

SDV Vol.28.1-2004

SDV. Sprache und Datenverarbeitung
International Journal for Language Data Processing
Heft 1/2004

Inhalt

Henk Zeevat und Manfred Stede

Introduction

Spenader, Jennifer

Using Simple Word-Alignment Measures to Study Discourse Particles

Alexandris, Christina and Stavroula-Evita Fotinea

Discourse Particles: Indicators of Positive and Non-Positive Politeness in the Discourse Structure of Dialog Systems for Modern Greek

Alonso, Laura ; Shih, Jennafer; Castellón, Irene; Padró, Lluís

An Analytic Account of Discourse Markers for Shallow NLP

Soffner, Martin

Tag questions and but in dialogue

Thomas, Kavita E.

Modelling Plan-Based and Expectation-Based “But” in Dialogue

Karagjosova, Elena

German *doch* as a marker of given information

Eckardt, Regine

Particles and Strategies

Schmitz, Hans-Christian und Bernhard Schröder

Updates with “eigentlich”

Introduction

Autor:

Henk Zeevat und Manfred Stede

Aufsatztitel: Introduction

Jahrgang:

28

Heft:

1 (2004)

Seiten:

5-7

Abstract:

Discourse particles are more and more investigated from a formal perspective. This is due to the fact that a proper formal understanding of them is related to a number of strategic goals. Discourse particles seem to challenge some foundational tenets of NL semantics and pragmatics as it is currently taught. A curious semantic property of particles is that they seem to stand in semantic opposition with the sentence without them, which seems to entail or implicate that the marked property is false. This goes directly against the compositional picture where the meaning of a complex is a function of the meaning of its parts (and not of their non-parts). A similar foundational question is whether particles can be regarded as presupposing whatever they express or whether they conventionally implicate that (and what that would mean). There are various systematic ways in which they seem to differ from other presupposition triggers and a proper understanding of that difference and of the way in which they do function could lead to deeper insights in the nature of presupposition and/or conventional implicature. Finally, particles are simple representatives of a class of grammaticalisation processes, and a better understanding promises progress in the understanding of grammaticalisation.

In computational linguistics, a better understanding of particles would contribute to the naturalness of automatically generated text. At the moment, intuition somehow tells us that a generated sentence is unnatural because some particle must be inserted. But it is very easy to come up with rules for inserting particles that make the text even less natural than it was without the particle. Hence, particle selection and placement have to be carefully synchronized with other generation decisions. In NL understanding, it is generally assumed that one would do better with particles as they would help in understanding what move the speaker or writer is making. This in turn has a strong bearing on inferring the discourse relation, the resolution of pronouns and presuppositions, and the meaning of the tense and aspect operators. Generation and understanding come together in spoken dialog systems, where indeed particles play a major role, since it is well known that in spoken language they are even more abundant than in writing. – All in all, progress with the particles (or their translation equivalents) seems to open up perspectives in linguistic theory and natural language processing that seem currently out of reach.

The papers in this volume, which grew out of a workshop held at the ESSLLI Summer School 2003 in Vienna, are representative of the formal interest in discourse particles, but also

address other issues. The most surprising new theme is the investigation of particles by means of empirical methods like corpora and experiments. The first five papers are instances of this new development, in which Spenader and Soffner are doing classical semantics/pragmatics with new methods, and Alexandris & Fotinea and Alonso et al. develop theoretical understanding in pursuit of computational goals. Spenader uses parallel corpora to test and refine various analyses about the Swedish particle „ju“ which is notorious for being hard to describe. The parallel corpus gives a variety of English lexical items or constructions that express different uses of „ju“. Alexandris and Fotinea give the theoretical results of applied research into politeness marking and speech acts, using both corpus investigation and experiments. While they aim for results that can be applied in fielded dialogue systems, they find that a group of particles have the double function of marking speech acts and expressing respect for the interlocutor.

Alonso, Castellon, Shih and Padro exploit the ease of recognition that particles provide for shallow processing of discourse structure. Annotation schemes are proposed and evaluated for drawing conclusions from particles and other discourse markers to the structure and intention of the utterances. Soffner develops an interesting methodology of corpus experimentation. The method consists in trying to find consequences from theoretical ideas that are observable in a corpus. The method is applied to the idea that „but“ always expresses denial of expectation. This would predict that „but“-clauses can never be marked as expected. This is tested on one such marking, by tantamount tag questions. (And then John sang, did he?) Thomas also studies „but“, and in particular reflects on the relationship between denial-of-expectation and frustrated-plan usages in different types of dialogs – where a frustrated-plan „but“ would cover cases like „John searched the entire building but didn‘t find his cat“, where arguably an expectation to be violated is at best well-hidden.

The last three papers use classical formal semantics/pragmatics methods and are targeted towards particular particles. They all represent a recent tendency in which real progress is made in the formal analysis of discourse particles. This seems due on the one hand to progress in the development of formal pragmatics and on the other to a wave of interest in these particles. Karagjosova gives the first working theory that unifies „doch“ and „DOCH“. This pair of particles seem to arise from each other (by accenting or deaccenting) and to be provided with opposite meanings: „doch“ marks the utterance as common ground between speaker and hearer, whereas „DOCH“ seems to indicate that the utterance is the negating part of the common ground. The crucial step is the adoption of the fine-grained distinction of Prince (1981) and Walker (1993) between evoked, inferable and brand-new. Eckhardt in a contributed squib presents a first look at particles like „dann“, „noch“, „sonst“ and „vielleicht“ (then, yet, in addition, maybe, but with particular meanings in the question environment) as they appear in questions. The question environment of particles is relatively unknown and a preliminary analysis is offered in terms of communication strategies (Büring 2003) and presupposition. Schmitz and Schröder provide a new theory of “eigentlich” (really) in which it blocks default inferences from the clause which contains “eigentlich”. The analysis is able to deal with an extensive number of examples (counterexamples seem to be entirely absent) and is formalized using Veltman’s (1996) default logic.

References

- Büring, D. (2003) On D-trees, Beans and B-accent. In *Linguistics and Philosophy* 26(5), 511-545.
- Prince, E. (1981) *Towards a taxonomy of given-new information*. In P. Cole (ed) *Radical Pragmatics*. New York.
- Walker, M. (1993) *Informational Redundancy and resource bounds in dialogue*. PhD thesis. University of Pennsylvania.
- Veltman, F. (1996) Defaults in Update Semantics. In: *Journal of Philosophical Logic* (26), 221-261.

Using Simple Word-Alignment Measures to Study Discourse Particles

Autor:

Spenader, Jennifer

Aufsatztitel: Using Simple Word-Alignment Measures to Study Discourse Particles

Jahrgang:

28

Heft:

1 (2004)

Seiten:

9-19

Abstract:

This work demonstrates that simple word-alignment measures from statistical machine translation can help in studying the contribution of discourse particles even though these lexical items often do not have recognizable translation equivalents. The aligned Europarl parallel corpora in Swedish and English was mined to extract examples of the use of the Swedish discourse particle *ju*. χ^2 tests were used to determine which unigrams and bigrams were significantly more frequent in the English translations of aligned areas where the particle appeared than in the rest of the corpus. Further, ϕ^2 tests were done on relevant significant co-occurring unigrams and bigrams to rank them in order of the strength of their relationship with the particle. These tests therefore offer a method by which intuitions obtained through introspective study of a small number of examples can be automatically corroborated with a large number of examples.

Discourse Particles: Indicators of Positive and Non-Positive Politeness in the Discourse Structure of Dialog Systems for Modern Greek

Autor:

Alexandris, Christina and Stavroula-Evita Fotinea

Aufsatztitel: Discourse Particles: Indicators of Positive and Non-Positive Politeness in the Discourse Structure of Dialog Systems for Modern Greek

Jahrgang:

28

Heft:

1 (2004)

Seiten:

22-33

Abstract:

In the present paper we propose that in Modern Greek a sub-group of discourse particles demonstrates a double function. Specifically, we propose that these discourse particles may function both as signals related to the structure of the discourse and as indicators of positive politeness, which is characteristic of Modern Greek (Sifianou 2001). Thus, this subgroup of discourse particles, which we will name “politeness markers” also constitutes a cultural element in Modern Greek dialogs. In contrast to the “politeness markers”, we additionally observe that another subgroup of discourse particles, referred to as “non-politeness markers”, function as signals indicating lack of positive politeness and pleasant attitude in Modern Greek dialogs. The behaviour of discourse particles was evaluated for the construction of a dialog system for Speech Technology applications in the domain of customer service.

An Analytic Account of Discourse Markers for Shallow NLP

Autor:

Alonso, Laura / Shih, Jennafer / Castellón, Irene / Padró, Lluís

Aufsatztitel: An Analytic Account of Discourse Markers for Shallow NLP**Jahrgang:**

28

Heft:

1 (2004)

Seiten:

35-46

Abstract:

We present a feature-based approach to the description of discourse markers (dms) oriented to automated discourse analysis for shallow Natural Language Processing (NLP). Dm describing features have been chosen based on previous work, descriptive adequacy and our concrete NLP needs and capacities. An organization of these features in dimensions of dm meaning has been inferred via data-driven techniques, and finally implemented in a computational dm lexicon.

Tag questions and but in dialogue

Autor:

Soffner, Martin

Aufsatztitel: Tag questions and but in dialogue

Jahrgang:

28

Heft:

1 (2004)

Seiten:

47-58

Abstract:

Das Ergebnis einer Korpusanalyse lässt eine Verwendung der Diskurspartikel *but* vermuten, die nicht unter eine der drei herkömmlichen Verwendungsweisen ‚semantic opposition‘, ‚denial of expectation‘ oder ‚concession‘ fällt. Dabei wird argumentiert, dass die Kontexte dieses speziellen Vorkommens eine bestimmte Bedingung nicht erfüllen, die von allen drei Analysen angenommen wird. Diese Untersuchung zielt darauf ab, genannte Analysen von *but* zugänglich zu machen für eine empirische Überprüfung an umfangreichem Datenmaterial natürlicher Sprache. Neben den Resultaten der Korpusanalyse steht also der Ansatz im Mittelpunkt, Korpusdaten für die Diskurspartikelanalyse heranzuziehen.

The result of a corpus survey suggests there is a usage of the discourse particle *but* that is covered by none of the three traditionally distinguished analyses ‘semantic opposition’, ‘denial of expectation’, nor ‘concession’: It is argued that contexts of that peculiar usage do not meet a particular condition that is necessary or assumed in the analyses. This study aims at making those semantic analyses of *but* accessible to an empirical analysis, based on a large collection of natural language data. So, besides the survey results, the paper focuses on the idea of employing corpus data for discourse particle analysis.

Modelling Plan-Based and Expectation-Based “But” in Dialogue

Autor:

Thomas, Kavita E.

Aufsatztitel: Modelling Plan-Based and Expectation-Based “But” in Dialogue

Jahrgang:

28

Heft:

1 (2004)

Seiten:

59-70

Abstract:

This paper addresses the distinction between plan-based and expectation-based “but”, focusing on dialogue-specific cross-turn “but” which relates material across speakers. We focus here on cases in which it signals denial of expectation (DofE) and concession following work by Lagerwerf (1998) and Knott (1999a). We will unify prior treatments (Thomas, 2003a; Thomas, 2003b) of plan-based and expectation-based DofE “but” and present a single algorithm that models both phenomena in the PTT (Poesio and Traum, 1998) Information State (IS, Matheson et al., 2000) model of dialogue. We also propose a novel treatment of concession in Task-Oriented Dialogue (TOD). We motivate this work by showing how it predicts speakers’ expectations and can be implemented in the IS model of dialogue to facilitate discourse understanding and natural language generation (NLG).

German *doch* as a marker of given information

Autor:

Karagjosova, Elena

Aufsatztitel: German *doch* as a marker of given information**Jahrgang:**

28

Heft:

1 (2004)

Seiten:

71-78

Abstract:

This paper investigates two uses of the German discourse particle **doch**, namely the cases in which it occurs in the Middle Field either accented or deaccented. We argue that none of these uses marks given information or common knowledge per se. More closely, we suggest that deaccented *doch* does not mark given information per se but as a result of the association of the whole proposition in its scope with the topic or Theme part of the sentence, and that accented DOCH restricts the alternatives evoked by the default sentence focus. We also argue that the common semantic denominator uniting the two uses of **doch** is the notion of contrast

which is differently manifested due to the prosodic difference: evoking an alternative proposition in the case of accented DOCH and a contrast between the givenness of the proposition in the scope of unaccented *doch* and its assumed or evident non-salience.

Particles and Strategies

Autor:

Eckardt, Regine

Aufsatztitel: Particles and Strategies

Jahrgang:

28

Heft:

1 (2004)

Seiten:

79-86

Abstract:

In this squib, I will draw attention to a family of particles that comprises at least *dann*, *noch*, *sonst*, and modal adverbs like *vielleicht*, as they occur in questions in German. I show that they serve to navigate through a strategy in the sense of Roberts and Büring. Following earlier work by Zeevat, it will be shown that these particles presuppose previous discourse. Particles like those discussed here are hence located between lexical presupposition triggers of the manage to do kind, and speech act associated particles like *ja* (Kratzer, 1999).

Updates with “eigentlich”

Autor:

Schmitz, Hans-Christian und Bernhard Schröder

Aufsatztitel: Updates with “eigentlich”

Jahrgang:

28

Heft:

1 (2004)

Seiten:

87-96

Abstract:

Das deutsche Wort “eigentlich” wird als Adjektiv, Modaladverb oder als Diskurspartikel verwendet. Wir analysieren die Bedeutung des Modaladverbs und des Diskurspartikels. Wir behaupten, dass “eigentlich” die Wahrheitsbedingungen einer Äußerung nicht ändert. Vielmehr verwendet ein Sprecher “eigentlich”, um Default-Inferenzen oder konversationelle Implikaturen zu blockieren, die anderenfalls vom Hörer vorgenommen würden. Wir bestimmen die Bedeutung des Modaladverbs im Rahmen von Veltmans Update-Semantik ([Veltman 1996]). Eine Langversion dieses Beitrags, die die Formalisierungen einschließt kann von <http://www.ikp.uni-bonn.de/ikpab/ikpab-nf08.pdf> heruntergeladen werden.

The German word “eigentlich” is used as an adjective, a modal adverb or a discourse particle. We analyse the meaning of the modal adverb and the discourse particle. We claim that “eigentlich” does not change the truth-conditions of an utterance. Rather a speaker uses “eigentlich” to block default inferences or conversational implicatures that might otherwise be drawn by the hearer. We specify the meaning of the modal adverb within Veltman’s framework of update semantics ([Veltman 1996]). A long version of this paper including formalizations can be found on <http://www.ikp.uni-bonn.de/ikpab/ikpab-nf08.pdf>.