COMPUTING MEANING

VOLUME 3

Studies in Linguistics and Philosophy

Volume 83



For other titles published in this series, go to
www.springer.com/series/6556

# COMPUTING MEANING
## Volume 3

edited by

HARRY BUNT
*Tilburg University, The Netherlands*

and

REINHARD MUSKENS
*Tilburg University, The Netherlands*

Springer

Library of Congress Control Number: 99056948

ISBN 978-1-4020-5956-8 (HB)
ISBN 978-1-4020-5957-5 (PB)
ISBN 978-1-4020-5958-2 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

*Printed on acid-free paper*

CONTENTS

HARRY BUNT AND REINHARD MUSKENS
Computing the Semantic Information in an Utterance 1

MASSIMO POESIO, UWE REYLE AND ROSEMARY STEVENSON Justified Sloppiness in Anaphoric Reference 11

AOIFE CAHILL, MAIREAD MCCARTHY, MICHAEL BURKE, JOSEF VAN GENABITH AND ANDY WAY
Deriving Quasi-Logical Forms from F-Structures for the Penn Treebank 33

HARRY BUNT
Semantic Underspecification: Which Technique for What Purpose? 55

ALEX LASCARIDES AND NICHOLAS ASHER
Segmented Discourse Representation Theory: Dynamic Semantics with Discourse Structure 87

RAQUEL FERNÁNDEZ, JONATHAN GINZBURG, HOWARD GREGORY AND SHALOM LAPPIN
SHARDS: Fragment Resolution in Dialogue 125

IVANA KRUIJFF-KORBAYOVÁ AND BONNIE L. WEBBER
Interpreting Concession Statements in Light of Information Structure 145

JAN VAN EIJCK
Context and the Composition of Meaning 173

MARC SWERTS AND EMIEL KRAHMER
Meaning, Intonation and Negation 195

MYROSLAVA DZIKOVSKA, MARY SWIFT AND JAMES ALLEN Customizing Meaning: Building Domain-Specific Semantic Representations from a Generic Lexicon 213

ARAVIND K. JOSHI, LAURA KALLMEYER AND MARIBEL ROMERO
Flexible Composition in LTAG: Quantifier Scope and Inverse Linking 233

FABRICE NAUZE AND MICHIEL VAN LAMBALGEN
Serious Computing with Tense 257

JAMES PUSTEJOVSKY, ROBERT KNIPPEN, JESSICA LITTMAN, ROSER SAURÍ
Temporal and Event Information in Natural Language Text 301

TIM FERNANDO
Finite-state Descriptions for Temporal Semantics 347

CLAIRE GARDENT AND KRISTINA STRIEGNITZ
Generating Bridging Definite Descriptions 369

KEES VAN DEEMTER AND EMIEL KRAHMER
Graphs and Booleans: On the Generation of Referring Expressions 397

JAN ALEXANDERSSON AND TILMAN BECKER
Efficient Computation of Overlay for Multiple Inheritance Hierarchies in Discourse Modeling 423

RICHARD CROUCH, ANETTE FRANK AND JOSEF VAN GENABITH
Linear Logic Based Transfer and Structural Misalignment 457

INDEX 473

HARRY BUNT AND REINHARD MUSKENS

# COMPUTING THE SEMANTIC INFORMATION IN AN UTTERANCE

## 1. Introduction

To compute the meaning of any given natural language expression is extremely hard. This is partly due to the structural complexity, variability, and flexibility of natural language, but also, and more importantly, to its pervasive ambiguity. It has been estimated that, due to the fact that individual words as well as their combination in a sentence can usually express a range of semantic concepts and relations, an ordinary sentence of average length can have several million possible meanings. So what do we mean by *'the meaning* of a given natural language expression'?

Language users are hardly ever aware of having to resolve an ambiguity, so in practice the understanding of a natural language expression does not mean choosing 'the right interpretation' among millions of possibilities. The crucial point is, of course, that as a language user we are never confronted with the task of computing the meaning of a sentence in splendid isolation. That happens only in the linguistic literature. In reality, natural language always occurs in a certain context. In a given context we are not talking about just anything, but we have a certain domain of discourse. This means that those word senses can be ruled out that do not belong to this domain, as well as those interpretations of structural ambiguities which express something that would be impossible or highly implausible in the domain.

While the fact that speakers and listeners do not struggle with ambiguity resolution may suggest that context information is sufficient to determine the intended meanings of natural language expressions, it is hard to believe that context information is really sufficient to exclude millions of potential sentence meanings and retain exactly one of them as *the meaning*. It seems more likely that readers and listeners disambiguate meanings to the extent that is required by the circumstances, and that speakers and writers, expecting this of their listeners and readers, formulate their utterances accordingly. In other words, a context



*H. Bunt and R. Muskens, (eds.), Computing Meaning, Volume 3*, 1–10.

comes with certain demands on the precision with which meanings should be computed, and so, in a given context what should be regarded as *the meaning* of a given sentence is something that contains a certain amount of ambiguity or incompleteness. In particular, the context of use determines the appropriate level of granularity for referring expressions, and thus for the permitted vagueness of reference. Frazier and Rayner (1990) report on empirical evidence supporting this intuition for the use of polysemous nouns, whose reference is shown to be often unresolved until information later in a sentence provides disambiguating information, in contrast to the case of homonyms, where an interpreter must make a choice that may turn out to be wrong, leading to garden path sentences. Poesio et al. (2006) provide evidence that under certain conditions a certain amount of 'sloppiness' is permitted in anaphoric reference. A consequence of this view is that a representation of the meaning of a given natural language expression in a given context is not a semantic representation in the classical Montagovian sense, being fully specified, complete, and unambiguous, but is an underspecified semantic representation, that leaves room for ambiguity, vagueness, and incompleteness.

Underspecified semantic representations (USRs) can be regarded not just as imperfect representations of meaning, waiting to become fully specified, but also as a way to salvage the adagio of compositionality: rather than saying that the semantic representation of a sentence represents its meaning, determined by the meanings of its components and its syntactic-semantic composition, we could say that a USR represents the meaning of a sentence *in sofar as determined by the meanings of its components and its syntactic-semantic composition*, or, in other words, a USR captures the semantic information contained in a sentence through its components and its syntactic-semantic composition. In the latter form, the compositionality assumption is no longer a thesis, that can be proved wrong or correct, but simply describes what one does when computing a USR: one computes the semantic information that a given sentence contains, and gives that a formal representation.

A slightly different way of looking at the interpretation of a sentence in context, i.e. of an *utterance*, is to observe that a human interpreter does not so much compute the contextually most appropriate meaning of the utterance, but computes the semantic information that the utterance contributes to the contextual knowledge that the interpreter already has. This view corresponds with the idea that understanding an utterance means trying to integrate the information that it conveys with the rest of one's knowledge, a classical notion in

artificial intelligence that is compatible with the approach to utterance meaning currently popular in dialogue studies, generally known as the *information-state update* or *context-change approach* (Smith & van Kuppeveld, 2003 Bunt, 2000; Traum & Larsson, 2003), and to some extent also with the basic ideas of dynamic semantics (Groenendijk & Stokhof, 1985; see also the discussion in Bunt, 1989 and Groenendijk & Stokhof, 1989). If utterance meanings are in general underspecified, as we just argued, then the way in which context models ('information states') are updated by utterance meanings is in general ambiguous, and incomplete... a consequence which researchers who follow this approach have so far not taken on board.

Returning to the view on meaning and underspecification which focuses on computing the semantic information in a given utterance, it may be noted that this view comes very close to that of modern approaches to semantic annotation. Traditionally, annotation is the enrichment of text with notes on some of its properties or background. In computational linguistics annotation has usually taken the form of labelling text elements with certain tags, such as part of speech tags. Semantic annotation is taking a somewhat different turn, where the annotations that are added to a text are supposed to be expressions in a formal language with a well-defined semantics (see Bunt & Romary, 2002; 2004). The reason for this is that semantic annotations are intended to support not only the retrieval of certain text elements, but *reasoning* as well. A clear case is presented by the annotation of the temporal information in a text. If, for instance, a question-answering system is asked

(1) What new products did Microsoft announce in the last quarter?

and the data for providing answers contain a newspaper item, dated 12 May 2006, and stating that:

(2) Microsoft announced its XP follow-up system yesterday at a press meeting in San Francisco.

then the answer to this question should include the new operating system, as a result of reasoning that *yesterday* in this case is the same as May 11, 2006; that May belongs to the second quarter of a year: and that *last quarter* refers to the second quarter of 2006, since the question was asked at a date belonging to the third quarter. It is therefore insufficient to simply tag temporal expressions as being temporal expressions, for instance. Questions that are input to the

system must also be time-stamped, and the time stamp should have a well-defined internal structure allowing the recognition of a month and year, which can then be compared to the creation dates of the documents in the database. Moreover, the date referred to by *yesterday* must be annotated not just as a date, but as the date on which the event occurred that is described in the corresponding sentence. Clearly, in order to support the fairly complex inferencing that is needed in such examples, the annotations have to meet syntactic and semantic requirements that go a long way beyond those of just labelling.

The interesting point is here that semantic annotations are developing into formal representations of some of the semantic information contained in the sentences in a text. That makes them formally comparable to underspecified semantic representations, the main difference being that annotations tend to focus on a particular type of semantic information, such as temporal information or semantic roles, whereas USRs are typically intended to capture *all* the semantic information in a sentence.

## 2. About this book

Following the present introductory chapter, the book continuess with four chapters concerned with aspects of ambiguity, vagueness, and underspecification. The chapter by Massimo Poesio, Uwe Reyle and Rosemary Stevenson, entitled *Justified sloppiness in anaphoric reference*, takes up the issue of the ambiguity that speakers and listeners allow in the meanings of what they say, focusing on the use of anaphoric expressions. They analyze a corpus of spoken dialogues to identify cases in which the addressee of an utterance containing an anaphoric pronoun does not appear to have enough evidence to resolve that pronoun, yet doesn't appear to find the pronominal use infelicitous. The two patterns of anaphoric use that were found to fit these conditions suggest three conditions under which *justified sloppiness* in anaphoric references is not perceived as infelicitous. Preliminary controlled experiments indicate that subjects do find anaphoric pronouns that satisfy the justified sloppiness conditions significantly easier to process than pronouns occurring in minimally different contexts in which these conditions are not satisfied.

In the second chapter of this group, by Aoife Cahill, Mairead McCarthy, Michael Burke, Josef van Genabith and Andy Way, *Deriving Quasi-Logical Forms from f-structures for the Penn Treebank*, the au-

thors show how the trees in the Penn-II treebank can be associated automatically with simple Quasi-Logical Forms (QLFs). Their approach is based on combining two independent strands of work: the first is the observation that there is a close correspondence between QLFs and LFG's f-structures (van Genabith and Crouch, 1996); the second is the development of an automatic f-structure annotation algorithm for the Penn-II treebank (Cahill et al., 2002a; Cahill et al., 2002b). The approach is compared with that of (Liakata and Pulman, 2002).

In the chapter, *Which underspecification technique for what purpose?*, Harry Bunt examines a number of techniques for underspecification in semantic representations, notably labels and holes, ambiguous constants, metavariables, dominance constraints, radical reification, stores, lists and disjunctions, and *in situ* quantified terms. These techniques are considered for their usefulness in dealing with a variety of linguistic phenomena and cases of incomplete input, which have motivated the use of underspecified semantic representations. It is argued that labels and constraints, and the use of ambiguous constants and variables, have nearly disjoint domains of application, and together cover a wide range of phenomena.

In the last chapter of this group, Alex Lascarides and Nicholas Asher motivate and describe Segmented Discourse Representation Theory (SDRT) as a dynamic semantic theory of discourse interpretation, using rhetorical relations to model the semantics/pragmatics interface. They describe the syntax and dynamic semantics of the SDRT language in which logical forms are represented, a separate but related language in which semantic underspecification is expressed as partial descriptions of logical forms, and a glue logic which uses commonsense reasoning to construct logical forms, relating the semantically underspecified forms that are generated by the grammar to their pragmatically preferred interpretations. The framework is applied to examples involving anaphora and other kinds of semantic ambiguities. Being concerned with the analysis of discourse, this chapter forms a bridge to the next three chapters, that are also concerned with the semantic interpretation of discourse and dialogue.

Raquel Fernández, Jonathan Ginzburg, Howard Gregory and Shalom Lappin present the main features of SHARDS, a semantically-based HPSG approach to the resolution of dialogue fragments. This implemented system interprets short questions ('sluices') and short answers. It provides a procedure for computing the content values of clausal fragments from contextual information contained in a discourse record of previously processed sentences.

Ivana Kruijff-Korbayova and Bonnie Webber describe an investigation into the sensitivity of discourse connectives to the Information Structure (IS) of the utterances they relate, in their chapter *Interpreting concession statements in light of information structure.* They illustrate this with an analysis of connectives signaling concession, distinguishing two senses – *denial of expectation* and *concessive opposition.* Their account thus refines earlier accounts that ignore IS. This work is part of a larger enterprise aimed at understanding what role(s) sentence-level IS plays in the interpretation of larger units of discourse.

Key ingredients in the description of discourse meaning are reference markers: objects in the formal representation that the discourse is about. It is well-known that reference markers are not like first-order variables; the received view is that reference markers are like the variables in imperative programming languages. However, in a computational semantics of discourse that treats reference markers as 'dynamically bound' variables, every noun phrase will get linked to a dynamic variable, so it will give rise to a marker index. In the chapter *Context and the composition of meaning*, Jan van Eijck addresses the question where these indices come from, and how they can be handled when combining (or 'merging') pieces of discourse. He argues that reference markers are better treated as indices into context, and presents a theory of context and context extension based on this view. In context semantics, noun phrases do not come with fixed indices, so the merge problem does not arise. This solves a vexing issue with coordination that causes trouble for all current versions of compositional discourse representation theory.

In the chapter *Meaning, intonation and negation*, Marc Swerts and Emiel Krahmer outline an approach to the study of meaning and intonation. The approach focusses both on what speakers can do, using production experiments, and on what hearers can do, using perception experiments. They show that such an experimental paradigm may yield interesting results from a semantic point of view, discussing the role intonation can play for the interpretation of negation phrases in natural language. Empirical evidence is presented for the existence of a set of prosodic differences between two kinds of negations, descriptive and metalinguistic ones. This distinction has been the subject of considerable debate in presupposition theory and also plays an important role in discussions about the division of labor between semantics and pragmatics. In general, it is argued that intonation gives rise to 'soft constraints', and that an optimality-theoretical framework may be suitable to model the relation between intonation and meaning.

Myroslava Dzikovska, Mary Swift and James Allen in their chapter *Customizing meaning: Building domain-specific semantic representations from a generic lexicon*, argue that language input to practical dialogue systems must be transformed into a semantic representation that is customized for use by the back-end domain reasoners, while at the same time one wants to keep front-end system components as domain-independent as possible for easy portability across multiple domains. They propose a transparent way to achieve domain-specificity from a broad-coverage domain-independent parser. They define a set of mappings from ontologies into domain-specific knowledge representations, and use these mappings both to customize the semantic representations output by the parser for the reasoners, and to specialize the lexicon to the domain – which improves parsing speed and accuracy. This method facilitates instances of semantic type coercion common in many domains by combining lexical representations with domain-specific constraints on interpretation.

The chapter by Aravind Joshi, Laura Kallmeyer and Mariel Romero addresses the problem of formulating constraints for relative quantifier scope, in particular in inverse linking readings where certain scope orders are excluded. They show how to account for such restrictions in the Tree Adjoining Grammar (TAG) framework by adopting a notion of 'flexible composition'. In the semantics used for TAG they introduce quantifier sets that group quantifiers that are 'glued' together in the sense that no other quantifier can scopally intervene between them. The flexible composition approach allows them to obtain the desired quantifier sets and thereby the desired constraints for quantifier scope.

The next three chapters are concerned with the expression of time in natural language. In the first of these, *Serious computing with tense*, Fabrice Nauze and Michiel van Lambalgen describe a comprehensive proposal for dealing with time and events as expressed in natural language. They argue that the simple davidsonian addition of time and event variables to predicates in the representation language is insufficient for reasoning about time and events, and argue that a theory of time and events should have great expressive power and be presentable in axiomatic form, so that it is entirely clear what it predicts and what it doesn't. They argue that the event calculus that has been developed in robotics (Shanahan, 1997) has all the desired properties. It allows one to formulate a goal and a causal theory of the domain. Based on the causal theory, a plan for reaching that goal can be inferred. In the version of the event calculus proposed by van Lambalgen and Hamm (2003), the inference mechanism is constraint logic programming with

negation as failure. The authors propose to apply this formalism to tense and aspect, since goals seem to play a prominent part there. For example, a profitable way to formulate the meaning of accomplishments is to specify a goal and a causal theory, which together yield a plan which achieves the goal when no unforeseen circumstances occur. This prevents the so-called 'imperfective paradox' to arise in case the goal is for some reason never achieved.

The second chapter in this group, by James Pustejovsky, Robert Knippen, Jessica Litman and Roser Saurí, *Temporal and event information in natural language text*, discusses the role that temporal information plays in natural language text, specifically in the context of question-answering systems. A descriptive framework is defined for examining the temporally sensitive aspects of natural language queries. The properties are investigated that a general specification language would need to have, in order to mark up temporal and event information in text. The language TimeML is presented, a rich specification language for event and temporal expressions in natural language text. The chapter shows the expressiveness of TimeML for a broad range of syntactic and semantic contexts, and demonstrates how it can play an important part in the development of more robust question-answering systems.

In the third of this group of chapters, *Finite-state descriptions for temporal semantics*, Tim Fernando outlines finite-state descriptions for temporal semantics through which to distinguish 'soft' inferences, reflecting manners of conceptualization, from more robust semantic entailments defined over models. Fernando argues that just what descriptions are built (before being interpreted model-theoretically) and how they are grounded in models of reality, explains upon examination why some inferences are soft and others are robust.

The next two chapters are both concerned with semantic aspects of language generation. Claire Gardent and Kristina Striegnitz in their chapter *Generating bridging definite descriptions* focus on the role that knowledge based reasoning plays in the generation of definite descriptions. Specifically, they propose an extension of Dale and Reiter's incremental algorithm which covers not only directly anaphoric descriptions, but also indirect and associative anaphora. Starting from a formalism independent algorithm, they further show how this algorithm can be implemented using description logic.

Kees van Deemter and Emiel Krahmer explore in their chapter *Graphs and booleans: On the generation of referring expressions* how a graph-theoretical perspective may be brought to bear on the generation

of *complex* referring expressions. The motivation for this exploration was that, if each of these types of referring expressions can be addressed using one and the same formalism, then this will make it easier to compare and, ultimately, to combine them into one unified algorithm. They sketch how relations, vague properties, and Boolean operators have been tackled by earlier algorithms, and ask how these algorithms can be formalised using a graph-theoretical approach. it is shown that most of the existing algorithms carry over without difficulty, through the technique of making implicit properties explicit in the knowledge base. However, in the case of one algorithm (which focusses on the generation of Boolean descriptions that also contain relational properties), this strategy turns out to be problematic. For this case, a new algorithm is presented, based on partitioning the target set, which can be implemented in a graph-theoretical formalism without difficulty.

The last two chapters of the book return to issues of underspecification and interpretation in context. In the first of these, *Efficient computation of overlay for multiple inheritance hierarchies in discourse modeling*, Jan Alexandersson and Tilman Becker note that default reasoning has been shown to be a convenient means for interpreting user utterances in context and introduce 'overlay', the combination of default unification and a scoring function where the latter is used for computing the degree of similarity between new and old information. In this work they continue their efforts for default unification of typed feature structures by giving an efficient algorithm for multiple inheritance hierarchies. The main contribution of this chapter is that, contrary to previous suggestions, most of the computation can be done on the type hierarchy. The scoring function is adapted accordingly.

The final chapter, by Dick Crouch, Anette Frank and Josef van Genabith is concerned with an application of underspecified semantic representations, namely ambiguity-preserving machine translation. In earlier work (van Genabith et al., 1998), the authors developed an approach where transfer takes place on the glue language meaning constructors of (Dalrymple et al., 1996); unfortunately, that approach was unable to deal with structural misalignment problems, such as embedded head switching, in a satisfactory way. This chapter proposes the use of a fragment of linear logic as a transfer formalism, and shows how it provides a more general and satisfactory solution to these problems.

## References

Bunt, H.: 1989. Dynamic Interpretation in Text and Dialogue. In H. Bouma and B. Elsendoorn, editors *Working Models of Human Perception*. New York: Academic Press. pp. 81–150.

Bunt, H.: 2000. Dialogue pragmatics and context specification. In H. Bunt and W. Black, editors *Abduction, Belief and Context in Dialogue. Studies in Computational Pragmatics*. Benjamins, Amsterdam. pp. 419–455.

Cahill, A., M. McCarthy, J. van Genabith and A. Way: 2002a. Automatic Annotation of the Penn Treebank with LFG F-Structure Information. In *Proc.LREC Workshop on Linguistic Knowledge Acquisition and Representation.*

Cahill, A., M. McCarthy, J. van Genabith and A. Way: 2002b. Evaluating Automatic F-Structure Annotation for the Penn-II Treebank in *Proceedings of the Treebanks and Linguistic Theories Workshop*, Sozopol, Bulgaria.

Dalrymple, M., Lamping, J., Pereira, F.C.N. and Saraswat, V.: 1996. Quantification, anaphora, and intensionality. *Journal of Logic, Language and Information* 6(3) 219–273.

Frazier, L. and K. Rayner: 1990. Taking on Semantic Commitments: Processing Multiple Meanings vs. Multiple Senses. *Journal of Memory and Language* 29, 181–200.

Genabith, J. van, and D. Crouch: 1996. Direct and Underspecified Interpretations of LFG f-Structures. In *Proceedings COLING 96*, Copenhagen, Denmark, pp. 262–267.

Genabith, J. van, Frank, A. and Dorna, M.: 1998. Transfer constructors. In Butt, M. and King, T. H., editors, *Proceedings of the LFG'98, Conference Brisbane, Australia*. CSLI Publications, pp. 190–205.

Groenendijk, J., and M. Stokhof: 1985. Dynamic Predicate Logic. ITLI Report, Amsterdam: ITLI.

Groenendijk, J., and M. Stokhof: 1989. Context and Information in Dynamic Semantics. In H. Bouma and B. Elsendoorn, editors *Working Models of Human Perception*. New York: Academic Press. pp. 457–486.

Liakata, M. and S. Pulman: 2002. *From trees to predicate-argument structures.* COLING'02, Proceedings of the Conference, Taipei.

Lambalgen, M. van, and F. Hamm: 2003. Moschovakis' notion of meaning as applied to linguistics. In M. Baaz and J. Krajicek, editors, *Logic Colloquium '01.* ASL Lecture Notes in Logic, Wellesley, MA.: A.K. Peters.

Shanahan, M.P.: 1997. *Solving the frame problem.* The M.I.T. Press, Cambridge MA.

Smith, R. and J. van Kuppevelt editors: 2003. *Current and new Directions in Discourse and Dialogue*. Dordrecht: Kluwer.

Traum, D., and S. Larsson: 2003. The Information State Approach to Dialogue Management. In R. Smith and J. van Kuppevelt, editors,: 2003, *Current and new Directions in Discourse and Dialogue*. Dordrecht: Kluwer, pp. 325–353.

MASSIMO POESIO, UWE REYLE AND ROSEMARY STEVENSON

# JUSTIFIED SLOPPINESS IN ANAPHORIC REFERENCE

## 1. Motivations

*The reason why I hate critics ... is that they write sentences like this:*

*.... Flaubert does not build up his characters, as did Balzac, by objective cultural description; in fact, so careless is he of their outward appearance that on one occasion he gives Emma brown eyes (14); on another, deep black eyes (15); and on another blue eyes (16).*

*I must confess that in all the times I read* Madame Bovary*, I never noticed the heroine's rainbow eyes. Should I have? Would you? Put it another way: IS THERE A PERFECT READER SOMEWHERE, A TOTAL READER?*

(From J. Barnes, *Flaubert's Parrot*, Picador, 1984, p. 74–76)

The quote above expresses an intuition shared by many computational semanticists: namely, that readers and listeners do not always seem to construct complete interpretations of everything they read or hear. The possibility that utterance meanings may occasionally be 'incomplete' or 'underspecified' may lead to a fundamental rethink of traditional ideas about how meaning is constructed, and has therefore generated a great interest in recent years, testified, e.g., by the collection (van Deemter and Peters, 1996), by a special issue of the *Journal of Semantics* on this topic, and by a number of workshops. Computational semanticists, primarily concerned with lexical and scopal underspecification, focused on developing logical characterizations of the type of interpretation that may be assigned to an utterance when its meaning remains underspecified. As a result, it has been shown that it is possible to provide a logical characterizations of the complete space of possible interpretations (see, e.g., (Alshawi and Crouch, 1992; van Eijck and Jaspars, 1996; Muskens, 1995; Pinkal, 1995; Poesio, 1991; Poesio, 1996; Reyle, 1993; Reyle, 1996) as well as the papers in van Deemter and Peters, 1996).

However, not much empirical evidence has yet been found supporting this intuition, except perhaps for the case of lexical underspecification (Frazier and Rayner, 1990; Copestake and Briscoe, 1995). On the contrary, there is a lot of evidence suggesting that a number of semantic



*H. Bunt and R. Muskens, (eds.), Computing Meaning, Volume 3*, 11–31.

interpretive processes take place immediately, just like syntactic interpretation does. (Well-known results concerning the incrementality of semantic interpretation are discussed in (Swinney, 1979) for lexical interpretation and (Tanenhaus et al., 1995) for anaphora resolution.) Our long-term goal is to examine the evidence for and against underspecification in anaphora resolution by identifying cases in which anaphoric expressions, and especially pronouns, are not completely interpreted, and determining the interpretation they receive. (We focus on anaphoric interpretation rather than scope disambiguation – the area that has motivated the most work on underspecification in computational semantics (Reyle, 1993; Poesio, 1994; Muskens, 1995; Pinkal, 1995) – because empirical evidence on anaphora resolution is much easier to collect.) Our research combines corpus analysis and more traditional psychological experimentation: using corpus analysis first to identify contexts in which pronouns may remain underspecified, then running controlled psychological experiments to verify whether indeed this is the case.

In this chapter we first discuss the results of a study of a corpus of task-oriented spoken conversations that led us to identify contexts in which an apparently ambiguous anaphoric expressions does not seem to result in a problem being signalled by the other conversational participant. We then propose a preliminary hypothesis concerning what these cases have in common, and finally discuss psychological experiments supporting our hypothesis.

## 2. Underspecification in Reference: Psychological Evidence

Whereas work on underspecification in computational syntax such as (Sturt and Crocker, 1996) is supported to some extent by empirical evidence – e.g., on the greater or lesser facility of certain syntactic reanalysis processes – the work on underspecification in computational semantics has been driven much less by psychological results, since such evidence was, until recently, pretty minimal. There is, however, an increasing number of studies pointing out cases in which aspects of semantic interpretation are not immediately resolved. These studies are discussed in some length in (Poesio, 1999), to appear as (Poesio, to appear); a useful summary of the psycological evidence is in (Sanford and Sturt, 2002).

The best-known results concerning semantic underspecification are the studies by (Frazier and Rayner, 1990). Frazier and Rayner observed that cases of homonymy like *pitcher* or *records* could be experimentally separated from cases of polysemy like *newspaper*: whereas words belonging to the first class would originate garden paths when subsequent context disconfirmed the preferred interpretation at the point the word (e.g., *record*) was encountered, as in (3d), no garden paths were observed for polysemous words like *newspaper* in (4d).

(3) a. *After they were scratched, the records were carefully guarded.*

b. *After the political takeover, the records were carefully guarded.*

c. *The records were carefully guarded after they were scratched.*

d. *The records were carefully guarded after the political takeover.*

(4) a. *Lying in the rain, the newspaper was destroyed.*

b. *Managing advertising so poorly, the newspaper was destroyed.*

c. *Unfortunately the newspaper was destroyed, lying in the rain.*

d. *Unfortunately the newspaper was destroyed, managing advertising so poorly.*

For the case of anaphoric reference, the evidence is mixed. On the one hand, it is known (see, e.g., Tanenhaus et al., 1995) that definite descriptions referring to objects in the visual situation are interpreted immediately. It is also known, however, that the case with anaphoric definite descriptions and other anaphoric expressions such as pronouns is more complex. Although the interpretation of pronouns begins immediately, anaphoric pronouns in ambiguous contexts (i.e., with multiple same-gender potential antecedents) remain uninterpreted until the end of the sentence (Gernsbacher and Hargreaves, 1988). And some evidence suggests that pronouns not referring to a focal entity are not immediately interpreted (Garrod and Sanford, 1985; Garrod et al., 1994). Garrod et al. (1994), for example, tested the effect of gender, focusing, and verb bias using an eye-tracker and materials like those in (5). A context paragraph was used to establish either a male or a female entity as focus, and to introduce a second entity whose gender either matched or didn't match that of the focused entity. Then a target

sentence was presented, containing either a masculine or a feminine pronoun, and a verb biased either towards the focused entity (*sank*) or towards the second entity (*jumped*).

(5) A dangerous incident in the pool
**Elizabeth**$_1$ / **Alexander**$_1$ *was an inexperienced swimmer and wouldn't have gone in if* **the male lifeguard**$_2$ *hadn't been standing by the pool. But as soon as she*$_1$*/he*$_1$ *got out of her*$_1$*/his*$_1$ *depth she*$_1$*/he*$_1$ *started to panic and wave her*$_1$*/his*$_1$ *hands about in a frenzy.*

a. *Within seconds, she sank into the pool.*
b. *Within seconds, he sank into the pool.*
c. *Within seconds, she jumped into the pool.*
d. *Within seconds, he jumped into the pool.*

First-pass reading times for the verb indicated that conflicts between the interpretation of the pronoun and the verb (as in (5c)) were only immediately detected, resulting in a slowed reading time for the verb, when both gender information and focus information converged on the interpretation of the pronoun; otherwise, the conflict was only detected later. According to Garrod et al., this suggested that unless both these conditions were satisfied, the interpretation of the pronoun was delayed.

## 3. Finding (Anaphoric) Underspecification in Corpora

### 3.1. *Methodology*

How is it possible to use a corpus to find cases in which aspects of the meaning of an utterance – in our case, an anaphoric expression – remain underspecified? Our method has been to analyze task-oriented conversations from the TRAINS corpus collected at the University of Rochester[1], and to look for cases in which (i) more than one potential antecedent of an anaphoric expression matches it in gender and number, (ii) no focusing principle we are aware of makes one of the interpretations preferred, and yet (iii) the recipient of the utterance is able to accomplish the task without signalling a problem. The assumption underlying this approach is that in task-oriented conversations, unlike so-called 'cocktail-party' situations, the participants need to signal when they didn't understand something, as S does in 24.5 in (6):

[1] See `http://www.cs.rochester.edu/research/speech/trains.html`

(6)
```
23.7    M: what would be faster
23.8     : to send
23.9     : an engine
23.10    : from Elmira
23.11    : to
23.12    : ... one of the boxcars
23.13    : or from
23.14    : Avon
24.1    S: well there's
24.2     : there's a boxcar
24.3     : already _at_ Elmira
         [3sec]
24.4     : and
24.5     : t / YOU MEAN TO GO TO CORNING
```

We identified several patterns of pronominal use that satisfy these conditions: three annotators (including two of the authors) agree that a use of a pronoun is ambiguous, and the absence of repair signals (such as *Sorry, what did you mean?*, or *I didn't understand*) indicates that the listener didn't seem to have a problem with the expression being ambiguous. We discuss these patterns in the rest of this section, and the conclusions that we drew from these cases in the next.

### 3.2. *References to Mereologies*

A first clear pattern emerging from the corpus is illustrated by the following example (from dialogue d91-3.1):[2]

(7)
```
3.1     M: can we .. kindly hook up
3.2      : uh
3.3      : engine E2 to the boxcar at ..
           Elmira
4.1     S: ok
5.1     M: +and+ send it to Corning
5.2      : as soon as possible please
```

[2] The same person plays the system's role in all TRAINS-91 dialogues; 8 different subjects play the manager role, and each of them is involved in two dialogues. The first part of the identification code of the dialogue says which speaker was involved; the second which dialogue it was - for example, d91-1.1 is the first dialogue for speaker 1, d91-1.2 is the second dialogue for speaker 1, d91-2.1 is the first dialogue for speaker 2, etc.

```
6.1     S: okay
           [2sec]
7.1     M: do let me know when it gets
           there
8.1     S: okay it'll /
8.2      : it should get there at 2 AM
9.1     M: great
9.2      : uh can you give the
9.3      : manager at Corning instructions
           that
9.4      : as soon as it arrives
9.5      : it should be filled with
           oranges
10.1    S: okay
10.2     : then we can get that filled
```

In this example, it's not clear whether the pronoun *it* in 5.1 refers to *the engine E2* which has been hooked up to *the boxcar at Elmira*, to the boxcar itself, or indeed whether that matters. It's only at utterance 9.5 that we get evidence that *it* probably referred to *the boxcar at Elmira*, since it is only boxcars that can be filled with oranges; yet, if anything, focusing theories would predict engine E2 to be the antecedent, since engine E2 is the direct object, the THEME, and comes first (Sidner, 1979; Stevenson et al., 1994; Grosz et al., 1995; Poesio and Stevenson, To appear).

Note that this type of pronoun use cannot be viewed as an example of vagueness, at least not according to the standard vagueness tests (Lakoff, 1970; Zwicky and Sadock, 1975). Whereas the ellipsis test, for example, would suggest that *a glove* in (8a) is indeterminate, because it's possible for John to have lost its left glove and for Bill to have lost its right one, the same test applied to *it* in (8b) suggests that this expression is ambiguous, in that it's not possible to interpret *Then, John should check if IT gets to Bath in time, and Bill should too* as meaning that John should check that the engine gets to Bath in time, whereas Bill should check that the boxcar gets there in time:

(8) a. *John lost a glove, and Bill did too*

b. *Let's hook the engine to the boxcar.*
*Then, John should check if IT gets to Bath in time, and Bill should too*

Another context also apparently leading to unproblematic use of pronouns with multiple matching antecedents is illustrated by (9). In this example, the pronoun *that* in 27.4 may refer either to the orange juice or to the tanker car in which the orange juice has been loaded:

(9)
```
26.1    S: okay
27.1    M: so then we'll
27.2     : ... we'll be in a position to
27.3     : load the orange juice into the
           tanker car
27.4     : ... and send that off
```

In order to characterize in a more systematic fashion the possible interpretations of *that* in 27.4 in this last example (and of *it* in (7), 5.1) we will borrow some notation from Link (1983). We will write $oj \oplus tc$ to indicate the object that has $oj$ and $tc$ as subparts, and $a \lhd b$ to say that $a$ is a mereological part of $b$.[3] With this notation, we can formalize the first and most obvious property of examples (7) and (9): namely, that actions like *hooking up* and *loading* are performed that create a new object $a \oplus b$ out of the potential antecedents $a$ and $b$ (e.g., $oj \oplus tc$ in (9)).

The second property of these examples is that four interpretations for the pronominal expression are possible. The complete list of the possible interpretations of *that* in (9), 27.4 is:

$$that = oj,\ tc,\ oj \oplus tc, \text{ or an indeterminate } x \lhd (oj \oplus tc)$$

This latter interpretation ($x \lhd (oj \oplus tc)$) is what has been called a P-UNDERSPECIFIED INTERPRETATION in (Poesio, 1999) – i.e., a 'disjunctive' interpretation that 'covers' all of the alternative interpretations, similar to those proposed for certain cases of lexical polysemy in (Copestake and Briscoe, 1995). We will hypothesize below that the existence of such an underspecified interpretation may be a further important property of these contexts.

The third property that these examples have in common is that both in situations involving attaching two objects together and in situations involving loading objects into other objects, all of the alternative interpretations of the anaphoric expression are equivalent as far as the plan (moving these objects to a new location) is concerned: after the two

[3] Link uses $\leq$ to indicate the ordering relation on parts for which we use the notation $\lhd$.

explicitly mentioned potential antecedents are joined, if one of them gets moved, the other one must be moved as well. E.g., in (9), 27.4, all interpretations of the instruction *send that off* will achieve the same result irrespective of how the pronoun is interpreted. We will write X $\sim$ Y to indicate that interpretation X is equivalent to interpretation Y for the purpose of the plan: i.e., we write

$$oj \sim tc$$

to indicate that from the point of view of the plan, interpreting the pronoun *that* as referring to the orange juice or the tanker car are equivalent. Similarly, in the case of (7), we will write $e \sim b$ to say that from the point of view of the plan, the interpretation of *it* in which it refers to engine E2 and that in which it refers to the boxcar are equivalent.

### 3.3. *Reference to plans*

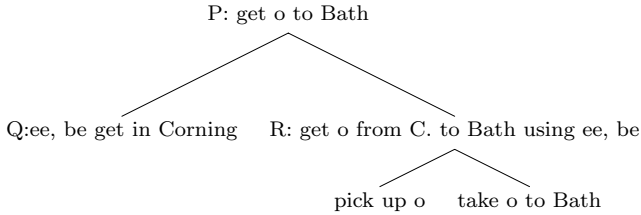
A second class of pronominal uses in the TRAINS dialogues satisfies our three conditions (ambiguity, no preferences for one interpretation, and no complaints). These are the uses of demonstratives such as *that* to refer to parts of a plan, as in the following example (from TRAINS dialogue d91-1.1):

(10)
```
11.6     : aha
11.7     : I see an engine and a boxcar both
             at Elmira
12.1    S: right
13.1    M: this looks like the best thing
             to do
13.2     : so we should get
13.3     : ... the eng / engine to picks up
             the boxcar
13.4     : and head for Corning
13.5     : 's that sound reasonable
14.1    S: sure
14.2     : that sounds good
15.1    M: and from Corning we'll pick up
             the oranges
15.2     : and um
15.3     : take them to Bath
15.4     : will it / that get m / me
```

```
15.5     : do you think that I can get  ..
           this all over to Bath by 8 o'clock
16.1    S: yeah
16.2     : that gets us to Bath at f / 5 AM
16.3     : so it's plenty of time
```

The demonstratives in question are *that* in 15.4 and *that* in 16.2. Roughly speaking, the structure of the plan at this point in the conversation can be represented as follows:

(11)




but from the transcript it is unclear whether these two demonstratives refer just to the portion of the plan consisting of the action of picking up the oranges and taking them to Bath (or perhaps just to one of the two actions, say the second), or to the larger plan which also includes the previous action of getting the engine to pick up the boxcar and heading for Corning. Our two other conditions are met as well: no focusing principle we are aware of suggests one interpretation over the other, and the other participant does not complain.

These examples do not look like cases of vagueness, either, as shown by the fact that (12) does not have the interpretation that John could agree to a plan, whereas Bill would agree to a subpart of it:

(12) *John agree to THAT PLAN/THAT, and Bill did too*;

The diagrammatic representation of the plan in (11) already hints at a further semantic property that plans share with the antecedents of the previously discussed class of pronominal use: plans, at least as classically viewed in the Artificial Inteligence literature, have a mereological structure as well. According to this view of plans, plans denote event or action types, and these in turn can be seen as sets of actions (of the appropriate type). So, for example, plan R in (11) denotes the set of events of getting oranges from Corning to Bath using the engine and the boxcar,

$$\mathrm{R} = \{e|e{:}\mathbf{get}(\mathrm{o,c,ba,ee,be})\}$$

where *o* refers to the oranges, *c* refers to Corning, *ba* refers to Bath, *ee* refers to the engine at Elmira, and *be* to the boxcar at Elmira. Plan P refers to the set of events of getting oranges to Bath:

$$\mathrm{P} = \{e' | \exists x, y, z\ e' : \mathbf{get}(\mathrm{o,x,ba,y,z})\}$$

Intuitively, set R is a subset of set P. This allows us to define a 'part-of' relation $\lhd$ between plans, to be interpreted as PLAN DECOMPOSITION, as follows. Let $\subseteq$ be the relation of inclusion between events; then R is a part of P iff for every event $e$ in R there is an event $e'$ in P such that $e \subseteq e$, as follows:

$$R \lhd P \equiv [\forall e \in R, \exists e' \in P e \subseteq e']$$

With this interpretation of plans, we can see that the potential antecedents of the ambiguous demonstratives are part of a mereological structure similar to that observed in the previous examples, and that there is a similar range of possibilities concerning the interpretation derived by the listener. In particular, a p-underspecified interpretation is available, as in the previous cases. One crucial difference is that in the case of reference to plans one interpretation does not seem available: this is the interpretation in which the pronoun refers to Q. This appears to be yet another instance of the so-called 'right frontier constraint' often discussed in the literature on references to abstract objects (Webber, 1991). As a result, the p-underspecified interpretation now appears to be further constrained, as well: the antecedent $z$ of the pronoun is not merely dominated by the supremum of the current plan, P; but it's also part of its right frontier:

$$z \lhd {}^{*}P,\ z \in RF(P)$$

What's more, these last examples have a further, and crucial, similarity to the mereology cases discussed earlier: in these cases, as well, we can say that the two possible interpretations of the relevant utterance (e.g., 15.4, *that gets us to Bath at 5 AM*) are equivalent for the purpose of the plan. This is because once the part of the plan being proposed by M in 13.1–13.5, Q, has been accepted by S (utterances 14.1–14.2), whether or not the plan as a whole (P) is going to work depends entirely on whether subplan R is going to work; so accepting R is essentially equivalent to accepting P as a whole:

$$P \sim R$$

## 4. The Justified Sloppiness Hypothesis

The discussion in the previous section makes it clear that the two patterns of anaphoric reference we have observed in the TRAINS dialogues have at least three aspects in common:

1. Both explicitly mentioned potential antecedents $x$ and $y$ are elements of an underlying mereological structure with summum $\sigma = x \oplus y$ which has been explicitly constructed (and made salient) in the dialogue ($\sigma = oj \oplus tc$ in (9), $\sigma = P$ in the case of (10));
2. the existence of this structure makes it possible to construct a p-underspecified interpretation in which the anaphoric expression is interpreted as denoting an element $z$ included in the mereological structure - i.e., part-of its summum $\sigma$:

$$\begin{array}{|l|}\hline x\ y\ \sigma\ z \\ \hline \ldots \\ \sigma = x \oplus y \\ z \lhd {}^{*}\sigma \\ \ldots \\ \hline \end{array}$$

3. All possible interpretations $(x, y, z, x \oplus y)$ are equivalent for the purposes of the plan.

This suggest the following preliminary hypothesis:

Ambiguous anaphoric expressions are not perceived as infelicitous provided that Conditions 1–3 hold.

This may be because if these three conditions hold, the speaker's sloppiness in using an anaphoric expression in an ambiguous context is not problematic; we will therefore use the term JUSTIFIED SLOPPINESS to indicate cases such as those discussed in the previous section, and refer to the hypothesis above as **Justified Sloppiness Hypothesis**, or JSH.

Of course, the fact that a p-underspecified interpretation exists does not mean that the listener will adopt it as its final interpretation; however, this possibility is what makes these examples interesting from an underspecification perspective.

## 5. Additional cases of justified sloppiness

### 5.1. *Events related by generation*

After identifying the cases discussed above, we discovered that similar examples of potentially ambiguous anaphoric expressions which, however, did not appear to be problematic for the reader had already been discussed by Schuster (1988), albeit not in connection with the question of whether anaphoric reference is underspecified or not.[4] Schuster analyzes two types of data: dialogues between an expert and a novice attempting to learn how to use the Emacs editor, and questionnaires asking subjects to indicate the preferred referents of anaphoric expressions referring to events. The transcripts of Emacs dialogues include several examples of reference to events, the most interesting among which for our purposes are examples like the following (our own indices):

(13) a. E: Do this: [$_1$ set a "mark" at some point (any old point) by [$_2$ typing <esc>-M]]. It will say "mark set". Try it.

b. E: <esc>-M will give set-mark. Try it.

Schuster uses such examples to argue for a representation of events in which an event may GENERATE another event, such as those developed by Goldman (1970) and Pollack (1986). Examples like those in (13) are used as diagnostics, in the sense that the possibility of using a single pronoun indicates that the two events are related by a generation relation:

> In both cases [our (13a) and (13b), NDR] the referent(s) of the pronoun *it* can be either "setting the mark" or "typing <esc>-M" or even both: "setting the mark by typing <esc>-M". ... we can claim that "typing <esc>-M at a given time" can generate "setting the mark at that given time" ... ((Schuster, 1988), p.9–10).

Schuster also observes that these references appear to be unproblematic; her explanation is:

> This relationship [generate] allows us to establish a connection between "typing <esc>-M" and "setting the mark" and it can be understood as one relationship ... When the pronoun *it* is used as is the case in both examples, neither of the two referents need to be specified because the generation relationship indicates that they are both related to each other ...

[4] These examples were pointed out to us by Bonnie Webber.

She also argues that if the generation relation is not properly established, such references may turn out to be ambiguous, as in the following example:

(14) Set the mark at the beginning of the region. Type <esc>-M and once you've done $that_1$, move to the end of the region.

These cases bear a considerable resemblance to those discussed in the previous section: two actions that are closely 'tied together' and as a result reference to the one becomes equivalent to a reference to the other. Using *tem* to indicate "typing <esc>-M" and *stm* for setting the mark, we can say again that the two interpretations for the discourse entity $z$ for *that* are equivalent:

$$tem \sim stm$$

The difference in this case is that instead of a 'part-of' relationships between the actions, as in the case of reference to plans discussed earlier, we have a much tighter relationship: generation makes the two actions almost into a single action. As a result, in this case we cannot even talk of 'sloppy' references: these references are perfectly accurate. We can still assume that the pronoun gets assigned an underspecified interpretation as discussed above, but this interpretation is almost not 'underspecified':

$$z \lhd^* (stm \oplus tem),\ \mathbf{generates}(stm, tem)$$

### 5.2. *An unclear case: References to spatial areas*

We are also aware of a few cases that cannot clearly be reconduced under our generalization, either because it isn't clear whether the reference was truly ambiguous or merely vague, or because of lack of evidence concerning whether readers truly find these references unproblematic. Poesio and Vieira (1998) report that human subjects do not agree on the interpretation of definite descriptions such as *the area* in the following example:

(15) *About 160 workers at a factory that made paper for the Kent filters were exposed to asbestos in the 1950s. Areas of the factory were particularly dusty where the crocidolite was used. Workers dumped large burlap sacks of the imported material into a huge bin, poured*

> *in cotton and acetate fibers and mechanically mixed the dry fibers in a process used to make filters. Workers described "clouds of blue dust" that hung over parts of the factory, even though exhaust fans ventilated the area.*

Three subjects were asked to indicate the antecedent of this description in the text. One subject indicated *parts of the factory* as the antecedent; another indicated *the factory*; and the third indicated *areas of the factory*.

In this example, again, we have an underlying mereological structure: both *parts of the factory* and *areas of the factory* are obviously included in the total area of the factory. There are, however, a few problematic issues in these examples. First of all, in this case there is no obvious equivalence between the three different interpretations. Furthermore, in this case one could argue that the object referred to – the area – is not in focus; to the extent that one can say that there is a focus in this text, it is most likely the factory. So, given the results of the experiments of Garrod et al., this is perhaps the example in which it is most likely that the reader did not even attempt to construct an interpretation for the anaphoric expression.

## 6. The Limits of Corpus Analysis

It is a good idea to stop at this point and think again about which questions we would like to see answered, and what we can expect from an analysis of corpus data like the one we just presented. With corpus analysis, we can identify contexts in which 'sloppy references' take place, and commonalities between the semantic interpretations produced in each case. But it is important to realize that most of the important questions cannot be answered this way.

In all of the examples we discussed the speaker is taking a risk, yet the listener does not signal a problem in understanding. An interesting question raised by this observation is whether the speaker is simply being sloppy, or he/she has done what Hobbs (1985) would call 'collapsing a complex theory in one of coarser granularity' – i.e., he/she is aware that there are two possible interpretations, but is also aware that the two interpretations are equivalent. This is one example of a question that cannot be answered by corpus analysis – or indeed, by any other technique we are aware of, because doing so would require reading the mind of the speaker.