The effect of conversation type on entrainment: Evidence from laughter

Bogdan Ludusan and **Petra Wagner**
Phonetics Workgroup, Faculty of Linguistics and Literary Studies & CITEC,
Bielefeld University, Germany
{bogdan.ludusan, petra.wagner}@uni-bielefeld.de

Abstract

Entrainment is a phenomenon that occurs across several modalities and at different linguistic levels in conversation. Previous work has shown that its effects may be modulated by conversation extrinsic factors, such as the relation between the interlocutors or the speakers' traits. The current study investigates the role of conversation type on laughter entrainment. Employing dyadic interaction materials in German, containing two conversation types (free dialogues and task-based interactions), we analyzed three measures of entrainment previously proposed in the literature. The results show that the entrainment effects depend on the type of conversation, with two of the investigated measures being affected by this factor. These findings represent further evidence towards the role of situational aspects as a mediating factor in conversation.

1 Introduction

An aspect frequently observed in conversation is the fact that interlocutors become more similar to each other during their interaction, a phenomenon called, among other terms, entrainment. It has been seen to occur for different linguistic levels (e.g., syntactic Branigan et al., 2000, lexical Brennan and Clark, 1996; Nenkova et al., 2008, acoustic Pardo, 2006; Levitan et al., 2015), but also with respect to non-verbal behaviour (Edlund et al., 2009). Moreover, entrainment effects can be seen both on the form level (adopting the same structures), and on the temporal level, through an increase in temporal co-ordination between interlocutors.

Different points of view on the mechanisms behind entrainment exist, with some viewing it as an automatic process (Pickering and Garrod, 2004), while others arguing that the occurrence of entrainment depends on social factors (Pardo, 2012). This latter viewpoint seems to be supported by studies finding that various conversation aspects (e.g., the role of the interlocutors in the conversation Beňuš et al., 2014; Reichel et al., 2018, their relation Menshikova et al., 2021) or individual factors (e.g., speaker traits Lewandowski and Jilka, 2019, native language Kim et al., 2011) may modulate or interact with entrainment.

Laughter is one of the most often encountered non-verbal vocalisations in spoken interaction (Trouvain and Truong, 2012a), having a wide range of roles in communication, including social (Glenn, 2003) and linguistic (Mazzocconi et al., 2020; Ludusan and Schuppler, 2022). Laughter has been found to be subject to entrainment effects. Interlocutors become more similar in their acoustic realization of laughter, as well as in the timing of their laughter productions (Trouvain and Truong, 2012b; Ludusan and Wagner, 2019). Laughter production may be affected by external factors, such as the gender of the speaker or the familiarity of the interlocutors (Smoski and Bachorowski, 2003). However, no evidence exists towards these factors modulating the amount of entrainment in laughter, with previous works investigating these aspects finding no effect of familiarity on entrainment measures (Trouvain and Truong, 2012b; Ludusan and Wagner, 2022).

We investigate here the effect of one conversation factor, namely the conversation type, on entrainment. We define by conversation type the nature of the interaction, considering it to be either task-based, in which the conversation partners have a specific task to solve during their interaction, or free dialogue, in which interlocutors chat freely about topics of their choice. In particular, we evaluate the role of conversation type (free dialogue vs. two different types of task-based dialogues) on three measures of laughter entrainment.

2 Materials

Materials from two corpora, the GRASS corpus (Schuppler et al., 2014) and the DUEL corpus (Hough et al., 2016) were used for the experiments.

Type	Class	Corpus	Duration [min]	#Dyads	Gender			Age	#Laughter events
					f-f	f-m	m-m		
free	GR	GRASS	769	13	4	4	5	30.5	2272
task	DA	DUEL	103	7	4	2	1	22.7	442
task	FS	DUEL	104	8	2	5	1	23.1	737

Table 1: Information on the data used in this analysis: conversation type (free dialogue or task-based), conversation class (DA/FS/GR), the source corpus (DUEL/GRASS), total duration, number of dyads included, gender composition of the dyads (f-f, f-m, or m-m), average age of the speakers, and number of produced laughter events.

The GRASS corpus (GR) contains both read materials and conversations between two persons. We employed here the latter subset of the corpus, in which the interlocutors (19 dyads), native speakers of Austrian German, were recorded chatting for one hour straight. The interlocutors knew each other beforehand, being either colleagues, friends, family members or couples. They were asked to chat about whichever subject(s) they desired, with some pairs simply continuing the discussion they had before the recording started. This resulted in spontaneous conversations including a wide variety of topics, such as about vacations, local issues, work, family or relationship problems and public figures. The materials were orthographically transcribed and annotated for conversational phenomena, including laughter (both laughs and speech-laughs).

The second corpus, DUEL, contains dyadic interactions between native speakers of three languages: French, German and Mandarin Chinese. Two different scenarios from the German part of the corpus were employed here: Dream Apartment (DA) and Film Script (FS). For the DA scenario, the interlocutors were told they had a large sum of money to design and furnish an apartment they would have to share. In the FS task, they were supposed to come up with the script for a film, based on an embarrassing moment, which could have been inspired from personal experience. The considered materials were recorded by 10 dyads/scenario (which differed between the two scenarios). The dyads were all students, the majority of them being colleagues/friends, but also some pairs consisting of strangers. The corpus was orthographically transcribed and annotated for laughter and other conversational phenomena.

In order to control for the effect the relation between interlocutors might have on entrainment, we did not consider in our analysis the recordings from the GRASS corpus that involved family members or couples (6 dyads). Similarly, we excluded those between strangers from the DUEL corpus (4 dyads). In this way, the dyads from both corpora were either colleagues or friends. Detailed information on the datasets considered in the analyses and their characteristics can be found in Table 1.

3 Methods

We investigated three measures previously employed in the study of laughter entrainment, all of which were computed at the dyad level. They included both temporal-related entrainment measures such as the amount of overlapping laughter produced by the interlocutors and the synchrony of the produced laughter, and form-related ones, namely the difference in maximum intensity between non-consecutive and consecutive laughter produced by the speakers in the dyad. We examined whether the results of these measures varied with the conversation type (free vs. task-based dialogue), while also considering a second analysis level, the conversation class (examining here three classes: GR, DA, FS).

The first measure, the amount of overlapping laughter, was inspired by the temporal alignment proposed by Trouvain and Truong (2012b) as a measure of laughter entrainment. A higher amount of overlapping laughter implies a higher level of entrainment. The measure was determined by counting all events in which the two interlocutors were laughing at the same time (we took into account any amount of overlap), as well as the total number of laughter events produced during the interaction. We then applied logistic regression models to test the differences between the various conditions (conversation type/class), by considering the odds of overlapping laughter, represented by the pair (overlapping laughter counts, total laughter counts - overlapping laughter counts) as dependent variable of the model and the condition as predictor.

For the synchrony measure, we applied the process described in Ludusan and Wagner (2019).

However, since we had recordings of different lengths within and across datasets, we did not split the recordings into a fixed number of bins. Instead, we used bins of equal duration – 90 seconds (15 minutes / 10 bins, as in Ludusan and Wagner 2019). We then counted the number of laughter events produced by each speaker in each bin and computed the synchrony, defined as the Spearman ρ correlation coefficient between the vectors composed of the binned laughter counts of the interlocutors in a conversation. Positive values of this measure represent entrainment. These first two measures were computed on the data from all 28 dyads included in the study.

The form-related measure characterizes the similarity of consecutive laughter pairs produced by the interlocutors in terms of maximum speech signal intensity (Ludusan and Wagner, 2022). The intensity was computed by means of the Praat software (Boersma and Weenink, 2020), employing a minimum pitch of 75 Hz and subtracting the microphone DC offset. The maximum value over each laughter event was then considered for this entrainment measure. Consecutive laughter pairs are composed of the laughter event of a speaker either overlapping with or followed within one second, by a laughter produced by their interlocutor (similar to the definition of antiphonal laughter in Smoski and Bachorowski 2003). We then compared the difference in intensity between the laughter events of a consecutive pair ($intD_C$) with the same measure computed between the events of non-consecutive laughter pairs ($intD_N$). Non-consecutive pairs were composed of a laughter event from a consecutive laughter pair, and a randomly sampled laughter produced by the interlocutor, except for the one in the same consecutive pair (see Ludusan and Wagner 2022 for more details). The measure was then defined as: $intD_N$ - $intD_C$, with positive values denoting entrainment. This measure was analyzed for 27 dyads, those which produced at least 5 consecutive laughter pairs (one all-male dyad from the GRASS subset was removed).

In addition to comparing these three measures across conversation types, we also determined whether the obtained values represent entrainment or not. For the intensity-based measure, a positive value significantly different from 0 denotes entrainment, and the opposite effect for negative values. For the overlapping laughter and the synchrony measures, we determined whether the dyads achieved entrainment, by comparing their value with those obtained for all pseudo-dyads, similarly to previous work on entrainment (e.g, Ramseyer and Tschacher, 2010). For each dyad in the investigated subset, we created pseudo-dyads, by putting together the speech of each speaker within the dyad with all other speakers in that subset, but the one from the same dyad. For each created pseudo-dyad, the two entrainment measures were computed and the average value across all pseudo-dyads was compared to the entrainment measure of the actual dyad. If the latter was significantly higher than the former, it represented entrainment, while a significantly lower value meant disentrainment.

Finally, there are characteristics which we could not control for in the analyzed data and which may influence laughter production and possibly, indirectly, its entrainment. Therefore, we examined any effect that dyad gender composition (two classifications: f-f/f-m/m-m or same/mixed-gender) or age (two measures: absolute age difference or average age of the dyad) may have on the entrainment measure.

For all analyses except for the ones pertaining to the overlapping laughter measure (which employed logistic regression), linear regression models were fitted with the respective measure values as dependent variable and the various factors investigated as predictors. In case the residuals of the fitted models were found to be not normally distributed (by means of a Shapiro-Wilk test), we applied a corresponding non-parametric method: either a Wilcoxon rank-sum test (for two groups), or a Kruskal-Wallis test (for three groups). To determine whether the studied measures show entrainment on each subset we compared them (either with the 0 level or with the value obtained for the pseudo-dyads) by means of t-tests or Wilcoxon tests (if the samples were not normally distributed). All statistical analyses were run using the appropriate functions of the R software (R Core Team, 2020).

4 Results

The values of the three investigated measures across the considered conversation types and classes are illustrated in Figure 1 and Figure 2, respectively.

In terms of percentage of overlapping laughter between the interlocutors, both conversation types showed entrainment (Figure 1, left panel),

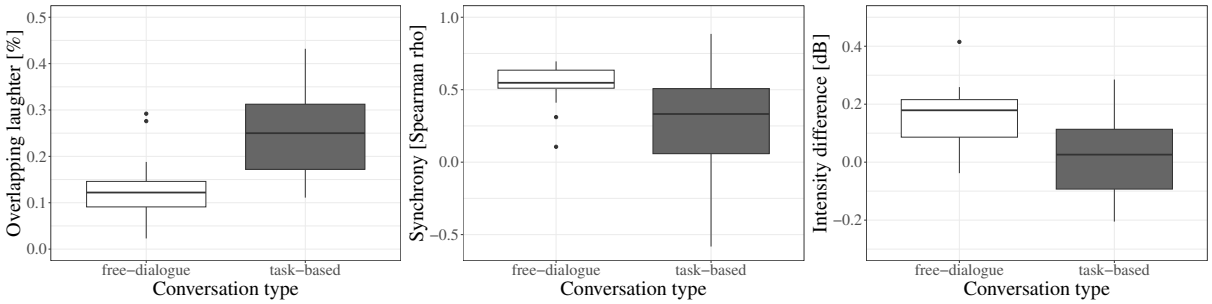

Figure 1: The results of the investigated entrainment measures, with respect to the considered conversation types: overlapping laughter (left panel), synchrony (middle panel) and form-related measure (right panel). The horizontal line represents the median value, the hinges of the boxes the first and third quartiles, and the whiskers going up to 1.5*IQR (inter-quartile range) from the hinges.

as revealed by Wilcoxon tests ($p = 2.4e^{-4}$ for free dialogues and $p = 6.1e^{-5}$ for task-based dialogues). We then investigated the effect of conversation type on entrainment, by using it as predictor in a logistic regression model (AIC = 252.3). The difference between the two types was found to be significant ($\beta = 0.740, z = 8.26, p < 2e^{-16}$). When looking at conversation classes (Figure 2, left panel), entrainment was observed for GR and for both classes included in the task-based data: DA ($t = 8.26, p = 1.7e^{-4}$) and FS ($t = 5.70, p = 7.3e^{-4}$). The ANOVA analysis of the logistic model fitted with the overlapping laughter odds as dependent variable and the class as independent variable (Akaike Information Criterion, AIC = 254.2), revealed a significant effect of class ($\chi^2 = 67.3, p = 2.4e^{-15}$). Moreover, the model showed that the differences between GR and each of the other two classes were significant: DA ($\beta = 0.752, z = 6.07, p = 1.3e^{-9}$) and FS ($\beta = 0.733, z = 7.08, p = 1.5e^{-12}$). No significant difference was found between the DA and FS. Lastly, we verified, by means of logistic regression, whether the age (mean or difference) of the conversation partners or the dyad composition (exact composition or same/mixed) may play a role in the production of overlapping laughter. All but the age difference showed a significant effect, although the fit of these models was worse than that of the models employing the conversation class or type as predictor (the best of these four models had an AIC of 296.5 – lower AIC represents a better model).

For the synchrony measure, we observed entrainment for both free and task-based dialogues (Figure 1, middle panel): $t = 9.32, p = 7.6e^{-7}$ and $t = 3.28, p = 0.005$, respectively. The difference between conversation types was not significant, as given by a Wilcoxon rank sum test ($p = 0.339$). At the level of conversation classes (Figure 2, middle panel), entrainment effects were observed only for FS ($t = 2.99, p = 0.020$), in addition to GR. A Kruskal-Wallis test showed no significant overall difference between conversation classes ($\chi^2 = 2.33, p = 0.312$), but pairwise differences were found between GR and DA, using a Wilcoxon test ($p = 0.024$). Additional Kruskal-Wallis tests revealed no significant effects of age or dyad gender composition.

The last measure, defined as the difference in maximum intensity between non-consecutive and consecutive laughter pairs (Figure 1, right panel), was found to entrain for free dialogues ($t = 4.92, p = 4.6e^{-4}$), but not for the task-based ones ($t = 0.44, p = 0.67$). A significant difference was observed between conversation types, as given by the ANOVA of the fitted linear model ($F = 7.96, p = 0.009$). A similar linear regression model, using the intensity difference as dependent variable and the conversation class as predictor was then fitted (Figure 2, right panel). The ANOVA analysis of this model revealed a significant overall effect of class ($F = 5.50, p = 0.011$), with the difference between GR and DA reaching significance ($\beta = -0.207, z = -3.30, p = 0.003$). None of the subsequent linear models, fitted with the gender make-up of the dyad and the age measures as predictors, showed a significant effect of these factors.

5 Discussion and conclusions

Our findings paint a complex relationship between the investigated entrainment measures and the different conversation types/classes considered here. We found entrainment across the various dialogues types/classes, and differences between types and some classes (overlapping laughter), entrainment

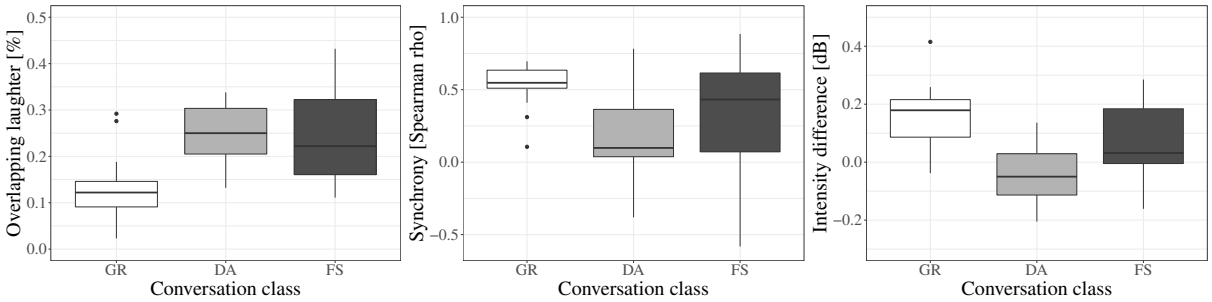

Figure 2: The results of the investigated entrainment measures, with respect to the considered conversation classes: overlapping laughter (left panel), synchrony (middle panel) and form-related measure (right panel). The horizontal line represents the median value, the hinges of the boxes the first and third quartiles, and the whiskers going up to 1.5*IQR (inter-quartile range) from the hinges.

across types, but not for all classes, and some differences between classes (synchrony), entrainment for one type only and differences between some classes (intensity measure). An effect of conversation type/class was observed when controlling for the relation between interlocutors, while other dimensions of variability between the different subsets used (age of interlocutors, gender composition of the dyad) had either no significant effect, or explained the differences in entrainment worse than the conversation type/class.

Another factor of variability may be the fact that the interlocutors in the analyzed corpora spoke different varieties of German and came from slightly different cultures. Yet, evidence from studies that examined laughter entrainment measures cross-linguistically/culturally (Ludusan and Wagner, 2019, 2022), showed no language/culture differences for more distant language pairs (German-Chinese and French-Chinese) than the ones here. One could assume, instead, that the observed differences stem from the fact that task-based interactions require a higher cognitive load, and previous studies have shown that a higher cognitive load may impede entrainment (Abel and Babel, 2017). However, our results did not show an inverse relation between the level of entrainment and the difficulty of the task. Some of the values of the studied measures revealed either the opposite tendency or similar trends between task-based and free dialogue interactions. These findings indicate that what is being captured by our conversation type factor differs from cognitive load.

The results obtained for the overlapping laughter measure, with the free dialogue/GR values being significantly lower than for the other cases, may seem surprising, especially considering that synchrony, another measure of temporal alignment, suggests rather the opposite. It might be that the overlap measure employed here is too strict. Since mirthful laughter, which is predominant in the FS data and partly in the DA recordings, is generally longer than social laughter, it is more likely that, when the conversation partner joins in laughing in response to a mirthful laughter, their laughter will overlap that of their interlocutor. A more appropriate measure could be one that takes into account also the interval immediately following the produced laughter, such as the antiphonal laughter definition of Smoski and Bachorowski (2003).

To conclude, our findings represent further evidence for entrainment not being a fully automatic process (Pardo, 2012), but that different factors (here, the conversation type) may influence it and should be taken into account when investigating this phenomenon. As future work, on the one hand, we would like to tease apart the effect of conversation type on entrainment from that potentially brought by laughter type, since the employed dialogues contain different types of laughter. One the other hand, our results raise further questions about the potential effect of conversation type on the entrainment of other levels. Thus, extending this investigation to conversation elements/linguistic levels previously shown to be subject to entrainment is highly desirable. This will shed further light on the role of entrainment in human communication and will also allow more realistic implementations of this phenomenon in spoken dialogue systems (e.g., Stoyanchev and Stent, 2009; Duplessis et al., 2017).

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) project number 461442180.

References

Jennifer Abel and Molly Babel. 2017. Cognitive load reduces perceived linguistic convergence between dyads. *Language and Speech*, 60(3):479–502.

Štefan Beňuš, Agustín Gravano, Rivka Levitan, Sarah Ita Levitan, Laura Willson, and Julia Hirschberg. 2014. Entrainment, dominance and alliance in supreme court hearings. *Knowledge-Based Systems*, 71:3–14.

Paul Boersma and David Weenink. 2020. Praat: doing phonetics by computer [Computer program]. Version 6.1.35, retrieved 1 December 2020 from http://www.praat.org/.

Holly P Branigan, Martin J Pickering, and Alexandra A Cleland. 2000. Syntactic co-ordination in dialogue. *Cognition*, 75(2):B13–B25.

Susan E Brennan and Herbert H Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6):1482–1493.

Guillaume Dubuisson Duplessis, Chloé Clavel, and Frédéric Landragin. 2017. Automatic measures to characterise verbal alignment in human-agent interaction. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pages 71–81.

Jens Edlund, Julia Hirschberg, and Mattias Heldner. 2009. Pause and gap length in face-to-face interaction. In *Proceedings of INTERSPEECH*, pages 2779–2782.

Phillip Glenn. 2003. Towards a social interactional approach to laughter. In *Laughter in Interaction*, pages 7–34. Cambridge University Press.

Julian Hough, Ye Tian, Laura de Ruiter, Simon Betz, Spyros Kousidis, David Schlangen, and Jonathan Ginzburg. 2016. DUEL: A multi-lingual multimodal dialogue corpus for disfluency, exclamations and laughter. In *Proceedings of LREC*, pages 1784–1788.

Midam Kim, William S. Horton, and Ann R. Bradlow. 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, 2(1):125–156.

Rivka Levitan, Štefan Beňuš, Agustín Gravano, and Julia Hirschberg. 2015. Acoustic-prosodic entrainment in slovak, spanish, english and chinese: A cross-linguistic comparison. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 325–334.

Natalie Lewandowski and Matthias Jilka. 2019. Phonetic convergence, language talent, personality and attention. *Frontiers in Communication*, 4:18.

Bogdan Ludusan and Barbara Schuppler. 2022. To laugh or not to laugh? the use of laughter to mark discourse structure. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 76–82.

Bogdan Ludusan and Petra Wagner. 2019. Laughter dynamics in dyadic conversations. In *Proceedings of INTERSPEECH*, pages 524–528.

Bogdan Ludusan and Petra Wagner. 2022. Laughter entrainment in dyadic interactions: Temporal distribution and form. *Speech Communication*, 136:42–52.

Chiara Mazzocconi, Ye Tian, and Jonathan Ginzburg. 2020. What's your laughter doing there? A taxonomy of the pragmatic functions of laughter. *IEEE Transactions on Affective Computing*, 13(3):1302–1321.

Alla Menshikova, Daniil Kocharov, and Tatiana Kachkovskaia. 2021. Lexical Entrainment and Intra-Speaker Variability in Cooperative Dialogues. In *Proceedings of INTERSPEECH*, pages 1957–1961.

Ani Nenkova, Agustín Gravano, and Julia Hirschberg. 2008. High frequency word entrainment in spoken dialogue. In *Proceedings of ACL-08: HLT, Short Papers*, pages 169–172.

Jennifer Pardo. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, 119(4):2382–2393.

Jennifer Pardo. 2012. Reflections on phonetic convergence: Speech perception does not mirror speech production. *Language and Linguistics Compass*, 6(12):753–767.

Martin J Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2):169–190.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Fabian Ramseyer and Wolfgang Tschacher. 2010. Nonverbal synchrony or random coincidence? How to tell the difference. In Anna Esposito, Nick Campbell, Carl Vogel, Amir Hussain, and Anton Nijholt, editors, *Development of Multimodal Interfaces: Active Listening and Synchrony. Lecture Notes in Computer Science, vol 5967*, pages 182–196. Springer.

Uwe Reichel, Stefan Beňuš, and Katalin Mády. 2018. Entrainment profiles: Comparison by gender, role, and feature set. *Speech Communication*, 100:46–57.

Barbara Schuppler, Martin Hagmüller, Juan Andres Morales-Cordovilla, and Hannes Pessentheiner. 2014. GRASS: the Graz corpus of Read And Spontaneous Speech. In *Proceedings of LREC*, pages 1465–1470.

Moria Smoski and Jo-Anne Bachorowski. 2003. Antiphonal laughter between friends and strangers. *Cognition and Emotion*, 17(2):327–340.

Svetlana Stoyanchev and Amanda Stent. 2009. Lexical and syntactic adaptation and their impact in deployed spoken dialog systems. In *Proceedings of Human Language Technologies: The 2009 Annual*

Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers, pages 189–192.

Jürgen Trouvain and Khiet Truong. 2012a. Comparing non-verbal vocalisations in conversational speech corpora. In *Proceedings of the LREC Workshop on Corpora for Research on Emotion Sentiment and Social Signals*, pages 36–39.

Jürgen Trouvain and Khiet P Truong. 2012b. Convergence of laughter in conversational speech: effects of quantity, temporal alignment and imitation. In *Proceedings of the International Symposium on Imitation and Convergence in Speech*, pages 37–38.