processing*, C. Zelinsky-Wibbelt (ed), 1–24. Berlin: Mouton de Gruyter.

Appendix

						Date		
			Langua	ge Surv	vey			
Child's Na	ime _			🗆 f	emale	□male		
Age in mo	onths_							
Date of Bi	rth _							
Does the child I	have a	any bro	others or sisters?		_			
∟yes			age:	⊔r	10			
			age:					
hich she / he at	ole to	produ	ce(♥):			Noune		
Actions			011	8	0	Nouns		
		•	on	<u> </u>	•	front	<u> </u>	•
give	9	•	away	9	0	back	ÿ	ø
fall	2	•	Relations betwe	een obj	ects	top (on top of)	9	•
hide	2	•	to	9	0	inside	9	•
hang	9	•	behind	9	0	outside	9	•
turn	9	•	in front of	9	0	bottom	9	•
turn over	2	•	in back of	9	0	(on the) edge	9	•
go/ get in	2	•	between	2	0	(in the) middle	9	•
take out	9	•	next to	2	0	side	9	•
tance out	g	•	in	9	0	underside	P	•
open	ď						Ø	•
open close		•	on	9	0	hole	8	
open close up	9 9 9	0	on under	9 9	0	hole	ď	
open close up down	2 9 9 9	•	on under underneath	9 9 9	0 0	Others		
open close up down upside down	9 9 9 9 9	0 0 0	on under underneath above	9 9 9 9	0 0 0	Others	9	0
open close up down upside down together	9 9 9 9 9 9	0 0 0 0	on under underneath above below	9 9 9 9 9	0 0 0 0	Others here! where?	9 9	0
open close up down upside down together apart	9 9 9 9 9 9 9	0 0 0 0 0	on under underneath above below through	9 9 9 9 9 9	0 0 0 0	hole Others here! where? there!	2 9 9 9	0 0

If the chilld produces two word combination, please provide some examples (e.g. Daddy home)

Thank you!

Blocking modal enrichment (tatsächlich)*

Hans-Christian Schmitz Fraunhofer FIT

I show that under the right circumstances recipients can carry out operations of modal enrichment. Such enrichment operations can be blocked. One means of blocking modal enrichment in German is using the adverb *tatsächlich* ('in fact'). I define *tatsächlich* as an enrichment blocker and support this definition by experimental data.

1. Introduction

Several forms of meaning enrichment have been discussed in the literature and subsumed under the terms 'impliciture' (Grice 1989; Bach 1994) and 'expliciture' (Carston 2002), among other terms.¹ The sentences in (1) are two well-known examples:

- (1) a. It's raining.
 - b. You are not going to die.

Example (1a) is from Perry (1998). Perry claims that the sentence is always interpreted as meaning that it is raining at some specific location, although no location is explicitly mentioned. The hearer must add this location in order to derive the sentence meaning as intended by the speaker. Example (1b) is from Bach (1994). Imagine a mother uttering (1b) to her son who is upset about a minor injury. Of course, the mother does not convey that her son his immortal (which would be the sentence's literal meaning) but that he will not die *from that particular injury*. In both examples, the literal meaning of the respective sentence has to be expanded.

^{*} Thank you to all test subjects who participated in the experiments. Thank you also to Benedikt Löwe for providing me with the paradigmatic time-example, and to Bernhard Fisseni, Eric Fuß, Uwe Kirschenmann, Joost Kremers, Manfred Kupffer, Charlotte Matheson, Cécile Meier, Jörg Meibauer, Ulrich Missberger, Katja Niemann, Christian Plunze, Bernhard Schröder, Magda Schwager, Markus Steinbach, Henk Zeevat, Ede Zimmermann and an anonymous reviewer for helpful comments and stimulating discussions. All mistakes are mine.

^{1.} For a comparison of the terms and the approaches connected to them, cf. Bach (2006).

In this paper, I will discuss one kind of meaning enrichment that has not gained much attention so far and that I call 'modal enrichment'. It is exemplified by (2):

(2) A: Wie spät ist es?
B: Es ist 5 nach 3, meine Uhr geht aber 5 Minuten vor.
'A: What time is it?
B: It's 5 past 3, but my watch is 5 minutes fast.'

As will be shown in the next section, B's answer is usually interpreted non-literally, meaning that it is 5 past 3 *by the speaker's watch* and that, therefore, it is in fact 3 o'clock. Hearers of the answer expand it's literal meaning by constructing and applying a modal operator (*by the speaker's watch*). Such an expansion is what I call 'modal enrichment'.

I claim that enrichment operations are conventionalised. They can be defined within pragma-semantic feature structures that represent both the literal meanings and expanded meanings of sentences. I will introduce such a feature structure for the answer of (2).

Enrichment operations are performed by recipients, not by speakers. There can be situations in which modal enrichment does not lead to the interpretation intended by the speaker. In such a situation, the speaker can avoid a misunderstanding by explicitly blocking enrichment. I will argue that in German such a blocking function is fulfilled by the word 'tatsächlich' ('in fact'). Within the proposed pragma-semantic feature structures, 'tatsächlich' can be defined as a pure enrichment blocker that affects the applicability of modal enrichment without changing the literal meaning of its host.

My analyses both of modal enrichment and of 'tatsächlich' are based on experimental data obtained from linguistically naïve test subjects. I presented them dialogues like (2) and asked questions like what time it is according to B's response. Answering these questions requires the interpretation of the examples. Thus, from the subjects's answers I can infer their interpretations of the examples and, in particular, I can infer whether they perform an operation of modal enrichment or not. All test subjects are speakers of German. Therefore, the experimental data only support an analysis of 'tatsächlich', not of its English counterpart 'in fact'. However, I assume that 'tatsächlich' and 'in fact' play similar roles.

The outline of this paper is as follows: in Section 2, I will discuss serveral examples that serve as evidence for modal enrichment, and I will introduce pragma-semantic feature structures for defining modal enrichment operations. In Section 3, I will add 'tatsächlich' to the examples and show how this effects interpretation. I will propose a definition of 'tatsächlich' as an enrichment blocker. Finally, in Section 4, I will discuss objections against my definition.

2. Modal enrichment

I asked 42 test subjects to write down what time it is, according to B's response in example (2), here repeated as (3):

(3) A: Wie spät ist es?
B: Es ist 5 nach 3, meine Uhr geht aber 5 Minuten vor.
'A: What time is it?
B: It's 5 past 3, but my watch is 5 minutes fast.'

The experiment, like all other experiments that I will refer to in this paper, took place in a classroom setting. The example was projected onto a wall, and it was read aloud twice. The test subjects noted down the time on their own. That is, the experiment was not a multiple choice experiment, I did not present a list of times to choose from.² This is the result: 36 subjects (85.7%) believed it to be 15:00, which is not the literal meaning of B's answer. Only six subjects (14.3%) believed it to be 15:05, which is the literal meaning of the answer. Thus, the vast majority interpreted B's answer non-literally.

Of those subjects that arrived at the non-literal interpretation I asked the subjects how they had arrived at this interpretation. They answered that they interpreted the example in the sense of '*By my watch* it's 5 past 3, but my watch is 5 minutes fast'. This entails that it is 15:00. The example was not invented. When I first heard it, I asked the speaker what he meant. He replied that he meant that it was 15:00. Thus, the subjects interpreted the example in the way that was intended by the original speaker.

The non-literal interpretation can be modelled as follows: in the second conjunct of the answer, the subjects identify a noun phrase that denotes a potential information source ('my watch'). They transform the semantic representation of the noun phrase into a modal operator (*by the speaker's watch*), apply this operator to the semantic representation of the first conjunct, and then infer what time it is. I call the operation of constructing and applying a modal operator an 'operation

^{2.} All test subjects of this and the other experiments referred to in this paper were firstyear students in linguistics at the University of Bonn or the University of Frankfurt. The experiments were carried out between 2005 and 2008. In this first experiment, 26 subjects were native speakers of German and 16 subjects were non-native speakers of German. There was no significant difference between these groups regarding their answers. In all other experiments – that is, in every experiment except this first one (consisting of two parts, namely the interpretation of example (3) and the interpretation of example (10) in Section 3 – all test subjects were native German speakers.

of modal enrichment. The subjects perform an operation of modal enrichment in order to grasp the meaning intended by the speaker.

The experiment was repeated serveral times with different test subjects. The resulting data are very robust:

- 43 test subjects 15:00: 94%; 15:05: 6%
- 42 test subjects 15:00: 85.7%; 15:05: 14.3%
- 17 test subjects 15:00: 82.3%; 15:05: 5.9%; 15:10: 5.9%; '?': 5.9%
- 44 test subjects 15:00: 90.9%; 15:05: 6.8%; 15:10: 2.3%

One might speculate that the non-literal translation is prompted by the presence of 'aber' ('but'). However, this is not the case: we can change the position of 'aber' (example (4a)), we can replace 'aber' with 'und' ('and', example (4b)), and we can even leave out the conjunction and divide the answer into two separate sentences (example (4c)) without changing the interpretations. The first part of Table 1 shows experimental results for the examples (4a)–(4c) which prove that the non-literal interpretation does not depend on 'aber'.

(4) A: Wie spät ist es?

'A: What time is it?'

- a. B: Es ist 5 nach 3, aber meine Uhr geht 5 Minuten vor.'B: It's 5 past 3, but my watch is 5 minutes fast.'
- b. B: Es ist 5 nach 3, und meine Uhr geht 5 Minuten vor.'B: It's 5 past 3, and my watch is 5 minutes fast.'
- c. B: Es ist 5 nach 3. Meine Uhr geht 5 Minuten vor.'B: It's 5 past 3. My watch is 5 minutes fast.'

One might also speculate that modal enrichment for time announcements is hard-wired, that is, that time announcements are always interpreted with respect

1 1		, , , , , , ,		
	15:00	15:05	15:10	·?'
example (4a), 48 test subjects:	85.4%	12.5%	2.1%	
example (4b), 22 test subjects:	81.8%	13.6%	4.6%	
example (4c), 48 test subjects:	81.3%	16.7%	2.1%	
example (5a), 70 test subjects:	37.6%	52.9%	8.6%	
example (5b), 17 test subjects:	5.9%	82.4%	5.9%	5.9%
example (5c), 21 test subjects:	31.8%	68.2%		
example (6), 35 test subjects:	34.3%	65.7%		

Table 1. Interpretations of examples (4a)-(4c), (5a)-(5c), (6)

to a given clock or watch. However, this is not the case either, the interpretation of the example remains influenced by both choice of words and sentence order. When we replace 'aber' ('but') or 'und' ('and') with 'denn' ('because', example (5a)) or 'obwohl' ('although', example (5b)), the interpretations change significantly. Test subjects also grasp different interpretations when we change the order of the answer sentences, as in example (5c). (Cf. the second part of Table 1.)

(5) A: Wie spät ist es?

'A: What time is it?'

- a. B: Es ist 5 nach 3, denn meine Uhr geht 5 Minuten vor.'B: It's 5 past 3, because my watch is 5 minutes fast.'
- b. B: Es ist 5 nach 3, obwohl meine Uhr 5 Minuten vorgeht.'B: It's 5 past 3, although my watch is 5 minutes fast.'
- c. B: Meine Uhr geht 5 Minuten vor. Es ist 5 nach 3.'B: My watch is 5 minutes fast. It's 5 past 3.'

Furthermore, interpretation can be influenced by accentuation. When we place a very strong, contrastive accent on 'ist' ('is') in the original example, the majority of test subjects favour the literal interpretation. (Cf. the third part of Table 1.).

(6) A: Wie spät ist es?B: Es IST 5 nach 3, meine Uhr geht aber 5 Minuten vor.'A: What time is it?B: It IS 5 past 3, but my watch is 5 minutes fast.'

Let me take stock: firstly, examples (3) and (4) show that it is possible to apply an operation of meaning expansion in order to construct a modal context for interpretation. Test subjects strongly tend to interpret the answers of theses examples non-literally; their interpretation demands an operation of modal enrichment. In principle, the subjects could interpret the answers literally and infer that it is 10 past 3 by the speaker's watch: 'It is 5 past 3. The watch is five minutes fast. Therefore, it is 10 past 3 according to the watch.' This interpretation, however, seldom occurs, and it is not the interpretation intended by the speaker. Secondly, examples (5) and (6) show that modal enrichment depends on linguistic criteria like word choice, sentence order and stress. The interpretations of theses examples differ significantly from the interpretations of examples (3) and (4). When we perform tests³ for comparing the experimental data – for instance, the interpretations of (4c) vs. the interpretations of (5c) – we get p-values < 0.001 as results.

Why do the test subjects perform an operation of meaning enrichment in interpreting (3) and (4)? A pragmatic explanation of the phenomenon might go

^{3.} Fisher tests for exact data, two sided.

as follows: meaning enrichment is triggered by the Gricean maxim of relevance. The test subjects perform an operation of modal enrichment in order to 'make' B's entire answer relevant. The information that the watch is five minutes fast is irrelevant unless it has an effect on the determination of the time. It does not have such an effect if 'It's 5 past 3' is interpreted literally, but it has such an effect if 'It's 5 past 3' is interpreted literally, but it s 5 past 3'.

Against this explanation one might argue that A asks for the time and not for technical details about B's watch. B's answer contains irrelevant information about the watch irrespectively of whether the answer is interpreted literally or non-literally. Thus, modal enrichment does not make the entire answer relevant; relevance cannot be a reason for choosing the non-literal interpretation.

Two replies: firstly, it is true that A does not ask for information about the watch. However, it is only when B's answer is interpreted non-literally that this information becomes crucial – that is, *relevant* – for determining the time. Secondly, it is only due to the information about the watch that A comes to know what B takes to be the time, and how it is that he has come to his conclusion. A can therefore estimate B's credibility and switch from the belief '*B believes that it is 3 o'clock*' to the belief '*It is three o'clock*'. This makes the information about the watch interesting (and relevant).

Even in examples (5a), (5c) and (6), about one third of the test subjects grasped the non-literal meaning, which means that this interpretation is in principle available (although not preferred by the majority). When several interpretations are available, criteria for identifying the best interpretation(s) are needed. If 'relevance' only refers to the information given by the answer – if, for instance, 'relevant' means nothing more than 'giving only information necessary for answering the question under discussion' (cf. Groenendijk (1999), Schmitz (2008a)) –, then it is not a sufficent criterion for explaining why example (4c) (two separate answer sentences) is interpreted non-literally while example (5c) (the same sentences in a different order) is interpreted literally. It seems as, in order to be an adequate criterion, 'relevance' must refer to rhetorical structure and human reasoning capacities. Schmitz & Fisseni (2011) discuss the explication of an adequate relevance concept in further detail.

So far, I have dealt with only one specific modal operator (*by the speaker's watch*) and therefore with only one operation of modal enrichment. Can other kinds of modal contexts be created by modal enrichment as well, or is example (3) a singular example?

(7) a. A: Was glaubt Paul, wann er kommt?B: Er kommt um 3 Uhr, er kommt aber wie immer eine Stunde später als er glaubt.

'A: When does Paul believe that he will arrive?B: He will arrive at 3 o'clock, but as always he will be one hour later than he believes he will be'.

b. A: Du hast doch mit Peter gesprochen. Wann kommt er?B: Er kommt um 3 Uhr, er verspätet sich aber wie immer um eine Stunde.

'A: You have talked to Peter. When will he arrive?B: He will arrive at 3 o'clock, but as always he will be one hour late.'

I presented the dialogues (7a) and (7b) to test subjects, and asked them to write down when they expected Paul and Peter to be 'here' ('Wann, glauben Sie, wird Paul/ Peter hier sein? ').⁴ I assume that if a test subject expected Paul or Peter to be here at 15:00 then B's respective answer was interpreted literally by the subject. Conversely, if the test subject expected Paul or Peter to be here at 16:00, the answer was interpreted non-literally in the sense of 'Paul believes that he will arrive at 3 o'clock' or 'Peter said that he will arrive at 3 o'clock'. As with example (3), I assume that it can be infered from the test subjects' answers whether they performed an operation of modal enrichment or not. Table 2 shows the results of the experiment: about half of the subjects performed an operation of modal enrichment in interpreting example (7a), more than 2/3 of the subjects performed an operation of modal enrichment in interpreting example (7b).

	14:00	15:00	16:00	·;·
example (7a) (41 test subjects)	4.9%	43.9 %	51.2%	
example (7b) (43 test subjects)	2.3%	20.9%	72.1%	4.7%

 Table 2. Interpretations of examples (7)

The examples have two interesting features: firstly, they show that in principle modal enrichment is possible with doxastic operators (like in example (7a)) and with indirect-speech operators (like in example (7b)). Thus, examples (3) and (4) are not the only cases in which modal enrichment can take place.⁵ Secondly, the examples show that whether a recipient will perform an operation of modal

^{4.} The sets of test subjects were distinct.

^{5.} One might claim that B's response in example (7a) can be interpreted as a constituent answer to A's question and that therefore (7a) is not a proper example for modal enrichment. I doubt that - a constituent answer would be rather just '3 o'clock'. In any case, B's response in example (7b) cannot be interpreted as a constituent answer to A's question. Therefore, at least (7b) is an additional example for modal enrichment.

enrichment or not cannot be anticipated in all cases. In both examples (7), B cannot be certain that A grasps one specific meaning of his answer. B is probably misunderstood.

- (8) Ich habe mich mit Cécile unterhalten, die sich vorher mit Eric unterhalten hat.
 - Was hat Cécile Dir erzählt?
 - Eric glaubt, dass wir uns im Club Voltaire treffen, aber Cécile verwechselt den Club Voltaire immer mit dem Club Rousseau.

'I have spoken to Cécile, who previously spoke to Eric.

- What did Cécile tell you?
- Eric believes that we will meet in Club Voltaire, but Cécile always mixes up Club Voltaire with Club Rousseau.'

The interesting feature of example (8) is that it already contains a modal operator (*Eric believes that*). In two experiments, I asked test subjects to write down in which club Eric thinks 'we' will meet ('In welchem Club, glaubt Eric, treffen wir uns? '). If a subject assumed that Eric believes that we will meet in Club Voltaire⁶ then he interpreted the answer literally. Conversely, if the subject assumed that Eric believes that we will meet in Club Rousseau then he did not interpret the answer literally but in the sense of 'According to Cécile, Eric believes that we will meet in Club Voltaire'. (Cécile always mixes up Club Voltaire with Club Rousseau. Thus, Eric believes that we will meet in Club Rousseau.) The results of the experiments are shown in Table 3: in the first experiment, the majority of test subjects interpreted the answer non-literally. In the second experiment about one third of the subjects interpreted the answer non-literally. This proves that the non-literal interpretation is available although not generally preferred.

	Club Voltaire	Club Rousseau
first experiment (17 test subjects)	47.1%	52.9%
second experiment (42 test subjects)	69.0%	31.0%
both experiments (59 test subjects)	62.7%	37.3%

Table 3. Interpretations of example (8)

Marginal note: the subjects were told that it was a linguistic hypothesis – and not the subjects themselves – that was being tested. However, it is likely that the subjects misconceived the task of interpreting example (8) as a quiz. This may be

^{6.} A famous club in Frankfurt.

the reason that the subjects chose one of the clubs as an answer, rather than simply choosing to answer 'I do not know'. I assume that in a real world setting most people would ask for clarification. This, however, does not affect my interpretation of the data: if a modal enrichment operation was not available in the context of the example, then the subjects would not have to ask for clarification but would simply grasp the literal meaning.

Another marginal note: obviously, the results of the two Voltaire-experiments differ significantly. I do not have an explanation for this difference.

Let me take stock: recipients can perform operations of meaning enrichment. One of these operations is the creation of a modal context for interpretation (*by the speaker's watch, ...*). A cooperative speaker wants to be understood. He must anticipate possible meaning enrichments which might be carried out by the recipient. A recipient wants to understand the speaker. He can only carry out meaning enrichments that can be anticipated, or intended by the speaker. It must be clear in advance what kinds of meaning enrichment operations can be carried out. The set of enrichment operations that can in principle be carried out is restricted. We can take these operations into account when we construct meaning representations for utterances.

$$\{\lambda w \text{ [it's-5-past-3]}(w), \lambda w \text{ [by-watch}(\lambda w' \text{ [it's-5-past-3}(w')])}(w) \} \subseteq \\ \varphi \left(\begin{bmatrix} \text{NORMAL} & \lambda w \text{ [it's-5-past-3}(w)] \\ \text{ENR} & \begin{bmatrix} \text{MODAL} & \left[\text{OP} & \left\langle \lambda Q \lambda w \text{ [by-watch}(Q)(w)], ... \right\rangle \end{bmatrix} \end{bmatrix} \right)$$
(1)

In formula 1, an abridged feature structure is used to describe the meaning of the first conjunct of B's answer in example (2)/(3), here again repeated as example (9):

(9) A: Wie spät ist es?
B: Es ist 5 nach 3, meine Uhr geht aber 5 Minuten vor.
'A: What time is it?
B: It's 5 past 3, but my watch is 5 minutes fast.'

For reasons of readability, the structure includes lambda-expressions, but it can be easily transformed into a proper HPSG-like feature structure without lambda-expressions.⁷ The value of the NORMAL-feature is a representation of the literal meaning of 'It's 5 past 3'. This is the normal semantic part of the meaning representation. By the ENRichment-feature possible meaning enrichment operations are specified. This is the pragmatic part of the meaning representation. Here, only the

^{7.} A further example is given by Schmitz (2008b).

operation of modal enrichment with the available modal OPerators is named. Since modal enrichment is not declared to be obligatory, the meaning representation is underspecified. By a function φ , we can compute the set of fully specified meaning representations of 'It's 5 past 3'. This set contains 'It's 5 past 3' (λw [it's-5-past-3](w)) and 'By my watch it's 5 past 3' (λw [by-watch($\lambda w'$ [it's-5-past-3](w)). To filter out all but one of these representations – that is, to unambiguously identify the meaning intended by the speaker – additional pragmatic criteria are needed.

To sum up: operations of (pragmatic) modal enrichment can influence the truth conditions of a sentence. I specify the truth-conditional content of both literal and non-literal interpretations within the same framework. If one urges me to distinguish semantics and pragmatics within this framework,⁸ then I claim semantics to describe the literal meanings of sentences and pragmatics to describe possible modifications of these literal meanings. (This is of course not every-thing that is to say about pragmatics.) Both the semantics and the pragmatics of a sentence, as understood here, are conventionalized.

3. Blocking modal enrichment

It can be that in a given situation an operation of meaning enrichment is not obligatory but merely optional for the recipient. This is the case when several interpretations can be derived and additional pragmatic criteria do not suffice to filter out all but one of these interpretations (cf. the examples (7) and (8) above). In such a situation, a speaker cannot be sure whether the recipient will perform this operation; and the recipient cannot be sure whether he should perform it or not. In order to assure that he is properly understood, the speaker must control which enrichment operations are performed by the recipient. I claim that some words are exclusively used for controlling meaning enrichment. In particular, I claim that the German adverb 'tatsächlich' ('in fact') serves the singular purpose of controlling modal enrichment.

(10) A: Wie spät ist es?
B: Tatsächlich ist es 5 nach 3, meine Uhr geht aber 5 Minuten vor.
'A: What time is it?
B: In fact, it's 5 past 3, but my watch is 5 minutes fast.'

^{8.} An anonymous reviewer asked me to.

I asked 42 test subjects – the same subjects that also interpreted example (3) – to write down what time it is according to B's answer of example (10). Example (10) differs from example (3) only with respect to the occurrence of 'tatsächlich'. The results are given in the first part of Table 4: only two subjects performed an operation of modal enrichment and interpreted the answer non-literally. 40 subjects (95.2%) did not perform an enrichment operation but interpreted the answer literally. Remember that the experiment with example (3) (answer without 'tatsächlich') yielded converse results, the clear majority of subjects (85.7%) performed an operation of modal enrichment. The pooled data of both experiments are highly significant regarding a correlation of the test subjects' interpretations and the occurrence of 'tatsächlich'. The one-sided Fisher-test⁹ yields a p-value of < 0.001.

	15:00	15:05	15:10	·;·
example (3), 42 test subjects:	85.7%	14.3%		
example (10), 42 test subjects:	4.8%	95.2%		
example (11a), 17 test subjects:	82.4%	5.9%	5.9%	5.9%
example (11c), 17 test subjects:		94.1%	5.9%	
example (12a), 35 test subjects:	34.3%	65.7%		
example (12b), 35 test subjects:	5.7%	94.3%		

Table 4. Interpretations of examples (3) and (10), (11), (12)

(11) A: Wie spät ist es?'A: What time is it?'

- a. B: Es ist 5 nach DREI, meine Uhr geht aber 5 Minuten vor.'B: It's 5 past THREE, but my watch is 5 minutes fast.'
- b. B: TATSÄCHLICH ist es 5 nach 3, meine Uhr geht aber 5 Minuten vor.'B: IN FACT, it's 5 past 3, but my watch is 5 minutes fast.'
- c. B: Es ist tatsächlich 5 nach DREI, meine Uhr geht aber 5 Minuten vor.'B: In fact, it's 5 past THREE, but my watch is 5 minutes fast.'

Is the change of interpretations really an effect of 'tatsächlich', or is it a focus effect? The original example (3), repeated as (11a), was neutrally accentuated with nuclear stress on 'drei' ('three'). In example (10) (= (11b)), 'tatsächlich' carries the strongest accent. Therefore, the examples do not only differ with respect to the occurrence of 'tatsächlich' but also in their stress patterns. In order to test whether this difference has an effect on the test subjects' interpretations, I repeated the

^{9.} This test, like every other test mentioned here, was the Fisher test for exact data.

experiment with the examples (11a) and (11c).¹⁰ Example (11a) is identical to the original example (3). The answer of (11c) contains 'tatsächlich' but, unlike (10)/ (11b), I read the answer with the nuclear stress on 'drei', like in the original example.¹¹ The results are given in the second part of Table 4: the data are still highly significant regarding a correlation of the interpretations and the occurrence of 'tatsächlich'; the one-sided Fisher-test yields a p-value of <0.001. The data are not significant regarding a correlation of the interpretations and the different stress patterns (example (11b) vs. example (11c)).

- (12) A: Wie spät ist es?'A: What time is it?'
 - a. B: Es IST 5 nach 3, meine Uhr geht aber 5 Minuten vor.'B: It IS 5 past 3, but my watch is 5 minutes fast.'
 - b. B: Es ist tatsächlich 5 nach DREI, meine Uhr geht aber 5 Minuten vor.'B: In fact, it's 5 past THREE, but my watch is 5 minutes fast.'

In the previous section it was proved that a strong, contrastive accent on 'ist' ('is') in the original example deters test subjects from performing an operation of modal enrichment. (Cf. example (6), repeated as (12a).) I conducted another experiment in order to compare the accentuation effect with the effect of 'tatsächlich'. I asked 35 subjects to write down what time it is according to the answers given in examples (12a) and (12b). The results are given in the third part of Table 4: as can be seen, the effect of 'tatsächlich' is much stronger than the effect of contrastive stress. The two-sided Fisher-test to compare the interpretations of the examples, and therefore to compare the different effects, yields a p-value of <0.01. Thus, it is not by accident, that the subjects stronger prefer the literal interpretation when 'tatsächlich' occurs.

Interim stock-taking: due to the occurrence of 'tatsächlich', example (10) is interpreted literally.

As long as we define meaning in terms of truth-conditions, 'tatsächlich' does not affect the literal meaning of the example sentence. Literally understood, 'Es ist 5 nach 3' and 'Tatsächlich ist es 5 nach 3' have the same truth conditions. We can define a function [[]]_N for translating the German sentences into literal meaning representations. By this function the sentences with and without 'tatsächlich' can

^{10. 17} sujects took part in this experiment.

^{11.} The example-dialogues were not pre-recorded but read aload 'live' to the test subjects. Nevertheless, I can be assured that the examples were realized with exactly the stress patterns given in (11).

be translated into different representations (cf. formula 2 and 3). However, these representations have to be equivalent (cf. formula 4).¹²

	[[Es ist 5 nach 3]] _N = λw [it's-5-past-3(w)]	(2)
[[Tatsächlich ist es 5 nach 3]] $_N$	=	λw [FACT ($\lambda w'$ [it's-5-past-3(w')]) (w)]	(3)
λw [it's-5-past-3(w)]	≡	λw [FACT ($\lambda w'$ [it's-5-past-3(w')]) (w)]	(4)

'Tatsächlich' does not have direct truth-conditional content, but it blocks the creation of a modal interpretation context. This function can be easily modelled within an approach that makes use of feature structures as introduced in the previous section. I add an APPLication-feature and set its value on 'blocked'. OPerators that are blocked cannot be applied to the NORMAL meaning representation. Thus, modal enrichment is inhibited.

$$\lambda w [it's-5-past-3] (w) \in$$

$$\lambda w [by-watch(\lambda w'[it's-5-past-3(w')])(w)] \notin$$

$$\varphi \left(\begin{bmatrix} NORMAL & \lambda w [FACT(\lambda w'[it's-5-past-3(w'))](w)] \\ BNR & \left[MODAL & \left[OP & \left\langle \lambda Q\lambda w [by-watch(Q)(w)], ... \right\rangle \right] \\ APPL & blocked \end{bmatrix} \right] \right)$$
(5)

To sum up: in example (10), 'tatsächlich' blocks modal enrichment. Therefore, the answer sentence cannot be non-literally interpreted, in the sense of 'By the speaker's watch it's 5 past 3'. That does not mean that 'By the speaker's watch it's 5 past 3' is false. It just means that this is not the interpretation intended by the speaker.

In the previous section, examples for enrichments with other modal operators were given. Can these enrichments be blocked by 'tatsächlich' as well?

- (13)a. A: Was glaubt Paul, wann er kommt? B: Tatsächlich kommt er um 3 Uhr, er kommt aber wie immer eine Stunde später als er glaubt.
 - 'A: When does Paul believe that he will arrive?
 - B: In fact, he will arrive at 3 o'clock, but as always he will be one hour later than he believes he will be?

^{12.} Within Hybrid Logic the FACT-operator can be defined as an operator that points to the world given as argument. Cf. Blackburn et al. (2001). (Thanks to Bernhard Schröder for poiting that out to me.)

- b. A: Du hast doch mit Peter gesprochen. Wann kommt er?
 B: Tatsächlich kommt er um 3 Uhr, er verspätet sich aber wie immer um eine Stunde.
 - 'A: You have talked to Peter. When will he arrive?
 - B: In fact, he will arrive at 3 o'clock, but as always he will be one hour late.'

The same test subjects who interpreted the examples (7a) and (7b) were asked to interpret the same examples with 'tatsächlich' ((13a) and (13b)) and to answer the questions 'When, do you think, will Paul be here? ' and 'When, do you think, will Peter be here? ' ('Wann, glauben Sie, wird Paul/ Peter hier sein? '). The results are given in Table 5: as can be seen, 'tatsächlich' blocks modal enrichment effectively. One sided Fisher-tests for proving the correlation of the interpretations and the occurrence of 'tatsächlich' yield p-values of < 0.001. The data are therefore highly significant regarding this correlation.

Table 5. Interpretations of examples (13)

	14:00	15:00	16:00	' ?'
example (7a)	4.9%	43.9%	51.2%	
example (13a) (with 'tatsächlich'))		90.2%	9.8%	
example (7b)	2.3%	20.9%	72.1%	4.7%
example (13b) (with 'tatsächlich')		79.1%	20.9%	

- (14) Ich habe mich mit Cécile unterhalten, die sich vorher mit Eric unterhalten hat.
 - Was hat Cécile Dir erzählt?
 - Tatsächlich glaubt Eric, dass wir uns im Club Voltaire treffen, aber Cécile verwechselt den Club Voltaire immer mit dem Club Rousseau.

'I have spoken to Cécile, who previously spoke to Eric.

- What did Cécile tell you?
- In fact, Eric believes that we will meet in Club Voltaire, but Cécile always mixes up Club Voltaire with Club Rousseau.'

The same 51 subjects who interpreted example (8) also interpreted the example with 'tatsächlich', namely (14). Again, they were asked to write down the club in which Eric thinks 'we' will meet ('In welchem Club, glaubt Eric, treffen wir uns?'). The results are given in table (6): as expected, 'tatsächlich' blocks modal enrichment. The data are (highly) significant regarding a correlation of interpretations and the occurrence of 'tatsächlich'. The one sided Fisher-test yields a p-value of <0.01.

	Club Voltaire	Club Rousseau	
example (8)	62.7%	37.3%	
example (14) (with 'tatsächlich')	86.4%	13.6%	

Table 6.	Interpretations	of exampl	le (14)
----------	-----------------	-----------	---------

The examples (7) and (8) from the first sections are ambiguous with respect to whether the recipient should perform an operation of modal enrichment or not. The same examples with 'tatsächlich' ((13) and (14)) are unambiguous since the readings based on modal enrichment are not available.

Conclusion: the experimental data prove that 'tatsächlich' ('in fact') can be used for blocking modal enrichment. This pragmatic function can be easily modelled within an account that makes use of feature structures of the kind introduced above. We do not have to assume that 'tatsächlich' has direct truth-conditional content.

4. Discussion

I claim that 'tatsächlich' has a mere pragmatic function that can have a truthconditional effect (it influences how a recipient interprets an utterance). Let us discuss some objections.

Objection 1: the time-example with 'tatsächlich' is unnatural; no one would give an answer like in example (10). (This is what some native German speakers claim.) Therefore, the experimental data are not significant. We could improve the naturalness of the example by adding a word like 'übrigens' ('by the way'):

- (15) A: Wie spät ist es? 'A: What time is it? '
 - a. B: Tatsächlich ist es 5 nach 3. Meine Uhr geht übrigens 5 Minuten vor.'B: In fact, it's 5 past 3. By the way, my watch is 5 minutes fast.'
 - b. B: Es ist 5 nach 3. Meine Uhr geht übrigens 5 Minuten vor.'B: It's 5 past 3. By the way, my watch is 5 minutes fast.'

Reply: the original example (3) is a real-world example. I did not want to change it, and, for proving the effect of 'tatsächlich', I could not add 'übrigens' only to example (10). Otherwise, it would not excluded the possibility that enrichment blocking is an effect of 'übrigens', not of 'tatsächlich'. However, we can compare (15a) and (15b) and see whether the improvement of naturalness influences the interpretations. It does not matter that now the example without 'tatsächlich' might be less natural since the data for the original version are clear. I asked 43 subjects what time it is according

to B's answer.¹³ The results, which are given in Table 7, are as before: the data are highly significant regarding a correlation of interpretations and the occurrence of 'tatsächlich' (p-value < 0.001);¹⁴ 'tatsächlich' blocks modal enrichment.

	15:00	15:05	
example (15a) (with 'tatsächlich'):	9.3%	90.7%	
example (15b):	74.4%	25.6%	

 Table 7. Interpretations of examples (15)

Objection 2: if 'tatsächlich' is nothing more than a modal enrichment blocker, than the double occurrence in example (16a) does not have an effect on interpretation. This might be right. However, there are cases, like example (16b), in which a second occurrence of 'tatsächlich' does have an effect on interpretation.

- (16) a. Tatsächlich ist es tatsächlich 5 nach 3. 'In fact, it's in fact 5 past 3.'
 - b. Tatsächlich ist es nur tatsächlich 5 nach 3. 'In fact, it's only in fact 5 past 3.'

Reply: (16a) has the same meaning as example (10) (with only one occurrence of 'tatsächlich'). Example (16b) has three salient features: firstly, 'tatsächlich' occurs twice and deleting one of the occurrences can change interpretation. Secondly, one occurrence of 'tatsächlich' is focussed and associated with the focus operator 'nur' ('only'). Thirdly, the first occurrence of 'tatsächlich' that is not associated with 'nur' takes scope over the entire sentence including the focus association.

(17) Es ist nur tatsächlich 5 nach 3.'It's only in fact 5 past 3.'

Let us start with the focus association (example (17)): literally interpreted, (17) means that it is 5 past 3 and that all modally enriched versions of this proposition – like the proposition that it is 5 past 3 by the speaker's watch¹⁵ – are false (not only not intended by the speaker). In other words: no alternative of the neutral operator FACT (meaning representation of 'tatsächlich') can be applied to the

^{13.} In fact, I performed the experiment with a slightly different example, namely 'It's 10 past 6 but my watch is 10 minutes fast.' This difference, however, should not matter in any respect.

^{14.} One sided Fisher test for exact data.

^{15.} Let it be given that the *by-watch*-operator is available for modal enrichment.

proposition that it is 5 past 3 without making it false. The set of FACT-alternatives contains the OPerators that are in principle available for modal enrichment. In interpreting the association between 'nur' ('only') and 'tatsächlich' we need not assume that 'tatsächlich' has direct truth-conditional content: it suffices when some of its focus alternatives have. Operators like *by the speaker's watch* have direct truth-conditional content.

(18) A: Hast Du mit Peter gesprochen?B: Ja. Es ist nur tatsächlich 5 nach 3. Aber ich habe gesehen, dass es auch auf seiner Uhr 5 nach 3 ist.

'A: Did you talk to Peter?

B: Yes. It's only in fact 5 past 3. But I saw that also by his watch it is 5 past 3.¹⁶

There can be contexts in which the entire sentence (17) is to be modally enriched. In (18), the sentence is most probably interpreted as meaning that *according to Peter* it is only in fact 5 past $3.^{17}$ (Peter is lying.) In (18), 'tatsächlich' does not block the modal enrichment of the entire sentence since it does not take scope over the entire sentence.

In (16b), an additional 'tatsächlich' occurs that takes scope over the entire sentence. By this additional 'tatsächlich', modal enrichment of the entire sentence is blocked. Therefore, example (16b) cannot be interpreted in the sense of 'According to Peter it is only in fact 5 past 3' (contrary to (18)). This is the reason why a second occurrence of 'tatsächlich' can have an effect on interpretation in some cases, for instance when the other 'tatsächlich' is associated with a focus operator like 'only' ('nur').

The meaning of (16b) can be respresented by the feature structure 6: the NORMAL meaning representation consists of a FUNCtor and ARGument. The FUNCtor is an abridged meaning representation of 'nur' ('only'). The ARGument contains a structured meaning representation as its NORMAL meaning representation and a specification of ENRichment operations for 'It's 5 past 3'. 'Only' and the structured meaning are represented according to Krifka (1992).¹⁸

^{16.} This is quite clumsy in English, but close to the German original.

^{17.} Of course, it is better German to say 'Es *sei* nur tatsächlich 5 nach 3.' In colloquial German, however, (18) is acceptable.

^{18.} Focus associations like that of example (16b) can not only be implemented in accordance with the structured meanings framework but also in accordance with alternative semantics (cf. Rooth 1992).



Objection 3: if 'tatsächlich' is nothing more than an enrichment blocker then in embedded sentences it is also nothing more than an enrichment blocker. It only makes sense to assume that 'tatsächlich' is an enrichment blocker in embedded sentences if enrichment is possible in embedded sentences. This is dubious: we cannot interpret example (19) as meaning that Benedikt believes that it is 5 past 3 by his watch and that he therefore believes that it is in fact 3 o'clock.

Benedikt glaubt, es sei 5 nach 3, seine Uhr gehe aber 5 Minuten vor.
 'Benedikt believes that it is 5 past 3 but his watch is 5 minutes fast.'¹⁹

Reply: I doubt that modal enrichment is never possible in embedded sentences.²⁰ Even if it were impossible, we would not have to assume that 'tatsächlich' has another meaning in these contexts. Instead, 'tatsächlich' would just be superfluous.

(20) Benedikt glaubt, es sei tatsächlich 5 nach 3, seine Uhr gehe aber 5 Minuten vor.

'Benedikt believes that it is in fact 5 past 3 but his watch is 5 minutes fast.'²¹

1 1 1	, , ,		
	15:00	15:05	15:10
example (19): example (20) (with 'tatsächlich'):	22.9% 2.9 %	74.3% 97.1%	2.9

Table 8. Interpretations of examples (19) and (20)

^{19.} The scope of the belief-operator ('Benedikt believes that') reaches over the whole sentence: 'Benedikt believes that: it is 5 past 3 but his watch is 5 minutes fast.' The German sentence is unambiguous with respect to the scope of the belief-operator.

^{20.} At least, modal enrichment of coordinated sentences seems to be possible: to my intuition, 'Wenn es 5 nach 3 ist und meine Uhr 5 Minuten vor geht, dann ist es tatsächlich 3 Uhr' ('If it is 5 past 3 and my watch is 5 minutes fast, then it is in fact 3 o'clock') can be interpreted as being true. This well-meaning interpretation, however, demands modal enrichment of the antecedent.

^{21.} The belief-operator has the same scope as in (19).

I performed an experiment, in which 35 test subjects wrote down what time Benedikt thinks it is according to (19) and (20) ('Wie spät, glaubt Benedikt, ist es?'). The results are given in Table 8: only a minority of subjects interpreted example (19) (without 'tatsächlich') non-literally which supports the claim that modal enrichment does rather not occur in this particular context. However, even fewer subjects interpreted example (20) non-literally. The tendency to interpret (20) non-literally is significantly weaker than the tendency to interpret (19) non-literally; the one sided Fisher-test yields a p-value of < 0.01. Thus, 'tatsächlich' fulfils its blocking function in (20) although this might not be crucial for interpretation.

Objection 4: we should prefer a more linguistic explanation of the 'tatsächlich'effect. Syntactic approach: we can assume that when modal enrichment takes place, the modal operator (by speaker's watch etc.) is an unarticulated constituent that, although unarticulated, appears in syntactic structure.²² We assume that there is a syntactic slot for such a constituent. In the 'tatsächlich'-examples, modal enrichment cannot take place, because the slot is already occupied by 'tatsächlich'. Thus, 'tatsächlich' is not essentially an enrichment-blocker but an ordinary modal adverb that fills a syntactic slot; blocking of modal enrichment comes out as an epiphenomenon. Semantic approach: we can transfer the slot from syntax to semantics by interpreting 'tatsächlich' as a type-shifting operator. We assume that it's-5-past-3 is not an expression of type $\langle s, t \rangle$ but of type $\langle \tau, \langle s, t \rangle \rangle$, where τ is the type of a modal operator like FACT. To make *it's-5-past-5* a proposition (of type $\langle s, t \rangle$), it has to be applied to a modal operator. In case the sentence does not contain a modal operator, one has to be derived from context. If the sentence contains 'tatsächlich', λR $[R (\lambda w [it's-5-past-3(w)])]$ is simply applied to the neutral FACT-operator (the semantic representation of 'tatsächlich': $\lambda Q[Q]$). We derive the meaning of the entire sentence which is represented as λw [it's-5-past-3(w)]. Again, enrichment blocking comes out as an epiphenomenon.

Reply: there are at least two reasons why these approaches do not work. Firstly, expressions of different types can be used for creating modal interpretation contexts. For the syntactic approach we have to assume that these expressions belong to the same syntactic category and therefore can occupy the same slot. It is not convincing that they belong to the same syntactic category. Secondly, example (8) ('Eric believes that we will meet in Club Voltaire, ...') proves that modal enrichment can take place even when a modal operator is already present, that is, when the syntactic or semantic slot for such an operator is filled.

^{22.} Cf. the discussion of the famous 'It's raining'-example (Perry (1998)) to which I already referred in the Introduction (example (1a)).

Thus, the more 'linguistic' approaches make the wrong predicition regarding the interpretation of examples like (8).

I presuppose that the operator denoted by 'Eric believes that' is of the same type as FACT and as operators that are applied via modal enrichment. One could try to defend the semantic approach by denying this presupposition and distinguishing the types of articulated modal operators on the one hand and unarticulated operators as well as FACT on the other hand: *we-will-meet-in-club-voltaire* is of type $\langle \tau, \langle s, t \rangle \rangle$ and *eric-believes-that* is of type $\langle \langle \tau, \langle s, t \rangle \rangle, \langle \tau, \langle s, t \rangle \rangle$. Thus, *eric-believes-thatwe-will-meet-in-club-voltaire* is of type $\langle \tau, \langle s, t \rangle \rangle$. This expression can be transfered to a proposition of type $\langle s, t \rangle$ by modal enrichment or by the application to FACT.

However, this solution is not convincing either: firstly, the distinction between articulated operators on the hand and unarticuled operators and FACT on the other hand is not immediately plausible but needs further motivation. Secondly, contrary to the idea of objection 4, the meaning of 'tatächlich' is no longer an 'ordinary' modal operator. Thirdly, the approach cannot deal with examples like (20) in which 'tatsächlich' occurs within the scope of another expression that denotes a modal operator. In this example, the FACT-operator denoted by 'tatsächlich' changes the type of the embedded *it's-5-past-3* to $\langle s, t \rangle$, so that *benedikt-believes-that* – being of type $\langle \langle \tau, \langle s, t \rangle \rangle, \langle \tau, \langle s, t \rangle \rangle \rangle$ – can not be applied. The entire sentence becomes uninterpretable.

Objection 5 – another attempt to find a semantic explanation: instead of defining 'tatsächlich' as a pragmatic operator for blocking modal enrichment, we can define it as a semantic operator for neutralising other modal operators. For all modal operators *R* and expressions *p* of type $\langle s, t \rangle$, by definition:

$$R(FACT(p)) \equiv FACT(p) \equiv p \tag{7}$$

Reply: it is far from simple to define the neutraliser, and it is not even promising because the operator would neutralise not only modal operators that are applied via enrichment but also explicitly mentioned operators. Thus, the *Benedikt believes* operator in example (20) – 'Benedikt believes that it is in fact (tatsächlich) 5 past 3' – would be neutralised, and the example would be interpreted as meaning that it is 5 past 3. That is clearly inapproriate.

5. Conclusion

To conclude: experimental data prove that recipients can carry out modal enrichment operations. A modal enrichment operation is an expansion of a sentence's literal meaning with a modal operator like, for instance, *by the speaker's watch* or *Peter said that*. In the examples discussed here, these modal operators were derived from the contexts of the to be enriched sentences.

Operations of modal enrichment can be blocked. Experimental data prove that one means of effectively blocking modal enrichment in German is the word 'tatsächlich' ('in fact'). I argued that 'tatsächlich' does not have direct truth-conditional content; its only function is to block modal enrichment operations.

Finally, I argued that in general operations of meaning enrichment must be conventionalized. They can be defined within pragma-semantic feature strutures that represent the literal and non-literal (expanded) meanings of sentences. Within such feature structures, the meaning of 'tatsächlich' as an enrichment blocker that does not affect literal sentence meanings can be elegantly defined.

References

- Bach, K. 1994. Conversational impliciture. Mind and Language 9: 124–162.
- Bach, K. 2006. Conversational impliciture. Granada Workshop on Explicit Communication, in honor of Robyn Carston, May 31-June 2, 2006. http://userwww.sfsu.edu/ækbach/Bach. ImplExpl.pdf> (Sept 11, 2009).
- Blackburn, P., de Rijke, M. & Venema, Y. 2001. Modal Logic. Cambridge: CUP.
- Carston, R. 2002. *Thoughts and Utterances. The Pragmatics of Explicit Communication.* Oxford: Blackwell.
- Grice, H.P. 1989. Studies in the Way of Words. Cambridge MA: Havard University Press.
- Groenendijk, J. 1999. The logic of interrogation. Classical version. In *SALT* IX: *Semantics and Linguistic Theory*, T. Matthews & D.L. Strolovitch (eds), 109–126. Ithaca NY: CLC Publications.
- Krifka, M. 1992. Compositional semantics for multiple focus constructions. In *Informations-struktur und Grammatik, Linguistische Berichte,* Special Issue 4, J. Jacobs (ed.), 17–53. Opladen: Westdeutscher Verlag.
- Perry, J. 1998. Indexicals, contexts and unarticulated constituents. http://www-csli.stanford. edu/æjohn/PHILPAPERS/context.pdf> (Sep 11, 2009).
- Recanati, F. 2004. Literal Meaning. Cambridge: CUP.
- Rooth, M. 1992. A theory of focus interpretation. Natural Language Semantics 1: 75-116.
- Schmitz, H.-C. 2008a. Accentuation and Interpretation. Basingstoke: Palgrave Macmillan.
- Schmitz, H.-C. 2008b. 'Eigentlich' again. In *Proceedings of SuB12*, A. Grønn (ed.), 567–581. Oslo: ILOS.
- Schmitz, H.-C. & Fisseni, B. 2011. Repairs for reasoning. To appear in *Repairs. The virture of going wrong*, P. Brandt & E. Fuß (eds). Berlin: de Gruyter.

The hepatitis called ...

Electrophysiological evidence for enriched composition*

Petra B. Schumacher

University of Mainz and Max Planck Institute for Psycholinguistics, Nijmegen

This paper reports electrophysiological correlates of enriched composition, i.e. when certain aspects of the interpretation of an utterance must be constructed on the basis of information outside the syntactic and lexical representations associated with the utterance itself. It investigates the processes underlying reference transfer, a process by which a salient property of an individual may be used to refer to this individual. On the basis of the experimental data from reference transfer, but also from previous research on thematically-driven event structure updating triggered by certain inferential relations, it is proposed that enriched composition is discernible on the basis of a particular event-related brain potential (ERP) signature: a late positivity.

1. Introduction

The meaning of an utterance does not exclusively depend on the combination of the lexical meanings of its constituents; rather, other information sources and additional combinatory operations may contribute to the overall composition of a felicitous interpretation. In principle, this view is taken for granted in the semantic and pragmatic literature; for example in the research of Jackendoff (1997), but also Pustejovsky (1995), a variety of phenomena are subsumed under the labels of "compositionality" or "enriched composition" to refer to interpretive processes that go beyond the information provided by a lexical entry. In the present

^{*} This research was carried out at the University of Marburg as part of a research project on referential processing funded by the Deutsche Forschungsgemeinschaft (DFG) (BU 1853/2-1). The manuscript was completed during a stay at the Max Planck Institute for Psycholinguistics in Nijmegen. I would like to thank my committed student assistants Elisabeth Dietz and Jona Sassenhagen for their assistance during data preparation and collection, as well as the organizers of the DGfS workshop (and this volume) for creating a forum for discussing experimental semantics/pragmatics.

paper, the real time processing of enriched composition is investigated through "reference transfer", a process by which a salient property of an individual may be used to refer to this person as illustrated in (1).

(1) The nurse told the doctor that the hepatitis had called a few minutes ago.

Together with previous findings from event structure updating, this paper proposes that enriched composition is reflected in a specific electrophysiological pattern: a late positivity. This positivity is here considered to reflect processes associated with the modification and updating of discourse representation structures that are crucially induced by pragmatic requirements.

In the next section, the phenomenon of enriched composition is introduced through examples from coercion and reference transfer, and a further connection is established to pragmatically driven inferencing, which requires the enrichment of previously introduced discourse representation structures. Section 3 provides the relevant background information on the electrophysiology of language, before empirical data on the processing of reference transfer are presented in Section 4 from a reading comprehension study in German, during which the electrical brain activity was recorded, and from a follow-up questionnaire that assessed the acceptability and attitude towards reference transfer. The important role of enriched composition and the implications of the data for the semantics-pragmatics debate are discussed in Section 5.

2. Enriched composition

A relatively well-investigated phenomenon of enriched composition is coercion, where a specific function (2b/3b – adopted from Jackendoff 1997) must be applied to the nominal or verbal domain in order to reach compositional well-formedness. For instance, complement coercion requires additional type-shifting from an object to an activity (e.g. yielding *Sean enjoyed reading the book*. in (2)) to satisfy the verb's semantic restrictions on the arguments it takes (i.e. the verb selects a complement that denotes an activity or event). Similarly, in aspectual coercion an additional iterativity function must be applied to fulfill the temporal requirements of an utterance as in (3), where the verb in isolation implies no duration but combined with the prepositional phrase that evokes a durative reading, the only felicitous interpretation is one of iteration (e.g. *Michelle sneezed repeatedly for an hour.*).

(2)	a.	Sean enjoyed the book.
	b.	Interpret NP as [activity performed on NP]

(3) a. *Michelle sneezed for an hour.*b. Interpret VP as [repetition of VP]

Crucially, the interpretation must be enriched beyond the lexical-semantic information expressed by the individual constituents and their syntactic combination. Hence, these types of coercion - in addition to metaphors, masscount alternations, adjective-noun modifications, and others (see Pustejovsky 1995; Jackendoff 1997) - are generally considered to represent evidence for extra-syntactic composition. For experimental evidence for the extra-syntactic nature of coerced composition see McElree, Traxler, Pickering, Seely & Jackendoff (2001), Piñango, Zurif & Jackendoff (1999), Traxler, Pickering & McElree (2002), among others. However, the exact nature of the additional operations is discussed controversially, and enriched composition is either viewed as an instantiation of particular semantic rules or elaborate qualia structures (cf. e.g. Copestake & Briscoe 1995; Pustejovsky 1995) or as the contribution of pragmatic principles that are geared towards an efficient and cooperative conversation (cf. e.g. Asher & Lascarides 1995; Nunberg 1995; Egg 2004). Ultimately, an extreme position might not be tenable, and both semantics and pragmatics might be identified as contributors to enriched composition, so that different subgroups of enriched composition might be formed. For instance, complement coercion as in (2) relies heavily on the intrinsic properties of the verb and its selectional requirement for an activity-denoting expression; at the same time, the specific nature of the activity depends on contextual information (e.g. Sean enjoyed reading/writing/criticizing the book.). In contrast, aspectual coercion as in (3) is not immediately triggered by the intrinsic properties of the verb, but by the combination of the verb with a particular adverbial. Furthermore, metaphoric expressions serve pragmatic and rhetorical purposes. Thus, enriched composition might be dissociable into different subtypes (see also Jackendoff 1997 for similarities and differences between different instances of enriched composition). Importantly, psycholinguistic investigation may provide new insights into this debate, and as I suggest in the general discussion, predictions for this semantics-pragmatics distinction can be tested experimentally.

Moving on to the phenomenon under investigation in the present paper, enriched composition is not only required to meet selectional criteria or to resolve conflicts arising from lexical-semantic mismatches, but also to achieve discoursepragmatic well-formedness, as is for instance necessary in reference transfer, which is illustrated in Nunberg's (1979) famous example in (4):

- (4) a. The ham sandwich is sitting at Table 20.
 - b. Interpret NP as [person contextually associated with NP]

Reference transfer involves an operation by which the meaning of one entity of a particular ontological type is transferred to an entity of a different type. For instance, a salient property of an individual may be used to refer to this particular individual

(e.g. a person having ordered a ham sandwich can under certain circumstances be referred to as *the ham sandwich*). This indicates that the lexical-semantic meaning associated with the object-denoting expression *the ham sandwich* does not suffice to felicitously interpret the utterance. The respective reference transfer function is exemplified in (4b) (from Jackendoff 1997). Again, there are semantic and pragmatic accounts for this type of enriched composition. The former focus on the selectional mismatch between the noun phrase and the predicate, and the resulting type-shifting operation is characterized as affecting the qualia structure.¹ Pragmatic accounts consider reference transfer as a means to satisfy Gricean maxims (i.e. brevity and clarity (e.g. Egg 2004)), and mutually shared contextual information supports enriched composition.

In particular, three criteria have been highlighted as prerequisites for successful reference transfer: (i) the salience or noteworthyness of the property denoting an individual, (ii) a functional correspondence between the source and the intended referent, and (iii) contextual support (cf. Jackendoff 1997; Nunberg 1995; Ward 2004). The following example indicates that the salience of the property is an important premise. Imagine you want to refer to a specific person out of a group of Dutch soccer fans who all wear orange shirts and you say The orange shirt gave me a drink., then this utterance is by no means sufficiently informative and the identification of the intended referent fails. Likewise, referring to somebody as the nose would either crash or implicate that there is a person who has a striking nose that distinguishes him or her from everyone else. This indicates that successful reference transfer rests upon the noteworthyness and prominence of the respective feature. Yet, reference transfer is very often used in specialized situations, where a straightforward correspondence is available between for example an illness or the symptoms displayed and a patient in the context of a hospital, or a dish and a customer in a restaurant setting. Here, stereotypical knowledge about situations and individuals and contextual information also ease the meaning transfer. In fact, reference transfer is facilitated when the property is salient in a given contextual setting. Hence, reference transfer happens relatively easily when a doctor refers to her patient as *the kidney*

(i) Die Drei Tenöre stehen im obersten Regal. [German] The Three Tenors are on the top shelf.

^{1.} A nice example illustrating that the selectional mismatch does not necessarily represent the trigger for type shifting is provided by Egg (2004):

Even though there is no selectional conflict between noun phrase and predicate, reference transfer (from the group of individuals to some sort of music storage medium) is required on the basis of world knowledge (i.e. what kind of things can and cannot be on shelves).

stone. However, when a professor refers to one of her students in the same way for no apparent reason, this is awkward and unsubstantiated. Of course, if the student called in sick because she was having a kidney stone removed, this property can be promoted in salience and might be used to refer to this particular student. In this case, salience outranks situational support. However, if at all, it is more likely for a professor to refer to a student by their thesis topic or the theme of their presentation (e.g. *Are you the coercion?*).

In this paper, the comprehension of reference transfer is investigated in cases where a person can be associated with a property that is salient in a specific situation and enters into a clear functional correspondence with the intended referent. In addition, world knowledge also facilitates interpretation in these particular cases.

Apart from examining the processes underlying the transfer of reference, a connection is drawn to a comparable phenomenon, the establishment of certain inferential links during the integration of indirect anaphors. Although these two phenomena are subject to distinct interpretive operations, I claim that they have something fundamental in common, and this is that they depend on pragmatic enrichment. In his work on inferential bridging, Clark (1975) distinguishes between necessary, probable and inducible parts and roles, signifying the decreasing predictability of an indirectly referring entity, as exemplified in (5).

(5) Yesterday, a Ph.D. student was shot/killed/found dead downtown. The press reported that <u>the pistol</u> was probably from army stocks.

In all three situations, the integration of *the pistol* depends on an inference-based relation with a previously introduced event. While the integration of the pistol is relatively easy when it is a necessary instrument (in the shooting-event) - which presumably allows mapping onto an implicit argument role in the relevant event representation – it requires the modification of the event structure representation in the other two cases (such that the *killing*-event is upgraded to a *shooting*-event, and so on). Thus complex inferencing brings about an additional updating of the event structure - triggered by an operation that requires the establishment of a dependency between an argument and a previously introduced event - and the enrichment of the discourse representation, which is the basis for drawing a connection between inferencing and reference transfer. The difference between these two phenomena is that reference transfer requires an ontological shift of some sort (e.g. object to individual, individual to object, substance to container, ...), while in complex inferencing, an event representation must be modified towards a more specific event representation (e.g. adding an extra thematic role). Both phenomena demand the modification of discourse representations, which ultimately yields a richer, new sort of discourse entity, and they are contingent on context information.

3. Previous electrophysiological evidence

In the present research, event-related brain potentials (ERPs) were recorded to investigate the processes underlying reference transfer. ERPs are voltage fluctuations that reflect the spontaneous electrical activity of the brain, which arises in response to a sensory or cognitive stimulus. This activity is measured in a non-invasive manner by means of electrodes applied to the scalp and it is particularly informative about the time course of processing. In addition to this high temporal resolution (latency relative to onset of stimulus event), ERP components are characterized by their polarity (negative- or positive-going voltage change), amplitude (magnitude of response), and topography (maximum activity relative to electrode position). Crucially, ERPs are *relative* measures, i.e. an ERP effect reflects the comparison of a critical condition with a minimally differing control condition (in the present case, comparing the reference transfer). For a more detailed description of the ERP methodology and its impact on language research across different domains, see Kutas & van Petten (1994) and Kutas, van Petten & Kluender (2006).

One of the earliest observed language-related ERP components is the N400, a negative deflection peaking around 400 ms after stimulus onset, whose amplitude varies as a function of semantic integration difficulties. It has been shown that the amplitude of the N400 increases the less plausible the integration of a lexical item is (e.g. (6) from Kutas & Hillyard 1980). A comparable N400 effect has been observed for world knowledge driven implausibility (e.g. (7) from Hagoort, Hald, Bastiaansen & Petersson 2004). The findings generally indicate that the N400 reflects the degree of plausibility and predictability of a certain event within a specific sentential or discourse context (cf. Kutas & Federmeier 2000 for an overview).

- (6) The pizza was too hot to <u>eat/drink/cry</u>.
- (7) Amsterdam is a city that is very <u>old/new</u>.

In the research on referential processing, the N400 has been identified as a marker of dependency formation. The more difficult the establishment of a dependency is, the more enhanced is the amplitude of the N400. For instance, establishing an inference-based relation (e.g. (8a)) yields a more pronounced N400 than establishing a coreference relation (e.g. (8b) from Burkhardt 2006). In the latter case, an identity relation can be formed resulting in mere mapping onto a previously introduced discourse referent; in contrast, in the case of inference-based integration in (8a), inferential knowledge must be recruited to establish a linking relation between *concert* and *conductor*. In addition to enhanced demands during dependency formation (reflected in a pronounced N400), a later positive deflection emerged for the inferential relation over the coreference relation reflecting processing costs from the introduction of an independent discourse referent (Burkhardt 2006).

- (8) a. Tobias visited a concert in Berlin. He said that <u>the conductor</u> was rather impressive.
 - b. Tobias visited a conductor in Berlin. He said that <u>the conductor</u> was rather impressive.

Within the pragmatic domain, ERP research has been more than scarce. There have been investigations on language comprehension within a wider textual or situational context that have mostly yielded N400 effects and have considered later positive-going effects anomaly detections. For instance, mismatches between incoming information and prior discourse knowledge (9) elicited an N400, which was taken as evidence for early use of contextual information (cf. van Berkum, Zwitserlood, Hagoort & Brown 2003). This is in line with the observation that the N400 reflects predictability and plausibility considerations. Mismatches between assumptions about the speaker and the content of an utterance (10) have also registered an N400 (cf. van Berkum, van den Brink, Tesink, Kos & Hagoort 2008 - but see Lattner & Friederici 2003 for a late positivity in a similar study). The absence of an N400-effect in (11) was interpreted as a temporary "semantic illusion",² and the emergence of a later positivity was taken as a marker of the subsequent anomaly detection (cf. Nieuwland & van Berkum 2005 - but see the discussion below for an interpretation of these data in terms of reference transfer and enriched composition).

- (9) As agreed upon, Jane was to wake her sister and her brother at five o'clock in the morning. But the sister had already washed herself, and the brother had even got dressed. Jane told the brother that he was exceptionally <u>quick/slow</u>.
- (10) My favorite colors are <u>pink</u> and apple green. [uttered by an inconsistent male voice].
- (11) A tourist wanted to bring his huge suitcase onto the airplane. However, because the suitcase was so heavy, the woman behind the check-in counter decided to charge the tourist extra. In response, the tourist opened his suitcase and threw some stuff out. So now, the suitcase of the resourceful tourist weighed less than the maximum twenty kilos. Next the woman told the tourist/the suitcase that she thought he looked really trendy. [...].

To a large extent, these investigations have fallen short of a pragmatic account of the underlying interpretative processes. In contrast, the present line of research

^{2.} This term is derived from the "Moses illusion", an experimental design in which participants failed to detect the inconsistency in *How many animals of each sort did Moses put on the ark*? (Erickson & Mattson 1981).
intends to examine how pragmatic principles, such as pragmatic enrichment, are operationalized. A promising starting point for such an endeavor is a recent finding from the processing of different inferential relations (exemplified in (5) above), where the likelihood of an instrument noun phrase following a particular context sentence was not reflected in N400-modulations (as might very well have been predicted on the basis of the research on lexical-semantic integration difficulties), but rather in a late positivity peaking between 550-700 ms³ (Burkhardt 2007). The late positivity was observed at the critical NP as a twoway contrast between the two less probable context events (e.g. killing and finding *dead*) compared to the likely inference (e.g. *shooting – the pistol*). This late positivity is considered to reflect processing costs arising from the integration of an unexpected instrument role which demands the enrichment of previously introduced information. Accordingly, the discourse representation must be updated and an earlier established event must be modified towards a more specific event (i.e. the finding dead-event is updated towards a shooting-event). The processes underlying this enrichment of the discourse representation structure might then be similar to those hypothesized for enriched composition during reference transfer: the modification of a discourse representation for the ham sandwich towards a discourse referent denoting the person contextually associated with the ham sandwich involves similar updating mechanisms where an entity of a certain ontological type must be recoded towards an entity of another type, and this should thus also elicit processing costs associated with the enriching of discourse representation structures, i.e. a late positivity.

Additional support for a pragmatic interpretation of the late positivity comes from the processing of metaphors (12) in comparison to literally interpretable utterances.

(12) He knows that <u>whiskey/power</u> is a strong intoxicant.

^{3.} Even though this positive deflection peaks around 600 ms after stimulus onset, I refrain from using the label "P600" to avoid confusion with a syntactic explanation of the underlying processes. As should become apparent soon, the late positivity has a distinct functional contribution which can be tied to pragmatic composition and should not be conflated with other positivities. Moreover, non-syntactic positivities have been reported elsewhere, for instance in response to non-literal meaning composition during metaphor comprehension (Coulson & van Petten 2002) or as a marker for the introduction of new discourse referents (see discussion of Burkhardt 2006 above). They have also been registered in response to semantic reversal anomalies – as in *The cat that fled from the mice* … (from van Herten, Kolk & Chwilla 2005) – and interpreted as processing costs from monitoring and the reassessment of prior processing decisions. See Bornkessel-Schlesewsky & Schlesewsky 2008 and Burkhardt 2007 for more detailed discussion.

Metaphors registered a biphasic N400-late positivity pattern (Coulson & van Petten 2002). The N400 reflects processing difficulties during initial integration with prior knowledge (most likely guided by associative reasoning), and the late positivity is indicative of non-literal composition required for proper metaphor interpretation.

4. Empirical evidence for enriched composition during reference transfer

4.1 ERP study

Methods

Participants

Twenty-four students (12 female; 19–26 years old; mean age: 22.5 years) from the University of Marburg participated in this study. All participants were right-handed, monolingual native speakers of German and reported normal or corrected-to-normal visual acuity. Data from one participant had to be excluded from the ERP analysis due to excessive artifacts.

Materials

Forty mini-discourses were constructed that consisted of a context and a target sentence each. The context sentence set up a particular situation to license the reference transfer by mentioning two individuals who generally represent prominent roles within this situation (e.g. a doctor and an assistant in the doctor's office in (13) below). The context sentence further included an explicit question for an individual (wer (hat X gemacht) – "who (did X)") to allow the enriched composition to take place right at the NP in the subsequent (target) sentence. The target sentence represented the answer to the question and was presented in two versions with (a) an NP denoting a salient property of an individual and requiring reference transfer (13a) or (b) an NP that clearly referred to an individual (13b).

(13) a. Reference Transfer Condition

Der Arzt fragt seine Helferin erneut, | wer so früh angerufen hat. | Die Helferin | antwortet, | dass | **die Hepatitis** | so | früh | angerufen | hat. *The doctor asks his assistant again who had called that early. The assistant responds that the hepatitis had called that early.*

b. Control Condition

Der Arzt fragt seine Helferin erneut, | wer so früh angerufen hat. | Die Helferin | antwortet, | dass | **die Therapeutin** | so | früh | angerufen | hat.

The doctor asks his assistant again who had called that early. The assistant responds that the therapist had called that early.

The critical NPs in the two conditions were matched for syllable length and frequency of occurrence (based on the database of the Leipziger Wortschatz: http://wortschatz.uni-leipzig.de) – e.g. *Hepatitis* – *Therapeutin* "hepatitis – therapist", *Gitarre* – *Sängerin* "guitar – singer", *Plombe* – *Gipser* "inlay – plasterer", *Braten* – *Texter* "roast – copywriter". While critical NPs in the reference transfer condition were primed by the context sentence, this was not necessarily the case for the NPs in the control condition. (However, the absence of N400-modulations in the ERP data below illustrate that this had no effect on the present investigation.).

The 80 critical items were interspersed with an additional 120 filler items, out of which 40 contained the critical NP used in the reference transfer condition, but this time following an appropriate, non-transfer inducing context introduced by a *what*-question (14). The remaining 80 filler items followed the same context-target sentence pattern, but also included other types of embedded structures (*why, when,* ...).

(14) Filler item [match for (13)]

Der Arzt fragt seine Helferin erneut, | was so viele Menschen beunruhigt. | Die Helferin | antwortet, | dass | **die Hepatitis** | so | viele | Menschen | beunruhigt.

The doctor asks his assistant again what (it is that) concerns so many people. The assistant responds that **the hepatitis** concerns so many people.

Each of the 200 items were followed by a sentence verification question that probed information from either context or target sentence (e.g. *Fragt der Arzt, wer so früh angerufen hat?* – "Does the doctor ask who had called that early?" (expected answer: yes); *Antwortet der Arzt?* – "Does the doctor answer?" (expected answer: no)). Participants had to respond by pressing a 'yes' or 'no' button on a response box. 'Yes' and 'no' responses were distributed evenly across conditions and blocks, and 50% of the participants used their right hand to respond and 50% their left hand.

Procedure

Participants sat comfortably in front of a computer monitor and were instructed to silently read the mini-discourses for comprehension and to answer a comprehension question after each mini-discourse. All stimuli were presented visually in the center of the computer screen in yellow letters against a blue back-ground. Each trial began with the presentation of three asterisks (for 300 ms) to fixate the participant's eyes at the center of the screen. This was followed by an inter-stimulus interval (ISI) of 200 ms. To assure that the context sentence, which introduced the licensing situation, was read carefully, it was presented in two chunks (main and embedded sentence) and participants had to press a button to move on to the next chunk once they had completely read the current chunk (with

an ISI of 150 ms). The target sentence was then presented in a segmented manner with a predetermined presentation rate (NPs for 550 ms, all other elements for 450 ms, with an ISI of 150 ms each) and participants did not have to press a button to move on to the next segment. The segmentation is indicated by vertical bars in the example discourses above. After a blank screen of 500 ms, a verification question was presented on the computer monitor and participants had to press a 'yes' or 'no' button to indicate their response. Response times were restricted to 4000 ms. Following an intertrial interval of 1000 ms, the next trial started.

Each session started with a practice block during which participants were familiarized with the procedure. The experimental session, which consisted of 200 pseudo-randomized trials, was carried out in five blocks of 40 trials with short breaks between blocks.

Data Analysis

The electroencephalogram was recorded from 24 Ag/AgCl scalp electrodes mounted in an elastic cap (*EasyCap*). The ground electrode was placed at position C2 (according to the standard electrode position system reported in Jasper, 1958). Recordings were referenced to the left mastoid and rereferenced offline to linked mastoids. They were further filtered offline with a 0.3–20 Hz bandpass filter. In order to control for ocular artifacts, horizontal and vertical eye movements were monitored by means of two sets of electrode pairs, placed above and below the participant's left eye and at the outer canthus of each eye. Electrode impedances were kept below 5 k Ω . All channels were amplified using a *BrainVision Brain-Amp* amplifier and recorded with a digitization rate of 250 Hz.

Statistical analyses were performed on the behavioral measure and the ERP data. For the verification task, error rates and mean reaction times were computed with the factor NP Type (2 levels: reference transfer and control condition) and the random factors participants (F_1) and items (F_2) . Average ERPs were time-locked to the onset of the critical NP (in bold in the example sentences) and computed per condition and participant, before grand averages were calculated over all participants. Trials that registered an incorrect or timed-out response (i.e. 4000 ms after presentation of verification question) or that contained ocular, amplifier-saturation or other artifacts were discarded prior to averaging (21.86% of the data points). For the statistical analysis of the ERP data, a repeated measures analysis of variance (ANOVA) was performed with the factor NP TYPE (2 levels: reference transfer and control condition). The analysis was carried out separately for lateral and midline electrodes. The lateral analysis included the factor regions of interest (ROI) (with 4 levels: left anterior (comprising the following electrodes: F3/F7/FT7/FC5), right anterior (F4/F8/FT8/FC6), left posterior (C3/CP5/P3/P7), right posterior (C4/CP6/ P4/P8)). The midline analysis included the factor ELECTRODE (with 5 midline

electrodes as separate levels: FZ, FCZ, CZ, CPZ, PZ). All analyses were carried out in a hierarchical manner on the mean amplitude value per condition in two temporal windows ranging from 300–500 ms (N400) and 650–800 ms (late positivity).

Results: Behavioral Data

Participants performed at ceiling level in the sentence verification task (92.3% correct for Reference Transfer and 93.3% for Control condition), and mean reaction times did not differ reliably between conditions (2005 ms (SD: 303 ms) for Reference Transfer and 1912 ms (SD: 276 ms) for Control condition). Statistical analyses yielded no significant differences between these two conditions for both error rates and mean reaction times (all F's < 1).

Results: ERP Data

Figure 1 presents the grand-average ERPs for the two critical conditions and reveals an enhanced late positivity for the Reference Transfer (solid line) contrasted with the Control condition (dotted line), but no differences within the N400 window. This was confirmed by the repeated measures ANOVA. The analyses in the window from 300–500 ms post-onset revealed no reliable effects for both the lateral and the midline analysis [all F's < 1]. In the range between 650–800 ms, the analysis revealed a marginally significant effect of NP Type [F(1,22) = 4.06,p < .06], as well as an interaction of NP Type × ROI [F(3,66) = 6.38, p < .003] in the analysis over lateral regions. The resolution of this interaction registered main effects of NP Type over the two posterior ROIs (left posterior: [F(1,22) = 16.54], p < .001], right posterior: [F(1,22) = 4.30, p < .05]). The analysis over midline electrode sites registered a main effect of NP Type [F(1,22) = 9.08, p < .01]. In addition, the comparison between the Control condition and the Filler condition from (14) yielded no significant differences, which indicates that the observed effect cannot be attributed to NP inherent properties, but arises from the underlying interpretative mechanisms required for the full composition of meaning.

Discussion

The current ERP data show a late positive deflection for the processing of reference transfer between 650 and 800 ms after the onset of the critical NP that has a posterior distribution, but no N400-differences between the reference transfer and the control condition. Assuming that the N400 reflects difficulties arising from lexical-semantic integration and referential processing, the absence of such an effect indicates that the reference transfer inducing NP can initially be integrated as easily as the control NP. This might be facilitated by either the lead-in question in the context sentence, which explicitly introduces a placeholder in discourse representation (corresponding to "who") onto which both NPs can be mapped, or by the contextual support and facilitating priming provided by the scene that is



Figure 1. Grand average ERPs recorded to the onset of the critical NP (onset at vertical bar) at 9 selected electrode positions for Reference Transfer (solid line) and Control condition (dotted line). Time course is plotted on horizontal axis (from 0.2 seconds before the critical NP until 1.2 seconds after its onset); voltage fluctuations (in μ V) are charted on vertical axis, with negative voltage going upwards.

set up in the context sentence. More importantly, the late positivity for the reference transfer condition indicates that extra processing demands are exerted during the interpretation of the respective NPs that dissociates them from the NPs in the control condition. Together with the findings from NP integration in complex inferences (Burkhardt, 2007), this finding suggests that enriched composition is reflected in a late positive ERP component.

4.2 Norming study

Following the ERP recordings, the participants were asked to complete a questionnaire to assess their attitude towards reference transfer. This questionnaire first elicited acceptability ratings and then assessed the participants' attitude towards and use of reference transfer constructions.

Methods

Participants

The questionnaire was completed by the same 24 participants mentioned in 4.1.

Materials

40 mini-discourses were randomly selected from the material used in the ERP experiment (20 Reference Transfers and 20 matching Controls) and assigned to two lists with 10 Reference Transfer conditions and 10 Control conditions each. 10 further filler passages were constructed that consisted of two sentences that clearly violated general coherence requirements and should be rated as unacceptable passages (e.g. the following incoherent mini-discourse: *The cobbler asks his client whether she had watched TV last night. The innkeeper yells that the restaurant will be closed.*). If participants had rated these filler items as adequate, they would have been excluded from further analysis. However, none of the participants had to be discarded on the basis of this criterion.

Procedure

Participants were asked to rate 30 passages on a 6-point-scale using paper and pencil, where '1' signified that the mini-discourse was not acceptable (i.e. the participant or anyone else would never say something like this) and '6' indicated that the mini-discourse was very acceptable (i.e. the participant or somebody else would say something like this). Next, following a debriefing, during which participants read an explanation and justification of reference transfer, they were asked to write down how they liked this way of referring and whether they used or were exposed to reference transfer in everyday conversation.

Data Analysis

Mean acceptability ratings were computed over participants (F_1) and items (F_2) and analyzed using a repeated measures ANOVA with NP TYPE (reference transfer, control) as within-group factor. In addition, the participants' assessment of their own attitude towards reference transfer entered the analysis (by participants) as a between-group factor. The self-assessment was classified by two independent raters. Data from one participant had to be excluded because he had not completed the entire questionnaire.

Results

The statistical analysis of the acceptability ratings revealed a main effect of NP Type in both the participant $(F_1(1,22) = 39.58, p < .001)$ and item analysis $(F_2(1, 19) = 127.98, p < .001)$ with significantly poorer ratings for the Reference Transfer items (M(ean) = 3.37, SE = 0.30) over the Control condition (M = 5.34, SE = 0.14). Regarding the reported usage of reference transfer, 12 participants reported that they used or were exposed to reference transfer constructions, 9 participants indicated that they did not use them, and 2 additional participants did not provide a clear statement about their usage of reference transfer. Note also that a number of the participants who denied using reference transfer stated that they found this way of referring rude and impolite. In contrast, participants who showed a positive attitude towards reference transfer felt that it represented an efficient means of reference in certain situations. Additional analysis that took the reported reference transfer behavior into consideration (I employ reference transfer vs. I do not use reference transfer) revealed that there was a trend towards an interaction between the participants' reported preference to make use of reference transfer and their respective acceptability ratings (p < .07 for the 21 participants who indicated a clear attitude). This was reflected in higher mean ratings of the Reference Transfer items in the group of participants that reported to employ this way of referring themselves (M = 3.98, SE = 0.41, ranging from 2.2 – 5.8) compared to those participants who reported no exposure to and usage of Reference Transfer (M = 2.78, SE = 0.43, range: 1-4.6).



Figure 2. Mean acceptability ratings for passages involving Reference Transfer and Control passages. Dark bars indicate mean ratings of participants who reported use of and exposure to reference transfer ("positive attitude"), light bars reflect mean responses of participants who reported no use of reference transfer ("negative attitude").

Discussion

The behavioral data from the rating study show that reference transfer is more marked in comparison to more direct referential expressions. This is not surprising, given that reference transfer requires additional operations and is constrained by factors such as salience and contextual licensing. In addition, the participants' general attitude towards reference transfer affects their rating behavior. However, a post-hoc comparison of the ERP data on basis of the self-assessment revealed no group effect in the online patterns, indicating that there is a separation between the underlying online processing and conscious offline evaluation patterns.

5. General discussion

The main goal of the present research was to investigate the processes underlying reference transfer. The ERP data revealed a late positivity for the reference transfer compared to a baseline condition indicating that the additional operation required for the transfer of meaning from a property to an individual engenders processing costs. This positivity is interpreted as a marker of enriched composition. Crucially, such a positivity has also been observed in cases of thematically-driven enrichment, where the anchoring of an entity in discourse representation leads to event structure modifications (Burkhardt 2007). In this latter case, inference-based knowledge is primarily utilized to establish a dependency between an entity and information already given in discourse representation (i.e. anchoring). Hence while the respective utterance can be properly interpreted in and of itself (e.g. The pistol was from army stocks.), the discourse representation is updated in order to reach coherence and to determine the intended meaning on the basis of the wider context. Contrary to this, enriched composition in reference transfer is essential to arrive at a proper interpretation of the expression itself - i.e. an isolated utterance such as The hepatitis called very early. represents a semantic anomaly. Despite the differences between these two phenomena, the electrophysiological evidence indicates that processes of enriched composition that serve to satisfy discourse-pragmatic principles are manifested in a late positivity. Driven by contextual information, the initial discourse representation is adapted by integrating an unexpected thematic role into a previously established event representation (and therefore altering the make up of this discourse representation) or by performing a type shifting operation (which also yields a new type of discourse referent). Independent evidence has shown that enhancing discourse complexity by introducing new discourse referents also results in a late positivity (cf. Burkhardt 2006; Kaan, Dallas & Barkley 2007).

In addition, neither the current investigation nor the study of complex inferences has revealed differences in the N400-signature. In previous work, inference-based dependencies registered a more enhanced N400 in comparison to coreferential dependencies, reflecting increased processing demands during inference-based referential integration (Burkhardt 2006). This suggests that the type of dependency (coreference or inference) influences the processes underlying the N400, but that varying degrees of inferential strength do not differentially affect the initial access to discourse representation (Burkhardt 2007); similarly, the contextual information in the current investigation facilitates initial access equally in the conditions tested. This is an indication that referential dependencies and associations are established prior to enriched composition.

These data are furthermore fully compatible with a similar study conducted in Dutch that also found a late positivity, but viewed reference transfer as a semantic illusion and anomaly (Nieuwland & van Berkum 2005 – see (11) above). In this investigation, stories were constructed in which an individual with a prominent characteristic (e.g. *a tourist with a suitcase*) interacted with another individual (e.g. *a woman at the check-in counter*) and was later referred to by this characteristic (e.g. *the woman told the suitcase* ...). The experimental material thus satisfied the constraints on reference transfer (salience, noteworthyness, correspondence, contextual support), contra the authors' claim that their items represent "severe anomalies" (p. 698) that cannot be resolved through a type-shifting function as suggested here. From the perspective of enriched composition pursued in the present work, their findings of no N400 and a late positivity converge with the present data and add further strength to an account of pragmatic enrichment.

Moreover, it has been a matter of debate, whether such enrichment is based on lexical-semantics or whether it is pragmatic in nature or a combined effort involving both domains. While some cases of meaning transfer and polysemy could be accounted for on the basis of underspecified lexical representations, the cases of transfer of reference investigated here clearly demand an explanation that reaches beyond lexical-semantics. A lexicalization of these relatively specialized meanings would result in a far-fetched inflation of the lexicon because any entity could potentially serve as a trigger for reference transfer if embedded in a specific situational context. Adopting this view, the late positivity indexes the processing of enriched composition, which goes beyond lexical-semantic encoding – and critically does so not only in the case of reference transfer, where a salient property is used to refer to its possessor, but also in complex inferences that require the updating of previously established event representations or during metaphor comprehension (Burkhardt 2007; Coulson & van Petten 2002).

Finally, an alternative interpretation of the positivity, namely that it marks general well-formedness violations, can be discarded on the basis of the offline data and the comparison with complex inferencing. First, the questionnaire data showed that even though the reference transfer items are more marked relative to the baseline passages, they are not generally rated as unacceptable, and that the use of reference transfer is considered a valid means of reference by at least a subgroup of the participants. Hence, a categorical well-formedness violation must be abandoned and reference transfer should not be regarded as sheer deviation. In fact the questionnaire data suggest that reference transfer is "conventionalized" to a certain extent, i.e. type-shifted expressions are commonly used in everyday conversation - consider for instance the typical use of The string bass called. to refer to a member of an orchestra by his/her instrument or the use of Ann put the wine on the table. where a substance denotes a container (see also Lakoff & Johnson 1980 on conventionalized meaning). Furthermore, some participants reported that they considered this kind of referential use impolite. Note however that a post-hoc comparison of the ERP data on the basis of this evaluation did not reveal distinct ERP patterns and hence an impoliteness explanation can be discarded as well. Second, the inferences illustrated in (5) above - at least with respect to the use of probable instruments (e.g. killing - the pistol) - represent well-formed continuations, disallowing an overall well-formedness account of the late positivity.

The present investigation may thus serve as the starting point for a line of research on pragmatic composition in general and reference transfer in particular - two areas of research that still give rise to numerous questions. In the present study, reference transfer was supported by contextual licensing, a functional correspondence between the property and the denoting individual, and the salience of the property. Future research should determine the exact role of these different criteria. A follow-up study indicates for instance that in the absence of contextual licensing, initial integration processes are encumbered (reflected in an enhanced N400), but enriched composition is still taking place (reflected in a late positivity). This suggests that enriched composition is more automatic and independent of contextual support than is traditionally assumed. Subsequent research should therefore determine what kind of information facilitates the identification of referents and which knowledge is needed to trigger a transfer of meaning (but also other processes of enriched composition). Moreover, the present findings provide an initial indication that ERPs could possibly be used to differentiate semantic from pragmatic processes. One might speculate at this point that enriched composition can be divided into enrichment that is guided by lexical-semantics (e.g. complement coercions which is driven by the

verb's lexical-semantic restrictions – but as mentioned above is also contingent on context) and enrichment guided by pragmatics (e.g. reference transfer that emerges from combinatory requirements beyond lexically encoded information). On the basis of the current research, these two types of enrichment should give rise to N400 and late positivity effects respectively.

In sum, on the basis of empirical data the present investigation supports the view that the construction of meaning is not necessarily strictly lexically driven. Rather, the construction of a proper interpretation may require the introduction of semantic content that is not explicitly expressed by the constituents of an utterance, but enhanced through enriched composition. That the computation of extra-lexical meaning is taxing for the language processor is demonstrated by the pronounced late positivity observed during the processing of noun phrases that demand reference transfer. Together with other findings from language comprehension, this positivity is taken to reflect pragmatically driven processing. The findings therefore indicate that electrophysiological measures may be utilized to dissociate semantic from pragmatic processes. Since the early days of languagerelated ERP research, the N400 has been particularly susceptible to lexical-semantic information (including associative links). Extending this view to referential processing, the N400 mirrors the ease of dependency formation (cf. e.g. Burkhardt 2006). The late positivity reflects processing costs associated with the modification of discourse representation structures, which are motivated by pragmatic principles supporting coherence and efficiency. A somewhat similar division of labor is discussed in the sentence comprehension architecture of Bornkessel-Schlesewsky and Schlesewsky (2008), where lexical-semantic processing precedes "generalized mapping", which in the case of reference transfer reflects a conflict between the qualia structure of the noun phrase and the mechanisms linking the argument to its predicate. As I have tried to argue above, such a conflict is resolved on the basis of pragmatic knowledge.

References

- Asher, N. & Lascarides, A. 1995. Lexical disambiguation in a discourse context. *Journal of Semantics* 12: 69–108.
- Bornkessel-Schlesewsky, I. & Schlesewsky, M. 2008. An alternative perspective on 'semantic P600' effects in language comprehension. *Brain Research Reviews* 59: 55–73.
- Burkhardt, P. 2006. Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language* 98(2): 159–168.
- Burkhardt, P. 2007. The P600 reflects cost of new information in discourse memory. *Neuroreport* 18(17): 1851–1854.

- Clark, H.H. 1975. Bridging. In *Theoretical Issues in Natural Language Processing*, B. Nash-Webber & R. Schank (eds), 188–193. Cambridge MA.: Yale University Mathematical Society Sciences Board.
- Copestake, A. & Briscoe, T. 1995. Semi-productive polysemy and sense extension. *Journal of* Semantics 12: 15–67.
- Coulson, S. & van Petten, C. 2002. Conceptual integration and metaphor: An event-related potential study. *Memory and Cognition* 30(6): 958–968.
- Egg, M. 2004. Metonymie als Phänomen der Semantik-Pragmatik-Schnittstelle. *Metaphorik* 6: 36–53.
- Erickson, T.D. & Mattson, M.E. 1981. From words to meaning A semantic illusion. Journal of Verbal Learning and Verbal Behavior 20(5): 540–551.
- Hagoort, P., Hald, L., Bastiaansen, M. & Petersson, K.M. 2004. Integration of word meaning and world knowledge in language comprehension 304(5669): 438–441.
- Jackendoff, R.S. 1997. The Architecture of the Language Faculty. Cambridge MA: The MIT Press.
- Jasper, H.H. 1958. The ten twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology* 10(2): 371–375.
- Kaan, E., Dallas, A.C. & Barkley, C.M. 2007. Processing bare quantifiers in discourse. Brain Research 1146: 199–209.
- Kutas, M. & Federmeier, K.D. 2000. Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science* 4(12): 463–470.
- Kutas, M. & Hillyard, S.A. 1980. Reading between the lines: Event-related brain potentials during natural sentence processing. *Brain and Language* 11(2): 354–373.
- Kutas, M. & van Petten, C. 1994. Psycholinguistics electrified: Event-related brain potential investigations. In *Handbook of Psycholinguistics*, M.A. Gernsbacher (ed.), 83–143. San Diego: CA Academic Press.
- Kutas, M., van Petten, C. & Kluender, R. 2006. Psycholinguistics electrified II: 1994–2005. In Handbook of Psycholinguistics, M. Traxler & M.A. Gernsbacher (eds), 659–724. New York NY: Elsevier.
- Lakoff, G. & Johnson, M. 1980. Metaphors We Live by. Chicago IL: Chicago University Press.
- Lattner, S. & Friederici, A.D. 2003. Talker's voice and gender stereotype in human auditory sentence processing – Evidence from event-related brain potentials. *Neuroscience Letters* 339: 191–194.
- McElree, B., Traxler, M.J., Pickering, M.J., Seely, R.E., & Jackendoff, R. 2001. Reading Time Evidence for Enriched Composition. *Cognition* 78 (1): B17–B25.
- Nieuwland, M.S. & van Berkum, J.J.A. 2005. Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research* 24(3): 691–701.
- Nunberg, G. 1979. The non-uniqueness of semantic solutions: Polysemy. *Linguistic and Philosphy* 3: 143–184.
- Nunberg, G. 1995. Transfers of meaning. Journal of Semantics 12: 109-133.
- Piñango, M.M., Zurif, E. & Jackendoff, R. 1999. Real-time processing implications of enriched composition at the syntax-semantics interface. *Journal of Psycholinguistic Research* 28 (4): 395–414.
- Pustejovsky, J. 1995. The Generative Lexicon. Cambridge MA: The MIT Press.
- Traxler, M.J., Pickering, M.J. & McElree, B. 2002. Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language* 47(4): 530–547.

- van Berkum, J.J.A., van den Brink, D., Tesink, C.M.J.Y., Kos, M. & Hagoort, P. 2008. The neural integration of speaker and message. *Journal of Cognitive Neuroscience* 20(4): 580–591.
- van Berkum, J.J.A., Zwitserlood, P., Hagoort, P. & Brown, C.M. 2003. When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research* 17(3): 701–718.
- van Herten, M., Kolk, H.H. J. & Chwilla, D.J. 2005. An ERP study of P600 effects elicted by semantic anomalies. *Cognitive Brain Research* 22: 241–255.
- Ward, G. 2004. Equatives and deferred reference. Language 80(2): 262-289.

The role of QUD and focus on the scalar implicature of *most**

Arjen Zondervan University of Utrecht

Where previous studies supported the effect of the contextual property of Question Under Discussion (QUD) and focus on the scalar implicature of *or*, this paper presents two experiments that replicate this effect with the scalar term *most*. Both experiments show that, while story and target sentence are kept constant, more scalar implicatures are calculated when the scalar term is in the focus (new information) part of the sentence. In the experiments, the focus is manipulated by an explicit QUD. It is shown that the effect also holds for sentential answers to yes/no-questions, and might even extend to scalar implicatures in questions themselves.

1. Introduction

One of the few things researchers in pragmatics agree on is that a sentence containing a scalar term sometimes ends up being interpreted with a scalar implicature (SI). Actually, the crucial part of the previous sentence itself is a scalar implicature: *sometimes* implies *not always*. It is an issue of heated debate by which procedure the absence of SIs in some cases comes about: Some authors claim the SI is calculated and consequently canceled, while others state it is not calculated in the first place. However, leaving the procedural discussion aside, we know there are two types of factors that determine the presence or absence of an SI in a sentence with a scalar term: structural, sentence-level factors, and contextual factors.

A striking example of the first type is the intuitive absence (or reversal) of SIs in downwards entailing (DE) environments, such as the antecedent of a conditional (see e.g. Chierchia et al. 2001). Other sentence-level factors that affect SIs are embedding under certain quantifiers, modals, or factive verbs. However, although many authors acknowledge its existence (see e.g. Levinson 1983: 115–116), little is

^{*} I would like to thank the anonymous reviewers for very useful comments on an earlier version of this paper.

known about the second type of factor that can make or break SIs, factors on the contextual, bigger-than-sentence level. Usually, it is mentioned that relevance in the context is necessary, without a further specification of which factors make up relevance. In Zondervan (2007) and Zondervan (in press), I propose a crude definition of a contextual factor that affects SI-calculation, the QUD Focus Condition (QUDFC), given in (1):

QUD Focus Condition for Scalar Implicatures (QUDFC)
 An SI will arise in a sentence iff the scalar term (with which the SI is associated) is in a constituent that answers the QUD of the context that the sentence is part of, and therefore is part of the focus.

This condition builds on theoretical work on SIs by van Kuppevelt (1996) and van Rooij & Schulz (2004). Van Kuppevelt proposes a model in which discourse structure is organized by an ongoing questioning process. The *topic* of a discourse unit (a sentence or a number of sentences) is determined by the explicit or implicit question it answers. This answer provides the *comment*. Van Kupplevelt says (van Kuppevelt 1996: 396): ¹

By definition, a topic T_p is *that which is being questioned* by means of a contextually induced explicit or implicit question Q_p . The corresponding comment C_p is provided by answer A_p . C_p is *that which is asked for* by Q_p . [his italics]

The topic T_p is the intension of the topic term of the question, e.g. in (2), it is the intension of *(the one) who is laughing*.

(2) Q: Who is (the one who is) laughing?A: Alan is laughing.

The topic of (2) is the set of possible extensions of this term, so in a domain with only Alan and Brian, this is (3), where *S* stands for *Situation*:²

(3) $T_1 = \{\langle S_1, \{Alan\} \rangle, \langle S_2, \{Brian\} \rangle, \langle S_3, \{Alan, Brian\} \rangle \}$

The comment is the extension of the topic term in the actual situation, e.g. {Alan} for the answer in (2). If the answer uniquely determines the topic extension, the topic T_p is closed off, as the necessary condition for topichood, the underdetermination of the topic extension, is no longer met.

^{1.} The subscript p is, as far as I can tell, meaningless.

^{2.} Notice that this is not a propositional account of questions and answers like Hamblin (1973), where a question is the set of all its possible (propositional) answers, but an individualistic one, where a question is the set of all possible term answers (Van Kuppevelt refers to e.g. Hausser 1983).

So the *comment* part of the sentence corresponds to the part that is the (information) focus, based on the distinction of focus and presupposition made by Jackendoff (1972): focus is the information assumed by the speaker not to be shared by the hearer, i.e. it is the new information. Van Kuppevelt claims SIs only arise if the scalar term is in this part of the sentence. He illustrates with an example with numerals, which have traditionally been assumed to be lower bounded by the semantics (e.g. *four* means *at least four*), and upper bounded by an SI negating higher numbers.³ Consider (4):

Q1 : Who has fourteen children?
 A1 : <u>Nigel</u>_{Comment} has fourteen children.
 ⟨Q2⟩ : ⟨How many children does he have?⟩
 A2 : He has <u>twenty</u>_{Comment}.

Van Kuppevelt claims that the answer A1 gives rise to the implicit question Q2 asking for the exact number. This can only be the case if *fourteen* still gives rise to an indeterminacy, which is impossible if its interpretation were *exactly fourteen* (so the interpretation with SI), but possible if its interpretation is *at least fourteen* (without the SI). Therefore, Van Kuppevelt concludes that the latter (i.e. semantic) meaning has to be the right meaning here, that is, no SI is calculated for *fourteen* in A1. However, in A2, where *twenty* is part of the comment, subquestioning is no longer possible, indicating that the exactly meaning does arise there.

Also, according to van Kuppevelt, A1 in (4) does give rise to another implicature: the term *Nigel* gives rise to the implicature that he is the only one who has fourteen children. This is because *Nigel* has comment status here. So van Kuppevelt concludes that SIs only arise in the comment: 'In other words, one of our criteria for implicature generation, and thus for scale activation, is that the inducing context must have comment function.' (p. 407). As van Kuppevelt's notion of *comment* is actually identical to that of information focus, we can rephrase the predictions of van Kuppevelt in terms of focus, which is done in the QUDFC in (1).

Very similar predictions are made by the account of SIs of van Rooij (2002), van Rooij & Schulz (2004) and Schulz & van Rooij (2006).⁴ They propose (different versions of) a covert exhaustivity operator (*exh*) to account for the calculation of

^{3.} The claim that numerals are upper bounded by SI is heavily challenged nowadays (see e.g. Carston (1998), but Van Kuppevelt's approach applies to all scalar items.

^{4.} A view that is somewhat in the middle between Van Kuppevelt and Van Rooij and Schulz is the one by Scharten (1997), who adopts Van Kuppevelt's discourse framework but adds exhaustive interpretation to it. See Carston (1998) for discussion.

SIs, building on an exhaustivity operator proposed by Groenendijk and Stokhof (1984), which is given in (5):⁵

(5) $exh^{GS} = \lambda W\lambda P\lambda w[W(P)(w) \land \neg \exists P'[W(P)(w) \land P'(w) \neq P(w) \land \forall x[P'(x)(w) \rightarrow P(x)(w)]]]$

In this formula, W is the denotation of the term answer and P stands for the property underlying the wh-question. When exh^{GS} is applied to the answer to a question, it picks out the minimal elements of the set of sets of the answer. Consider for instance the example in (6):

(6) Q: Who came to the party?A: John came.

Assuming that we are in a domain with three individuals (say John, Bill and Mary), exh^{GS} picks out the set of sets {{j}} from the GQ-meaning of *John* (which corresponds to {{j},{j,m},{j,b},{j,b,m}}. This gives us the exhaustive interpretation that *only* John came to the party (of the three relevant people). Van Rooij shows that this mechanism can account for many SIs. If for instance the answer to the question in (6) contains a disjunction, the *not and* SI follows from applying *exh* to the answer. See (7):

(7) Q: Who came to the party?A: John or Bill came.

Here, *exh* picks out the set of sets {{j},{b}} from the GQ meaning of *John or Bill*, crucially excluding the set {j,b}. So the SI *it's not the case that both John and Bill came to the party* is derived directly by applying exh^{GS} to the answer.

However, *exh* only picks out the minimal set of sets of (the GQ-meaning) of the term answer, and it does nothing to the question predicate. So if the scalar term is in the question predicate, as in Q1 in (4) above, no SI arises. *Exh* will pick out the set of sets that contains only Nigel, leaving the semantic meaning of the question predicate (*has at least fourteen children*). So the exhaustivity account of van Rooij makes very similar predictions to the account of van Kuppevelt with respect to the relation of SIs and focus. These predictions are formulated in the QUDFC in (1).

^{5.} Van Rooij (2002), van Rooij & Schulz (2004) and Schulz & van Rooij (2006) take this exhaustivity operator as a starting point and propose several improvements. E.g. Van Rooij (2002) proposes the operator exhR which takes the relevance ordering (induced by the question) into account, van Rooij & Schulz (2004) propose eps1 and eps2, which can account for differences in epistemic force of SIs, and in Schulz & van Rooij (2006), a dynamic version of exh is introduced (exhdyn) to account for some data that were problematic for exhGS. These different operators make the same predictions for the simple cases that are discussed in this paper, so for ease of exposition I discuss the most basic exhaustivity operator, exhGS.

What is crucial about the theories of van Kuppevelt and van Rooij and Schulz, is that they both propose that every declarative sentence should be considered as an answer to a question. This question can be explicitly stated, but often it is implicit in the context. I will adopt the notion Question Under Discussion (QUD) for this contextual question (see e.g. Roberts 1996). There is a close correspondence between this QUD and the focus structure of a sentence: the focus part of a sentence corresponds to the questioned position of the question (Rooth's 1996 *Question-Answer Congruence for Focus*). This is why a sentence-level property like focus structure, is actually reflecting a contextual property, namely that of the QUD. Both van Kuppevelt and van Rooij & Schulz claim that the focus structure of the sentence, which in their theories is a crucial factor for SIs, depends on the question the sentence is an answer to.

The QUDFC makes the prediction that one and the same sentence will trigger an SI in one case, but not in another, depending on its focus structure. For instance, the scalar term *or* will trigger an SI in (8), but not in (9):

- (8) QUD: What/which things does John have?A: John has <u>A or B_F</u>
- (9) QUD: Who/which man has A or B?A: John_F has A or B.

In (8) and (9), the answer to the question is the same sentence, in which *or* is in a constituent in object position. However, only in (8) is the QUD questioning the object, making *A* or *B* the focus part of the answer. Therefore, the QUDFC predicts the SI (that John does not have A and B) only to arise in (8), and not in (9). In Zondervan (2007) and Zondervan (in press) I present a number of experiments that were set up to test these predictions. For instance in Experiment 2 of Zondervan (2007), 36 participants read items of two conditions of which examples (translated from Dutch) are given in (10) and (11).

(10) Condition 1: non-focus condition

Harry and Hermione were going to a party of Joost, a friend of theirs. Joost had asked both of them to bring bread or potato chips. He told them that he would be forever grateful to the one who brings bread or chips. Harry brought bread. He also brought chips. Hermione is usually very reliable, but this time she totally forgot to bring anything at all.

- A: "Who brought bread or chips?"
- B: "Harry brought bread or chips."

(11) Condition 2: focus condition

Harry is going to a party of a (female) friend of his. He promised to bring some snacks for the party. She was hoping he brought nuts, because she forgot to buy them. Harry brought bread. He also brought potato chips. He had considered bringing nuts, but he figured it wouldn't be necessary.

- A: "What did Harry bring?
- B: "Harry brought bread or chips."

In both conditions, Harry brought both bread and chips, but B's answer contains the scalar term or. Participants were told that B knew what happened, but A did not. They were asked to judge whether B's answer was true, relative to the story. The rationale behind the setup was that if participants would calculate the SI of or in B's answer (Harry did not bring bread and chips), they would judge the sentence untrue, while if they would not calculate the SI, the sentence would be considered true. I will discuss the possible shortcomings of this setup in the discussion of Experiment 1 below. The QUDFC predicts that since A's question (the explicit QUD) caused bread or chips to be in the focus only in Condition 2, this condition would receive more 'untrue' answers than Condition 1. This is indeed what happened: in the non-focus condition, SIs were calculated in 55% of the cases, while in the focus condition, the SI-rate went up to 73%, which was a significant difference. In another experiment, similar results were found when the two stories differed only in QUD, while everything else was kept constant over the two conditions. Two other experiments showed that the same results could be obtained when the QUD was not explicitly given, but implicit in the context. These experiments relied on other cues to mark focus, such as the focus sensitive operator only, and intonational cues. See Zondervan (in press) for a short overview of these experiments.⁶

The experiments discussed above all tested the QUDFC on the scalar term *or*, and with a wh-question as QUD. The goal of this work is to check whether the effect of QUD and focus on SIs extends to other SI-triggers, and whether it also applies if the QUD is a yes/no-question. Therefore, two experiments were conducted with a different scalar term: the quantifier *most*, which is normally associated with the SI *not all*. To check the applicability of the QUDFC to yes/no-questions, the experiments contained both wh-questions and yes/no-questions. Experiment 2 was a control experiment for Experiment 1, in which effects of the wording of the task were controlled for.

^{6.} An anonymous reviewer suggested to include some comments on how these data contribute to the debate on whether SIs are default inferences or context-dependent inferences. On the one hand, the difference between the two conditions shows that SIs definitely depend on the context, but on the other hand the observation that still a reasonable number of SIs is calculated in the non-focus case, suggests that the SIs might be more default than predicted by some contextualists' theories.

2 Experiment 1

2.1 Setup and items

The setup of the experiment was similar to that of Experiment 2 of Zondervan (2007) discussed above. Again, the Truth Value Judgment Task (TVJT) paradigm (Crain & Thornton, 1998) was adopted. Participants were presented with a story, followed by a dialogue between speaker A and speaker B, and were asked to judge speaker B's answer true or false. In the stories of the test items the stronger scalar item *all* held, while in the target sentence the weaker *most* was used. An 'untrue' answer therefore indicated that the participant calculated the SI. An example of a test item is given in (12). The experiment was conducted in Dutch, so (12) is a translation.

- (12) Five people were present at the drinks of the Celtic language studies program at the University. Several drinks were available.
 Sander is a student. He drank beer.
 Tom is also a student. He also drank beer.
 Eric is a professor. He drank wine.
 Martin is a professor too. He drank apple juice.
 Frans is a student. He drank beer.
 - A: "What did most students drink?"
 - B: "Most students drank beer."

In all stories, 5–7 people were introduced, either by name as in (12), numbered (*The first student… The second student*), or paired (*John… His wife…*). The people in the story were always divided into two groups (in this example students and professors). The QUD was always about the biggest group (in this example the students), which always consisted of 3–5 people. The other group was only introduced to make the QUD more natural, by introducing a contrast set for both the restrictor of *most* (in this case *students*), and for the object (*beer*).⁷ The story was the same in all conditions, but the following dialogue varied over conditions.

- (i) Three students came to the drinks.
 - John drank beer. Bill drank beer. Mary drank beer.
 - A: "What did most students drink?"
 - B: "Most students drank beer."

^{7.} If the other group had not been introduced, all items would have looked like (i):

This way, items would have become very similar and boring, and the critical inference would become quite obvious.

The dialogues of the two wh-conditions for the story in (12) are given in (13) and (14):

- (13) Condition 1 (wh non-focus)
 - A: "What did most students drink?"
 - B: "Most students drank beer.
- (14) Condition 2 (wh focus)
 - A: "How many of the students drank beer?"
 - B: "Most students drank beer."

Just like in the experiments with *or*, the same target sentence was used in both conditions, with different QUDs. Only the QUD in Condition 2 makes *most* the focus part of the answer, so the QUDFC predicts more SIs to arise there than in Condition 1. Condition 2 is slightly different from the focus condition in the experiments with *or*, as here it is the scalar *most* itself that is the focus, while in the *or*-experiments, it was the whole constituent *A or B*. This was unavoidable, as the question *Who drank beer*? (which would make the whole constituent *most students* the focus) would have made the target sentence *Most students drank beer* partly true: it was indeed the students that drank beer, not the professors, and partly false (because of the SI). To avoid confusion over this, the *How many* QUD was used.

The two conditions above were included to replicate the results found for *or* with *most*. However, another goal of this experiment was to test the QUDFC with yes/no-QUDs. Therefore, two more conditions were included, in which an explicit yes/no-question was asked by speaker A. An example of the dialogues of the first yes/no condition is given in (15):

- (15) Condition 3 (yes/no sentential answer)
 - A: "Did most students drink beer?"
 - B: "Most students drank beer."

The same target sentence was used as in Condition 1 and Condition 2, but here it is a sentential affirmative answer to the yes/no-question (an implicit 'yes'). This sentence itself has no part that contains new information (or in Jackendoff's 1972 terms, no information not shared by the hearer, in this case speaker A), so it is focus-less.^{8,9} This corresponds to the intuition that there is no specific part of the sentence that is questioned by the yes/no-question (see Section 4 for discussion).

^{8.} The only new information provided by the answerer is the implicit 'yes'.

^{9.} An anonymous reviewer pointed out that that (neutral) yes/no-questions can have so-called verum focus (with the truth values as the set of semantic alternatives).

As the whole sentence is non-focus, *most* is in a non-focus part and the QUDFC therefore predicts a lower SI-rate than in the focus wh-condition (Condition 2).

The second yes/no condition was included to test how *most* was interpreted in the yes/no-question itself. In this condition, only speaker A's question was given, and participants were asked to answer the question themselves by clicking 'yes' or 'no'. I will call this the *yes/no polar answer* condition. An example is given in (16):

- (16) Condition 4 (yes/no polar answer)
 - A: "Did most students drink beer?"
 - □ yes
 - 🛛 no

Although the QUDFC in its current shape makes no predictions about SIs in questions, it is usually assumed that SIs do not (or at least to a lesser extent) arise in questions (see Noveck et al. 2002), but I will return to this issue in the General Discussion. For now, we expect to find less SIs in this condition than in the focus declarative condition (Condition 2).

2.2 Design

Four lists were created, each contained two test items per condition, on a total of 16 stories per list. To avoid effects of particular lexical items in the stories, every list contained a different condition of that story. For instance (13) above was assigned to List 1, (14) to List 2, and so on. Every participant gave judgments on one list and lists were distributed evenly over participants. The fillers consisted of comparable stories and dialogues with target sentences with *most*. In the filler stories, *most* was true irrespective of SI (e.g. 2 out of 3, 3 out of 4, 4 out 5), or *most* was false (e.g. 1 out of 3, 2 out of 5, 0 out of 4). All QUD-types (wh about subject, wh about object, yes/no) were represented in the fillers in such a way that overall the QUD-types were evenly distributed over the stories. Test items and fillers were divided per list in a semi-randomized order with the following restrictions: there were never two consecutive test items of the same condition, never more than two consecutive test items, never more than two consecutive items with the same QUD-type, and the first test item was of a different condition for each of the four lists.

2.3 Participants and procedure

35 adult native speakers of Dutch were recruited via e-mail, and filled out a webbased experiment on their own computers. In the instructions, participants were told that they would read a story followed by a dialogue between speaker A, who did not know what happened in the story, and speaker B, who did know what happened. Speaker A would ask speaker B a question about the story and the participant would have to judge whether speaker B's answer was true or not, by clicking the 'true' or 'false' button, while the story and the dialogue remained on the screen. Participants were also told that after some stories, only a question was given by a speaker who did not know what happened in the story. In that case participants were asked to answer the question by clicking the 'yes' or 'no' button. With every story, participants could add comments in a textbox under the 'true'/'false' or 'yes'/'no' buttons. Participants were reminded that there were no 'right' or 'wrong' answers, and that they should follow their intuition. There was no time limit, but most participants completed the experiment within 15 minutes, with an average of 11 minutes.

A number of participants was excluded because of an unavoidable problem with the test items of Condition 2, repeated here for convenience:

- (14) Condition 2 (wh focus)
 - A: "How many of the students drank beer?"
 - B: "Most students drank beer."

Some participants seemed to have judged this answer 'untrue' not because of an SI, but because they felt the speaker should have provided the exact number of students that drank beer (in this case 'three') as an answer, instead of a proportion. This problem could however be checked for by looking at the fillers and participants' comments. If participants ruled out (14) because no exact number was provided, they should have also done so for the fillers in which *most* was true irrespective of SI (e.g. a situation with 4 out of 5). Therefore, participants were excluded if they judged one or more of these true *most* fillers with a *How many* QUD false, or if they commented in the textbox that an exact number was asked (or both). Five participants were excluded based on these criteria. Three participants commented about an exact number in the textbox, but answered 'true' to the true *most* fillers with a *How many* QUD, so they were kept in. No participants were excluded based on error rates on the fillers, as all of the participants' accuracy rate on the fillers was at least 75%. After exclusion the judgments of the remaining 30 participants were analyzed.

2.4 Results and discussion

Results of the wh-QUD conditions

The percentages of 'false' answers, signaling SIs, are given in Table 1 for the two wh-QUD conditions.

Condition 1 (wh non-focus)	Condition 2 (wh focus)	
42%	52%	

Tabl	e 1.	SI-rates	for w	h-QUD	conditions	of]	Experiment	1
------	------	----------	-------	-------	------------	------	------------	---

Statistical analysis of the data reveals a significant difference between the SI-rates of C1 and C2 (F(1,29) = 5.12, p = 0.031), just like was found in the experiments with or.¹⁰ If we look at the absolute numbers, the difference between the focus wh-condition and the non-focus wh-condition appears smaller than the difference we found for *or*. However, if we consider the distribution of participants based on their relative behavior on Condition 1 (non-focus) and Condition 2 (focus), given in Table 2 below, we see a clear pattern.

	,		1
more SIs on focus	SIs on both	SIs on neither	more SIs on non-foc
23% (7)	27% (8)	47% (14)	3% (1)

Table 2. Distribution of subjects based on behavior on wh-QUD conditions of Exp. 1

From this table it is immediately clear that more participants made a difference between the two conditions in the direction that was predicted by the QUDFC (7 participants calculated more SIs in the focus condition than the non-focus condition), than in the opposite direction (only one participant calculated more SIs in the non-focus condition). However, we see that a big group of participants (74%) makes no distinction between the two conditions and calculates SIs in both or neither of the conditions. This group is even bigger than in the experiments with *or*, in which it was typically around 50%.

As I already discussed in Zondervan (in press), the fact that many participants make no distinction between the two conditions, could be due to a problem with the used paradigm, the TVJT. In this paradigm, the participant already knows what happened in the story when he is asked to interpret the sentence with the scalar term. This is contrary to the normal situation in which an SI arises, where the hearer does not know what happened and draws an SI based on the speaker's utterance that contains a weaker scalar term. In that spontaneous situation, there is no discrepancy between what happened and the used term, because the hearer does not know what happened. In the TVJT however, the participant is asked to reconstruct how she would have interpreted the sentence (including drawing the SI or not) *if she would not have known what had happened*. It is very unlikely that a participant will do this. Much more likely is that a participant will consider the fit of the target sentence to the story. As the critical items all contain a weaker

^{10.} I compared C1 and C2 for easy comparison to the experiments with or, but in line with the predictions made, C2 should be compared to the pooled data of C1, C3 and C4. This difference is also significant: F(1,29) = 7.43, p = 0.011. This also avoids correction for multiple comparisons.

scalar term (*most* or *or*), where a stronger one would have also been appropriate (*all* or *and*), there is a discrepancy between sentence and story in all the test items. Therefore, participants might rule out the target sentence based only on the use of an underinformative item, without having actually calculated the SI. So what the TVJT is actually measuring might not be the presence or absence of an SI in a certain condition, but the acceptability of a weaker form in a certain condition. Of course one can reason that in environments that typically trigger SIs, weaker forms will be less acceptable than in non-SI environments, but this relation is rather indirect.

Furthermore, this discrepancy between target sentence and story is present in both conditions. Even if participants feel that the weaker form is more acceptable in one of the two conditions, they will try to answer *consistently* throughout the experiment, so as soon as they notice the discrepancy, they will either consistently reject or consistently accept the target sentence. The fact that we find even more participants exhibiting this behavior in this experiment than in previous experiments with *or*, might be caused by the smaller item-filler ratio of this experiment (1:1), which was the result of having more conditions in a short experiment. However, despite the fact that this strategy of judging all marked items the same seems to be dominant, a significant proportion of participants (23%) still differentiate between the conditions, drawing more SIs in the focus condition, confirming that the effect of QUD and focus on SIs is also present with the scalar term *most*.

Results of the yes/no-QUD conditions

The percentages of 'false' answers to the yes/no sentential condition and 'no' answers to the yes/no polar condition are given in Table 3.

Condition 3 (yes/no sentential)	Condition 4 (yes/no polar)	
42%	40%	

Table 3. SI-rates for yes/no conditions of Experiment 1

It is clear that both yes/no conditions pattern with the non-focus wh-condition, rather than with the focus wh-condition. The SI-rate on the yes/ no sentential condition seems to confirm that the QUDFC does indeed extend to yes/no-questions. After a yes/no-question, an affirmative sentential answer is focus-less, therefore *most* is not in the focus constituent. Consequently, we find the same SI-rate we found after a wh-question in which another part of the

sentence was questioned. The SI-rate in the yes/no polar condition, representing the SI-rate in the yes/no-question itself, is also lower than in the focus declarative condition, as we predicted based on previous research on SIs in questions. I will return to these results more elaborately in the General Discussion. First I will present a control experiment that was conducted to investigate the effects of the wording of the task.

3. Experiment 2

One of the critical properties of the TVJT that is often overlooked is the question what exactly participants were asked to judge the target sentence on. The TVJT is widely used in acquisition, where the experimenter uses a puppet to utter the target sentences, and asks the child to judge the puppet's behavior. Usually, children are asked whether the puppet 'said it well', or whether what it said was 'right or wrong'. In the experiments presented above, I explicitly asked participants to judge whether the sentence was 'true' or 'false', a question often asked in versions of the TVJT designed for adults. This choice was partly based on the problem I sketched in Section 2.4., that there is always a discrepancy between story and target sentence, as a weaker item is used instead of a more informative item. When asked for 'right' or 'wrong' instead of 'true' or 'false', participants might be even more likely to reject the target sentence based on this discrepancy alone, instead of based on the interpretation of the target sentence. However, when considering the effect of QUD on the target sentence, asking for truth might also have a downside. When asked whether a sentence is true or false, participants could be more likely to ignore the question the sentence is an answer to, in an effort to be more accurate about the truth of the sentence itself. They might think they are asked to judge the sentence itself, and ignore the fit to the question. This is of course undesirable if we are interested in the effects of the question on the interpretation of the sentence. Experiment 2 was set up to control for this. The experiment was an exact copy of Experiment 1, with the only difference that instead of judging 'true' or 'false', participants were asked to judge the answer 'right' or 'wrong'.

3.1 Setup, items, design and procedure

The setup, items and design were exactly the same as Experiment 1. The procedure only differed from Experiment 1 in that participants were instructed to judge speaker B's answer 'right' or 'wrong' instead of 'true' or 'false'. The 'true' and 'false' buttons were replaced by 'right' and 'wrong' buttons.

3.2 Participants

43 participants were recruited by email. None of them had participated in Experiment 1. Again, participants who were suspected of answering 'wrong' to *How many* QUDs, based on the expectation of an exact number instead of a proportion, were excluded on the same criteria as in Experiment 1. A total of 8 subjects were excluded based on these criteria. The fact that this number is slightly higher than in Experiment 1 (5 subjects) is not surprising in the light of what was suggested above: when asked for 'right' or 'wrong', subjects are more likely to reject the sentence based on a bad fit to the question, than when asked for 'true' or 'false'. Two more subjects were excluded because their accuracy scores on the fillers was lower than 75%. Again, most participants completed the experiment within 15 minutes. After exclusion, the judgments of the remaining 33 participants were analyzed.

3.3 Results and discussion

The SI-rates on the four conditions are summarized in Table 4.

Condition 1	Condition 2	Condition 3	Condition 4
(wh non-focus)	(wh focus)	(yes/no sentential)	(yes/no polar)
48%	64%	44%	38%

Table 4. SI-rates for the four conditions of Experiment 2

The difference in SI-rate between the two wh-conditions is again significant (F(1,32) = 6,48 p = 0.016).¹¹ So just like in Experiment 1, more SIs are calculated when the scalar term is in a focus constituent. The distribution of participants on the wh-QUD conditions is summarized in Table 5.

Table 5. Distribution of subjects based on behavior on wh-QUD conditions of Exp. 2

more SIs on focus	SIs on both	SIs on neither	more SIs on non-foc
24% (8)	39% (13)	33% (11)	3% (1)

This distribution reveals the same pattern as in Experiment 1: More participants make a distinction between the two conditions in the direction that the

^{11.} Again, also the difference between C2 and the pooled data of C1, C3 and C4 is significant: F(1,32) = 10.66, p = 0.003. (See ftn. 3)

QUDFC predicts (8 participants) than the other way around (1 participant), and the majority of the participants (72%) is consistent over conditions. Overall, SI-rates are slightly higher than in Experiment 1 (with the exception of the yes/no polar condition), but a between-subjects analysis revealed no significant difference between the two experiments (p = 0.626). We can conclude that Experiment 2 replicated the results of Experiment 1 with a different version of the TVJT.

4. General Discussion

Experiment 1 replicated the effects of QUD and focus found in earlier experiments with or for another scalar term, most. In a TVJT in which the story and the target sentence was kept constant over conditions, manipulation of the explicit wh-QUD had an effect on SI-rate. More SIs were calculated when the scalar term was in the focus, than when it was not. The data do not support the strong version of the QUDFC in (1), as a considerable percentage of SIs was calculated in the non-focus condition. However, this might be due to the experimental paradigm that was used, which in combination with the tested inference and the relative small number of filler items, was vulnerable to a strategy of judging items of both conditions the same. Still, a significant proportion of the participants distinguished between the two conditions in the direction the QUDFC predicts, and hardly any participants the other way around. The results of the yes/no sentential condition showed that if the same target sentence follows a yes/no-question, the SI-rates are similar to those of the non-focus wh-condition rather than to the focus wh-condition. This is in accordance with the view that an affirmative sentential answer to a yes/no-question is as a whole focus-less. Finally, the SI-rate observed in the yes/no-questions itself also patterned with the non-focus whcondition, and not with the focus wh-condition, confirming earlier studies that indicate SIs are less likely to arise in questions.

Experiment 2 controlled for a possible effect of wording of the task of Experiment 1. The results did not reveal a significant difference between the two ways of wording the task: judging 'true' or 'false' vs. judging 'right' or 'wrong'. The results of Experiment 1 were replicated, further strengthening the claim that QUD and focus have an effect on SIs.

One aspect of the comparison between wh-questions and yes/no-questions is worth elaborating on. Above I stated that an affirmative sentential answer to a yes/no-question is as a whole focus-less. This conclusion might be a bit too hasty. Jackendoff (1972: CH6) claims that questions themselves also have a focus structure. For wh-questions, this is relatively simple: the wh-phrase is the focus, and the rest of the sentence is non-focus (Jackendoff calls this *presupposition*, which is not to be confused with the use of that word in modern pragmatics). This follows straightforwardly from the generalization that was already referred to in the introduction: question and answer share the same focus structure (see e.g. Rooth 1996). However, with yes/no-questions, things are a bit more complicated. According to Scharten (1997), there are two types of yes/no-questions: neutral and topicalized. Neutral yes/no-questions are the type of questions in which no part has any special intonation, and which cannot be rephrased as wh-questions. An example of a neutral yes/no-question is (17):¹²

(17) Did you clean up your room?

Topicalized yes/no-questions on the other hand, are actually wh-questions in disguise. A part of the question is intonationally marked, and that part corresponds to the wh-phrase in the corresponding wh-question. Scharten gives example (18), and example (19) is from Jackendoff (1972):

- (18) Did JOHN take your books?≈ Who took your books?
- (19) Did Maxwell kill the judge with a HAMMER?≈ What did Maxwell kill the judge with?

These topicalized yes/no-questions clearly have a focus structure, namely the same as their wh-equivalents. In (18), *John* is the focus, and in (19) *hammer*. It therefore seems reasonable to assume that a sentential answer to a topicalized yes/no-question, just like an answer to a wh-question, inherits the focus structure of the question, e.g. as in (20):

- (20) A: "Did Maxwell kill the judge with a HAMMER?"
 - B: "(Yes,) he killed him with <u>a hammer</u>"

If this is the case, we can no longer be sure that the sentential answer to the yes/ no QUD in Experiment 1 and 2 was as a whole focus-less. As the dialogues were presented in written form, participants could have superimposed all kinds of intonational patterns on the yes/no-question, possibly turning it into a topicalized yes/ no-question, as in (21a–d):

- (21) a. Did most students drink BEER?
 - b. Did MOST students drink beer?
 - c. Did most STUDENTS drink beer?
 - d. Did most students DRINK beer?

^{12.} Of course, intonation can change this neutral yes/no-question into a topicalized yes/ no-question. But with normal intonation, this question is a good example of a neutral yes/ no-question.

This is problematic, because now we can no longer assume *most* was non-focus in the answer. If for instance participants read the question as in (21b), this would have caused *most* in the answer to be the focus, and the QUDFC would predict an SI in that situation. As all target sentences in the experiment (test items and fillers) contained *most*, it is not even unlikely that participants read the question with extra emphasis on *most*, so as a topicalized yes/no-question. The only way to get rid of this problem is to set up a control experiment in which intonation disambiguates the yes/no-question, which I will for now leave as a suggestion for further research.

If we take the claim that topicalized yes/no-questions have a focus structure seriously, this leads to another interesting question: Does the QUDFC hold for questions? In other words: Will more SIs arise in questions if a scalar term is in the focus? There is no way to investigate this question with wh-questions, as there the scalar term is always non-focus (unless it is part of the wh-phrase), but consider for instance the paradigm with topicalized yes/no-questions in (22):¹³

 (22) Story: John has a and b.
 Non-focus condition: Does JOHN have a or b? yes/no focus condition: Does John have A OR B? yes/no

If the QUDFC extends to questions, more SIs should be observed in the second condition, and it should be answered negatively more often.

Summarizing the discussion about the focus structure of yes/no-questions: adopting the idea of topicalized yes/no-questions leads to interesting new predictions about the presence of SIs in questions. If these predictions are supported by experimental results, the dominant view in the literature that SIs do not arise in questions, might be reconsidered.

5. Conclusion

The results of the two experiments presented in this paper support the idea that one of the contextual properties that governs SI-calculation is the property of QUD and focus. The experiments showed the effect of this property can also be observed with the scalar term *most*, and can be extended to sentential answers to

^{13.} An interesting third condition might be Does John have a OR b?, in which only the scalar itself is the focus.

yes/no-questions. Finally, an outlook was given on how the effect of QUD and focus might even extend to SIs in questions themselves, an environment in which SIs are traditionally assumed not to arise.

References

- Carston, R. 1998. Informativeness, relevance, and scalar implicature. In *Relevance Theory: Applications and Implications* [Pragmatics & Beyond New Series 37, R. Carston & S. Uchida (eds), 179–236. Amsterdam: John Benjamins.
- Chierchia, G., Crain, S., Guasti, M.T., Gualmini, A. & Meroni, L. 2001. The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In *BUCLD 25 Proceedings*, A.H.-J. Do et al. (eds), 157–168. Sommerville MA: Cascadilla Press.
- Crain, S. & Thornton R. 1998. Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics. Cambridge MA: The MIT Press.
- Groenendijk, J. & Stokhof, M. 1984. Studies on the Semantics of Questions and the Pragmatics of Answers. Ph.D. dissertation, University of Amsterdam.
- Hamblin, C.L. 1973. Questions in Montague English. Foundations of Language 10: 41-53.
- Hausser, R.R. 1983. Surface compositionality and the semantics of mood. In *Speech Act Theory and Pragmatics*, J.R. Searle, F. Kiefer & M. Bierwisch (eds), 71–95. Dordrecht: Reidel.
- Jackendoff, R. 1972. Semantic Interpretation in Generative Grammar. Cambridge MA: The MIT Press.
- van Kuppevelt, J. 1996. Inferring from topics. Scalar implicatures as topic-dependent inferences. *Linguistics and Philosophy* 19: 393–443.
- Levinson, S.C. 1983. Pragmatics. Cambridge: CUP.
- Noveck, I., Chierchia, G., Chevaux, F., Guelminger, R. & Sylvestre, E. 2002. Linguistic-pragmatic factors in interpreting disjunctions. *Thinking and Reasoning* 8(4): 297–326
- Roberts, C. 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics [Papers in Semantics, OSU Working Papers in Linguistics 49]. Columbus OH: Department of Linguistics, the Ohio State University.
- van Rooij, R. 2002. Relevance implicatures. Ms, ILLC, Amsterdam. <http://semanticsarchive. net/Archive/WIyOWUyO/Implicfinal.pdf>
- van Rooij, R. & Schulz, K. 2004. Exhaustive interpretation of complex sentences. Journal of Logic, Language and Information 13(4): 491–519.
- Rooth, M. 1996. Focus. In *The Handbook of Contemporary Semantic Theory*, S. Lappin (ed.), 271–297. London: Blackwell.
- Schulz, K. & van Rooij, R. 2006. Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Linguistics and Philosophy* 29: 205–250.
- Scharten, R. 1997. Exhaustive Interpretation: A Discourse-Semantic Account. Ph.D. dissertation, Katholieke Universiteit, Nijmegen.
- Zondervan, A. 2007. Effects of question under discussion and focus on scalar implicatures. In Proceedings of the Fifth Semantics in the Netherlands Day, M. Kluck & E.J. Smits (eds), 39–52. Groningen: CLCG.
- Zondervan, A. 2009. Experiments on QUD and focus as a contextual constraint on scalar implicature calculation. In *Semantics and Pragmatics. From Experiment to Theory*, U. Sauerland & K. Yatsushiro (eds), 94–110. Basingstoke: Palgrave.

Index

A

ambiguity 21–22, 39, 63, 68, 74, 135 asymmetric coordination 5, 7 availability hypothesis 49

B

background knowledge 4, 82, 90, 157, 172 bring-in strategies 157 bridging 5, 203 but 12, 100–108, 115–116, 123–125

С

canonicality 155-159, 165, 167-168, 170-172 child language development 12, 19, 69, 79, 101, 106, 115, 116, 117 compositional meaning 5-6, 13, 102, 103, 106, 113, 120 compositionality 199 conjunction 5, 45, 100-108, 110, 113–119, 122, 139, 143, 180 contextualism 2, 43 conversational competence 11, 19, 21, 38-40 cooperative principle 21 coordination 5, 7, 101–103, 105-107, 109-110, 116, 126

D

deictic gestures 163, 166, 168, 170 discourse 8, 10, 23, 25, 39, 45, 63–67, 72, 75, 80, 82–86, 89–90, 94–97, 124, 130, 135, 142, 157, 200–215, 217, 222–223 disjunction 100–108, 110, 113–119, 124, 140, 224 downward entailment 12, 112, 120, 139–141, 147

E

enriched composition 5-6, 13, 199-202, 205-207, 211, 214-217 enriched proposition 55 enrichment 1-7, 12-13, 44-45, 51, 53, 134, 177-197, 200, 203, 206, 214-217 enrichment blocking 13, 191, 195 entailment 75, 129-130, 135, 137, 140, 147 event-related brain potential (ERP) 11, 199 event structure 199-200, 203, 214 experimental pragmatics 1-4, 6, 9-11, 48, 173 experimental semantics 2-3, 10, 101, 199 explicature 2, 4-5, 44, 51

F

familiarity 30, 102, 155–156, 158, 172 feature structures 178, 189, 191, 197 felicity conditions 2, 8, 72, 75 focus 2–3, 7–8, 12–13, 31, 40, 53, 59, 79–97, 132, 144, 154, 167, 187, 192–193, 202, 221–238 focus particles 80–90, 93–97 *follow-in* strategies 157 free enrichment 44–45, 51

G

generalized conversational implicature (GCI) 4, 102, 113, 114, 117–119, 124–126 gestural behavior 12, 151, 157, 161–162, 165, 168, 171 Gricean maxims 1, 12, 19–21, 24–25, 27, 35–37, 131, 182, 202

I

iconic gestures 161, 166–167, 170 implicature 1, 3–5, 7, 10, 12–13, 21, 25, 43-55, 101, 103-105, 111-125, 129, 131-132, 136-138, 140, 147-148, 221, 223 independence hypothesis 49 indirect speech acts 6, 22 in fact 12, 177, 178, 186-197 inference 21, 44, 55, 121, 130, 132, 136, 203-204, 206, 214-215, 227, 235 information structure 6, 79 informativeness 90, 137, 148 intuitive content 2 interactionism 5 interfaces 3, 5, 11, 138 intimated 55-57, 59 irony 2, 6, 10, 45-47 isomorphism 64, 68, 72, 75

L

language acquisition 2–3, 7, 10, 63, 117, 172 language development 19–20, 23–24, 27–30, 40, 79, 124, 151–152, 154, 157 language disorders 19, 21, 37–38 late positivity 13, 199–200, 205–207, 210–211, 214–217 lexical meaning 103, 106, 136, 217 literal meaning 12, 43, 45, 60, 177–179, 182, 185, 188, 196, 206 logic ambiguity 129–137, 142, 143, 147, 223 lower bound 10, 129–137, 142–143, 147, 223

Μ

manipulative gestures 163, 167, 170–171 materialistic factors 151–152 metaphor 2, 6, 58, 172, 206–207, 215 minimalism 2 modal enrichment 12, 177–183, 185–197 modal operator 178, 179, 184, 186, 189, 195, 196 modality 5, 35, 64, 71

N

N400 11, 204–208, 210, 215–217 negation 2–3, 12, 63, 65–66, 68–72, 76, 108, 110, 120, 123, 133, 141, 143, 146 Neogriceans 1, 132–133 neuropragmatics 10 neurosemantics 10 non-truth-conditional 101, 103–104, 109, 124 numerals 12, 129–134, 136–140, 147–148, 223

0

object 20, 24–25, 27, 83–84, 87–89, 94–97, 151–157, 161–164, 167–172, 200, 202–203, 225–229 object knowledge 151, 154, 156–157 objects perception *or* 101–104, 110–113, 125 *only* 79–87, 96–97

Р

particularized conversational implicature (PCI) 4, 113, 114, 118, 122–126 pointing saccade 169 polarity 63–64, 69, 129–130, 139–140, 143–146, 148, 204 pragmatics 1–4, 6, 9–12, 20–21, 24, 30, 38, 44, 48–55, 101–109, 113–114, 116,

119, 122–126, 138, 151, 153, 171-173, 186, 199-201, 217, 221, 236 pragmatic competence 6, 19, 48, 56 pragmatic development 3, 106 pragmatic impairment 1, 9, 20-37 pragmatic intrusion 2, 4–5, 44 pragmatic intuitions 46–48, 54 pragmatic meanings 101, 103, 124-125 pragmatic processing 10, 205, 216-217 preschool children 28, 30, 40,90 processing of semantics 11, 204, 216-217 processing of pragmatics 10, 205, 216-217 pseudo-scalar implicature 101 psycholinguistics 2, 131, 199

Q

question under discussion (QUD) 13, 66, 67, 72–76, 217, 221–238

R

reaction times 10, 11, 209–210 reference transfer 5–6, 13, 199–217 relevance theory 4, 7, 54, 122, 132, 134 response uniformity 12, 101–102, 105, 112–113, 119–120, 124, 126

S

salience 202–203, 214–216 scalar implicatures 2–4, 7, 10, 12–13, 55, 96, 118, 122–123, 125, 129, 131–134, 221–222 scope 5, 12, 63–64, 68–72, 80, 83, 85–86, 88–89, 94, 107–108, 118, 124, 136, 138, 147, 192–194, 196 semantics 1–6, 10–12, 25, 94, 101–103, 105–107, 109, 113, 115–116, 119, 122, 124–126, 129, 133–134, 136, 151–153, 158, 171, 186, 193, 195, 199–201, 215–216, 223 semantic-pragmatic disorder (spd) 22, 23, 25 semantic-pragmatic interface 2, 3, 11 semantic meaning 101-103, 105–106, 113–114, 116–117, 120, 123-126, 202, 224 semantic processing 11, 204, 216-217 semiotic character 152–153 sentence interpretation 82, 86,90 sentence meaning 44-46, 49, 53, 55, 59, 177 speech acts 6, 8-9, 22 some 3-4, 7-8, 51, 63-70, 72, 82, 121-123, 131-134, 147-148 specific language impairment 9-11, 19, 37, 40 subset principle 68, 75 symbol 151-152, 172

Т

tatsächlich ('in fact') 12, 177, 178, 186–197 topic maintenance 23 truth-conditional meaning 101, 103–104, 115, 124 truth-Value Judgment Task 101

U

underdeterminacy 2, 5 underinformativ sentences 7, 78, 90–94, 232 upper bound 10, 129, 130, 132–137, 140–142, 145–148, 223 upward entailment 12, 128, 130, 141, 147 utterance-meaning 43, 45, 50, 53, 56

W

what-is-said 12, 43–47, 49–59 what-is-implicated 49, 55

Y

yes/no-questions 221, 226, 228, 232, 235–238

Linguistik Aktuell/Linguistics Today

A complete list of titles in this series can be found on the publishers' website, www.benjamins.com

- 179 TUBINO BLANCO, Mercedes: Causatives in Minimalism. Expected July 2011
- 178 GALANI, Alexandra, Glyn HICKS and George TSOULAS (eds.): Morphology and its Interfaces. ix, 352 pp. + index. Expected May 2011
- 177 BISKUP, Petr: Adverbials and the Phase Model. ix, 230 pp. + index. Expected April 2011
- 176 YUASA, Etsuyo, Tista BAGCHI and Katharine BEALS (eds.): Pragmatics and Autolexical Grammar. In honor of Jerry Sadock. xxv, 335 pp. + index. *Expected April 2011*
- 175 MEIBAUER, Jörg and Markus STEINBACH (eds.): Experimental Pragmatics/Semantics. 2011. x, 240 pp.
- 174 LOMASHVILI, Leila: Complex Predicates. The syntax-morphology interface. 2011. xi, 190 pp.
- 173 **SAPP, Christopher D.:** The Verbal Complex in Subordinate Clauses from Medieval to Modern German. 2011. ix, 225 pp.
- 172 JUNG, Hakyung: The Syntax of the *Be*-Possessive. Parametric variation and surface diversities. xi, 264 pp. + index. *Expected April 2011*
- 171 SLEEMAN, Petra and Harry PERRIDON (eds.): The Noun Phrase in Romance and Germanic. Structure, variation, and change. 2011. vii, 283 pp.
- 170 HUNTER, Tim: Syntactic Effects of Conjunctivist Semantics. Unifying movement and adjunction. xi, 179 pp. + index. *Expected April 2011*
- 169 SÁNCHEZ, Liliana: The Morphology and Syntax of Topic and Focus. Minimalist inquiries in the Quechua periphery. 2010. xiii, 242 pp.
- 168 FELDHAUSEN, Ingo: Sentential Form and Prosodic Structure of Catalan. 2010. xiii, 285 pp.
- 167 MERCADO, Raphael, Eric POTSDAM and Lisa deMena TRAVIS (eds.): Austronesian and Theoretical Linguistics. 2010. vii, 379 pp.
- 166 BRANDT, Patrick and Marco GARCÍA GARCÍA (eds.): Transitivity. Form, Meaning, Acquisition, and Processing. 2010. vii, 308 pp.
- 165 **BREUL, Carsten and Edward GÖBBEL (eds.):** Comparative and Contrastive Studies of Information Structure. 2010. xii, 306 pp.
- 164 ZWART, Jan-Wouter and Mark de VRIES (eds.): Structure Preserved. Studies in syntax for Jan Koster. 2010. xxiii, 395 pp.
- 163 KIZIAK, Tanja: Extraction Asymmetries. Experimental evidence from German. 2010. xvi, 273 pp.
- 162 BOTT, Oliver: The Processing of Events. 2010. xvii, 383 pp.
- 161 HAAN, Germen J. de: Studies in West Frisian Grammar. Edited by Jarich Hoekstra, Willem Visser and Goffe Jensma. 2010. x, 384 pp.
- 160 MAVROGIORGOS, Marios: Clitics in Greek. A minimalist account of proclisis and enclisis. 2010. x, 294 pp.
- 159 BREITBARTH, Anne, Christopher LUCAS, Sheila WATTS and David WILLIS (eds.): Continuity and Change in Grammar. 2010. viii, 359 pp.
- 158 **DUGUINE, Maia, Susana HUIDOBRO and Nerea MADARIAGA (eds.):** Argument Structure and Syntactic Relations. A cross-linguistic perspective. 2010. vi, 348 pp.
- 157 FISCHER, Susann: Word-Order Change as a Source of Grammaticalisation. 2010. ix, 200 pp.
- 156 DI SCIULLO, Anna Maria and Virginia HILL (eds.): Edges, Heads, and Projections. Interface properties. 2010. vii, 265 pp.
- 155 SATO, Yosuke: Minimalist Interfaces. Evidence from Indonesian and Javanese. 2010. xiii, 159 pp.
- 154 HORNSTEIN, Norbert and Maria POLINSKY (eds.): Movement Theory of Control. 2010. vii, 330 pp.
- 153 CABREDO HOFHERR, Patricia and Ora MATUSHANSKY (eds.): Adjectives. Formal analyses in syntax and semantics. 2010. vii, 335 pp.
- 152 GALLEGO, Ángel J.: Phase Theory. 2010. xii, 365 pp.
- 151 **SUDHOFF, Stefan:** Focus Particles in German. Syntax, prosody, and information structure. 2010. xiii, 335 pp.
- 150 EVERAERT, Martin, Tom LENTZ, Hannah de MULDER, Øystein NILSEN and Arjen ZONDERVAN (eds.): The Linguistics Enterprise. From knowledge of language to knowledge in linguistics. 2010. ix, 379 pp.
- 149 AELBRECHT, Lobke: The Syntactic Licensing of Ellipsis. 2010. xii, 230 pp.
- 148 HOGEWEG, Lotte, Helen de HOOP and Andrej MALCHUKOV (eds.): Cross-linguistic Semantics of Tense, Aspect, and Modality. 2009. vii, 406 pp.
- 147 GHOMESHI, Jila, Ileana PAUL and Martina WILTSCHKO (eds.): Determiners. Universals and variation. 2009. vii, 247 pp.
- 146 GELDEREN, Elly van (ed.): Cyclical Change. 2009. viii, 329 pp.
- 145 WESTERGAARD, Marit: The Acquisition of Word Order. Micro-cues, information structure, and economy. 2009. xii, 245 pp.
- 144 PUTNAM, Michael T. (ed.): Towards a Derivational Syntax. Survive-minimalism. 2009. x, 269 pp.
- 143 ROTHMAYR, Antonia: The Structure of Stative Verbs. 2009. xv, 216 pp.
- 142 NUNES, Jairo (ed.): Minimalist Essays on Brazilian Portuguese Syntax. 2009. vi, 243 pp.
- 141 ALEXIADOU, Artemis, Jorge HANKAMER, Thomas McFADDEN, Justin NUGER and Florian SCHÄFER (eds.): Advances in Comparative Germanic Syntax. 2009. xv, 395 pp.
- 140 ROEHRS, Dorian: Demonstratives and Definite Articles as Nominal Auxiliaries. 2009. xii, 196 pp.
- 139 HICKS, Glyn: The Derivation of Anaphoric Relations. 2009. xii, 309 pp.
- 138 SIDDIQI, Daniel: Syntax within the Word. Economy, allomorphy, and argument selection in Distributed Morphology. 2009. xii, 138 pp.
- 137 **PFAU, Roland:** Grammar as Processor. A Distributed Morphology account of spontaneous speech errors. 2009. xiii, 372 pp.
- 136 KANDYBOWICZ, Jason: The Grammar of Repetition. Nupe grammar at the syntax-phonology interface. 2008. xiii, 168 pp.
- 135 LEWIS, William D., Simin KARIMI, Heidi HARLEY and Scott O. FARRAR (eds.): Time and Again. Theoretical perspectives on formal linguistics. In honor of D. Terence Langendoen. 2009. xiv, 265 pp.
- 134 ARMON-LOTEM, Sharon, Gabi DANON and Susan D. ROTHSTEIN (eds.): Current Issues in Generative Hebrew Linguistics. 2008. vii, 393 pp.
- 133 MACDONALD, Jonathan E.: The Syntactic Nature of Inner Aspect. A minimalist perspective. 2008. xv, 241 pp.
- 132 BIBERAUER, Theresa (ed.): The Limits of Syntactic Variation. 2008. vii, 521 pp.
- 131 **DE CAT, Cécile and Katherine DEMUTH (eds.):** The Bantu–Romance Connection. A comparative investigation of verbal agreement, DPs, and information structure. 2008. xix, 355 pp.
- 130 KALLULLI, Dalina and Liliane TASMOWSKI (eds.): Clitic Doubling in the Balkan Languages. 2008. ix, 442 pp.
- 129 **STURGEON, Anne:** The Left Periphery. The interaction of syntax, pragmatics and prosody in Czech. 2008. xi, 143 pp.
- 128 TALEGHANI, Azita H.: Modality, Aspect and Negation in Persian. 2008. ix, 183 pp.
- 127 **DURRLEMAN-TAME, Stephanie:** The Syntax of Jamaican Creole. A cartographic perspective. 2008. xii, 190 pp.
- 126 SCHÄFER, Florian: The Syntax of (Anti-)Causatives. External arguments in change-of-state contexts. 2008. xi, 324 pp.
- 125 **ROTHSTEIN, Björn:** The Perfect Time Span. On the present perfect in German, Swedish and English. 2008. xi, 171 pp.
- 124 IHSANE, Tabea: The Layered DP. Form and meaning of French indefinites. 2008. ix, 260 pp.
- 123 STOYANOVA, Marina: Unique Focus. Languages without multiple wh-questions. 2008. xi, 184 pp.
- 122 OOSTERHOF, Albert M.: The Semantics of Generics in Dutch and Related Languages. 2008. xviii, 286 pp.
- 121 **TUNGSETH, Mai Ellin:** Verbal Prepositions and Argument Structure. Path, place and possession in Norwegian. 2008. ix, 187 pp.
- 120 ASBURY, Anna, Jakub DOTLAČIL, Berit GEHRKE and Rick NOUWEN (eds.): Syntax and Semantics of Spatial P. 2008. vi, 416 pp.
- 119 FORTUNY, Jordi: The Emergence of Order in Syntax. 2008. viii, 211 pp.
- 118 JÄGER, Agnes: History of German Negation. 2008. ix, 350 pp.
- 117 HAUGEN, Jason D.: Morphology at the Interfaces. Reduplication and Noun Incorporation in Uto-Aztecan. 2008. xv, 257 pp.
- 116 ENDO, Yoshio: Locality and Information Structure. A cartographic approach to Japanese. 2007. x, 235 pp.
- 115 PUTNAM, Michael T.: Scrambling and the Survive Principle. 2007. x, 216 pp.
- 114 LEE-SCHOENFELD, Vera: Beyond Coherence. The syntax of opacity in German. 2007. viii, 206 pp.

- 113 EYTHÓRSSON, Thórhallur (ed.): Grammatical Change and Linguistic Theory. The Rosendal papers. 2008. vi, 441 pp.
- 112 AXEL, Katrin: Studies on Old High German Syntax. Left sentence periphery, verb placement and verbsecond. 2007. xii, 364 pp.
- 111 EGUREN, Luis and Olga FERNÁNDEZ-SORIANO (eds.): Coreference, Modality, and Focus. Studies on the syntax–semantics interface. 2007. xii, 239 pp.
- 110 ROTHSTEIN, Susan D. (ed.): Theoretical and Crosslinguistic Approaches to the Semantics of Aspect. 2008. viii, 453 pp.
- 109 **CHOCANO, Gema:** Narrow Syntax and Phonological Form. Scrambling in the Germanic languages. 2007. x, 333 pp.
- 108 REULAND, Eric, Tanmoy BHATTACHARYA and Giorgos SPATHAS (eds.): Argument Structure. 2007. xviii, 243 pp.
- 107 CORVER, Norbert and Jairo NUNES (eds.): The Copy Theory of Movement. 2007. vi, 388 pp.
- 106 DEHÉ, Nicole and Yordanka KAVALOVA (eds.): Parentheticals. 2007. xii, 314 pp.
- 105 HAUMANN, Dagmar: Adverb Licensing and Clause Structure in English. 2007. ix, 438 pp.
- 104 **JEONG, Youngmi:** Applicatives. Structure and interpretation from a minimalist perspective. 2007. vii, 144 pp.
- 103 WURFF, Wim van der (ed.): Imperative Clauses in Generative Grammar. Studies in honour of Frits Beukema. 2007. viii, 352 pp.
- 102 BAYER, Josef, Tanmoy BHATTACHARYA and M.T. Hany BABU (eds.): Linguistic Theory and South Asian Languages. Essays in honour of K. A. Jayaseelan. 2007. x, 282 pp.
- 101 KARIMI, Simin, Vida SAMIIAN and Wendy K. WILKINS (eds.): Phrasal and Clausal Architecture. Syntactic derivation and interpretation. In honor of Joseph E. Emonds. 2007. vi, 424 pp.
- 100 SCHWABE, Kerstin and Susanne WINKLER (eds.): On Information Structure, Meaning and Form. Generalizations across languages. 2007. vii, 570 pp.
- 99 MARTÍNEZ-GIL, Fernando and Sonia COLINA (eds.): Optimality-Theoretic Studies in Spanish Phonology. 2007. viii, 564 pp.
- 98 PIRES, Acrisio: The Minimalist Syntax of Defective Domains. Gerunds and infinitives. 2006. xiv, 188 pp.
- 97 HARTMANN, Jutta M. and László MOLNÁRFI (eds.): Comparative Studies in Germanic Syntax. From Afrikaans to Zurich German. 2006. vi, 332 pp.
- 96 LYNGFELT, Benjamin and Torgrim SOLSTAD (eds.): Demoting the Agent. Passive, middle and other voice phenomena. 2006. x, 333 pp.
- 95 VOGELEER, Svetlana and Liliane TASMOWSKI (eds.): Non-definiteness and Plurality. 2006. vi, 358 pp.
- 94 ARCHE, María J.: Individuals in Time. Tense, aspect and the individual/stage distinction. 2006. xiv, 281 pp.
- 93 PROGOVAC, Ljiljana, Kate PAESANI, Eugenia CASIELLES and Ellen BARTON (eds.): The Syntax of Nonsententials. Multidisciplinary perspectives. 2006. x, 372 pp.
- 92 BOECKX, Cedric (ed.): Agreement Systems. 2006. ix, 346 pp.
- 91 BOECKX, Cedric (ed.): Minimalist Essays. 2006. xvi, 399 pp.
- 90 DALMI, Gréte: The Role of Agreement in Non-Finite Predication. 2005. xvi, 222 pp.
- 89 VELDE, John R. te: Deriving Coordinate Symmetries. A phase-based approach integrating Select, Merge, Copy and Match. 2006. x, 385 pp.
- 88 MOHR, Sabine: Clausal Architecture and Subject Positions. Impersonal constructions in the Germanic languages. 2005. viii, 207 pp.
- 87 JULIEN, Marit: Nominal Phrases from a Scandinavian Perspective. 2005. xvi, 348 pp.
- 86 COSTA, João and Maria Cristina FIGUEIREDO SILVA (eds.): Studies on Agreement. 2006. vi, 285 pp.
- 85 MIKKELSEN, Line: Copular Clauses. Specification, predication and equation. 2005. viii, 210 pp.
- 84 PAFEL, Jürgen: Quantifier Scope in German. 2006. xvi, 312 pp.
- 83 SCHWEIKERT, Walter: The Order of Prepositional Phrases in the Structure of the Clause. 2005. xii, 338 pp.
- 82 QUINN, Heidi: The Distribution of Pronoun Case Forms in English. 2005. xii, 409 pp.
- 81 **FUSS, Eric:** The Rise of Agreement. A formal approach to the syntax and grammaticalization of verbal inflection. 2005. xii, 336 pp.
- 80 BURKHARDT SCHUMACHER, Petra: The Syntax–Discourse Interface. Representing and interpreting dependency. 2005. xii, 259 pp.
- 79 SCHMID, Tanja: Infinitival Syntax. Infinitivus Pro Participio as a repair strategy. 2005. xiv, 251 pp.

- 78 **DIKKEN, Marcel den and Christina TORTORA (eds.):** The Function of Function Words and Functional Categories. 2005. vii, 292 pp.
- 77 ÖZTÜRK, Balkız: Case, Referentiality and Phrase Structure. 2005. x, 268 pp.
- 76 STAVROU, Melita and Arhonto TERZI (eds.): Advances in Greek Generative Syntax. In honor of Dimitra Theophanopoulou-Kontou. 2005. viii, 366 pp.
- 75 DI SCIULLO, Anna Maria (ed.): UG and External Systems. Language, brain and computation. 2005. xviii, 398 pp.
- 74 HEGGIE, Lorie and Francisco ORDÓÑEZ (eds.): Clitic and Affix Combinations. Theoretical perspectives. 2005. viii, 390 pp.
- 73 CARNIE, Andrew, Heidi HARLEY and Sheila Ann DOOLEY (eds.): Verb First. On the syntax of verbinitial languages. 2005. xiv, 434 pp.
- 72 FUSS, Eric and Carola TRIPS (eds.): Diachronic Clues to Synchronic Grammar. 2004. viii, 228 pp.
- 71 GELDEREN, Elly van: Grammaticalization as Economy. 2004. xvi, 320 pp.
- 70 AUSTIN, Jennifer R., Stefan ENGELBERG and Gisa RAUH (eds.): Adverbials. The interplay between meaning, context, and syntactic structure. 2004. x, 346 pp.
- 69 KISS, Katalin É. and Henk van RIEMSDIJK (eds.): Verb Clusters. A study of Hungarian, German and Dutch. 2004. vi, 514 pp.
- 68 **BREUL, Carsten:** Focus Structure in Generative Grammar. An integrated syntactic, semantic and intonational approach. 2004. x, 432 pp.
- 67 MIŠESKA TOMIĆ, Olga (ed.): Balkan Syntax and Semantics. 2004. xvi, 499 pp.
- 66 GROHMANN, Kleanthes K.: Prolific Domains. On the Anti-Locality of movement dependencies. 2003. xvi, 372 pp.
- 65 MANNINEN, Satu Helena: Small Phrase Layers. A study of Finnish Manner Adverbials. 2003. xii, 275 pp.
- 64 BOECKX, Cedric and Kleanthes K. GROHMANN (eds.): Multiple Wh-Fronting. 2003. x, 292 pp.
- 63 BOECKX, Cedric: Islands and Chains. Resumption as stranding. 2003. xii, 224 pp.
- 62 CARNIE, Andrew, Heidi HARLEY and MaryAnn WILLIE (eds.): Formal Approaches to Function in Grammar. In honor of Eloise Jelinek. 2003. xii, 378 pp.
- 61 SCHWABE, Kerstin and Susanne WINKLER (eds.): The Interfaces. Deriving and interpreting omitted structures. 2003. vi, 403 pp.
- 60 TRIPS, Carola: From OV to VO in Early Middle English. 2002. xiv, 359 pp.
- 59 DEHÉ, Nicole: Particle Verbs in English. Syntax, information structure and intonation. 2002. xii, 305 pp.
- 58 DI SCIULLO, Anna Maria (ed.): Asymmetry in Grammar. Volume 2: Morphology, phonology, acquisition. 2003. vi, 309 pp.
- 57 **DI SCIULLO, Anna Maria (ed.):** Asymmetry in Grammar. Volume 1: Syntax and semantics. 2003. vi, 405 pp.
- 56 **COENE, Martine and Yves D'HULST (eds.):** From NP to DP. Volume 2: The expression of possession in noun phrases. 2003. x, 295 pp.
- 55 COENE, Martine and Yves D'HULST (eds.): From NP to DP. Volume 1: The syntax and semantics of noun phrases. 2003. vi, 362 pp.
- 54 **BAPTISTA, Marlyse:** The Syntax of Cape Verdean Creole. The Sotavento varieties. 2003. xxii, 294 pp. (incl. CD-rom).
- 53 ZWART, Jan-Wouter and Werner ABRAHAM (eds.): Studies in Comparative Germanic Syntax. Proceedings from the 15th Workshop on Comparative Germanic Syntax (Groningen, May 26–27, 2000). 2002. xiv, 407 pp.
- 52 SIMON, Horst J. and Heike WIESE (eds.): Pronouns Grammar and Representation. 2002. xii, 294 pp.
- 51 GERLACH, Birgit: Clitics between Syntax and Lexicon. 2002. xii, 282 pp.
- 50 **STEINBACH, Markus:** Middle Voice. A comparative study in the syntax-semantics interface of German. 2002. xii, 340 pp.
- 49 ALEXIADOU, Artemis (ed.): Theoretical Approaches to Universals. 2002. viii, 319 pp.
- 48 ALEXIADOU, Artemis, Elena ANAGNOSTOPOULOU, Sjef BARBIERS and Hans-Martin GÄRTNER (eds.): Dimensions of Movement. From features to remnants. 2002. vi, 345 pp.
- 47 BARBIERS, Sjef, Frits BEUKEMA and Wim van der WURFF (eds.): Modality and its Interaction with the Verbal System. 2002. x, 290 pp.
- 46 PANAGIOTIDIS, E. Phoevos: Pronouns, Clitics and Empty Nouns. 'Pronominality' and licensing in syntax. 2002. x, 214 pp.

- 45 ABRAHAM, Werner and Jan-Wouter ZWART (eds.): Issues in Formal German(ic) Typology. 2002. xviii, 336 pp.
- 44 TAYLAN, Eser Erguvanlı (ed.): The Verb in Turkish. 2002. xviii, 267 pp.
- 43 FEATHERSTON, Sam: Empty Categories in Sentence Processing. 2001. xvi, 279 pp.
- 42 ALEXIADOU, Artemis: Functional Structure in Nominals. Nominalization and ergativity. 2001. x, 233 pp.
- 41 ZELLER, Jochen: Particle Verbs and Local Domains. 2001. xii, 325 pp.
- 40 HOEKSEMA, Jack, Hotze RULLMANN, Victor SÁNCHEZ-VALENCIA and Ton van der WOUDEN (eds.): Perspectives on Negation and Polarity Items. 2001. xii, 368 pp.
- 39 **GELDEREN, Elly van:** A History of English Reflexive Pronouns. Person, *Self*, and Interpretability. 2000. xiv, 279 pp.
- 38 MEINUNGER, André: Syntactic Aspects of Topic and Comment. 2000. xii, 247 pp.
- 37 LUTZ, Uli, Gereon MÜLLER and Arnim von STECHOW (eds.): Wh-Scope Marking. 2000. vi, 483 pp.
- 36 GERLACH, Birgit and Janet GRIJZENHOUT (eds.): Clitics in Phonology, Morphology and Syntax. 2001. xii, 441 pp.
- 35 HRÓARSDÓTTIR, Thorbjörg: Word Order Change in Icelandic. From OV to VO. 2001. xiv, 385 pp.
- 34 REULAND, Eric (ed.): Arguments and Case. Explaining Burzio's Generalization. 2000. xii, 255 pp.
- 33 PUSKÁS, Genoveva: Word Order in Hungarian. The syntax of A-positions. 2000. xvi, 398 pp.
- 32 ALEXIADOU, Artemis, Paul LAW, André MEINUNGER and Chris WILDER (eds.): The Syntax of Relative Clauses. 2000. vi, 397 pp.
- 31 SVENONIUS, Peter (ed.): The Derivation of VO and OV. 2000. vi, 372 pp.
- 30 BEUKEMA, Frits and Marcel den DIKKEN (eds.): Clitic Phenomena in European Languages. 2000. x, 324 pp.
- 29 MIYAMOTO, Tadao: The Light Verb Construction in Japanese. The role of the verbal noun. 2000. xiv, 232 pp.
- 28 HERMANS, Ben and Marc van OOSTENDORP (eds.): The Derivational Residue in Phonological Optimality Theory. 2000. viii, 322 pp.
- 27 RŮŽIČKA, Rudolf: Control in Grammar and Pragmatics. A cross-linguistic study. 1999. x, 206 pp.
- 26 ACKEMA, Peter: Issues in Morphosyntax. 1999. viii, 310 pp.
- 25 **FELSER, Claudia:** Verbal Complement Clauses. A minimalist study of direct perception constructions. 1999. xiv, 278 pp.
- 24 REBUSCHI, Georges and Laurice TULLER (eds.): The Grammar of Focus. 1999. vi, 366 pp.
- 23 GIANNAKIDOU, Anastasia: Polarity Sensitivity as (Non)Veridical Dependency. 1998. xvi, 282 pp.
- 22 ALEXIADOU, Artemis and Chris WILDER (eds.): Possessors, Predicates and Movement in the Determiner Phrase. 1998. vi, 388 pp.
- 21 KLEIN, Henny: Adverbs of Degree in Dutch and Related Languages. 1998. x, 232 pp.
- 20 LAENZLINGER, Christopher: Comparative Studies in Word Order Variation. Adverbs, pronouns, and clause structure in Romance and Germanic. 1998. x, 371 pp.
- 19 JOSEFSSON, Gunlög: Minimal Words in a Minimal Syntax. Word formation in Swedish. 1998. ix, 199 pp.
- 18 ALEXIADOU, Artemis: Adverb Placement. A case study in antisymmetric syntax. 1997. x, 256 pp.
- 17 BEERMANN, Dorothee, David LEBLANC and Henk van RIEMSDIJK (eds.): Rightward Movement. 1997. vi, 410 pp.
- 16 LIU, Feng-hsi: Scope and Specificity. 1997. viii, 187 pp.
- 15 ROHRBACHER, Bernhard Wolfgang: Morphology-Driven Syntax. A theory of V to I raising and prodrop. 1999. viii, 296 pp.
- 14 ANAGNOSTOPOULOU, Elena, Henk van RIEMSDIJK and Frans ZWARTS (eds.): Materials on Left Dislocation. 1997. viii, 349 pp.
- 13 ALEXIADOU, Artemis and T. Alan HALL (eds.): Studies on Universal Grammar and Typological Variation. 1997. viii, 252 pp.
- 12 ABRAHAM, Werner, Samuel David EPSTEIN, Höskuldur THRÁINSSON and Jan-Wouter ZWART (eds.): Minimal Ideas. Syntactic studies in the minimalist framework. 1996. xii, 364 pp.
- 11 LUTZ, Uli and Jürgen PAFEL (eds.): On Extraction and Extraposition in German. 1996. xii, 315 pp.
- 10 CINQUE, Guglielmo and Giuliana GIUSTI (eds.): Advances in Roumanian Linguistics. 1995. xi, 172 pp.
- 9 GELDEREN, Elly van: The Rise of Functional Categories. 1993. x, 224 pp.

- 8 FANSELOW, Gisbert (ed.): The Parametrization of Universal Grammar. 1993. xvii, 232 pp.
- 7 ÅFARLÍ, Tor A.: The Syntax of Norwegian Passive Constructions. 1992. xii, 177 pp.
- 6 BHATT, Christa, Elisabeth LÖBEL and Claudia Maria SCHMIDT (eds.): Syntactic Phrase Structure Phenomena in Noun Phrases and Sentences. 1989. ix, 187 pp.
- 5 GREWENDORF, Günther and Wolfgang STERNEFELD (eds.): Scrambling and Barriers. 1990. vi, 442 pp.
- 4 **ABRAHAM, Werner and Sjaak De MEIJ (eds.):** Topic, Focus and Configurationality. Papers from the 6th Groningen Grammar Talks, Groningen, 1984. 1986. v, 349 pp.
- 3 ABRAHAM, Werner (ed.): On the Formal Syntax of the Westgermania. Papers from the 3rd Groningen Grammar Talks (3e Groninger Grammatikgespräche), Groningen, January 1981. 1983. vi, 242 pp.
- 2 EHLICH, Konrad and Jürgen REHBEIN: Augenkommunikation. Methodenreflexion und Beispielanalyse. 1982. viii, 150 pp. With many photographic ills.
- 1 **KLAPPENBACH, Ruth (1911–1977):** Studien zur Modernen Deutschen Lexikographie. Auswahl aus den Lexikographischen Arbeiten von Ruth Klappenbach, erweitert um drei Beiträge von Helene Malige-Klappenbach. (Written in German). 1980. xxiii, 313 pp.