Towards the Comprehension of Human Understanding

T. J. Visser
Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente, Enschede, The Netherlands
t.j.visser@student.utwente.nl

ABSTRACT
This paper presents the results of a research on the effectiveness of semantic relations as an indicator for understanding, the presence of acknowledgement grounding acts, in a two person dialogue. Semantic relations can easily be detected using WordNet. Three types of relations are used: least-common ancestor, i.e. words are semantically related if they can be described by one more general term; word relation finding, i.e. two words are semantically related if they are synonyms, antonyms, ‘siblings’, etc. and what-question analysis. The theory was implemented and verified over the Simple Dialogue Corpus. In 56.90% of the cases, the model did predict correctly. The results show that semantic relatedness, based on found relations in WordNet, indeed does work as an indicator for understanding in some cases, but that in more complex cases it is not possible to detect semantic relations in the proposed way.

Keywords
Grounding, semantic relations, implicit acknowledgements

1. INTRODUCTION
In the every day life, people take part in many conversations for many different reasons. But what all conversations have in common is that the participants want to be heard and understood by the other participants. The message itself is important, but knowing that it was understood correctly is too. It is the task of the other participants to let the speaker know that they understood him. Thus, taking part in a conversation seems to consist of more than just uttering words. The participants are working together on making information shared. This sharing of information in a dialogue is described by Clark and Schaefer [4] as grounding: one participant presents new information and the other participants may or may not acknowledge the receipt of this information. After the information is acknowledged, the information is not only known by all participants, but also all participants know that the others know it. Information in such a state, mutual knowledge [3], is said to be in the common ground (see Figure 1), which will, consequently, change during the conversation as new pieces of information are added or existing information is changed or removed.

In this paper, an approach to the analysis of grounding in a two person dialogue is presented. The approach uses semantic relations between utterances in a dialogue to determine the comprehension among the interlocutors. The underlying hypothesis is that when an utterance is highly semantically related to the previous utterance, in which new information was presented, the utterance acknowledges the successful transferal of information. The goal of the research was to investigate the efficiency of rather simple semantic analysis techniques for the detection of this kind of understanding.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

12th Twente Student Conference on IT January 22, 2010, Enschede, The Netherlands
Copyright 2010, University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.

Figure 1: Common ground (CG) in a dialogue with participants A and B.

In section 2, the theoretical context of the research will be discussed. Several important aspects and definitions that will warrant successful transferal of the paper’s content will be introduced. Section 3 will present a novel theory of semantic relatedness and grounding based on the study of real world dialogues from the TRAINS ‘93 corpus and a gathered corpus of simple dialogues (the Simple Dialogue Corpus, SDC). In section 4 the implementation of the theory in a computational model is described. The model was then verified by letting the implementation annotate the SDC, of which the results can be found in section 5. In section 6 these results are discussed. Section 7 scratches the surface of the application of the theory in multiparty conversations. Finally, section 8 will conclude the paper by stating the most important findings.

2. UNDERSTANDING AND GROUNDING
To understand is to ground the information that is presented in the course of the conversation. Since understanding is the object of the research, the process of grounding information has to be studied. This grounding process is however not the only grounding that takes place while conversing. Clark shows that a conversation consists of four levels and that grounding occurs at all of them [2]. He presents the following action ladder:
Level 4  Proposal and consideration
Level 3  Signalling and recognition, or meaning and understanding
Level 2  Presentation and identification
Level 1  Execution and attention

Each level is described as two opposed, but corresponding, actions, which have to be executed by the two interlocutors. To reach a state of understanding, the actions on all levels up to and including level 3 have to be grounded. Consider dialogue excerpt 1.

A1 : Do you still have that book I loaned you?
B1 : Sorry, I’m a slow reader.

(dialogue excerpt 1 (SDC dialogue 59))

Every utterance contains actions on all four levels. Lower level actions are required for higher level actions (upward causality) and a higher level action is evidence for the corresponding actions on all lower levels (downward evidence). The action ladder can be filled in with the actions that A performs while uttering A1:

Level 4  A is proposing B to answer the question
Level 3  A is asking B if he still has that book
Level 2  A is presenting to B a signal composed of the uttered words
Level 1  A is executing for B’s perception the articulation of “Do you still...”

In this time interval, A is in the process of proposing, asking, presenting and executing things. At the same time, B has the following action ladder:

Level 4  B is considering A’s proposal to answer the question
Level 3  B is recognising A’s question
Level 2  B is identifying A’s signal as composed of the uttered words
Level 1  B is attending to A’s articulation of “Do you still...”

B’s response (B1) fits A1 and thereby shows that B attended to, identified, recognised and considered A1. B did not only understand A1, but also chose to cooperate with A and answer the question. Thus, actions up to and including level 3 have been grounded. Following from the downward evidence property, level 3 and lower are also grounded. What makes B’s response fitting is that it answers A’s question. Two-utterance structures like this are called adjacency pairs [13], in this case a QUESTION-ANSWER pair. Other examples of adjacency pairs are GREETING followed by GREETING, COMPLIMENT followed by DOWNPLAY and REQUEST followed by GRANT [8].

If B would have responded with “So you’re asking me if I have that book?”, he would not have answered the question. But it still is evidence of understanding: the meaning of A1 (level 3) was grounded. In this paper, when talking about grounding, the grounding at level 3 or 4 is meant.

The grounding process itself, regardless of its level, can be described in various ways [16]. The grounding model that was used as a basis for the theory in this paper is by Traum [15],[17]. He developed his model with computational processing in mind, making it more suitable in the context of this research. In [16], Traum compares his model with the original contribution model by Clark and Schaefer, stating that the main deficiency of the latter is that it is difficult to tell the state of information.

Traum extended the contribution model by introducing a set of Grounding Acts [14]. Every grounding act affects the state of one piece of information, which is called a Discourse Unit (DU). Examples of these grounding acts are the presentation of a new DU (initiate), the acknowledgement of it (acknowledgement), signalling misunderstanding (request repair) and the absence of an acknowledgement (cancel). One utterance can contain multiple grounding acts and thus can modify the grounding state of more than one DU.

The relation between utterance, grounding act and discourse unit can be seen in dialogue excerpt 1. Utterance A1 presents new information (A wants to know if B still has that book) and therefore contains the initiate grounding act. In B1, B also presents new information, i.e. that he is a slow reader, but it also answers A’s question and thereby acknowledges the information from A1. Ergo, B1 contains both an initiate and an acknowledgement grounding act, each for a different discourse unit.

Once added to the common ground, the information can support the grounding of other information. Therefore, information that is contained in the common ground when the conversation begins also plays a role in the grounding that takes place during the conversation. Two interesting parts of the initial common ground are the communal common ground and the personal common ground [10], [2]. The communal common ground contains information that is assumed to be mutual knowledge based on experience of different communities, including information on language use and habits. The content of the communal common ground is based on the communities that one or more of the interlocutors are member of and the mutual knowledge of their membership. The personal common ground consists of mutual knowledge based on experience with the other individual, e.g. personal motives, background and ideas. Information from both parts of the common ground will support the grounding of new information during the conversation.

3. GROUNDING AND SEMANTICS

The conducted research has focussed on the acknowledgement grounding act and how it could be predicted based on semantic relatedness. Two utterances are semantically related if one or more words from the one utterance are related to one or more words from the other utterance. In dialogue excerpt 1 for example, A1 and B2 are semantically related because book and reader are related words. The semantic relatedness of two utterances u and v, u preceding v in the dialogue, indicates that v is a fitting response to u. Fitting with respect to the dialogue context (the previous utterance) and the common ground, or more precisely: the language part in the communal common ground that underlies the notion of semantic relatedness. A fitting utterance is presumed to be an acknowledgement.

For simplicity reasons, it was assumed that there is always information to be acknowledged.

B1 is an example of an implicit acknowledgement, opposed to explicit acknowledgements, which are of the form “Yeah”, “Ok”, “Hmhm” etc. Explicit acknowledgements can be detected using rather simple keyword matching techniques, but are not detectable by means of semantics. Implicit acknowledgements on the other hand are not detectable by keyword matching but do contain semantical content. The research therefore aimed at detect-
Table 1: Types of evidence of understanding, ordered in increasing strength

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Continued attention</td>
<td>B shows that he is continuing to attend and therefore remains satisfied with A’s presentation.</td>
</tr>
<tr>
<td>2</td>
<td>Demonstration</td>
<td>B demonstrates all or part of what he has understood A to mean.</td>
</tr>
<tr>
<td>3</td>
<td>Acknowledgement</td>
<td>B nods or says “uh huh”, “yeah”, or the like.</td>
</tr>
<tr>
<td>4</td>
<td>Initiation of relevant next contribution</td>
<td>B starts in on the next contribution that would be relevant at a level as high as the current one.</td>
</tr>
<tr>
<td>5</td>
<td>Display</td>
<td>B displays verbatim all or part of A’s presentation.</td>
</tr>
</tbody>
</table>

ing implicit acknowledgements. These implicit semantic acknowledgements correspond to types 2, 4 and 5 of evidence of understanding by Clark and Shaefer (see table 1).

For the detection of relations between words, WordNet, a large English lexical database, was used. WordNet contains more than 150,000 nouns, verbs, adjectives and adverbs, which are connected by semantic and lexical pointers. The semantic pointers include hypernymy (‘fruit’ is a hypernym of ‘apple’), hyponymy (the opposite of hypernymy) and meronymy (‘finger’ is a meronym of ‘hand’). Two words are semantically related if one or more paths in WordNet can be found.

Semantic relations between words differ by their underlying path(s) in WordNet. A path in WordNet is a sequence of pointers between words, e.g. a path between ‘apple’ and ‘food’ exist via ‘edible fruit’ and ‘fruit’ (all three hypernymy pointers). Properties of a path include the length, the type of the pointers that were traversed and the distance from the WordNet root. Using these properties, it is possible to define which semantic relations indicate an acknowledgement and distinguish them from relations that do not.

3.1 The TRAINS ‘93 Dialogues

To discover what kind of semantic relations indicate acknowledgement, the TRAINS ‘93 corpus [5] was studied for their co-occurrence. The corpus contains transcriptions of dialogues in which a user collaborates with a planning assistant over an audio link to accomplish some task involving manufacturing and shipping goods in a railroad system. The planning assistant has information on the Trains world that the user does not, but needs in order to solve the problem. The user has to come up with a plan to solve the problem but needs the planning assistant in order to schedule the required actions (e.g. moving an engine from one city to another).

The first observation that was made during the analysis was the lack of implicit acknowledgements (see table 2). This became clear in the first stage of analysis, where all utterances were annotated with their grounding acts. As a consequence, the corpus did not contain any empirical proof to back the theory and specify the semantic relations that co-occur with implicit acknowledgements.

There were however two interesting cases of implicit acknowledgements in the three studied dialogues. The first case (see dialogue excerpt 2) contains an acknowledgement that can only be pointed out as one with the knowledge of which task the user has to perform. He had to determine the maximum number of boxcars of oranges that he could get to Bath in seven hours. This information, together with the fact that the orange warehouse is in Corning, is evidence for the fact that utt16 is indeed an acknowledgement. Both utt14 and utt16 discuss actions that are part of a shared task. The fact that the user (u) with utt16 proceeds to another action is a signal to the planning assistant (s) that the user has enough information to execute the action discussed in utt14, the information from utt15 is thus understood.

utt14 : u : <noise> <sil> and how long would it take <sil> to get from Corning to Bath
utt15 : s : uh <sil> two hours.
utt16 : u : how long would it take <sil> to load the oranges from the warehouse into the engine

[dialogue excerpt 2 (TRAINs ‘93 dialogue d92a-3.1)]

This example shows the present and usage of an underlying structure in the form of a set of actions and could well have the same function as the model of semantic relatedness that was envisioned as a result of the research.

The second interesting implicit acknowledgement also assumes mutual background information. In this case, it is the map with the railroad system that both the user and the system have, showing two routes from Dansville to Bath: through Corning and through Avon. This knowledge enables the user to use utt24 as an acknowledgement. He received the system’s information on the travel time from Dansville through Corning to Bath and wants to compare it with the route via Avon. The key to understanding this efficient implicit acknowledgement is again an underlying structure: the TRAINS world map. This observation raises the question of how that map for an average dialogue would look like. Perhaps it is similar to WordNet. But would the map analogy still work in a dialogue context that is not as well defined and structured as is the case with the TRAINS ‘93 corpus?

utt18 : u : how long would it take <sil> from Dansville <sil> to Bath
utt19 : u : + going + the Corning <sil> Bath way (.)
utt23 : s : it’ll take uh three hours to take the boxcars from Dansville to Bath <sil> through Corning
utt24 : u : what about <sil> Avon

[dialogue excerpt 3 (TRAINs ‘93 dialogue d92a-1.2)]

Let us not forget the reason why the above examples are interesting cases and not the average acknowledgement. The fact that the corpus almost exclusively contained explicit acknowledgements is reason to question the relevance of the research. This observation was however put into perspective by the results of a research by Brennan and Ohaeri [1]. In their research on impoliteness of electronic conversations they discovered a difference in the use of acknowledgements in face-to-face and written dia-

1 What ‘pointer’ is referring to is often also called ‘relation’. To avoid confusion with the other kind of semantic relations, the term ‘pointer’ will be used.
Table 2: Fragment of the analysis of dialogue d92-1.2 from the TRAINS ‘93 corpus.

<table>
<thead>
<tr>
<th>Utterances</th>
<th>Discourse Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>u: you have to go to Bath &lt;sil&gt; and the boxcars &lt;sil&gt; there's three available &lt;sil&gt; at Dansville</td>
<td>DU1 initiate</td>
</tr>
<tr>
<td>s: okay</td>
<td>DU2 ack</td>
</tr>
<tr>
<td>u: and there's &lt;sil&gt; two available at mira</td>
<td>DU3 init</td>
</tr>
<tr>
<td>s: that's right</td>
<td>DU4 ack</td>
</tr>
<tr>
<td>u: there’s already &lt;sil&gt; two at Bath</td>
<td>DU4 init</td>
</tr>
<tr>
<td>s: yes there are two at Bath</td>
<td>DU4 ack init</td>
</tr>
<tr>
<td>u: okay</td>
<td>DU4 ack</td>
</tr>
</tbody>
</table>

3.3 Semantic relations

Three types of semantic relations have been specified based on observations from the SDC: least-common ancestor, word relations and what-question analysis. Each semantic relation will be discussed by defining its characteristics and a metric for its weight. The characteristics will describe the relation by means of the underlying path in WordNet. A metric for its weight is necessary to express that some words are less semantically related than others.

Before the types of semantic relations are introduced, some internals of WordNet are discussed. The pointers between words that were introduced earlier are, to be more precise, between synsets of words, and not between the words themselves. A synset represents one sense of a word. If a word has more than one synset, its meaning is ambiguous. When talking about a semantic relation between two words, the actual relation is between at least one synset of one word to at least one synset of the other word.

3.3.1 Least-Common Ancestor

The least, or lowest, common ancestor (LCA) of two words is the most specific word that both words fall under. For example, the LCA of ‘apple’ and ‘pear’ is ‘edible fruit’. The LCA of a set of words \( W \) can be found by searching the upward pointers \( \text{head}(\bigcap_{w \in W} \text{ancestors}(w)) \). It is assumed that \( \text{ancestors}(w) \) returns the set of ancestors of word \( w \), sorted on the distance in ascending order. Ancestors can be found in WordNet by traversing the upward pointers\(^2\). The presence of a LCA for \( W \) indicates semantic relatedness of the words in \( W \), which can be measured as

\[
\text{compactness}(W) = \sum_{w \in W} \frac{\text{dist}(w, \text{LCA}(W))}{|W|}
\]

\(^2\)The upward pointers are hypernym, hypernym instance, holonym member, holonym part and holonym substance.
in which $\text{dist}(w_1, w_2)$ is number of pointers between $w_1$ and $w_2$ in WordNet. When the mean distance increases, the words are less semantically related. In practice, a LCA for any arbitrary set of words can be found using WordNet: ‘entity’, the root of the tree. This is however no evidence for semantic relatedness. This can be prevented by restricting the LCA search to words that are a certain minimum distance away from the WordNet root and not too far away from the words in $W$. In the computational model, the minimum distance from the root was set at four, the maximum ancestor search distance at two, and $|W| = 2$ (two words, one from each utterance), as was determined during experiments.

### 3.3.2 Word relations

A more general way of finding a semantic relation between two words is the finding of a relation in WordNet between those two words. Words can be related by a path between them, or by the sharing of one or more synsets. In [7], definitions of three relations types of varying strength are given: the extra-strong, strong and medium-strong relation. These three relations were used in the research. The extra-strong relation exists only between two literal repetitions of a word and has the highest weight of all relations.

Three kinds of strong relations exist. The first occurs when two words share at least one synset. The second occurs when there is a horizontal pointer (e.g. synonymy, antonymy) between one synset of the one word to one synset of the other word. The third strong relation occurs when there is any kind of pointer between a synset of each word. The weight of a strong relation is lower than that of a extra-strong relation, but higher than a medium-strong relation.

A medium-strong relation occurs between two words when a valid path with a length between two and five pointers exists between synsets of these words. The validity of the path depends on the ‘shape’ of the path, i.e. the pointers that were traversed. Some sequences of pointers indicate no semantic relatedness at all. Therefore, Hirst and St-Onge [7] set up two rules that define the set of valid paths: No other direction may precede an upward pointer and at most one change of direction is allowed, except when a horizontal pointer is traversed between the transition from an upward to a downward direction.

The weight of a medium-strong is given by the following formula:

$$\text{weight} = C - \text{path length} - k \times \text{direction changes}$$

$C$ is the upper weight bound and $k$ is the direction change penalty, thus the longer the path and the more changes of direction, the lower the weight. During research, the following weight parameters were used: extra-strong relation weight: 30, strong relation weight: 20, medium-strong maximum weight ($C$): 15, direction change penalty ($k$): 2. These values were determined during test runs of the implementation on the SDC.

### 3.3.3 What-question analysis

To experiment with semantic analysis in more specific cases, a special type of word relation finding was used with what-question utterances. Based on the intuition that what-questions ask for further specification of what was said earlier, a third rule was added to the two rules by Hirst and St-Onge: all pointers must be downwards. Given three consecutive utterances $u, v, w$, where $v$ is a what-question, it can be said that a downward relation from $u$ to $w$ indicates that $v$ is an acknowledgement of $u$ and $w$, thus a sensible answer, is an acknowledgement of $v$.

In dialogue excerpt 4, the presence of a downward relation from cake to Cheesecake, backs this intuition.

$A1$: I made a cake yesterday.
$B1$: What kind of cake?
$A2$: Cheesecake, it was really good.

If it could be proved that there is any merit to this kind of specific analysers, further research on this topic would be interesting. Unfortunately, the simple dialogue corpus contained only one what-question. Based on the conducted research, nothing can be said about this analyser.

### 4. IMPLEMENTATION

To enable verification of the theory that was presented in section 3, a computational model based on the theory was created. It is written in Java and allows both dialogues from the SDC and manual input. For all utterances in the input it will calculate the semantic relatedness. Before the utterances are analysed, a preprocessor will prepare the utterances for analysis. The analysis itself is done by three analysers: LcaAnalyser (section 3.3.1), PathAnalyser (section 3.3.2) and WhatQuestionAnalyser (section 3.3.3). Figure 4 contains an overview of the implementation.

#### 4.1 Utterance preprocessing

The function of utterance preprocessing is to improve the quality of analysis. This is done in two ways: increasing semantic content by discovering word phrases and decreasing the semantic noise by filtering stop words. First, the utterance is converted to a list of words. Then all consecutive word pairs are combined into one word (with a space in between) and added to the end of the list. For example, the utterance

“I like green tea”

will result in

“green, tea, i like, like green, green tea”

Then, all words are looked up in WordNet. If a word is not found, it is stemmed and looked up again. If a word still can not be found, it is removed from the list. Continuing with the example, the final content word list is

“green, tea, i like, green tea”

and will be used as representation of the utterance during analysis. Every word in the final list is linked to its corresponding synsets in WordNet, which may be distributed over different part of speeches if the word can be interpreted as more than one part of speech.

---

[Figure 2: Outline of the implementation](#)
In addition to the standard preprocessing, when-questions have an extra content word added: ‘time’. This is because when-questions often do not contain content words (e.g. ‘When is it?’) and the answer presumably contains a time indication (e.g. ‘tomorrow’, ‘last week’), which is semantically related to ‘time’.

4.2 Analysis

The analysis of the utterances in search for semantic relations is done by three separate analysers. The Lca-Analyser and PathAnalyser award each pair of consecutive utterances in a dialogue with a semantic relatedness score. This semantic relatedness score is based on the relations that the analysers found. The WhatQuestionAnalyser is almost identical to the PathAnalyser, except for that it searches for relations (with the restrictions discussed above) between the utterance before and the utterance after the question. The analysers search for relations between words that are compatible: they have to be of the same part of speech. Analysis of words with different parts of speeches is useless, because WordNet contains almost no pointers between them. The relatedness analysis is executed for every possible pair of words, consisting of one word from each utterance. The combined words are analysed first and when a relation is found, the words that form the combination are removed from the word list to prevent duplicate analysis. The number and characteristics of the found semantic relations determine the semantic relatedness score of the utterance pair.

The semantic relatedness score that the LcaAnalyser gives to utterance \( u_1 \) and \( u_2 \) is equal to

\[
\text{score} = \frac{|R|}{(|u_1| + 1) \times (|u_2| + 1) \times \sum_{r \in R} \text{weight(r)}}
\]

where \(|R|\) is the set of found LCA relations, \(|u_i|\) is the number of content words in utterance \( i \) and \( \text{weight}(r) \) is the weight function as defined in section 3.3.1. The semantic relatedness score that the PathAnalyser and WhatQuestionAnalyser give to utterance \( u_1 \) and \( u_2 \) is defined by

\[
\text{score} = \frac{\text{sweight}(r)}{(|u_1| + 1) \times (|u_2| + 1) + ((30 \times |R|) - \text{sweight}(r))}
\]

where \( \text{sweight}(r) \) refers to the sum of all relation weights, described in section 3.3.2. The result of the WhatQuestionAnalyser is not interpreted as the semantic relatedness of \( u_1 \) and \( u_2 \). Instead, it is used as a measure for the relatedness of \( u_1 \) and \( u_3 \), the utterance containing the question, uttered directly after \( u_1 \) and before \( u_2 \), and the relatedness of \( u_3 \) and \( u_2 \).

The final semantic relatedness score of two utterances is determined by the average of the scores of the individual analysers.

5. RESULTS

The performance of the model was measured using the Simple Dialogue Corpus. The corpus was annotated both manually and by the model and then compared (see Table 4). Its overall accuracy is 56.9%, with a precision of 55.9% and a recall of 93.1%. The analysis of the SDC took 1:02 minutes. The current implementation is not efficient nor optimised for speed, and the possibilities of significantly reducing the processing time are numerous.

All utterances, except the first utterance of each dialogue was annotated with the model’s judgement. Its judgement, a score of 0.00 or higher, indicates the semantic relatedness as specified by the earlier introduced theory.

An utterance with a semantic relatedness score of 0.00 was marked as a ‘non-acknowledgement’. For all scores greater than 0.00, the model’s outcome was interpreted as the utterance being an ‘acknowledgement’.

The following section will thoroughly discuss the results of the SDC analysis and the observations that can be made.

6. DISCUSSION

All utterances can be divided into several classes based on the correctness of and grounds for the verdict. Besides the two cases in which the model predicted correctly (true-positive and true-negative), the following classes can be distinguished between:

Type 1 Utterances correctly predicted as an acknowledgement, but (partly) for the wrong reasons;

Type 2 Utterances wrongly predicted as a non-acknowledgement, because

a) semantic evidence from earlier utterances is needed;

b) they lack content words;

c) semantic relation, while present, were not detected.

Type 3 Utterances wrongly predicted as an acknowledgement, because relations were found;

The classes will be discussed by illustrating each of them with an example from the SDC, followed by an analysis of the underlying problems and their solutions.

6.1 Type 1

This first class contains all utterances that were correctly marked as acknowledgement, but (partly) for the wrong reasons. In this class, two types of ‘wrong reasons’ can be distinguished between: part-of-speech (PoS) ambiguity and meaning ambiguity.

Part-of-speech ambiguity occurs when a word can be interpreted as more than one PoS, e.g. ‘ship’ can be a noun or a verb. The model searches for semantic relations for all PoS’es of a word, while only one PoS is correct. Dialogue excerpt 5 contains PoS ambiguity, causing the model to find incorrect relations.

1: When I was young I used to pick blueberries all the time.

B1: Did you live near a forest then?

[dialogue excerpt 5 (SDC dialogue 58)]

The model gave B1 a relatedness value of 0.0812, based on two paths that were found in WordNet. The first path between did and time (via ‘cause’ and ‘determine’), the second from did to pick (via ‘cause’ and ‘provoke’). Due to the PoS ambiguity of time, the word could also be interpreted as a verb, enabling the finding of the first path. The, by intuition, correct semantic relation, which was not found, is between the picking of blueberries and living near the forest and doing it all the time.

The PoS ambiguity could be solved by adding a PoS-tagger as a preprocessor to the model that would annotate the words in the utterances with their PoS before the semantic analysis. As a consequence, less incorrect WordNet paths will be found. On the other hand, errors in the tagger will result in wrong model outcomes. This is, however,
with state of the art PoS-research reporting accuracies of 95-98% [12], only a minor concern.

The second type of ambiguity is meaning ambiguity, which occurs when a word has more than one meaning. The current implementation of the model uses all meanings (synsets) to find relations. One example of this can already be found in the dialogue excerpt above, where pick was interpreted as the act of provoking, instead of as taking hold and removing a blueberry from a bush. Dialogue excerpt 6 also contains meaning ambiguity.

A1 : I’ll have a glass of strawberry lemonade. 
What’s your favorite?
B1 : I’m more of a tea person.  
[dialogue excerpt 6 (SDC dialogue 52)]

The model rated the semantic relatedness of B1 w.r.t. A1 at 0.0536, based on its findings in WordNet. The least-common ancestor analysis succeeded for lemonade and tea, finding mutual parent ‘beverage’, which is the intuitive relation and thus the desired result. It also found a relation due to ambiguity between person and favorite by interpreting the latter as ‘darling’ instead of ‘favorite [drink]’. This problem can be solved by using word disambiguation techniques [7]. These would presumably solve the ambiguity of pick in dialogue excerpt 5. Resolving favorite as ‘favorite [drink]’ is much harder. These techniques rely on the semantics of the context of the ambiguous word. Solving this problem in the dialogue context is likely to be hard.

The SDC contains 31 utterances with PoS ambiguity and 7 utterances with meaning ambiguity. This is 40 % of the true-positives. The majority of the errors in this class can be solved easily, by means of a PoS tagger, and will most likely not influence the performance in a negative way. In almost all cases, a path was also found using the correct PoS for the ambiguous word.

6.2 Type 2a

The current implementation of the model works with utterance pairs, the semantic relatedness of an utterance is calculated in relation to the utterance preceding it (except for what-questions). However, in nine cases, an utterance could only be detected as an acknowledgement by means of semantics when looking at more than one utterance back.

A1 : Which countries did you visit during the holiday? 
B1 : I toured through Greece, Turkey and Moldavia. 
A2 : All during last holiday? 
[dialogue excerpt 7 (SDC dialogue 51)]

In dialogue excerpt 7, the model annotated A2 with a semantic relatedness value of 0.0 as it could not find any relations between B1 and A2. It is, however, an acknowledgement and we understand that A’s question is a response to the many countries B has been visiting. In order to be able to detect this kind of relations, much more than just the semantics have to be taken into account. Using only semantics, this problem could be solved by looking back more than one utterance, i.e. measuring the semantic relatedness of A2 and the combination of A1 and B1.

While this will work well in the above situation, it could introduce new errors in a dialogue where, for example, one person ignores something the other person said and keeps talking about the same subject, as is the case in dialogue excerpt 8.

A1 : I made a cake yesterday. 
B1 : What kind of cake? 
A2 : A fruit cake? 
B2 : Okay that’s nice. Did you have some friends over or something? 
A2 : I love baking.  
[dialogue excerpt 8 (SDC dialogue 204)]

With the suggested ad hoc solution, A2 will be marked as an acknowledgement, while this is not the case. A more sophisticated solution is needed and will possibly need to take the distance between utterances into account or focus on the influence of the utterance of one person between two of the other. Further research on this topic would be interesting.

6.3 Type 2b

The method proposed in this paper relies on the meaning of the words in a utterance, it requires meaningful words to function properly. When an utterance lacks meaningful words, e.g. ‘That is right’ or ‘Yeah’, the system will not be able to detect any relations.

A1 : Do you have plans for this weekend? 
B1 : No, not yet. 
A2 : Ok.  
[dialogue excerpt 9 (SDC dialogue 91)]

In dialogue excerpt 9, the model was not able to correctly predict both B1 as well as A2 as an acknowledgement because of the lack of content words in both utterances. In these cases, a model based on lexical analysis would suffice and give much better results.

The amount of utterances that fall inside this class says something about the applicability of the conducted research. The more “meaningless” utterances an average dialogue contains, the less applicable the theory is. Only four of the 75 false-negatives, i.e. utterances that where wrongly marked as non-acknowledging, did not contain any content words, indicating that the research is indeed relevant. This is in accordance with the results by Brennan and Ohaeri [1].

6.4 Type 2c

The utterances that were not detected as acknowledgements although semantic relations were present, can be divided into two groups: analysis shortcomings and complex cases. An example of the latter is the holiday example (dialogue excerpt 7). Some examples of the first group will follow to illustrate the problems that have to be solved in order to improve performance.

Dialogue excerpt 10 exposes one of the limitations of WordNet. Because WordNet consists of distinct nets for every
word type with relatively few relations between those nets, the relation between healthy and apple was not found. Intuitively, this relation should be present, relating apple, possibly via ‘fruit’, to healthy. The separateness between word types is slowly disappearing as WordNet updates will contain more relations between the networks.

A1: Did you watch TV last night?
B1: Yeah, I saw a great movie.

[dialogue excerpt 11 (SDC dialogue 62)]

In dialogue excerpt 11, the model was unable to find any relation from B1 to A1. In WordNet, no eligible path exists between TV and movie, while they are even of the same word type. WordNet lacks this, seemingly trivial, relation, as is the case with more word pairs like this. The solution to this problem seems to be the extension of WordNet with more relations and while it will indeed reduce the number of undetected semantic relations, it will most likely lead to an increase of the number of Type 1 and Type 3 results.

The size of this class combined with the Type 1 results indicates the feasibility of acknowledgement detection using semantic analysis as proposed in this paper. 13 utterances were not detected as acknowledgements because of WordNet under-prediction. These should be detectable after appropriate WordNet updates. What remains is a group of 49 utterances containing complex semantic relations, which most likely cannot be solved using the approach from this paper. Correct prediction of these cases will require the extension of the model with more sophisticated semantic understanding. This is a challenging task and would be very interesting for follow-up research.

6.5 Type 3

This result class is the opposite of Type 1 and covers the last set of interesting results. Again, this class can be divided into two separate groups: analysis shortcomings and complex cases. The first group consists of cases were the evidence of misunderstanding is too subtle and was confused with evidence for understanding.

A1: Did you watch TV last night?
B1: Yeah, I saw a great movie.
A2: There was this brilliant show last night.

[dialogue excerpt 12 (SDC dialogue 221)]

In dialogue excerpt 12, the model wrongly predicted that A2 was an acknowledgement. The cause of this kind of errors lies at the very root of the model: it is built to detect understanding by means of semantic similarities and not misunderstanding based on the semantic differences. But in this case not the present semantic similarity (between movie and show), but the also present small semantic difference, is the only evidence for the misunderstanding. This could be solved by inverting the approach of the model and search for semantic differences, which would affect all results completely. A more obvious and less substantial solution is to extend the model with a comparison of the semantic relatedness between 1) the utterance (A2) and the one before (B1), and 2) relatedness of the utterance (A2) and the preceding utterance by the same person (A1). If 2) is higher than 1), the utterance probably is not an acknowledgement.

The complex cases that fall inside this type are characterised by the need of additional analysis of aspects other than the semantical aspect, e.g. the analysis of the syntactic structure.

A1: I like apples but not only apples, I like bananas, strawberries... etc.
B1: I don't like apples either.

[dialogue excerpt 13 (SDC dialogue 338)]

Both utterances are about apples and a degree of liking them. Thus, the model predicts that B1 is an acknowledgement. Only with an understanding of the negation don't in combination with either, it can be detected that B misunderstood A. Further research on the extension of the semantic approach with syntactical concepts or a comparison with a model with a syntactical starting point would be interesting.

Inside this result class are 7 utterances, which is 35% of all non-acknowledgement utterances.

7. MULTIPARTY CONVERSATION

So far, the theory was discussed with respect to the context of a two person dialogue. In this context the goal was to annotate every utterance with if it is an acknowledgement or not. In multiparty conversations this is not sufficient, as it is also important to know who was acknowledged.

Finding out who is talking to whom is however proven to be hard [6]. A short introductory research into the applicability of the proposed theory in multiparty conversations was conducted to see whether further investigations would be interesting.

A1: [vocalsound] You said previously that you there’s um microphone inside an
B1: Embedded.
C1: Yeah, this is microphone array, in fact.

[dialogue excerpt 14 (AMI meeting IS1003d)]

For this research the meeting corpus collected during the AMI project was used [9]. Utterances starting with “yeah” were taken as the starting point because they are most likely to be an acknowledgement and individually addressed [6]. With a small modification in the model implementation, the analysers not only analysed consecutive utterance pairs, but all possible present-acknowledge utterance pairs. Then, after comparison of the semantic relatedness scores, the most likely combination can be pointed out.

In dialogue excerpt 14 the model analysed A1-C1 and B1-C1 and concluded that C1 is an acknowledgement of A1, based on the reoccurrence of microphone. Examples like

<table>
<thead>
<tr>
<th>Result class</th>
<th>TP</th>
<th>FP</th>
<th>TN</th>
<th>FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>38</td>
<td>-</td>
<td>-</td>
<td>9</td>
</tr>
<tr>
<td>Type 2a</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>Type 3c</td>
<td>-</td>
<td>7</td>
<td>-</td>
<td>62</td>
</tr>
</tbody>
</table>

Table 5: Size of the result classes and the distribution of their utterances over the error types (true positive, false positive, true negative and false negative).
this show that further research into acknowledgement resolving in multiparty conversation could be interesting.

8. CONCLUSIONS

In this paper, a model for the detection of understanding in a two person dialogue was presented. The main hypothesis was that semantic relatedness can be used to detect understanding, by means of the acknowledgement grounding act. Observations of the TRAINS '93 corpus show that this method is not feasible for dialogues in that corpus and for spoken dialogues in general as well. The results from the automatic analysis of the SDC show that the proposed theory does work for written dialogues, but also that it has its limitations. A minority of the errors can be solved by additional preprocessing (e.g. part-of-speech tagging and word disambiguation) but 49 false-negatives can only be solved by more complex semantical analysis, which lies in the direction towards full semantic understanding by a machine. This involves, for example, the comprehension of more subtle processes like alignment, i.e. the convergence on the lexical (word choice) and other levels, and the priming effects, i.e. earlier stimuli affecting the interlocutors utterances, which are also evidence for understanding and misunderstanding [11].

Efforts to decrease the size of one result class (see table 5) will increase the size of others. Increasing the sensitivity of the semantic analysis, enabling the finding of more far stretched relations, will decrease the number of type 2c errors, but will simultaneously increase the size of the type 3 class. By adding disambiguation techniques, the type 1 class will shrink, but it could also lower the number of true positives. Large improvements will therefore most likely require the addition of other techniques.

A1 : Shall we go to the cinema tomorrow?
B1 : I’m sorry, I have a girlfriend.

[dialogue excerpt 15 (based on SDC dialogue 314)]

The results have also shown that WordNet is capable of modelling only a small part of the communal ground. Understanding that you are being rejected by B (see dialogue excerpt 15) will remain a human privilege for some time to come.

9. ACKNOWLEDGEMENTS

Thanks to all who helped me build the corpus of simple dialogues: Martijn Weghorst, Rieks op den Akker, Jasper Kuperus, Jorrit Posthuma, Sanne Stienstra, Maartje van de Vrugt, Yori Kamphuis, Marja Doesburg, Marthe Visser, Hannah Visser, Johan Visser and all others who did not enter their name. I would also like to thank my fellow students and supervisors who supplied me with valuable comments.

10. REFERENCES