

Entailed feedback: evidence from a ranking experiment

Marcin Włodarczak

Bielefeld University
Bielefeld, Germany

mwłodarczak@uni-bielefeld.de, {harry.bunt, v.petukhova}@uvt.nl

Harry Bunt

Tilburg University
Tilburg, the Netherlands

Volha Petukhova

Tilburg University
Tilburg, the Netherlands

Abstract

In this paper we investigate relationships between entailment relations among communicative functions and dominance judgements in an annotation task in which participants are instructed to rank utterance functions in terms of their importance. It is hypothesised that on average entailed functions should be assigned lower ranks than entailing functions. Preliminary results of an experiment are reported for positive auto-feedback functions which are argued to be entailed by backward-looking functions such as *Confirm*.

1 Introduction

Since communication is a multi-faceted process in which participants must monitor and manage several aspects of their behaviour simultaneously, utterances which they produce are often multifunctional. Moreover, some of the utterance functions might entail or implicate other functions (Bunt, 2009a). In particular, backward-looking functions (Allen and Core, 1997), such as *Confirm*, *Answer* or *Agreement*, are claimed to entail positive feedback about some earlier utterance.

At the same time, it might be argued that some of the utterance functions have a priority over its other functions because achieving some communicative goals is of greater importance to the speaker in a given context.

The paper explores the hypothesis that entailed and implicated functions could be expected to be ranked lower than entailing functions. Specifically, preliminary results for entailed auto-feedback are reported.

2 Multidimensional Tagsets

Multifunctionality of utterances suggests that in an annotation task each utterance should be allowed to be labelled with more than one tag. This process is greatly facilitated if tags are organised into clusters such that only one tag for a cluster can be assigned to an utterance (Allen and Core, 1997; Clark and Popescu-Belis, 2004; Popescu-Belis, 2005; Popescu-Belis, 2008). Such clusters are commonly referred to as dimensions.

A conceptually-motivated definition of a dimension was provided by Bunt (2006). He defines it as an aspect of participating in dialogue which (1) can be addressed by means of dialogue acts that have communicative functions specific for this purpose, (2) can be addressed independently of other aspects. The first criterion requires that the proposed aspects of communication correspond to observable dialogue phenomena. The second requires that the dimensions be orthogonal.

3 Semantic types of multifunctionality

Bunt (2009a) distinguishes the following semantic forms of multifunctionality: *independent*, *entailed*, *implicated* and *indirect*.

3.1 Independent multifunctionality

Two or more communicative functions are independent when each is expressed by some features of a segment. An example could be “thank you” spoken with cheerful intonation and high pitch, which might signal both gratitude and goodbye.

From a point of view of information-state update approaches, such as DIT (Bunt, 1994), in which a dialogue act corresponds to a context update operation, indirect multifunctionality could be interpreted as independent update operations of addressee’s information state, one for each function.

3.2 Entailed multifunctionality

Entailed multifunctionality occurs when preconditions of one communicative function logically imply preconditions of another function.

Such relations usually occur between communicative functions of which one is a specification of the other (e.g. *Warning* and *Inform*), and hold between functions in the same dimension. In terms of context update, the update operation of the entailed function is subsumed by the update operation of the entailing function. Such entailed functions are, therefore, semantically vacuous.

There is, however, a less trivial case of entailment between functions, namely between *auto-feedback functions*, which provide information about the speaker's processing of some previous utterance (Bunt, 2009b), and *backward-looking functions*, such as *Answer*, *Confirm* or *Accept Request*, which "indicate how the current utterance relates to the previous discourse" (Allen and Core, 1997). Obviously, responding to some earlier utterance of the communication partner implies positive processing of the utterance being responded to (or at least speaker's belief that this was the case). Importantly, entailed feedback should be seen as a real source of multifunctionality since it involves an update of speaker's assumptions about the processing of a previous utterance by himself and his partner.

3.3 Implicated multifunctionality

Implicated functionality is found when one of the functions of a segment occurs by virtue of a conversational implicature. It is, therefore, context-dependent and intentional, and corresponds to an additional context update operation. An example is positive feedback implicated by shifting to a new but related topic.

3.4 Indirect multifunctionality

Indirect multifunctionality is a result of an indirect speech act. It is argued, however, that in information-state update approaches many of the indirect speech acts can be analysed in terms of conditional dialogue acts. For example, an utterance such as "Do you know what time it is?" is analysed as "Please tell me what time it is if you know." It remains an empirical question whether all indirect acts could be analysed in this way.

4 Ranked Annotation System

Ranked multidimensional dialogue act annotation was proposed by Włodarczak (2009). It assumes that while utterances are multifunctional, in a given context accomplishing some of the speaker's goals is of greater importance than accomplishing some other goals, and, hence, some utterance functions might dominate other functions.

The relative prominence of communicative functions was modelled by means of a *greater or equal prominence* relation, where the term *prominence* denotes the significance of a communicative function relative to other functions of the same utterance. It is assumed that prominences of any two functions of the same utterance are comparable, i.e. it is possible to decide whether one of the functions is more prominent than the other or whether they are equally prominent. Since more than one function is allowed to have the same prominence, the relation in question imposes a *non-strict linear order* on the set of functions of an utterance. As indicated above, the ordering of functions is viewed here from the speaker's point of view, i.e. it reflects the hierarchy of speaker's communicative goals.

Importantly, the above framework offers more flexibility than similar approaches (e.g. *Dominant Function Approximation*, Popescu-Belis (2008)) by allowing more than one highest ranking function, and more than two different ranks.

5 Ranking and types of multifunctionality

Given that entailment relations between communicative functions of utterances are formulated in terms of function preconditions and context update operations, and prominence relations are defined in terms of hierarchy of communicative goals, a connection between the two could be stipulated. Namely, it could be expected that in an annotation task in which participants are instructed to rank communicative functions of each utterance, independent functions, expressed by segment features and possibly corresponding to independent, highest-ranking communicative goals, should be in most cases assigned the same rank. On the other hand, entailed and implicated functions, corresponding to subordinate goals (and, at least for intra-dimensional entailment, semantically vacuous), could be expected to be ranked lower than functions expressed explicitly by utterance features. Similarly, entailed auto-feedback

should in most cases be ranked lower than the entailing backward-looking functions.

6 Experiment

Two experiments were conducted to find out whether the backward-looking *Confirm* function is ranked higher than the entailed positive auto-feedback. In the first experiment, annotators were asked to order functions assigned to segments with respect to their relative prominence. In the second, the annotators were first asked to assign the (possibly multiple) applicable communicative functions to pre-defined segments and then assign a prominence rank to each of them.

6.1 Corpus and tagset

HCRC Map Task Corpus was used in both experiments¹. Map task dialogues are so-called instructing dialogues in which one participant navigates another participant through a map. The total duration of the data selected for the experiments was equal to 4 minutes and 43 seconds.

The tagset chosen for the experiment was the DIT⁺⁺ dialogue act taxonomy (Bunt, 2009b). It consists of ten dimensions related to managing the task domain (*Task/Activity*), feedback (*Allo-* and *Auto-feedback*), time requirements (*Time Structuring*), problems connected with production of utterances (*Own* and *Partner Communication Management*), attention (*Contact Management*), discourse structure (*Discourse Structuring*) and social conventions (*Social Obligations Management*).

The data was segmented in multiple dimensions according to the approach presented in (Geertzen et al., 2007). 136 functional segments were identified. For the first experiments the data was annotated by two expert annotators. Full agreement was established on segmentation and annotation level beforehand. Specifically, ten segments were labelled with a *Confirm* tag.

6.2 Participants, task and procedure

The experiments were performed by naive annotators. The annotators were four undergraduate students. They had been introduced to the annotation scheme and the underlying theory while participating in a course on pragmatics during which they were exposed to approximately three hours of lecturing and a few small annotation exercises on data

other than map task dialogues.

All annotators accomplished both tasks individually, having received the materials (transcriptions and sound files) in electronic form. Time for both tasks was not limited. To encourage high quality of annotations the students were motivated by an award of 10% of the total grade for the pragmatic course.

In both experiments the ordering was done by assigning each function a numerical value from the set of subsequent natural numbers, starting from “1” as the most dominant function. More than one function could be assigned the same numerical value.

Since in the second experiment the same dialogue material was used, a two week break was made between the experiments to avoid the annotators being biased by the pre-annotated data from the first experiment.

As pointed out in section 3, entailed auto-feedback is a source of true multifunctionality. Annotators were, therefore, asked to include it in their annotations.

6.3 Results and discussion

Table 1 presents inter-annotator agreement about functions assigned ranks of one, two and three for each annotator pair and for each experiment calculated using Cohen’s kappa (Cohen, 1960). In the first experiment the mean kappa values for the ranks of one, two and three were equal to 0.64, 0.62 and 0.85 respectively, which indicates a high degree of agreement. For the second task the respective values are substantially lower (0.42, 0.27 and 0.58 respectively) but it should be borne in mind that these scores indicate agreement on a joint task of annotation and ranking, and that the annotators only had limited annotation experience.

In the first experiment the annotators reached nearly perfect agreement about ranking the entailed auto-feedback function lower than the entailing *Confirm* function. In the second task the entailed feedback was ranked higher in three cases.² Additionally, there were some cases in which the annotators did not annotate the entailed feedback at all, which in itself might indicate that it is treated as less prominent by inexperienced annotators. Table 2 gives inter-annotator agreement regarding the relative ranks of *Confirm* and entailed

¹Detailed information about the project can be found at <http://www.hcrc.ed.ac.uk/maptask/>

²All these cases came from one annotator, and correspond to 27% of all *Confirms* identified by this participant.

Annotators	Ranking of pre-annotated functions			Annotation and ranking (joint task)		
	Rank 1	Rank 2	Rank 3	Rank 1	Rank 2	Rank 3
1 & 2	0.63	0.59	0.84	0.53	0.32	0.76
1 & 3	0.76	0.75	0.92	0.38	0.17	0.57
1 & 4	0.62	0.54	0.85	0.43	0.29	0.57
2 & 3	0.53	0.53	0.82	0.51	0.19	0.57
2 & 4	0.59	0.55	0.83	0.34	0.21	0.61
3 & 4	0.69	0.71	0.86	0.34	0.44	0.42

Table 1: Cohen’s kappa scores for two rating experiments per annotator pair

Annotators	Experiment 1	Experiment 2
1 & 2	1.00	0.29
1 & 3	0.89	0.19
1 & 4	0.74	0.37
2 & 3	0.89	0.93
2 & 4	0.74	0.60
3 & 4	0.89	0.48

Table 2: Cohen’s kappa scores for relative ranks assigned to auto-feedback and *Confirm* for two experiments per annotator pair

auto-feedback for each experiment and each annotator pair. Mean kappa values were equal to 0.86 and 0.48 for the pre-annotated and not pre-annotated data respectively³.

7 Conclusions

A strong tendency was found for entailed positive feedback to be ranked lower than the entailing *Confirm* function by naive annotators. This was true both for dialogues in which functions were pre-annotated by experts and those in which annotators assigned functions to pre-defined segments themselves.

Although the low number of analysed items does not allow to draw definite conclusions, the results suggest that entailment relations might be a major factor influencing relative prominences of communicative functions, with entailed functions being perceived as less prominent than entailing functions.

Additionally, inter-annotator agreement about ranking of pre-annotated functions was found to be very high, with fair to moderate agreement in the joint task of annotation and ranking.

³Since three annotators failed to rank functions of some utterances in the first experiment, kappa values are not precisely equal to one.

References

- James Allen and Mark Core. 1997. *DAMSL: Dialogue Act Markup in Several Layers (Draft 2.1)*. Technical Report, Multiparty Discourse Group, Discourse Resource Initiative
- Harry Bunt. 1994. Context and Dialogue Control. *Think Quarterly* 3(1).
- Harry Bunt. 2006. Dimensions in Dialogue Act Annotation. *Proceedings of LREC 2006*, Paris.
- Harry Bunt. 2009a. Multifunctionality and multidimensional dialogue semantics. *Proceedings of Dia-Holmia 2009*, Stockholm.
- Harry Bunt. 2009b. The DIT++ Taxonomy for functional dialogue markup. *Proceedings of the EDAML@AAMAS, Workshop "Towards a Standard Markup Language for Embodied Dialogue Acts"*, Budapest.
- Alexander Clark and Andrei Popescu-Belis. 2004. Multi-level Dialogue Act Tags. *Proceedings of SIGDIAL'04 (5th SIGdial Workshop on Discourse and Dialogue)*, Cambridge, MA.
- Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Education and Psychological Measurement*, 20: 37–46.
- Jeroen Geertzen, Volha Petukhova and Harry Bunt. 2007. A Multidimensional Approach to Utterance Segmentation and Dialogue Act Classification. In: *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, Antwerp, pp. 140–149.
- Andrei Popescu-Belis. 2005. *Dialogue Acts: One or More Dimensions?* ISSCO Working Paper n. 62, University of Geneva.
- Andrei Popescu-Belis. 2008. Dimensionality of dialogue act tagsets: an empirical analysis of large corpora. *Language Resource and Evaluation* 42 (1).
- Marcin Włodarczak. 2009. *Ranked multidimensional dialogue act annotation*. MA thesis, Adam Mickiewicz University, Poznan.