Dialogue act classification is a laughing matter

Vladislav Maraev
University of Gothenburg
vladislav.maraev@gu.se

Bill Noble
University of Gothenburg
bill.noble@gu.se

Chiara Mazzocconi
Aix-Marseille University
chiara.mazzocconi@live.it

Christine Howes
University of Gothenburg
christine.howes@gu.se

Abstract

In this paper we explore the role of laughter in attributing communicative intents to utterances, i.e. detecting the dialogue act performed by them. We conduct a corpus study in adult phone conversations showing how different dialogue acts are characterised by specific laughter patterns, both from the speaker and from the partner. Furthermore, we show that laughs can positively impact the performance of Transformer-based models in a dialogue act recognition task. Our results highlight the importance of laughter for meaning construction and disambiguation in interaction.

1 Introduction

Laughter is ubiquitous in our everyday interactions. In the Switchboard Dialogue Act Corpus (SWDA, Jurafsky et al., 1997a) (US English, phone conversations where two participants that are not familiar with each other discuss a potentially controversial subject, such as gun control or the school system) non-verbally vocalised dialogue acts (whole utterances that are marked as non-verbal, 66% of which contain laughter) constitute 1.7% of all dialogue acts. Laughter tokens[1] make up 0.5% of all the tokens that occur in the corpus. Laughter relates to the discourse structure of dialogue and can refer to a *laughable*, which can be a perceived event or an entity in the discourse. Laughter can precede, follow or overlap the laughable, and the time alignment between them depends on who produces the laughable, the form of the laughter, and the pragmatic function performed (Tian et al., 2016).

Bryant (2016) shows how listeners are influenced towards a non-literal interpretation of sentences when accompanied by laughter. Similarly, Tepperman et al. (2006) shows that laughter can act as a contextual feature for determining the sincerity of an utterance, e.g. when detecting sarcasm.

Nevertheless there is a dearth of research exploring the use of laughter in relation to different dialogue acts in detail, and therefore little is known about the role that laughter may have in facilitating the detection of communicative intentions.

Based on previous work and the corpus study presented in this paper, we argue that laughter is tightly related to the information structure of a dialogue. In this paper, we investigate the potential of laughter to disambiguate the meaning of an utterance, in terms of the dialogue act it performs. To do so, we employ a Transformer-based model and look into laughter as a potentially useful feature for the task of *dialogue act recognition* (DAR). Laughs are not present in a large-scale pre-trained models, such as BERT (Devlin et al., 2019), but their representations can be learned while training for a dialogue-specific task (DAR in our case). We further explore whether such representations can be additionally learned, in an unsupervised fashion, from dialogue-like data, such as a movie subtitles corpus, and if it further improves the performance of our model.

The paper is organised as follows. We start with some brief background in Section 2. In Section 3 we observe how dialogue acts can be classified with respect to their collocations with laughs and discuss the patterns observed in relation to the pragmatic functions that laughter can perform in dialogue. In Section 4 we report our experimental results for the DAR task depending on whether the model includes laughter. We further investigate whether non-verbal dialogue acts can be classified as more specific dialogue acts by our model. We conclude with a discussion and outlining the directions for further work in Section 5.

[1] Switchboard Dialogue Act Corpus does not include speech-laughs.

2 Background

2.1 Laughter

Laughter does not occur only in response to humour or in order to frame it. It is crucial in managing conversations in terms of dynamics (turn-taking and topic-change), at the lexical level (signalling problems of lexical retrieval or imprecision in the lexical choice), but also at a pragmatic (marking irony, disambiguating meaning, managing self-correction) and social level (smoothing and softening difficult situations or showing (dis)affiliation) (Glenn, 2003; Jefferson, 1984; Mazzocconi, 2019; Petitjean and González-Martínez, 2015).

Moreover Romaniuk (2009) and Ginzburg et al. (2020) discuss how laughter can answer or decline to answer a question, and Mazzocconi et al. (2018) explore laughter as an object of clarification requests, signalling the need for interlocutors to clarify its meaning (e.g., in terms of what the "laughable" is) to carry on with the conversation.

2.2 Dialogue act recognition

The concept of a dialogue act (DA) is based on that of the speech act (Austin, 1975). Breaking with classical semantic theory, Speech Act Theory considers not only the propositional content of an utterance but also the actions, such as *promising* or *apologising*, it carries out. Dialogue acts extend the concept of the speech act, with a focus on the interactional nature of most speech. DAMSL (Core and Allen, 1997), for example, is an influential DA tagging scheme where DAs are defined in part by whether they have a *forward-looking* function (expecting a response) or *backward-looking* function (in response to a previous utterance).

Dialogue act recognition (DAR) is the task of labelling utterances with the dialogue act they perform, given a set of dialogue act tags. As with other sequence labelling tasks in NLP, some notion of context is helpful in DAR. One of the first performant machine learning models for DAR was a Hidden Markov Model that used various lexical and prosodic features as input (Stolcke et al., 2000).

Recent state-of-the-art approaches to dialogue act recognition have used a hierarchical approach, using large pre-trained language models like BERT to represent utterances, and adding some representation of discourse context at the dialogue level (e.g., Ribeiro et al., 2019; Mehri et al., 2019). However Noble and Maraev (2021) observe that without fine-tuning, standard BERT representations perform very poorly on dialogue, even when paired with a discourse model, suggesting that certain utterance-internal features missing from BERT's textual pre-training data (such as laughter) may have an adverse effect on dialogue act recognition.

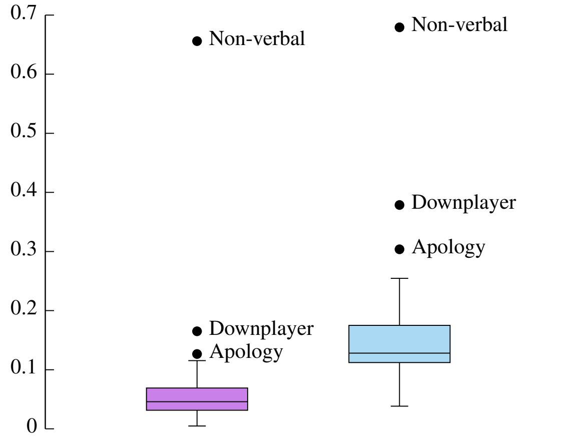


Figure 1: Box plots for proportions of dialogue acts which contain laughs in SWDA. On the left: proportion of DAs containing laughter, on the right: proportion of DAs having laughter in one of the adjacent utterances.

3 Laughs in the Switchboard Dialogue Act Corpus

In this section we analyse dialogue acts in the Switchboard Dialogue Act Corpus according to their collocation with laughter and provide some qualitative insights based on the statistics.

SWDA is tagged with a set of 220 dialogue act tags which, following Jurafsky et al. (1997b), we cluster into a smaller set of 42 tags.

The distribution of laughs in different dialogue acts has a rather uniform shape with a few outliers (Figure 1). The most distinct outlier is the *Non-verbal* dialogue act which is misleading with respect to laughter, because utterances only containing a single laughter token fall into this category. However isolated laughs can serve, for example, to acknowledge a statement, to deflect a question, or to show appreciation (Mazzocconi, 2019). We will further conjecture on this class of DAs in Sec. 4.5.

3.1 Method

Let us illustrate our comparison schema using the other two outliers, *Downplayer* (make up 0.05% of all utterances) and *Apology* (0.04%), comparing them with the most common dialogue act in SWDA – *Statement-Non-Opinion* (33.27%). We consider laughter-related dimensions of an utterance, and

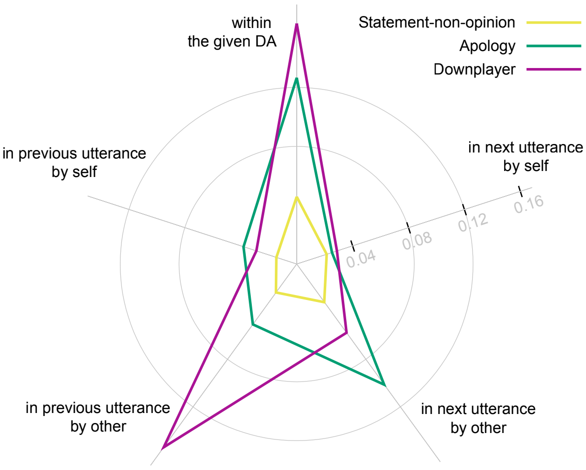


Figure 2: Comparison between the pentagonal representations of laughter collocations of dialogue acts.

create 5-dimensional (pentagonal) representations of DAs according to them. Each dimension's value is equal to the proportion of utterances of a given type which contain laughter:

- ↑ current utterance;
- ↖ immediately preceding utterance by the same speaker;
- ↗ immediately following utterance by the same speaker;
- ↙ immediately preceding utterance by the other speaker;
- ↘ immediately following utterance by the other speaker.

For instance, (1) is an illustrative example of the phenomenon shown in Figure 2.

(1) [2]	**A:**	I'm sorry to keep you waiting #<laughter>.#	*Apology*
	B:	#Okay# <laughter>. /	*Downplayer*
	A:	Uh, I was calling from work	*Statement (n/o)*

We show the representations of all dialogue acts on Figure 3. We believe that such a depiction helps the reader form impressions about similarities between DAs based on their laughter collocations and notice the ones that stand out in some respects.

To further assess the similarity between dialogue acts based on their collocations with laughs we factorise their pentagonal representations into 2D space using singular value decomposition (SVD). We can see that dialogue acts form some distinct clusters. The resulting plot is shown in Figure 6 in Appendix A.1. Let us now proceed with some qualitative observations.

[2]Overlapping material is marked with hash signs.

3.2 Observations

Laughter and modification or enrichment of the current DA We observe a higher proportion of laughter accompanying the current dialogue act (↑) when the laughter is aimed at modifying the current dialogue act with some degree of urgency to smooth or soften it (*Action-directive*, *Reject*, *Dispreferred answer*, *Apology*), to contribute to its enrichment stressing the positive disposition towards the partner (*Appreciation*, *Downplayer*, *Thanking*), or to cue for the need to consider a less probable meaning, therefore helping in non-literal meaning interpretation (*Rhetorical question*).

While *Apology* and *Downplayer* have rather distinct and peculiar patterns (Fig. 6) discussed in more detail below, we observe *Dispreferred answers*, *Action directives*, *Offers/Options/Commits* and *Thanking* to constitute a close cluster when considering the decomposed values of the pentagonal used for DA representation.

Laughter for benevolence induction and laughter as a response The patterns observed in relation to the preceding and following turns reflect the multitude of functions that laughter can perform in interaction, stressing the fact that it can be used both to induce or invite a determinate response (dialogue act) from the partner (*Downplayer*, *Agree/Accept*, *Appreciation*, *Acknowledge*) as well as being a possible answer to specific dialogue acts (e.g. *Apology*, *Offers/Options/Commits*, *Summarise/Reformulate*, *Tag-question*).

A peculiar case is the one of *Self-talk*, often followed by laughter by the same speaker. In this case the laughter may be produced to signal the incongruity of the action (in dialogue we normally speak to others, not to ourselves), while at the same time function to smooth the situation, for instance, when having issues of lexical retrieval, as in (2), or some degree of embarrassment from the speaker, when questioning whether a contribution is appropriate or not, as in (3).

(2)	**A:**	Have, uh, really, -	
	A:	what's the word I'm looking for,	*Self-talk*
	A:	I'm just totally drawing a blank <laughter>.	*Statement (n/o)*
(3)	**B:**	Well, I don't have a Mexi-, -	*Statement (n/o)*
	B:	I don't, shouldn't say that,	*Self-talk*
	B:	I don't have an ethnic maid <laughter>.	*Statement (n/o)*

Apology and Downplayer It is interesting to comment on the parallelisms of laughter usage in

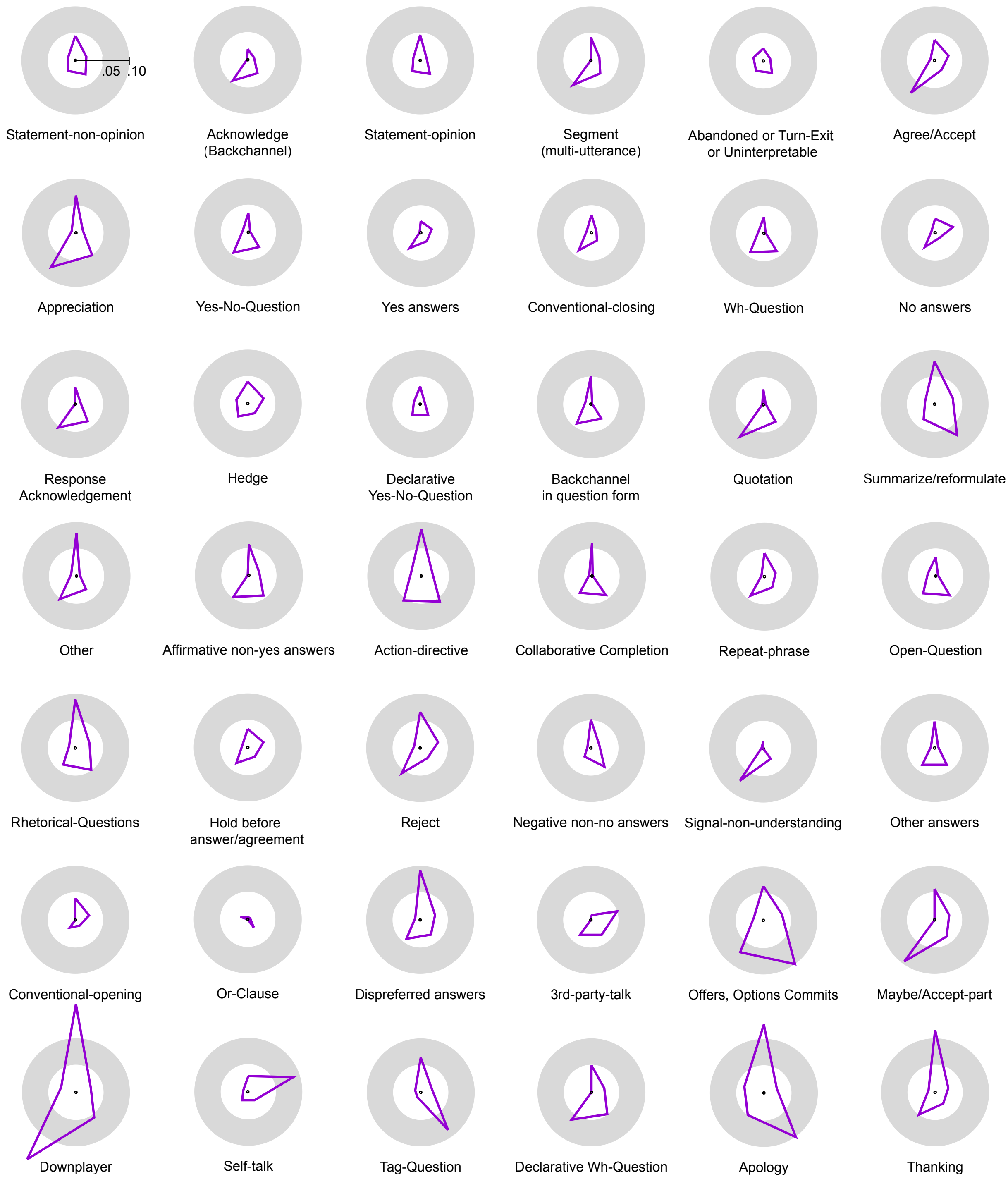


Figure 3: Pentagonal representation of dialogue acts: proportions of utterances which include laughter. Dimensions: ↑ current utterance; ↖ immediately preceding utterance by the same speaker; ↗ immediately following utterance by the same speaker; ↙ immediately preceding utterance by the other speaker; ↘ immediately following utterance by the other speaker. DAs are ordered by their frequency in SWDA (left-to-right, then top-to-bottom).

relation to *Apology* and *Downplayer*, represented in Fig. 2 in contrast to *Statement-non-opinion*, in as much as their graphic representations are more or less mirror-images of each other and show how the dialogue acts are linked by the pragmatic functions laughter can perform in dialogue.

In both *Apology* and *Downplayer* we observe a rather higher proportion of occurrences in which the dialogue act is accompanied by laughter (↑) in comparison to other DAs (Fig. 3). In the case of *Apology*, laughter can be produced to induce benevolence from the partner (Mazzocconi et al., 2020), while in the case of *Downplayer* the laughter can be produced to reassure the partner about some situation that had been appraised as discomforting (classified as *social incongruity* by Mazzocconi et al., 2020) and somehow signal that the issue should be regarded as not important (Romaniuk, 2009; **?**), as in (4).

(4) **A:** I don't, I don't think I could do that <laughter>. # *Statement (n/o)*
B: Oh, it's not bad at all. *Downplayer*
A: It's, it's a beautiful drive. *Statement (n/o)*

The interesting mirror-image patterns observable in the lower part of the graph can therefore be explained by considering the relation between the two dialogue acts. We observe cases in which an *Apology* is accompanied by a laughter, and then followed by a *Downplayer*, showing that the laughter's positive effect was attained and successful. This allows us to explain both the bottom left spike (↙) observed for *Downplayer* (often preceded by an utterance by the partner containing laughter) and the bottom right spike (↘) observed for *Apology* (often followed by an utterance by the partner containing laughter). In example (1) both the apology and the downplayer are accompanied by laughter, while in (5) a typical example of a laughter accompanying an *Apology* is followed by a *Downplayer*.

(5) **B:** I'm sorry <laughter>. # *Apology*
A: That's all right. / *Downplayer*
B: You, you were talking about, uh, uh, *Summarise*

We now turn to the question of whether our qualitative observations of patterns between laughs and dialogue acts can be used to improve a dialogue act recognition task.

4 The importance of laughter in artificial dialogue act recognition

4.1 Data

We perform experiments on the Switchboard Dialogue Act Corpus (SWDA, 42 dialogue act tags), which is a subset of the larger Switchboard corpus, and the dialogue act-tagged portion of the AMI Meeting Corpus (AMI-DA). AMI uses a smaller tagset of 16 dialogue acts (Gui, 2005).

Preprocessing We make an effort to normalise transcription conventions across SWDA and AMI. We remove disfluency annotations and slashes from the end of utterances in SWDA. In both corpora, acronyms are tokenised as individual letters. All utterances are lower-cased.

Utterances are tokenised using a word piece tokeniser (Wu et al., 2016) with a vocabulary of 30,000. We add a special laughter token to the vocabulary and map all transcribed laughter to that token. We also prepend each utterance with a speaker token that uniquely identifies the corresponding speaker within that dialogue.

Switchboard	**AMI Corpus**
Dyadic	Multi-party
Casual conversation	Mock business meeting
Telephone	In-person & video
English	English
Native speakers	Native & non-native speakers
2200 conversations	171 meetings
1155 in SWDA	139 in AMI-DA
400k utterances	118k utterances
3M tokens	1.2M tokens

Table 1: Comparison between Switchboard and the AMI Meeting Corpus

4.2 The model

To test the effectiveness of BERT for DAR, we employ a simple neural architecture with two components: an encoder that vectorises utterances, and a sequence model that predicts dialogue act tags from the vectorised utterances (Figure 4). Since we are primarily interested in comparing different utterance encoders, we use a basic RNN as the sequence model in every configuration.[3] The RNN takes the encoded utterance as input at each time step, and its hidden state is passed to a simple linear classification layer over dialogue act tags. Conceptually, the encoded utterance represents the context-agnostic features of the utterance, and the hidden state of the RNN represents the full discourse context.

As a baseline utterance encoder, we use a word-level CNN with window sizes of 3, 4, and 5, each with 100 feature maps (Kim, 2014). The model uses 100-dimensional word embeddings, which are initialised with pre-trained gloVe vectors (Pennington et al., 2014). For the BERT utterance encoder, we use the $\text{BERT}_{\text{BASE}}$ model with hidden size of 768 and 12 transformer layers and self-attention heads (Devlin et al., 2018, §3.1). In our implementation, we use the un-cased model provided by Wolf et al. (2019).

4.3 Experiment 1: Impact of laughter

In the first experiment we investigated whether laughter, as an example of a dialogue-specific signal, is a helpful feature for DAR. Therefore, we

[3] We have experimented with LSTM as the sequence model, but the accuracy was not significantly different compared to RNN. It can be explained by the absence of longer distance dependencies on this level of our model.

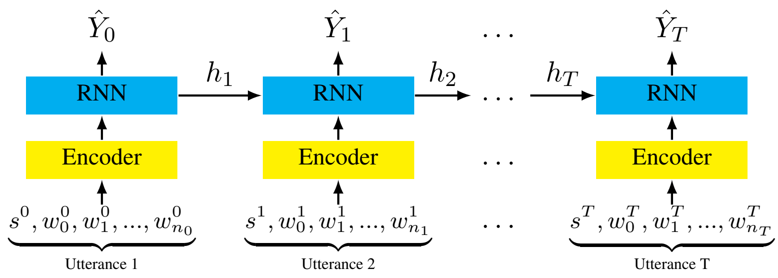


Figure 4: Simple neural dialogue act recognition sequence model

	SWDA		**AMI-DA**	
	F1	acc.	F1	acc.
`BERT-NL`	36.48	76.00	44.75	68.04
`BERT-L`	36.75	76.60	43.37	64.87
`CNN-NL`	36.95	73.92	38.00	63.18
`CNN-L`	37.59	75.40	37.89	64.27
Majority class	0.78	33.56	1.88	28.27

Table 2: Comparison of macro-average F1 and accuracy depending on using laughter on the training phase.

train another version of each model: one containing laughs (`L`) and one with laughs left out (`NL`), and compare their performances in DAR task. Table 2 compares the results from applying the models with two different utterance encoders (`BERT`, `CNN`).

`BERT` outperforms the `CNN` on AMI-DA. On SWDA, the two encoders are more comparable, though `BERT` has a slight edge in accuracy, suggesting that it relies more heavily on defaulting to common dialogue act tags. On SWDA, we see small improvements in accuracy and macro-F1 for models that included laughter. For AMI-DA, the effect of laughter is small or even negative – the impact of laughter on performance becomes more clear in the disaggregated performance over different dialogue acts. Indeed, laughter improves the accuracy of the model even on some dialogue acts in which laughter occurs rarely in the current and adjacent utterances (see Figure 7 in Appendix A).

Confusion matrices (Figure 5) provide some food for thought. Most of the misclassifications fall into the majority classes, such as *sd* (Statement-non-opinion), on left edge of the matrix. However, there are some important exceptions, such as *rhetorical questions*, that are misclassified as other forms of questions due to their surface question-like form. Importantly, laughter helps to classify rhetorical questions correctly, this is because in a conversation it can be used as a device to cancel seriousness or reduce commitment to literal meaning (Ginzburg et al., 2015; Tepperman et al., 2006)

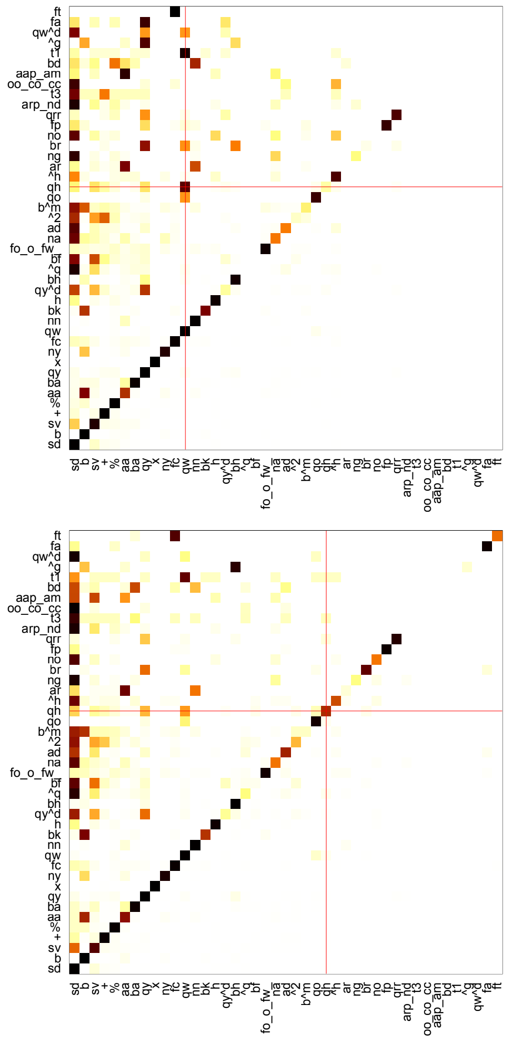


Figure 5: Confusion matrices for `BERT-NL` (top) vs `BERT-L` (bottom); SWDA corpus. Solid lines show classification improvement of rhetorical questions.

Therefore, questions, like the one we show in example (6), are easier to disambiguate with laughter.

(6) **B:** Um, as far as spare time, they talked about, *Statement (n/o)*
B: I don't, + I think, *Statement (n/o)*
B: who has any spare time `<laughter>`? *Rhetorical Quest.*
A: `<laughter>`. *Non-verbal*

4.4 Experiment 2: laughter and pre-training

As previously noted, training data for BERT does not include features specific to dialogue (e.g. laughs). We therefore experiment with a large and more dialogue-like corpus constructed from OpenSubtitles (Lison and Tiedemann, 2016) (350M tokens, where 0.3% are laughter tokens). We used a manually constructed list of words frequently used to refer to laughter in subtitles and replaced every occurance of one of these words with the special laughter token. We then collected every English-language subtitle file in which at least 1% of the utterances contained laughter (about 11% of the total). Because utterances are not labelled with speaker in the OpenSubtitles corpus, we randomly assigned a speaker token to each utterance to maintain the format of the other dialogue corpora.

The pre-training corpus was prepared for the combined masked language modelling and next sentence (utterance) prediction task, as described by Devlin et al. (2018).

We analyse how pre-training affects BERT's performance as an utterance encoder. To do so, we consider the performance of DAR models with three different utterance encoders: i) `FT` – pre-trained BERT with DAR fine-tuning; ii) `RI` – randomly initialised BERT (with DAR fine-tuning); iii) `FZ` – pre-trained BERT without fine-tuning (frozen during DAR training). For the pre-trained (`FT`, `FZ`) conditions we perform two types of pre-training: i) `OSL` – pre-training on the portion of OpenSubtitles corpus ii) `OSNL` – same as `OSL`, but with all the laughs removed. We fine-tune and test our models on the corpora containing laughs (`L`).

We observe that dialogue pre-training improves performance of the models. Fine-tuned models also perform better than the frozen ones because the latter provide less opportunities for the encoder to be trained for the specific task.

Including laughter in pre-training data improves F1 scores in most cases, except for the SWDA in the fine-tuned condition. The difference is especially pronounced for AMI-DA corpus in the fine-tuned condition (4.97 p.p. difference in F1). The question of relevance of movies subtitle data for either SWDA or AMI-DA can be a subject for further study, including the types of laughs in the corpora. It might be the case that nature of AMI-DA is congruent with those of movie subtitles, since participants in AMI-DA basically are role-playing being in a focus group rather than being involved in a natural dialogue. People might produce laughs in places only where they intuitively expected by them to be produced (i.e. humour related), just as in scripted movie dialogues.

	SWDA		AMI-DA	
	F1	acc.	F1	acc.
`BERT-L-FT`	36.75	76.60	43.37	64.87
`BERT-L+OSL-FT`	41.42	76.95	48.65	68.07
`BERT-L+OSNL-FT`	43.71	77.09	43.68	64.80
`BERT-L+OSL-FZ`	9.60	57.67	17.03	51.03
`BERT-L+OSNL-FZ`	7.69	55.29	16.99	51.46
`BERT-L-RI`	32.18	73.80	34.88	60.89
Majority class	0.78	33.56	1.88	28.27
SotA	-	83.1[4]	-	-

Table 3: Comparison of macro-F1 and accuracy with further dialogue pre-training.

4.5 Experiment 3: Laughter as a non-verbal dialogue act

In this experiment, following the observations regarding the misleading character of *Non-verbal* dialogue acts, we looked at the predictions that the model would give this class of dialogue acts if it wasn't aware of the *Non-verbal* class. To do so, we mask the outputs of the model where the desired class was *Non-verbal* and do not backpropagate these results. We used the `BERT-L-FT` for this experiment. After training we tested the resulting model on the test set containing 659 non-verbal dialogue acts, 413 of which contain laughter.

For 314 (76%) of such dialogue acts the model has predicted the *Acknowledge (Backchannel)* class and for 46 (11%) – continuations of the previous DA by the same speaker. The rest were classified as either something uninformative (the *Abandoned or Turn-Exit or Uninterpretable* class) or, from manual observation, clearly unrelated.

Acknowledge (Backchannel) can cover some uses of laughter, for instance, to show to the interlocutor acknowledgement of their contribution, implying the appreciation of an incongruity and inviting continuation, functioning simultaneously as a continuer and assessment feedback (Schegloff, 1982), as in example (7).

(7) (We mark continuations of the previous DA by the same speaker with a plus, and indicate misclassified dialogue acts with a star. Laughs shown in bold constitute *Non-verbal* dialogue acts)

[4] Kozareva and Ravi (2019)

B:	Everyone on the boat was catching snapper, snappers except guess who.	*Statement (n/o)*
A:	<laughter> It had to be you.	*Summ./reform.*
B:	<laughter> I ca-, I, -	*Uninterpretable*
A:	Couldn't catch one to save your life. Huh.	*Backchannel**
B:	That's right,	*Agree/Accept*
B:	I would go from one side of the boat to the other,	*Statement (n/o)*
B:	and, uh,	+
A:	**<laughter>.**	*Backchannel*
B:	the, uh, the party boat captain could not understand, you know,	+
B:	he even, even he started baiting my hook <laughter>,	*Statement (n/o)*
A:	**<laughter>.**	*Backchannel*
B:	and holding, holding the, uh, the fishing rod.	+
A:	How funny,	*Appreciation*

Nevertheless, these two cases clearly cannot account for all the examples discussed in the literature (e.g. standalone uses of laughter as signal of disbelief or negative response to a polar question Ginzburg et al., 2020) and above in Sec. 3.2. Future models will therefore require a manual assignment of meaningful dialogue acts to standalone laughs.

5 Discussion

The implications of the results obtained are twofold: showing that laughter can help a computational model to attribute meaning to an utterance and help with pragmatic disambiguation, and consequently stressing once again the need for integrating laughter (and other non-verbal social signals) in any framework aimed to model meaning in interaction (Ginzburg et al., 2020; Maraev et al., 2021).

Our results provide further evidence (e.g. Torres et al. (1997); Mazzocconi et al. (2021)) for the fact that non-verbal behaviours are tightly related to the dialogue information structure, propositional content and dialogue act performed by utterances. Laughter, along with other non-verbal social signals, can constitute a dialogue act in itself conveying meaning and affecting the unfolding dialogue (Bavelas and Chovil, 2000; Ginzburg et al., 2020).

In this work we have shown that laughter is a valuable cue for DAR task. We believe that in our conversations laughter is informative about interlocutors' emotional and cognitive appraisals of events and communicative intents. Therefore, it should not come as a surprise that laughter acts as a cue in a computational model.

On the question of laughter impact on the dialogue act recognition (DAR) task, this study found that laughter is more helpful in SWDA corpus than in AMI-DA. Due to the nature of interactions over the phone, SWDA dialogue participants can not rely on visual signals, such as gestures and facial expressions. Our results support the hypothesis that in SWDA, vocalizations such as laughter are more pronounced and therefore more helpful in disambiguating dialogue acts. This may also explain why our best models perform better on SWDA: more of the information that interlocutors and dialogue act annotators rely on is present in SWDA transcripts, whereas AMI-DA annotators receive clear instructions to pay attention to the videos (Gui, 2005). This finding is consistent with that of Bavelas et al. (2008) who demonstrate that in face-to-face dialogue, visual components, such as gestures, can convey information that is independent from what is conveyed by speech.

Laughter can be used to mark the presence of an incongruity between what is said and what is intended, coined as *pragmatic incongruity* by Mazzocconi et al. (2020). In those cases laughter is especially valuable for disambiguating between literal and non-literal meaning, as we have shown for rhetorical questions, a task which is still a struggle for most NLP models and dialogue systems.

There is abundant room for further study of how laughter can help to disambiguate communicative intent. Stolcke et al. (2000) showed that the specific prosodic manifestations of an utterance can be used to improve DAR. With respect to laughter, the form (duration, arousal, overlap with speech) can be informative about its function and position w.r.t. the laughable (Mazzocconi, 2019). Incorporating such information is crucial if models pre-trained on large-scale text corpora are to be adapted for use in dialogue applications.

Acknowledgments

Maraev, Noble and Howes were supported by the Swedish Research Council (VR) grant 2014-39 for the establishment of the Centre for Linguistic Theory and Studies in Probability (CLASP). Mazzocconi was supported by the "Investissements d'Avenir" French government program managed by the French National Research Agency (reference: ANR-16-CONV-0002) and from the Excellence Initiative of Aix-Marseille University—"A*MIDEX" through the Institute of Language, Communication and the Brain.

References

2005. Guidelines for Dialogue Act and Addressee Annotation Version 1.0.

John Langshaw Austin. 1975. *How to do things with words*, volume 88. Oxford university press.

Janet Bavelas, Jennifer Gerwing, Chantelle Sutton, and Danielle Prevost. 2008. Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, 58(2):495–520.

Janet Beavin Bavelas and Nicole Chovil. 2000. Visible acts of meaning: An integrated message model of language in face-to-face dialogue. *Journal of Language and social Psychology*, 19(2):163–194.

Gregory A Bryant. 2016. How do laughter and language interact? In *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11)*.

Mark G Core and James F Allen. 1997. Coding Dialogs with the DAMSL Annotation Scheme. In *Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines*, pages 28–35, Boston, MA.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805 [cs]*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Jonathan Ginzburg, Ellen Breitholtz, Robin Cooper, Julian Hough, and Ye Tian. 2015. Understanding laughter. In *Proceedings of the 20th Amsterdam Colloquium*, pages 137–146.

Jonathan Ginzburg, Chiara Mazzocconi, and Ye Tian. 2020. Laughter as language. *Glossa: a journal of general linguistics*, 5(1).

Phillip Glenn. 2003. *Laughter in Interaction*. Cambridge University Press, Cambridge, UK.

Gail Jefferson. 1984. On the organization of laughter in talk about troubles. In *Structures of Social Action: Studies in Conversation Analysis*, pages 346–369.

D Jurafsky, E Shriberg, and D Biasca. 1997a. Switchboard dialog act corpus. *International Computer Science Inst. Berkeley CA, Tech. Rep*.

Daniel Jurafsky, Liz Shriberg, and Debra Biasca. 1997b. Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual.

Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. *arXiv:1408.5882 [cs]*.

Zornitsa Kozareva and Sujith Ravi. 2019. ProSeqo: Projection Sequence Networks for On-Device Text Classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3894–3903, Hong Kong, China. Association for Computational Linguistics.

Pierre Lison and Jorg Tiedemann. 2016. OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*, page 7.

Vladislav Maraev, Jean-Philippe Bernardy, and Christine Howes. 2021. Non-humorous use of laughter in spoken dialogue systems. In *Linguistic and Cognitive Approaches to Dialog Agents (LaCATODA 2021)*, pages 33–44.

Chiara Mazzocconi. 2019. *Laughter in interaction: semantics, pragmatics and child development*. Ph.D. thesis, Université de Paris.

Chiara Mazzocconi, Vladislav Maraev, and Jonathan Ginzburg. 2018. Laughter repair. *Proceedings of SemDial*, pages 16–25.

Chiara Mazzocconi, Vladislav Maraev, Vidya Somashekarappa, and Christine Howes. 2021. Looking at the pragmatics of laughter. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 43.

Chiara Mazzocconi, Ye Tian, and Jonathan Ginzburg. 2020. What's your laughter doing there? A taxonomy of the pragmatic functions of laughter. *IEEE Trans. on Affective Computing*.

Shikib Mehri, Evgeniia Razumovskaia, Tiancheng Zhao, and Maxine Eskenazi. 2019. Pretraining Methods for Dialog Context Representation Learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3836–3845, Florence, Italy. Association for Computational Linguistics.

Bill Noble and Vladislav Maraev. 2021. Large-scale text pre-training helps with dialogue act recognition, but not without fine-tuning. In *Proceedings of the 14th International Conference on Computational Semantics (IWCS)*, pages 166–172, Groningen, The Netherlands (online). Association for Computational Linguistics.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Cécile Petitjean and Esther González-Martínez. 2015. Laughing and smiling to manage trouble in french-language classroom interaction. *Classroom Discourse*, 6(2):89–106.

Eugénio Ribeiro, Ricardo Ribeiro, and David Martins de Matos. 2019. Deep Dialog Act Recognition using Multiple Token, Segment, and Context Information Representations. *arXiv:1807.08587 [cs]*.

Tanya Romaniuk. 2009. The 'clinton cackle': Hillary rodham clinton's laughter in news interviews. *Crossroads of Language, Interaction, and Culture*, 7:17–49.

Emanuel A Schegloff. 1982. Discourse as an interactional achievement: Some uses of 'uh huh'and other things that come between sentences. *Analyzing discourse: Text and talk*, 71:93.

Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. 2000. Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. *Computational Linguistics*, 26(3):339–373.

Joseph Tepperman, David Traum, and Shrikanth Narayanan. 2006. " yeah right": Sarcasm recognition for spoken dialogue systems. In *Ninth International Conference on Spoken Language Processing*.

Ye Tian, Chiara Mazzocconi, and Jonathan Ginzburg. 2016. When do we laugh? In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 360–369.

Obed Torres, Justine Cassell, and Scott Prevost. 1997. Modeling gaze behavior as a function of discourse structure. In *First International Workshop on Human-Computer Conversation*. Citeseer.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and Jamie Brew. 2019. HuggingFace's Transformers: State-of-the-art Natural Language Processing. *arXiv:1910.03771 [cs]*.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.

A Supplementary materials

A.1 Collocations of laughs and dialogue acts

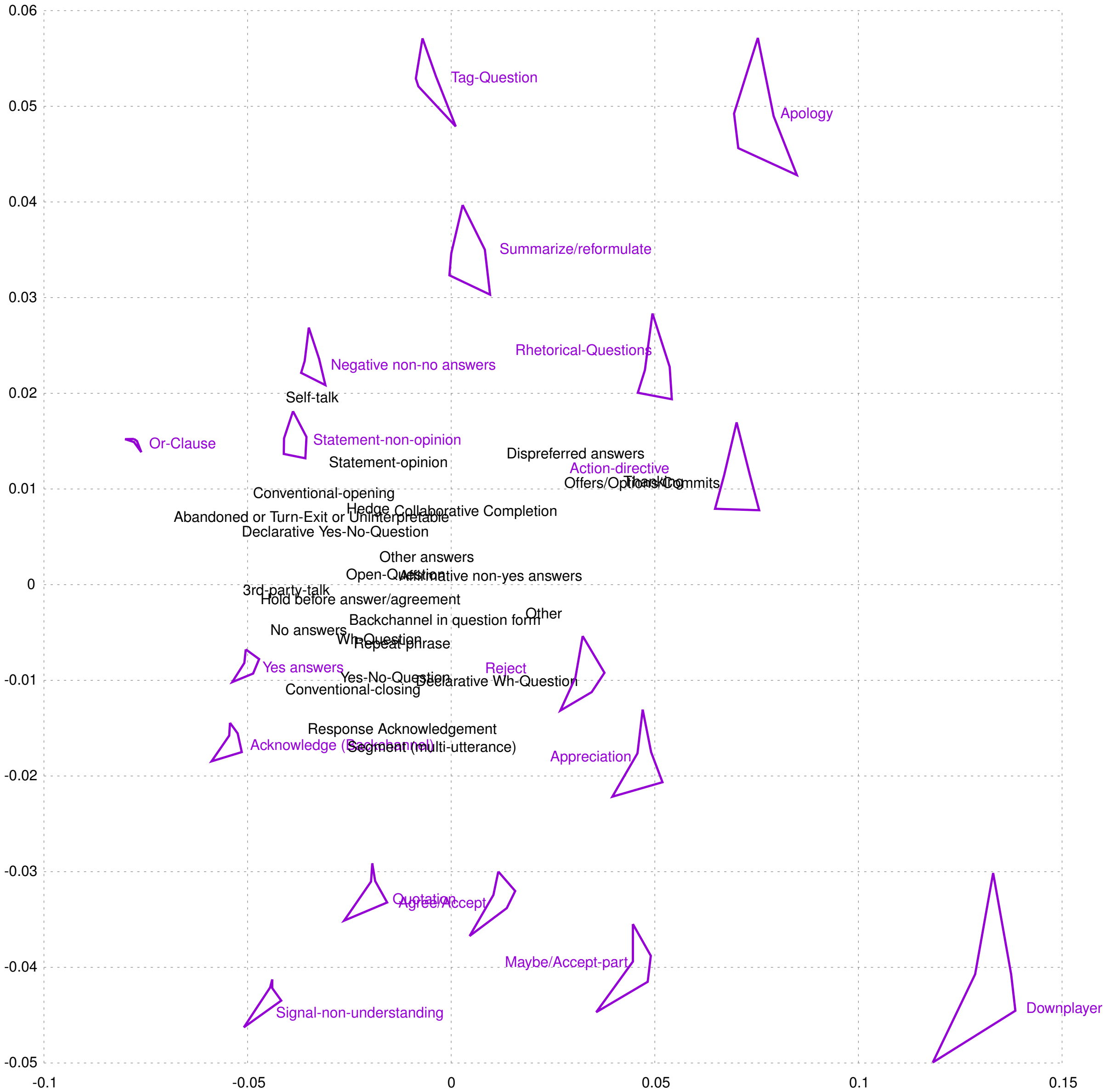
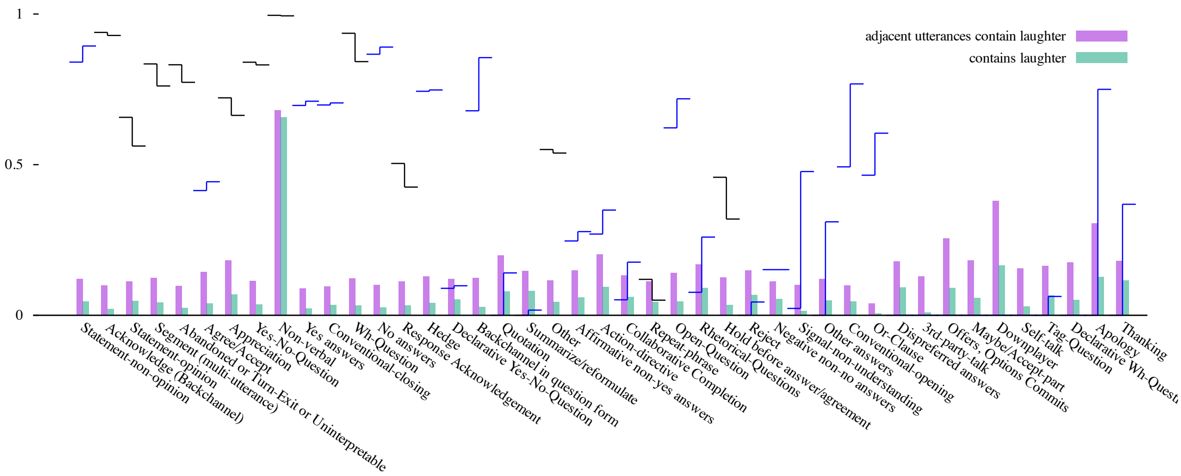


Figure 6: Singular value decomposition of pentagonal representations of dialogue acts. For a selection of dialogue acts (in purple) we depict their pentagon representations.

A.2 Model performance in DAR task



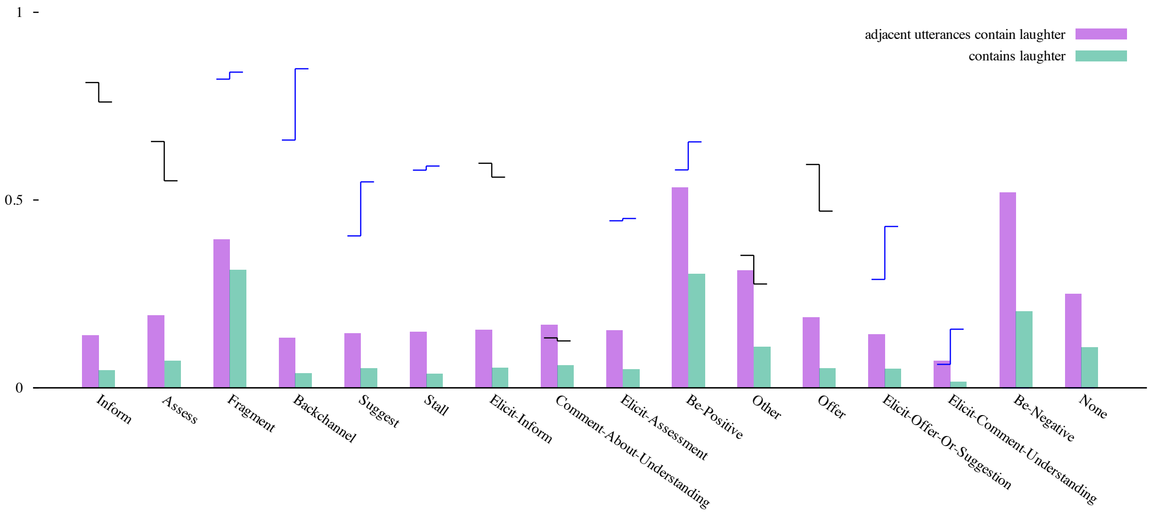


Figure 7: Change in accuracy for each dialogue act (BERT-NL vs BERT-L). Positive changes when adding laughter (BERT-L) are shown in blue. Vertical bars indicate how often dialogue act is associated with laughter. Top chart: SWDA, Bottom chart: AMI-DA.