Towards a Multidimensional Semantics of Discourse Markers in Spoken Dialogue

Volha Petukhova and Harry Bunt

v.petukhova@uvt.nl; harry.bunt@uvt.nl

Abstract

The literature contains a wealth of theoretical and empirical analyses of discourse marker functions in human communication. Some of these studies address the phenomenon that discourse markers are often multifunctional in a given context, but do not study this in systematic and formal ways. In this paper we show that the use of multiple dimensions in distinguishing and annotating semantic units supports a more accurate analysis of the meaning of discourse markers. We present an empirically-based analysis of the semantic functions of discourse markers in dialogue. We demonstrate that the multiple functions, which a discourse marker may have, are automatically recognizable from utterance surface-features using machine-learning techniques.

1   Introduction

Discourse markers are key indicators of discourse structure, and have been shown to be useful devices for (a) segmenting discourse into meaningful units, and (b) identifying relations between these units. The determination of the meanings of discourse markers is often crucial for understanding the communicated message.

Discourse markers have been studied for their role in the organization of discourse structure in larger texts ([12], [16]), in argumentative dialogues ([6]), in interviews ([15], [10]) and in dialogues that are highly interactive in nature and are characterized by rapid turn switching among participants, such as task-oriented dialogues ([9]) or meeting conversations ([14]).

The research reported in this paper regards the use of discourse markers in spoken dialogue. In dialogue, discourse markers play an important role in establishing boundaries between dialogue units and in indicating the communicative functions of such units (see e.g. [14], [9], [10]).

(1) *A1: it ties you on in terms of the technologies and the complexity that you want*
    *A2: **like** for example voice recognition*
    *A3: **because** that you might need to power a microphone and other things*
    *A4: **so** thats one constraint there*

In example (1) discourse markers are used by the speaker to indicate the steps in a sequence of arguments: he makes a statement (Inform); then he provides an example for this statement (Inform Exemplify); he justifies his choice (Inform Justification); and he draws a conclusion (Inform Conclude).

An important goal of studies of dialogue structure is to explore possible meanings and functions of discourse markers in dialogue as reflected in observable utterance features (prosodic, syntactic, lexical), to enable their successful recognition and classification.

One aspect of the meaning of discourse markers is that they may not only have a variety of semantic functions, but that they may also have several functions simultaneously – their *multifunctionality* (see [11], [1], [15] among others). This paper introduces a formal and systematic, empirically-based approach to the study of the multifunctionality of discourse markers. We show how the multifunctionality of discourse markers can be described systematically by using a multidimensional model of the interpretation of communicative behaviour in dialogue (Section 2). Section 3 introduces the analysed data and features. We illustrate the multifunctionality of discourse markers in some detail for the example of *and*, as one of the most frequently used and ambiguous dialogue markers. We provide the results of statistical and machine-learning experiments on the automatic recognizability of discourse marker meanings, and give an overview of the observed multifunctionality of markers that occur in our data (Section 4). Conclusions and perspectives for further research are outlined in the final Section 5.

2   The notion of multifunctionality

The multifunctionality of discourse markers has been described first by Schiffrin in [15]. She distinguishes between (1) ideational structure, with relations between propositions, e.g. a cohesion relation, a topic relation or a functional relation; (2) action structure, which describes the organisation and constraints on the use of speech acts; (3) exchange structure, which is 'the outcome of decision procedures by which speakers alternate sequential roles and define those alternations in relation to each other'. Schiffrin argues that discourse markers may simultaneously have roles within each of these three structures, e.g. the discourse marker *and* may 'coordinate ideas' and

'continue a speaker's action'. However, the multifunctionality of discourse markers in this study escaped extensive and formal description.

Hovy in [11] states that each discourse marker signals either (1) a 'semantic interpropositional relation', e.g. CAUSE or PART-OF, or (2) 'interpersonal intentions' (a communicative purpose of an utterance), e.g. to inform someone about a fact, or to instruct someone to do something, or (3) both. Moreover, according to Hovy each discourse marker 'articulates a rhetorical relation', such as Elaboration or Presentational-Sequence. Hovy argues that there are several parallel simultaneous structures that underlie coherent discourse and argues that an adequate description of discourse requires at least four distinct structural analyses: semantic, interpersonal/goal-oriented, attentional/thematic, and rhetorical.

This approach seems to apply very well to the analysis of the meaning of discourse markers in dialogue. Discourse markers may have various communicative purposes (also called communicative functions) in dialogue with respect to the underlying task or goal, attention, topic or arguments, turn management, etc. We only want to add that discourse markers may have various communicative functions *simultaneously*. For example, if the speaker wants to provide additional or more detailed information about something that he mentioned before, he can signal the relation between the two pieces of information by using discourse markers such as 'and', 'also', 'moreover'. The discourse marker signals an elaboration relation and the communicative purpose of the whole utterance, which contains the discourse marker, is Inform with the rhetorical function Elaborate. Additionally, the discourse marker is used here to show that the speaker wishes to continue in the speaker role (Turn Keep function).

In our analysis by 'multifunctionality' we mean the phenomenon of having multiple meanings simultaneously, which are related to the multiple purposes that an utterance may have in communication.

There are different forms of multifunctionality. Allwood in [1] claims that if an utterance is multifunctional, 'its multifunctionality can be sequential and simultaneous'. Bunt in [5] examines this claim using empirical data from several dialogue annotation experiments and concludes that sequential multifunctionality disappears if we take sufficiently fine-grained dialogue units into account (so-called 'functional segments' rather than turns). A *functional segment* is defined as 'a smallest(possibly discontinuous) stretch of communicative behaviour that has one or more communicative functions' ([8]). It was shown in [5] that even if we consider fine-grained units of communicative behaviour we do not get rid of simultaneous multifunctionality; and the minimum number of functions that one segment may have

in dialogue is 1.3 on average. The number of functions grows rapidly if we take forms of multifunctionality into account such as implicated and entailed functions, feedback levels, and indirect functions.

It is noticed in [5] that pragmatically implicated functions, e.g. an expression of thanks also expressing positive feedback, are a true source of multifunctionality. Logically entailed functions, such as the positive feedback on understanding that is entailed by answering a question or accepting an offer, can also be important. For the purpose of this study, however, we left such forms of multifunctionality out of consideration. This has two reasons. First, we believe that discourse markers as such do not signal any implicated or entailed functions. Second, since we want to operate on the basis of observable utterance features (prosodic and linguistic) that can be extracted automatically from raw data, to investigate how dialogue participants express and recognise the intended and explicitly indicated multiple functions of dialogue utterances.

In the next section we first describe the relevant aspects of the semantic framework that we will use to study the multiple meanings of discourse markers in dialogue in a systematic fashion.

3   Semantic framework

The semantic framework of Dynamic Interpretation Theory (DIT, [3]) takes a multidimensional view on dialogue, in the sense that it views participation in a dialogue as being engaged in several activities simultaneously, such as trying to advance a task that motivates the dialogue, providing communicative feedback, taking turns, and so on. Communicative behaviour is interpreted in terms of bundles of update operations on participants' information states (or 'contexts'); such update operations consist of a semantic (referential, propositional, or action-related) content and a communicative function, which specifies what an addressee is supposed to do with the semantic content in order to update his information state [5]. Consider the following dialogue fragment:

(2) *A1: that's why i think the option of the kinetic thing*
     A2: which basically means as long as you shake it like an automatic watch
     *D1: -1.78[1] **but***
     D2: are people gonna wanna shake their movie controller?

---

[1]Here and further in text figures given in brackets indicate the token duration in seconds; figures without brackets indicate silence pauses between tokens in seconds.

Utterance (D1) is multifunctional, since it indicates that (1) the speaker wants to have the turn by interrupting the previous speaker (signalled by 'but' overlapping A3); (2) the speaker interpreted and evaluated the utterances A1 and A2 successfully; and (3) the speaker encountered a problem in applying the information from the previous utterances (due to the adversative meaning of 'but') – he probably does not agree with the previous claim or needs some clarification, which is indeed expressed in D2. Thus, as the example shows, the various functions of 'but' are related to different 'dimensions' of the interaction [4], such as the allocation of the speaker role and the processing of previous utterances.

In DIT the information which can be addressed is divided into: the domain or task (*Task*), feedback on the processing of previous utterances by the speaker (*Auto-feedback*) or by other interlocutors (*Allo-feedback*), managing difficulties in the speaker's utterance production (*Own-Communication Management*) or that of other interlocutors (*Partner Communication Management*), the speaker's need for time to continue the dialogue (*Time Management*), establishing and maintaining contact (*Contact Management*), the allocation of the next turn (*Turn Management*), the way the speaker is planning to structure the dialogue (*Dialogue Structuring*), and attention for social aspects of the interaction (*Social Obligations Management*).

It was observed in DIT that some utterances have communicative functions that can be applied to any kind of semantic content (*general-purpose (GP) functions*). In particular, they can be applied not only to content information concerning a certain task or domain, but also to information concerning the communication, e.g. an Inform like *'First of all we need to discuss the project finances'* is used to introduce a new topic into the discussion. *Dimension-specific* (DS) functions, by contrast, are applicable only to information concerned with a specific dimension of communication, e.g. using the utterance *'Let me see'* the speaker indicates that he needs some time to do something in preparation of continuing the dialogue (Stalling act). The phenomenon of general-purpose functions means that, when a stretch of communicative behaviour has a GP function, its full functional characterization requires in addition also the specification of the dimension that is addressed, so we get characterizations like Feedback Question and Task Suggestion.

We found that discourse markers are used (i) as 'preface' of a range of GP functions, in particular Informs of various rhetorical kinds; (ii) as indicators of dialogue acts with a DS function, e.g. of topic shifts; and (iii) as full-blown dialogue acts (without explicit semantic content), e.g. as a Turn Take act. This means that discourse markers can have two kinds of meanings:

as a dialogue act, i.e. as a context update operator, and as an element that contributes to the determination of the communicative function of a dialogue act with either a GP- or a DS-function.

The DIT framework supports a 'multidimensional' semantics by relating context update operators to different compartments of structured context models which include, besides information states of the usual kind (beliefs and goals related to a task domain), also a dialogue history, information about the agent's processing state, beliefs about dialogue partners' processing states, information and goals concerning the allocation of turns, and so on, relating to the various 'dimensions' that dialogue acts belong to. The interpretation of a multifunctional stretch of communicative behaviour corresponds to updating the context models of the communicating agents in multiple ways, combining the effects of each of the component functions. For example:

(3) *B1: what anybody wants to add about what they don't like about remote controls*
     A1:0.48 **and** *you keep losing them*

Since it answers B's Set-Question B1, utterance A1, which includes the discourse marker *and*, updates the context model of participant B with the information that[2]: (1) *A believes that B wants to know which elements of a given set have a given property*; (2) *A believes that B believes that A knows which elements of that set have that property*; (3) *A believes to know which elements of that set have that property*; and (4) *A believes that B made the turn available.* Thus, the simultaneous performance of the turn management and feedback acts through the use of A1, in particular of *and*, constitutes the multidimensional interpretation of what A says.

4 Data analysis and classification experiments

4.1 Corpus data and automatically extracted features

For the empirical semantic analysis of discourse markers we selected three meetings from the AMI meeting corpus.[3] Our data contain 17,335 words which form 3,897 functional segments with an average length of 4.4 words. Average turn length is 7.7 segments.

All the features that we used were low-level features, automatically extracted both from the transcriptions and from sound files. In this respect our analysis differs from those where manually labelled intonational information is used such as tones and pitch accents ([10]), or automatically derived

---

[2]For a formal representation of updates in participants' information state see [13].

[3]Augmented Multi-party Interaction (http://www.amiproject.org/).

syntactic information ([14], [16]) such as parts-of-speech. Our aim was to discover how well a classifier can perform if no other external knowledge sources are available and only the output of the speech recogniser is accessible. The features relate to *prosody*, *word occurrence*, and *collocations*.

Prosodic features were minimum, maximum, mean, and standard deviation of *pitch* (F0 in Hz), *energy* (RMS), *voicing* (fraction of locally unvoiced frames and number of voice breaks), *speaking rate* (number of syllables per second) and segment *duration*.[4] Word occurrence is represented by a bag-of-words vector[5] indicating the presence or absence of words in the segment. As lexical features bi- and trigram models were constructed.

4.2 Dialogue act tagset and notes on segmentation

The training data was manually annotated using the DIT[++] annotation scheme.[6] The DIT[++] taxonomy is multilayered and multidimensional, with a conceptually well-defined set of communicative functions, supporting dialogue act annotation in multiple dimensions simultaneously.

The utterances were segmented per dimension following the approach in [8], which leads to a more accurate analysis of human communication than the segmentation of dialogue in single sequences of utterances, since utterances may contain functional 'holes' due to protractions, repairs, restarts, etcetera. Moreover, participants may interrupt each other and talk simultaneously, and utterances may also spread over several turns. A meaningful unit in this approach is a *functional segment* as defined in Section 3. An example of per-dimension segmentation is given in Figure (1).

4.3 Results: multifunctionality of discourse markers

We do not aim in this paper to discuss all discourse markers which occur in our corpus data, we rather demonstrate the minimal multifunctionality of the most frequently used discourse markers in dialogue and discuss the case of *and* in more detail.

Table 1 lists some discourse markers (DMs) identified in the studied corpus with their absolute frequency in the corpus, gives an overview of their

---

[4]We examined both raw and normalized versions of these features. Speaker-normalized features were normalized by computing z-scores ($z = (X\text{-}\ mean)/standard\ deviation$), where *mean* and *standard deviation* were calculated from all functional segments produced by the same speaker in the dialogues. We also used normalizations by the first and prior speaker turn.

[5]With a size of 1,640 entries.

[6]For more information about the tagset and the dimensions that are distinguished, please visit: `http://dit.uvt.nl/`

Speaker	Dimension	
A1		We're aiming a fairly young market
	Task	Inform
B1		Do you think then we voice should really consider voice recognition
	Task	Propositional Question
	Auto-Feedback	Pos.execution A1
	Turn M.	Turn-Assign to A
B2		What What do you think Craig
	Task	Set-Question
	Turn M.	Turn-Take ... Turn-Assign to C
C1		Well did you not say it was the adults that we're going for
	Auto-Feedback	pos.execution B2 neg.execution B1 ... Propositional Question to A1
	Turn M.	Turn-Accept Turn-Assign to A

Figure 1: *Example of per-dimension segmentation.*

observed mutlifunctionality by indicating the average number of communicative functions in our dialogues, and lists the observed communicative functions. Note that all DMs serve more than one communicative func-

DM	Occurrence	Multi-functionality	GP	Auto Feedback	Allo Feedback	Turn M.	Time M.	DS	Contact M.	OCM
and	214	2.6	elaborate suggest exemplify explain	pos. evaluate pos. execute	pos. execute	take grab keep	stall	topic shift		
because	37	1.2	justify explain			keep	stall			
but	87	1.9	disagree correction explain warning	pos. evaluate neg. execute	neg. execute	take grab keep	stall	topic shift		retract
I mean	41	2	elaborate			keep				retract
Like/such as	70	1.7	exemplify			keep	stall			
oh	31	2	answer	neg./pos. execute		grab accept	pause	topic shift		error signal
so	226	2	conclude suggest elaborate	pos. execute		take grab keep release	stall	topic shift preclose open	indicate	retract
then	45	1.9	instruct elaborate suggest	pos. execute		keep grab take	stall	topic shift		
well	63	2.1	disagree correction explain	pos. execute	neg. execute	grab take accept keep	stall	topic shift		retract
you know	84	2.3			check elicit	give keep	stall		check	retract

Table 1: *Distribution and observed multifunctionality of discourse markers.*

tion. *And* is the most multifunctional discourse marker in our corpus and *because* the least multifunctional one. *Because* mostly prefaces Informs with the rhetorical functions Justify or Explain, and only in 2.4% of all cases is used to simultaneously perform Turn Keeping and Stalling acts. All discourse markers, except *'you know'*, preface GP functions in Task or other dimensions (often in Discourse Structuring and Feedback) and may perform

164

dialogue acts addressing other dimensions simultaneously, as we will illustrate for *and* below. This pattern is observed for 50.7% of all studied DMs. A discourse marker may also perform full-fledged dialogue acts addressing more than one dimension simultaneously (as in example 2). This is often the case for Turn Management in combination with Feedback, Time Management, Discourse Structuring and Own Communication Management (27.7% of all discourse markers are observed to be used in this way). It was noticed that at most 3 dialogue acts are performed by one discourse marker in a given context, e.g. feedback, turn and time management acts. A third pattern of DM use, which was observed in 18.2% of cases, is as a single dialogue act, e.g. a turn taking act or a feedback act. In the rest (3.4%) discourse markers are part of general purpose functions and do not perform a dialogue act on their own.

And is one of the most frequently used discourse markers in our corpus. The corpus contained 470 occurrences of *and*, where about 54.5% is not used as a discourse marker and the rest of 45.5% as discourse marker. Differentiating between *and* as non-DM and DM is important for segmentation purposes. Used in clause-initial position or as an autonomous segment, *and* as DM so to speak brackets segments and helps define their boundaries. We investigated the prosodic and durational differences, and differences in surrounding material between the two basic uses of *and*, and performed machine-learning recognition experiments.

Experiments using the RIPPER rule inducer [7] showed that the two uses of *and* are well recognized automatically. An accuracy score of 80.6% was obtained on unsegmented data. Baseline score, in this case the percentage of the most frequent class (non-DMs), was 54.5%. There are significant mean differences ($p < .05$) for both raw and speaker-normalized features in terms of duration (DMs are almost twice as long as non-DMs: 289-327ms and 173-217ms respectively); initial pause (there is no or a negligible small pause before non-DMs, and initial pauses before DMs range between 59 and 228ms); mean pitch (*and* as DM has higher mean pitch: > 12Hz). Preceding and following tokens as features also have high information gain. *And* as DM is often so to speak backed up between *um, uh, so, then* and *also*.

As a discourse marker, *and* may have various and multiple communicative functions (see Table 1). *And* may signal that the upcoming speech is adding new information or used in explanatory sequences [15], like Inform Explain in (4):

(4) *A1: like you said a problem was how many components are in there*
     A2: um (0.4)
     *A3: 0.28 **and** (0.12) the power is basically a factor of that*

*A4: 0.55 um (0.47) **and** (0.32) this affects you in terms of the size of your device*

*A5: 0.59 um (0.26) **and** (0.16) that would have some impact*

And can also mark the transition from one discussion topic to another by introducing topic shifts, for example:

(5) *A1: you could group plastic and rubber, but it might be easier to use one*

D1: -0.29 mm-hmm

*A2: 0.74 um (0.32)**and** (0.2) the other components are logic chips*

Classification task	BL	Ripper
Task	23.5	51.2*
Auto-Feedback	96.2	99.8*
Turn Management	53.9	82.6*
Time Management	92	99.2*
Discourse Structuring	98.5	92.5
Partner Communication Management	99.7	99.8

Table 2: *Overview of accuracy on the baseline (BL) and the classifier on the prediction of communicative functions of 'and' in different dimensions obtained using 10-fold cross-validation experiments. * significant at $p <$ .05, one-tailed z-test.*

In 57% of all cases *and* is used as a marker of speaker continuation (turn keeping) as illustrated in A3, A4 and A5 in (4). *And* in A2 in (5) also has a positive allo-feedback function related to the utterance D1.

We trained the RIPPER rule inducer to automatically classify the communicative functions of *and* in several dimensions. The results are shown in Table 2. The classifier outperforms the baseline, which for all tasks is the percentage of the majority class (e.g. Elaboration in Task), except for the classification of Discourse Structuring and Partner Communication Management functions, for which there were not enough instances in the training set.

As for features, for the prediction of the *Task* dimension, the bag-of-words feature representing word occurrence in the segment and word collocation features were important. For example, Inform Elaborate is often signalled by focusing adverbs like *especially, mainly, additionally*, etc, or contains relative pronouns like *who, whose, which* and *that*. The presence of some expressions of comparision was noted in Exemplifications, such as *one of, rather than, like,comparing, by contrast, similar*, etc. The most frequent words that occurred in Suggestions were *maybe, might, better, should, could/can, probably* and *let's*; and Discourse Structuring functions are marked by *next, then, other* and *further*. For all other dimensions prosodic features were more important than the surrounding lexical material. For example, for Turn Management functions duration, initial pause

166

and mean pitch are key features. Important is also the information about the current and the previous speaker. Speaker switch is an important sign of *and* in the Auto-Feedback dimension. Stallings were characterized by a long duration (about 585ms) and long initial pause (365ms), and a pause after (100ms). They also often were preceded and followed by *um* and *uh*.

5   Conclusions and future work

To summarize we can conclude that discourse markers are truly multifunctional dialogue units. The analysis of discourse markers as important instruments for the understanding of dialogue and its computational modelling can only benefit from a multidimensional approach to the exploration of their complex meaning.

We showed that discourse markers may simultaneously serve several communicative functions in different dimensions. They are good indicators of (plentiful) general-purpose communicative functions, such as informs, elaborations of various kinds, suggestions, warnings, disagreements, etc., mostly in relation to the task that underlies the dialogue, but they are also frequently used to create or maintain the conditions for successful interaction (indicating dialogue control acts). Our investigations showed that discourse markers may have communicative functions in all the dimensions distinguished in DIT, except perhaps in Social Obligations Management. We noted the importance of discourse markers for segmenting dialogue into meaningful units, since they so to speak bracket functional segments. Binary automatic classification of DM vs non-DM was successfully performed. The automatic recognition of the various and multiple communicative functions of discourse markers is even more important. Our automatic classification experiments showed that machine learning techniques can be profitably used in the automatic recognition of multiple meanings of dialogue markers.

For future work, we intend to extend the studies reported here in two directions. First, we plan to collect more annotated data containing a richer set of discourse markers and sufficient numbers of instances per class, so that we may increase the accuracy of the classifier for further evaluation of our model on unmarked examples (see [16]). Furthermore, since AMI meetings are face-to-face interactions and video recordings are available, we plan to consider other modalities besides speech audio, e.g. hand and head gestures, for better understanding of discourse markers functions in support of adequate computational dialogue modelling.

Acknowledgments

References

[1] Allwood, J. (1992). *On dialogue cohesion*. Gothenburg Papers in Theoretical Linguistics 65. Gothenburg University, Department of Linguistics.

[2] Allwood, J. (2000). *An activity-based approach to pragmatics*. H. Bunt and W. Black (eds.) Abduction, Belief and Context in Dialogue, Amsterdam: Benjamin, pp.47-81.

[3] Bunt, H. (2000). *Dynamic Interpretation and Dialogue Theory*. M.M. Taylor, D.G. Bouwhuis and F. Neel (eds.) The Structure of Multimodal Dialogue, Vol 2., Amsterdam: John Benjamins, pp. 139–166.

[4] Bunt, H. (2006). *Dimensions in Dialogue Act Annotation*. Proceedings LREC 2006, Genova.

[5] Bunt, H. (2007). *Multifunctionality and Multidimensional Dialogue Act Annotation*. E. Ahlsen et al. (ed.) Communication - Action - Meaning. Gothenburg, pp. 237 – 259.

[6] Cohen, R. (1984). *A computational theory of the function of clue words in argument understanding*. Coling-ACL 1984, Standford, pp. 251–258.

[7] Cohen, W.W. (1995). *Fast effective rule induction*. In Proceedings of the 12th International Conference on Machine Learning (ICML'95), pp. 115–123.

[8] Geertzen, J. Petukhova, V. and Harry Bunt. (2007). *A Multidimensional Approach to Utterance Segmentation and Dialogue Act Classification*. Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue, Antwerp, pp. 140–149.

[9] Heeman, P.A. and Allen, J.F. (1999). *Speech repairs, intonational phrases and discourse markers: Modelling speakers utterances in spoken dialogue*. Computational Linguistics, 12(3): 1–45.

[10] Hirschberg, J. and Litman, D. (1993). *Empirical studies on the disambiguqtion of cue phrases*. Computational Linguistics, 25(4): 501–530.

[11] Hovy, E.H. (1995). *The multifunctionality of discourse markers*. Proceedings of the Workshop on Discourse Markers, Egmond-aan-Zee, The Nertherlands.

[12] Mann, W. and Thompson, S. (1988). *Rhetorical structure theory: toward a functional theory of text organisation*. The MIT Press, Cambridge, MA.

[13] Morante, R. (2007). *Computing Meaning in Interaction*. PhD Thesis, Tilburg University.

[14] Popescu-Belis, A. and Zufferey, S. (2006). *Automatic identification of discourse markers in multiparty dialogues*. Working paper 65, ISSCO, University of Geneva.

[15] Schiffrin, D. (1987). *Discourse Markers*. Cambridge: University Press.

[16] Sporleder, C. and Lascarides, A. (2008). *Using Automatically Labelled Examples to Classify Rhetorical Relations: An Assessment*. Natural Language Engineering, 14.03: 369–416.