DISFLUENCY AND LAUGHTER ANNOTATION IN A LIGHT-WEIGHT DIALOGUE MARK-UP PROTOCOL

Julian Hough*, Laura de Ruiter**, Simon Betz** and David Schlangen*

Bielefeld University, *Dialogue Systems Group, **Workgroup Phonetics and Phonology
julian.hough@uni-bielefeld.de

ABSTRACT

Despite a great deal of research effort, disfluency and laughter annotation is still an unsolved problem, both in terms of consensus for a general applicable system, and in terms of annotation agreement metrics. In this paper we present a new annotation scheme within a light-weight mark-up for spontaneous speech. We show, despite the low overhead required for understanding the annotation protocol, it allows for good inter-annotator agreement and can be used to map onto existing disfluency categorization, with no loss of information.

Keywords: Disfluency annotation, laughter, German corpora, inter-annotator agreement, spontaneous speech

1. INTRODUCTION

Annotating spontaneous speech material is always constrained by a trade-off between time and effort on the one hand and coverage on the other. Here we develop a system for annotating spontaneous speech that reduces effort while increasing coverage of disfluency and laughter phenomena.

The low effort required is due to (1) a minimalistic and comprehensive vocabulary to be learned by the annotator and (2) the application directly on the transcription text *at transcription time*. This makes the scheme easy to learn and handle for non-experts who can facilitate the markup of spontaneous speech phenomena on the fly while transcribing. Further analysis and correction is left to interested experts who can focus on their own research questions more directly than in existing annotation systems where more expert work for identification and labeling of these phenomena might be necessary from the outset. Analyses of agreement confirm that non-experts indeed have little trouble in comprehension and application of this system.

Despite the existence of numerous annotation systems, some of which even focus on spontaneous speech phenomena (see §2 for an overview), these systems never cover all the phenomena potentially of interest to disfluency and dialogue researchers. Our system encompasses a light-weight way of covering all disfluency and laughter phenomena and the potential to be mapped onto existing annotation schemes.

2. EXISTING DISFLUENCY AND LAUGHTER ANNOTATION SCHEMES

There is a plethora of existing speech annotation schemes, some of which have been developed especially to capture disfluency phenomena. They vary in terms of practical implementation, the use of category labels and whether or not they mark non-verbal or paralinguistic events.

[15]'s thesis and the ensuing Switchboard disfluency annotation manual [9] are perhaps the most well known for disfluency mark-up. Other more general schemes, such as that used in the German Verbmobil corpus for task oriented dialogue annotation, have various conventions, covering a wide range of spontaneous speech phenomena, including, but not explicitly focusing on, disfluencies [2].

Technically, some schemes annotate disfluencies on a separate tier, for instance [4], [2] or [10]), while others use an in-text annotation on the transcription tier, for instance [9], [12], [13], with more recent schemes often doing this by means of an XML-style annotation [12], [1]. Several schemes require annotators to assign disfluency category labels like "repair", "insertion" or "stutter" from a pre-specified set (vocabulary) as they go through the data ([12], [2], [1], [4]), so training costs can be substantial.

3. A LIGHT-WEIGHT DISFLUENCY AND LAUGHTER ANNOTATION TECHNIQUE

In an effort to both improve the ease of annotation for non disfluency experts, and allow subsumption of existing categorical approaches, here we propose a more light-weight approach than those mentioned above which does not reduce the mark-up information. It combines transcription and annotation in one pass, so that turn segmentation and word transcription can be aided by directly observing their inter-

Table 1: Exemplary DF annotations and their equivalent labels in [1] and [8].

Annotation	[1]	[8]
(Ich + ich) will *(I + I) want*	<repetition> <rm> Ich </rm><rs> ich </rs> </repetition>	Covert repair
(nicht verwinkelt so dass +) und breit *(not contorted so that+) and wide*	<restart><rm> nicht verwinkelt so dass </rm><rs/></restart> und breit	Fresh start
die (<p="Küche">Krü-</p> + Küche) *the (kri- + kitchen)*	die <sot> Krü- </sot> Küche	Phonetic repair

action with disfluencies. Transcription and annotation can be conducted based on perception without lengthy training: transcribers do not have to learn categorical labels for the phenomena because such analyses can be carried out afterwards with the use of automatic search, dependent on the given interest of the researcher.

The advantage of avoiding category labels in the annotation process is that the annotated data is more versatile and more 'theory-neutral', allowing researchers to map it onto different existing classification schemes should they so wish, but not constraining them to do so. Labels like "repetition" or "replacement" (used for instance by [1]) can easily be derived from our bracketed annotation, which is similar to [15]. Other categories like those suggested by [8] can also be derived from our mark-up.

Table 1 provides some examples of how disfluencies annotated according to our scheme can be be mapped onto other classification schemes. While the bracketing of our mid-utterance repair disfluencies is based on [15], with reparandum, interregnum and repair phases available, with a strictly right-branching mark-up of chaining disfluencies such as in (4), we emphasize in our annotation manual the distinction between abandonment of turns (whereby annotators will simply transcribe an utterance final '-') as in (1) and (2) below, and turn-initial fresh starts marked with no repair phase such as in (3) below and in Table 1. Note that while not a focus of this paper, we account for phonetic variants of lexical items with `<v="..">..</v>` tags.

(1) Oder das <v="ist">is'</v> im -
Or that is in the -

(2) Ja eigentlich <v="wäre es"> wär's </v> cool in der Küche <v="einen">'n</v> kleinen Tisch zu haben wo man -
Yes actually it would be cool to have a small table in the kitchen where you -

We include filled pauses, marked simply by a `{F }` bracketing and other fillers simply use `{ }`. We also include laughed speech with simple XML-style tags spanning the affected speech `<laughter>...</laughter>` and a `<laughterOffset/>` tag for the often audible deep inhalation of breath after laughed speech or a laughter bout marked `<laughter/>` (see (3)).

(3) (Und mit einem +) mit vielleicht Sachen die nicht <laughter> auseinander brechen </laughter> <laughterOffset/> -
(And with a +) with perhaps things that don't fall apart -

For partial words, we encourage transcribers to guess the complete form of the word where possible, again using a simple tag `<p="..">..-</p>`, as below:

(4) (<p="Wohnzimmer">Wohn-</p> + . {ja also} (die + (die + das)) {F äh} ... Wohnzimmer)
<p="living room">liv-</p> yes well the the the uh living room -

4. INTER-ANNOTATOR AGREEMENT

Category	Agreement	κ_{free}
reparandum	0.9477	0.8954
repair	0.9677	0.9353
filled pause	0.9968	0.9937
laughed speech	0.9558	0.9117

Table 2: Inter-annotator agreement scores for disfluent word types using κ_{free}

We test our annotation scheme on a corpus of dyadic interactions between German speakers, the Dream Apartment (DAP) corpus [7], which in contrast to existing corpora used to studies disfluencies

Class	SWBD % of words	DAP % of words
Reparandum words	5.16	5.51
Partial words*	0.75	1.10
Filled pauses*	1.12	1.81
Laughed words*	0.45	6.06

Table 3: The frequency of disfluent and laughed words in Switchboard (SWBD) and the Dream Apartment (DAP) corpora. Starred categories indicate a significantly different frequency between the two corpora.

like [6],[14] or [13], is a relatively domain-general corpus of face-to-face interactions. The DAP consists of 9 dialogues of 15 minutes in length. In the task, participant pairs were instructed to discuss their ideal apartment they could jointly design such that they could describe it to an architect. They are given a substantial budget of 500,000 Euros. The familiarity of the subjects varied with 2 of the pairs being strangers and the others varying in familiarity. All participants were students.

To test agreement we use one transcript and 3 annotators: one was the second author, while the other two were non experts. We compared the inter-annotator agreement of words being part of different disfluency and laughed speech elements using the marginal-free multi-rater metric κ_{free} [11]– we use this metric as other multi-rater agreement measures like Fleiss' κ suffer from an assumption annotators know a priori how many cases they should assign to each category, which is not the case here.

The results shown in Table 2 are both interesting and encouraging. Filled pauses and repair phase words have very good agreement, while the lower reparandum word agreement shows a deviation in the way annotators perceive the extent of repairs, and consequently their discourse effects– see [5] for a similar finding. The lower agreement for laughed speech segmentation is not detrimental, as it is still good enough to provide search terms for subsequent stand-off annotation.

5. USE CASE 1: DISFLUENCY AND LAUGHED SPEECH RATES IN PHONE AND FACE-TO-FACE CONVERSATIONS

One of the benefits of our scheme is that it is directly compatible with established schemes, including the Switchboard disfluency annotation mark up [9]. We can therefore directly compare the rates of the disfluency and laughter phenomena in the DAP with Switchboard.

In Switchboard we use the held-out data for disfluency detection (all files named sw4[5-9]* in the Penn TB III release: 52 transcripts, 6.5K utterances, 49K words) marked up according to the scheme in [9]. The DAP is smaller with fewer, but longer, dialogues (9 transcripts, 4.1K utterances, 20k words).

The proportion of reparandum words in each corpus was not significantly different ($\chi^2_{(1)}$=3.568,p=0.06) however the proportion of filled pauses, laughed words and partial words of the total word tokens was significantly lower in the Switchboard corpus.

The most striking difference is in the proportion of laughed words. We hypothesize this may have been due to the difference in topics between the DAP and Switchboard, and also due the familiarity of the participants. Upon inspection there were many opportunities for laughables based primarily on the incongruity of the situation of being students with a vast amounts of spending money.

6. USE CASE 2: DISFLUENCY AND LAUGHTER INTERACTION

A second use case we investigate for our annotation protocol is a qualitative analysis of the interaction of laughed speech, laughter and disfluency in the DAP corpus. In Figure 1 three extracts from the corpus with our mark-up scheme are shown with their English translations. In Example 1 we see some evidence to explain the high frequency of laughed speech described above being due to the topic of conversation. Here the subject of intimacy and privacy with one's partner in the dream apartment is not being fully addressed but is jointly laughed at. A filled pause is employed after the joint laughter.

In Example 2, disfluency and laughed speech interact again on the same topic as in Example 1, with a chaining replacement repair directly following the laughter. In Example 3, a laughable is taken up by both participants in response to a self-answered question by A who mocks her own lack of intelligibility in her explanation to B. Following this the turn is immediately held by A by use of a filled pause.

While these few examples do not provide a thorough analysis of the interactions, we hope they illustrate that Conversation Analysts and dialogue theorists may also benefit from our simple mark-up.

Example 1: Laughter and laughed speech (joint):

A	okay also (wenn wir + wenn wir) <v="eine">'ne</v> Küche haben ein {F ähm } Wohnzimmer zusammen haben und {F äh } ein [Badezimmer] das ist ja so wie in <v="einer">'ner</v> WG [ ] <breathing/> brauchen wir auf jeden Fall <breathing/> jeder so grosse Zimmer dass jeweils unsere <laughter> Partner die dann ja auch noch zu [Besuch kommen] </laughter> <breathing/> {F ähm } auch noch Platz finden
A-en	*okay so (when we + when we) have a kitchen, a {F uh } livingroom and {F uh } a [bathroom] together which is the way it is in a shared flat [ ], we must definitely have big rooms so when our <laughter> partners come [to visit] </laughter> <breathing/>{F um} they have a place to go.*
B	[mhm] / [ja] / [<laughter>sehr richtig</laughter>]
B-en	*[mhm] /[yes] / [<laughter>very true</laughter>]*

Example 2: Chaining substitution repair after laughed speech:

A	dann hat jeder genug Privatsphäre .. mit seinem <laughter> Partner </laughter> / (und die Küche + (und die + {F ähm } (und die + ... und das Wohnzimmer))) ist quasi so ... mittig
A-en	*then everyone has some privacy ... with their <laughter> partner </laughter> / (and the kitchen + (and the + {F um } (and the + ... and the living room))) is kind of ... central*

Example 3: Laughter at embarrassment of disagreement followed by turn hold filled pause:

A	und vom Wohnzimmer kannst du halt in die Küche gehen / <v="verstehst du">verstehste</v> das ? / [okay] nee / [gut] / {F ä:hm }
A-en	*and from the livingroom can you sort of go to the kitchen / do you understand ? / [okay] no / [good] / {F u:m }*
B	[nee] / [<laughter/>] <laughterOffset/>
B-en	*[no] / [<laughter/>] <laughterOffset/>*

Figure 1: Examples of the interaction between disfluency and laughter in the dream apartment corpus

7. CONCLUSION

We present a light-weight and reliable protocol for disfluency and laughter annotation which is currently being used in the DUEL (DisflUencies, Exclamations and Laughter in dialogue) project [3]. It is both fast and easy to use for non-experts, and subsumes existing schemes. Stand-off timing information for detailed investigation into phenomena of interest is derivable from this automatically using the MINT tools [7] software, among others. We have shown two use cases of the scheme, one which allows direct comparability to other schemes, and one which allows fast mark-up of dialogues at transcription time for quantitative and qualitative analysis.

8. REFERENCES

[1] Besser, J., Alexandersson, J. 2007. A comprehensive disfluency model for multi-party interaction. *SigDial 2007* volume 8 182–189.

[2] Burger, S., Weilhammer, K., Schiel, F., Tillmann, H. G. 2000. Verbmobil data collection and annotation. In: *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer 537–549.

[3] Ginzburg, J., Tian, Y., Amsili, P., Beyssade, C., Hemforth, B., Mathieu, Y., Saillard, C., Hough, J., Kousidis, S., Schlangen, D. 2014. The Disfluency, Exclamation and Laughter in Dialogue (DUEL) Project. *Proceedings of the 18th SemDial Workshop (DialWatt)* Herriot Watt University, Edinburgh. 176–178.

[4] Hedeland, H., Schmidt, T. 2012. Technological and methodological challenges in creating, annotating and sharing a learner corpus of spoken german. *Multilingual Corpora and Multilingual Corpus Analysis* 14, 25.

[5] Hough, J., Purver, M. Dec. 2013. Modelling expectation in the self-repair processing of annotat-, um, listeners. *Proceedings of the 17th SemDial Workshop (DialDam)* Amsterdam. 92–101.

[6] Kohler, K. J. 1996. Labelled data bank of spoken standard german: the kiel corpus of read/spontaneous speech. *ICSLP 96* volume 3. IEEE 1938–1941.

[7] Kousidis, S., Pfeiffer, T., Schlangen, D. 2013. Mint. tools: Tools and adaptors supporting acquisition, annotation and analysis of multimodal corpora. *Interspeech 2013*.

[8] Levelt, W. J. 1983. Monitoring and self-repair in speech. *Cognition* 14(1), 41–104.

[9] Meteer, M. W., Taylor, A. A., MacIntyre, R., Iyer, R. 1995. *Dysfluency annotation stylebook for the switchboard corpus*. University of Pennsylvania.

[10] Moniz, H., Batista, F., Mata, A. I., Trancoso, I. 2014. Speaking style effects in the production of disfluencies. *Speech Communication* 65, 20–35.

[11] Randolph, J. J. 2005. Free-marginal multirater kappa (multirater k [free]): An alternative to fleiss' fixed-marginal multirater kappa. *Online Submission*.

[12] Rodríguez, L. J., Torres, I., Varona, A. 2001. Annotation and analysis of disfluencies in a spontaneous speech corpus in spanish. *ISCA 2001*.

[13] Schiel, F., Heinrich, C., Barfüsser, S. 2012. Alcohol language corpus: the first public corpus of alcoholized german speech. *Language resources and evaluation* 46(3), 503–521.

[14] Schmidt, T., Hedeland, H., Lehmberg, T., Wörner, K. 2010. Hamatac–the hamburg maptask corpus.

[15] Shriberg, E. E. 1994. *Preliminaries to a theory of speech disfluencies*. PhD thesis.