

The independence of dimensions in multidimensional dialogue act annotation

Volha Petukhova and **Harry Bunt**
Tilburg Center for Creative Computing
Tilburg University, The Netherlands,
{v.petukhova,h.bunt}@uvt.nl

Abstract

This paper presents empirical evidence for the orthogonality of the DIT⁺⁺ multidimensional dialogue act annotation scheme, showing that the ten dimensions of communication which underly this scheme are addressed independently in natural dialogue.

1 Introduction

Studies of human dialogue behaviour indicate that natural dialogue utterances are very often multifunctional. This observation has inspired the development of multidimensional approaches to dialogue analysis and annotation, e.g. (Allen & Core, 1997), (Larsson, 1998), (Popescu-Belis, 2005), (Bunt, 2006). The most frequently used annotation scheme that implements this approach is DAMSL (Allen and Core, 1997), which allows multiple labels to be assigned to utterances in four layers: Communicative Status, Information Level, Forward-Looking Function (FLF) and Backward-Looking Function (BLF). The FLF layer is subdivided into five classes, including (roughly) the classes of commissive and directive functions, well known from speech act theory. The BLF layer has four classes: Agreement, Understanding, Answer, and Information Relation. These nine classes, also referred to as ‘dimensions’, form mutually exclusive sets of tags; no further motivation is given for the particular choice of classes.

Popescu-Belis (2005) argues that dialogue act tagsets should seek a multidimensional theoretical grounding and defines the following aspects of utterance function that could be relevant for choosing

dimensions (1) the traditional clustering of illocutionary forces in speech act theory into five classes: Representatives, Commissives, Directives, Expressives and Declarations; (2) turn management; (3) adjacency pairs; (4) topical organization in dialogue; (5) politeness functions; and (6) rhetorical roles.

Structuring an annotation scheme by grouping related communicative functions into clusters makes the structure of the schema more transparent. Such clusters or ‘dimensions’ are usually defined as a set of functions related to the same type of information, such as Acknowledging Understanding and Signalling Non-understanding, or Dialogue Opening and Dialogue Closing. Bunt (2006) shows that this does not always lead to a notion of dimension that has any conceptual and theoretical significance, and argues that some of the function classes of DAMSL do not constitute proper dimensions.

In particular, a theoretically grounded multidimensional schema should provide an account of the possible multifunctionality of dialogue utterances. In (Bunt, 2006); (Bunt and Girard, 2005) a dimension in dialogue act analysis is defined as *an aspect of participating in dialogue* which can be addressed:

- *by dialogue acts which have a function specifically for dealing with this aspect;*
- *independently of the other dimensions.*

The independence of dimensions, required by this definition, has the effect that an utterance may have a function in one dimension independent of the functions that it may have in other dimensions, and helps to explain why utterances may have multiple functions. Moreover, it leads to more manageable and

more adaptable annotation schemas (compared to, for instance, DAMSL and its derivatives), since it allows annotators to leave out certain dimensions that they are not interested in, or to extend the schema with additional dimensions; and it allows restricting or modifying the set of tags in a particular dimension without affecting the rest of the schema.

Based on the above definition and extensive theoretical and empirical studies, 10 dimensions are defined in the DIT⁺⁺ dialogue act annotation scheme¹: the domain or task/activity (*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, OCM*) or that of other interlocutors (*Partner Communication Management, PCM*); 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, SOM*).

This paper investigates the independence of these ten dimensions. In Section 2 we discuss the notion of independence of dimensions and how it can be tested. Section 3 reports test results and Section 4 draws conclusions.

2 Independence of dimensions

We define two dimensions D1 and D2 in an annotation scheme to be independent iff (1) an utterance may be assigned a value in D1 regardless of whether it is assigned a value in D2; and (2) it is not the case that whenever an utterance has a value in D1, this determines its value in D2.²

Dependences between dimensions can be determined empirically by analyzing annotated dialogue data. Dimension tags which always co-occur are nearly certainly dependent; zero co-occurrence scores also suggest possible dependences. Besides co-occurrence scores, we also provide a statistical analysis using the phi coefficient as a measure of

¹For more information about the scheme and its dimensions please visit <http://dit.uvt.nl/>

²See Petukhova and Bunt (2009) for a more extensive discussion.

relatedness. The phi measure is related to the chi-square statistic, used to test the independence of categorical variables, and is similar to the correlation coefficient in its interpretation.

If a dimension is not independent from other dimensions, then there would be no utterances in the data which address only that dimension. We therefore also investigate to which extent it happens that an utterance addresses only one dimension. We also investigate whether a dimension is addressed only in reaction to a certain other dimension. For example, the *answer* dimension as defined in DAMSL cannot be seen as independent, because *answers* need *questions* in order to exist. The test here is to examine the relative frequencies of pairs <dimension tag, previous dimension tag>.

To sum up, we performed four tests, examining:

1. the relative frequency of *communicative function co-occurrences* across dimensions;
2. *the extent of relatedness between dimensions* measure with the phi coefficient;
3. for all dimensions whether there are utterances *addressing only that dimension*;
4. the relative frequency of pairs of *dimension* and *previous dimension*.

3 Test results

Since different types of dialogue may have different tag distributions, three different dialogue corpora have been examined:

- The DIAMOND corpus³ of two-party instructional human-human Dutch dialogues (1,408 utterances);
- The AMI corpus⁴ of task-oriented human-human multi-party English dialogues (3,897 utterances);
- The OVIS corpus⁵ of information-seeking human-computer Dutch dialogues (3,942 utterances).

All three corpora were manually segmented and tagged according to the DIT⁺⁺ annotation scheme.

³For more information see Geertzen, J., Girard, Y., and Morante R. 2004. The DIAMOND project. Poster at CATALOG 2004.

⁴Augmented Multi-party Interaction (<http://www.amiproject.org/>)

⁵Openbaar Vervoer Informatie System (Public Transport Information System) <http://www.let.rug.nl/vannoord/Ovis/>

within	Task	Auto-F.	Allo-F.	Turn M.	Time M.	DS	Contact M.	OCM	PCM	SOM
Task	-	0.05(67.9)	0(24.9)	10.2(97.5)	1.4(2.4)	1.4(1.5)	0(0.4)	5.1(69.6)	0(0.1)	0(0.7)
Auto-F.	0.7(78.9)	-	0(0)	9.1(98.7)	0.6(1.4)	0.3(1.2)	0(20.2)	0(0.7)	0(65.0)	0(0.7)
Allo-F.	0(24.9)	0	-	59.2(94.8)	1.2(35.7)	0(2.1)	0(1.2)	0(7.9)	0.6(0.7)	0(0.3)
Turn M.	50.2(76.0)	3.5(66.2)	5.6(19.4)	-	8.0(42.9)	1.2(3.9)	0.1(13.8)	25(99.6)	0.2(1.0)	0.2(0.5)
Time M.	28.2(13.4)	0.5(11.3)	2.8(7.8)	96.9(98.6)	-	0.7(1.7)	0(0)	2.5(83.2)	0(0.5)	0(0)
DS	28.3(92.2)	0.4(58.3)	0(29.1)	22.6(87.5)	4.2(4.9)	-	0(25.0)	0(3.7)	0(0)	3.2(12.5)
Contact M.	0(2.4)	0(97.1)	0(1.6)	18.2(98.8)	0(0)	0(2.4)	-	0(0.3)	0(0)	0(0)
OCM	75.5(82.2)	0(0.8)	0(2.5)	82.9(96.9)	3.4(7.8)	1.3(3.9)	0(13.5)	-	0(0.9)	0.2(0.6)
PCM	0(11.8)	0(65.0)	4.9(11.8)	12.2(79.1)	0(12.2)	0(0)	0(0)	0(0)	-	0(0)
SOM	0(0.7)	0(80.0)	0(10.0)	6(90.0)	0(0)	10.0(30.0)	0(0)	0(2.0)	0(0)	-

Table 1: Co-occurrences of communicative functions across dimensions in AMI corpus expressed in relative frequency in % implicated and entailed functions excluded and included (in brackets).

The test results presented in this section are similar for all three corpora.

The co-occurrence results in Table 1 show no dependences between dimensions, although some combinations of dimensions occur frequently, e.g. time and turn management acts often co-occur. A speaker who wants to win some time to gather his thoughts and uses Stalling acts mostly wants to continue in the sender role, and his stalling behaviour may be intended to signal that as well (i.e., to be interpreted as a Turn Keeping act). But stalling behaviour does not always have that function; especially an extensive amount of stallings accompanied by relatively long pauses may be intended to elicit support for completing an utterance.

It is also interesting to have a look at co-occurrences of communicative functions taking implicated and entailed functions into account (the corpora were reannotated for this purpose). An implicated function is for instance the positive feedback (on understanding and evaluating the preceding utterance(s) of the addressee) that is implied by an expression of thanks; examples of entailed functions are the positive feedback on the preceding utterance that is implied by answering a question, by accepting an invitation, or by rejecting an offer.

Co-occurrence scores are higher when entailed and implicated functions are taken into account (the scores given in brackets in Table 1). For example, questions, which mostly belong to the Task dimension, much of the time have an accompanying Turn Management function, either releasing the turn or assigning it to another dialogue participant, allowing the question to be answered. Similarly, when accepting a request the speaker needs to have the turn, so communicative functions like Accept Re-

quest will often be accompanied by functions like Turn Take or Turn Accept. Such cases contribute to the co-occurrence score between the Task and Turn Management dimensions.

Table 1 shows that some dimensions do not occur in combination. We do not find combinations of Contact and Time Management, Contact and Partner Communication Management, or Partner Communication Management and Discourse Structuring, for example. Close inspection of the definitions of the tags in these pairs of dimensions does not reveal combination restrictions that would make one of these dimensions depend on the others.

Table 2 presents the extent to which dimensions are related when the corpus data are annotated with or without taking implicated and entailed functions into account, according to the calculated phi coefficient.

No strong positive (phi values from .7 to 1.0) or negative (-.7 to -1.0) relations are observed. There is a weak positive association (.6) between Turn and Time Management (see co-occurrence analysis above) and between OCM and Turn Management (.4). Weak negative associations are observed between Task and Auto-feedback (-.5) when entailed and implicated functions are not considered; between Task and Contact Management (-.6); and between Auto- and Allo-feedback (-.6) when entailed and implicated functions are included in the analysis. The weak negative association means that an utterance does not often have communicative functions in these two dimensions simultaneously. Some negative associations become positive if we take entailed and implicated functions into account, because, as already noted, dialogue acts like answers, accepts and rejects, imply positive feedback.

Dimensions	Task	Auto-F.	Allo-F.	Turn M.	Time M.	Contact M.	DS	OCM	PCM	SOM
Task		.1	.3	.06	-.4	-.6	.03	-.03	-.1	.04
Auto-F.	-.5		-.6	.1	-.3	.2	-.02	-.02	-.1	.04
Allo-F.	-.2	-.03		.09	-.1	-.2	.03	-.01	-.02	-.01
Turn M.	-.03	-.04	.14		.6	.04	-.06	.02	.02	-.03
Time M.	-.4	-.06	.14	.6		-.1	-.02	.04	-.03	-.02
Contact M.	-.05	-.006	-.003	.001	-.007		.04	-.01	-.04	-.03
DS	-.2	-.02	-.01	-.01	-.02	-.002		-.01	-.01	.2
OCM	.01	-.05	.02	.4	-.03	-.006	-.003		-.03	-.007
PCM	-.1	-.01	.01	-.006	.01	-.001	-.005	-.01		-.003
SOM	-.1	-.01	-.007	-.02	-.02	-.001	.05	-.007	-.003	

Table 2: Extent of relation between dimensions for AMI corpus expressed in the Phi coefficient (implicated and entailed functions excluded (white cells) and included (grey cells)).

The third independence test, mentioned above, shows that each dimension may be addressed by an utterance which does not address any other dimension. The Task dimension is independently addressed in 28.8% of the utterances; 14.2% of the utterances have a function in the Auto-Feedback dimension only; for the other dimensions these figures are 0.7% - Allo-Feedback; 7.4% - Turn Management; 0.3% - Time Management; 0.1% - Contact Management; 1.9% - Discourse Structuring; 0.5% - OCM; 0.2% - PCM; and 0.3% - SOM.

within	Task	Auto-F.	Allo-F.	Turn M.	Time M.	Contact M.	DS	OCM	PCM	SOM
Task	21.2	27.4	27.7	20	32.5	0	7.1	16.4	15.2	32.1
Auto-F.	15	24.4	25	21.4	15.4	27.8	12.3	7.5	22.7	12.8
Allo-F.	0.4	1.3	5.8	0.5	0.5	0	0.8	0.4	0	0
Turn M.	14.3	4.7	0	8.5	8.2	0	8.5	2.2	7.6	6.4
Time M.	22.2	18.3	16.7	23.5	15	0	36.5	47.1	37.9	19.2
Contact M.	0	0.1	0	0.2	0	27.8	0	0	0	0
DS	2	2	0	0.5	0.5	27.8	5.2	0.4	0	0
OCM	7.7	8.3	5.8	7.7	11.2	0	0	7	0	0
PCM	0.4	0.4	0	0	0.08	0	0	0.2	0	0
SOM	0.1	0.3	0	1.2	0.08	0	0.8	0	0	6.4

Table 3: Overview of relative frequency (in%) of pairs of dimension and previous dimensions by previous utterances observed in AMI data, per dimension, drawn from the set of 5 pairs from the dialogue history.

We finally investigated the occurrences of tags given the tags of the previous utterances, taking five previous utterances into account. Table 3 shows no evidence of dependences across the dialogue history. There are some frequent patterns, for example, retractions and self-corrections often follow hesitations because the speaker, while monitoring his own speech and noticing that part of it needs revision, needs time to construct the corrected part.

4 Conclusions

In this paper we investigated the independence of the dimensions defined in the DIT++ dialogue act

annotation scheme, using co-occurrences matrices and the phi coefficient for measuring relatedness between dimensions.

The results show that, although some dimensions are more related and co-occur more frequently than others, on the whole the ten DIT++ dimensions may be considered to be independent aspects of communication.

Acknowledgments

This research was conducted as part of ISO project 24617-2: Semantic annotation framework, Part 2: Dialogue acts, and sponsored by Tilburg University.

References

- James F. Allen and Mark G. Core. 1997. Draft of DAMSL: Dialog Act Markup in Several Layers.
- Jens Allwood. 2000. An activity-based approach to pragmatics. In Bunt, H., and Black, W. (eds.) *Abduction, Belief and Context in Dialogue; Studies in Computational Pragmatics*, pp. 47–80. Benjamins, Amsterdam.
- Harry Bunt and Yann Girard. 2005. Designing an open, multidimensional dialogue act taxonomy. In Gardent, C., and Gaijfe, B. (eds.) *Proc. 9th Workshop on the Semantics and Pragmatics of Dialogue*, pp. 37–44.
- Harry Bunt. 2006. Dimensions in dialogue annotation. In *Proceedings of LREC 2006*.
- Mark G. Core and James F. Allen. 1997. Coding dialogues with the DAMSL annotation scheme. In *Working Notes: AAI Fall Symposium on Communicative Action in Humans and Machines*, pp. 28–35.
- Staffan Larsson. 1998. Coding Schemas for Dialogue Moves. *Technical report from the S-DIME project*.
- Volha Petukhova and Harry Bunt. 2009. Dimensions in communication. *TiCC Technical Report 2009-002*, Tilburg University.
- Andrei Popescu-Belis. 2005. Dialogue Acts: One or More Dimensions? *ISSCO Working Paper 62, ISSCO*.