THE EFFECT OF EXPLICIT TRAINING ON THE PROSODIC PRODUCTION OF L2 SARCASM BY DUTCH LEARNERS OF ENGLISH

Laura Smorenburg[1], Joe Rodd[1] & Aoju Chen[1, 2]

[1]Utrecht University, [2]Max Planck Institute for Psycholinguistics
B.J.L.Smorenburg@uu.nl; J.J.E.Rodd@uu.nl; Aoju.Chen@uu.nl

ABSTRACT

Previous research [9] suggests that Dutch learners of (British) English are not able to express sarcasm prosodically in their L2. The present study investigates whether explicit training on the prosodic markers of sarcasm in English can improve learners' realisation of sarcasm. Sarcastic speech was elicited in short simulated telephone conversations between Dutch advanced learners of English and a native British English-speaking 'friend' in two sessions, fourteen days apart. Between the two sessions, participants were trained by means of (1) a presentation, (2) directed independent practice, and (3) evaluation of participants' production and individual feedback in small groups. L1 British English-speaking raters subsequently evaluated the degree of sarcastic sounding in the participants' responses on a five-point scale. It was found that significantly higher sarcasm ratings were given to L2 learners' production obtained after the training than that obtained before the training; explicit training on prosody has a positive effect on learners' production of sarcasm.

Keywords: Prosody, sarcasm, L2, training, rating

1. INTRODUCTION

Sarcasm, a commonly used rhetorical device, refers to the phenomenon whereby a speaker uses words to express the opposite of the literal meaning of those words. Given identical wording, this difference between sincerity and sarcasm must be signalled in other ways: context and prosody. Adults can detect sarcasm using prosody alone when context is not available [6, 8], and detect sarcasm more effectively when both context and prosody are available compared to context alone [6].

Annoli et al. [2] propose two categories of sarcasm: firstly 'scornful sarcastic irony', and secondly 'bantering sarcastic irony'. 'Scornful sarcastic irony' is believed to be associated with a low mean pitch and little pitch variability [7], but a bantering variety of 'sarcastic irony' with high pitch, high pitch variability, increased loudness and reduced speech rate [4, 10]. In this study we consider the bantering type of sarcasm.

Previous research on prosodic expression of sarcasm in L2 suggests that both intermediate and advanced Dutch learners of (British) English are not able to use prosody to express sarcasm satisfactorily in their L2 [9]. Specifically, the L2 speakers' prosodic expression of sarcasm was evaluated more negatively by L1 British English-speaking raters than L1 Dutch speaking raters [9]. This finding suggests that prosodic encoding of sarcasm may not be acquired easily without formal instructions, and that it may differ between languages, in line with [8], who found that listeners were successful at identifying the intended manner of speaking (e.g. sarcastic, sincere) in their native language, but not in foreign languages.

A question that arises in the context of L2 teaching is whether explicit training in the English prosodic markers of sarcasm can positively influence learners' accuracy in producing it. This question of explicit prosodic instruction has not been extensively addressed. [12] examined the effectiveness of prosodic versus segmental training on L1 Italian learners of German and found prosodic training more effective at improving pronunciation, rated for its likeness to native production. [13] found that a short training intervention improved the perception of English sarcasm by L1 Chinese learners of English. In light of these limited findings we hypothesise that explicit training will improve prosodic encoding of sarcasm in L2 (Hypothesis 1).

Previous work [9] also found an effect for sentence type in the ratings that sarcastic production received. They predicted that sentences with non-prosodic markers of sarcasm (tag-questions and sentences containing particles such as *gosh*) would receive higher ratings because the presence of the non-prosodic marking may compensate for limited use of prosody. This was only partially borne out: whilst sentences with particles received high ratings, tag questions received low ratings. This result remains difficult to explain. We therefore seek to further test for differences across sentence types, and maintain the hypothesis that structures containing

non-prosodic markers of sarcasm will receive higher ratings than those that do not (Hypothesis 2).

2. METHOD

Two production experiments were conducted to elicit sarcastic production from Dutch learners of English. The production experiments were conducted individually in week one (pre-test) and week three (post-test). In week two, the participants were trained to produce sarcasm prosodically in English. Subsequently, L1 English-speakers rated the productions collected in the pre- and post-tests on a five-point scale of "not sarcastic" to "very sarcastic".

2.1 The pre- and post-tests

2.1.1 Participants

Twelve L1 Dutch-speakers (9 females, 3 males, mean age = 21.3, SD = 1.7) participated in the pre- and post-tests and in the training. All were second- or final-year undergraduate students of *English Language and Culture* at Utrecht University. Eleven of the participants were specialised in British English; one studied American English. All participants were estimated to have an advanced level (C1/C2 on the Common European Framework of References for Languages [4]) of English proficiency. They were paid for their participation.

2.1.2 Task

In both the pre- and the post-tests, the participants' task was to respond to the remarks of a native British English-speaking 'friend' about fictional people and situations in a sarcastic manner in simulated telephone conversations, following [9].

2.1.3 Test items

The pre- and post-tests had different lists of remark-response-background sets. The response sentences represented three different sentence types: what-exclamations (*What a terrible shame!*), simple declarative sentences (*That's very scary.*) and sentences with negative question tags (*They were gracious guests, weren't they?*). Responses within each sentence type were controlled to ensure maximum comparability in length and syntactic complexity. Three L1 British English-speakers, who had no connection to the investigation, were consulted to ensure equal acceptability of the response sentences as sarcastic or sincere remarks on lexical and syntactic levels. The remark sentences were constructed carefully such that they created a convincing situation for the participants to respond in a sarcastic manner. The remarks were recorded by the second author, an L1 British English-speaker. Additional background information was provided textually in some cases to set the common ground between the participants and the English 'friend', which contributed to the naturalness of the telephone conversations.

Eight sets were included for each type of responses in each test. The test items were partially adapted from previous studies [1, 9, 6, 7, 11].

2.1.4 Procedure

The participants were tested individually, in a sound-attenuated booth at the labs of the Utrecht Institute of Linguistics, Utrecht University. First, each participant was instructed (with standardised written instructions in English) to make his responses *sound* sarcastic and was alerted to the fact that the sentences were lexically ambiguous, so they would have to rely on prosody. Second, the participant was asked to summarise the task to the experimenter (first author), to ensure that the instructions had been understood correctly. Third, the participant did six practice trials in the presence of the experimenter. In each trial, the background information (if present), the transcription of the remark, and the response that the participant was to make were displayed on a computer screen using Microsoft PowerPoint. The pre-recorded remark played automatically as the participant advanced through the trials. The participant could choose to hear the remark again by clicking a button with the mouse or to move on to the next trial. Two semi-randomised stimulus orders were created, in which no trials of the same type occurred twice in a row. The participants were randomly assigned to these orders. The participant's responses together with the remarks were recorded using a ZOOM 1 digital recorder.

The participants' responses were subsequently extracted using Praat [5]. In the case of more than one attempt at producing the same response, the final attempt was chosen.

2.2 The training

Training was given to all the participants in three phases on the same day. The materials used in the training sessions were a small selection of items used in the pre-test, recorded by two native speakers of British English unconnected to the investigation. Firstly, the first author gave a presentation to the participants and their classmates (who did not participate in this study) on how sarcasm is

expressed prosodically in British English. After explaining key prosodic concepts, the students were told that sarcasm in English is marked by slower, louder speech, with a wider pitch range. They were told to place strongest emphasis on the most significant content word according to the context (usually the subject or adjective/adverb), with a delayed peak. Example recordings and pitch contours from the native speakers were presented and discussed throughout the presentation. The participants were then given the opportunity to practice individually for approximately 45 minutes. They did this by listening to the production of the native speakers using Praat. They were given a booklet containing transcriptions and pitch contours for these recordings. They then tried producing these themselves, using Praat to record their own production and plot the pitch contours. They were asked to repeat the sentences until their pitch contours were similar to the sample sentences in terms of overall shape and pitch range.

Finally, the participants were given feedback in groups of two or three by the authors. One sentence per sentence type was randomly selected so that each participant received feedback on all sentence types. Their production was recorded, played back, and visualised using Praat so that detailed suggestions could be made on how to improve their prosodic production.

2.3 The rating experiment

2.3.1 Participants

Twelve L1 British English speakers (3 males, 9 females, mean age = 20.0, SD = 2.2) were recruited as raters at the University of Leeds. None of the raters had any experience of Dutch, some had very limited experience of other non-English Germanic languages. All were brought up and educated in the UK as monolingual English speakers, and were considered by the authors to speak fairly standard British English. They were paid for their participation.

2.3.2 Task

The raters were asked to rate how sarcastic the responses produced by the Dutch learners of English in the pre- and post-tests sounded on a five-point scale, from "not sarcastic" to "very sarcastic". They heard only the responses, without the accompanying remarks and background information. There were 576 sentences in total to be evaluated. Six raters (group A) rated the pre-test productions of half of the learners, and the post-test productions of the other half. The other six raters (group B) rated the remaining items. This means that each item was rated by six raters.

2.3.4 Procedure

The rating task was set up in ZEP, an experimental control software package [14]. The raters were given written instructions, informing them that they were going to listen to simulated telephone conversation between several non-native speakers of English and an English friend. They were told that non-native speakers of English were trying to sound sarcastic in all their responses.

The raters entered their responses by clicking radio-buttons on a computer screen. They had the option to listen to the recording up to two additional times by clicking a "listen" button. When they were happy with their response, they clicked "next".

In addition to the main rating scale ("not sarcastic" to "very sarcastic"), the raters were asked to evaluate their own certainty about the score that they had awarded. We intend to make use of this additional data in future work.

3. STATISTICAL ANALYSIS AND RESULTS

Statistical analysis was conducted in R. The distribution of the average scores was revealed to be significantly non-normal by a Shapiro-Wilk test ($W = 0.98$, $p = 6.94e\text{-}08$), with significant skew and kurtosis. The assumptions of conventional statistical analysis methods were therefore not met, and robust alternatives were used [15, 16].

3.1 Inter-rater agreement

Since the twelve raters did not all evaluate the same material, the inter-rater agreement was calculated separately for the two groups. Rater group A (mean rating = 2.90, SD = 0.61) showed a robust Cronbach's alpha [16] value of .74, (with 7.99% downweighting). Rater group B (mean rating = 2.87, SD = 0.61) had a robust Cronbach's alpha value of .69 (with 3.14% downweighting). Both of these alpha coefficients are acceptable.

3.2 Mixed-design repeated measures ANOVA

A repeated measures analysis of variance (ANOVA) on trimmed-means [15] was conducted with one dependent variable: the sarcastic-sounding scores and two within subject independent variables: TEST-SESSION and SENTENCE TYPE.

3.3.1 Effect of training

The main effect of TEST-SESSION was highly significant (Q = 27.23, p < .001). The training improved the learners' production of sarcasm significantly, from a mean pre-test rating of 2.69 (SD 0.44) to a mean post-test rating of 3.08 (SD 0.54), as shown in Figure 1, where dots represent outliers beyond 1.5 * IQR.

3.3.2 Effect of Sentence type

The main effect of SENTENCE TYPE was highly significant (Q = 14.18, p < .001). That is, the learners scored significantly higher for some sentence types than others. These results are depicted in Figure 1.

3.3.3 Interaction of TEST-SESSION × SENTENCE TYPE

The interaction of TEST-SESSION × SENTENCE TYPE was significant (Q = 3.77, p =.023), demonstrating that the relative ratings received by the sentence types differed between TEST SESSIONS: the participants showed more improvement for declaratives than for what-exclamations and tag questions. Post-hoc pairwise comparisons using Wilcox's dependent variation of Yuen's trimmed-means comparison method [15] revealed that these effects were significant for all pairs (between test sessions within each sentence type).

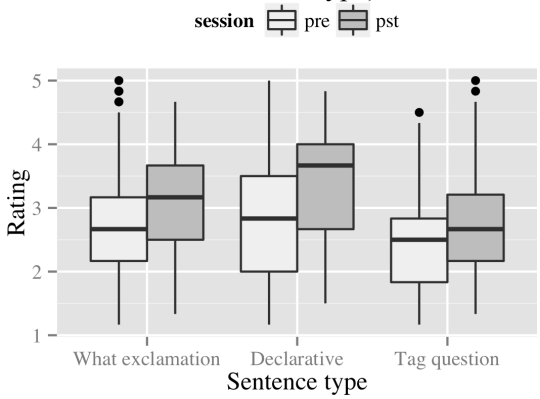

Figure 1: Ratings by sentence type and test session

4. DISCUSSION AND CONCLUSIONS

The results show clearly that explicit training in the prosodic features of utterances in a second language is effective, extending previous findings of a benefit of prosodic training on general native-likeness of pronunciation [12]. Specifically, we demonstrate that focused training is also effective in improving prosodic production on a functional level, the production of convincing sarcasm.

Our results have a useful implication for language teaching; and training of the kind we provided could be further developed into a computerised program, allowing participants to receive identical training and follow the training at their own pace.

5. ACKNOWLEDGEMENTS

This study is supported by an Aspasia grant (015-007-013) to the third author from the NWO (Netherlands Organisation for Scientific Research). We thank Diantha de Jong for her earlier work that served as the foundation for the methodology applied here.

We are indebted to the colleagues at the Department of Linguistics and Phonetics, School of Languages, Cultures and Societies, University of Leeds for hosting the rating experiments. We thank Saskia Verstegen for help with the annotation of the production data, and we thank the participants in both Utrecht and in Leeds for their valuable help.

6. REFERENCES

[1] Ackerman, B. P. 1983. Form and Function in Children's Understanding of Ironic Utterances. *Journal of Experimental Child Psychology* 35, 487-508.

[2] Anolli, L., Ciceri, R., Infantino, M. G. 2002. From "blame by praise" to "praise by blame": Analysis of vocal patterns in ironic communication, *International Journal of Psychology* 37(5), 266-276.

[3] Attardo, S., Eisterhold, J., Hay, J., Poggy. I. 2003. Multimodal markers of irony and sarcasm. *Humor: International Journal of Humor Research* 16 (2), 243–260.

[4] Beeker, A., Canton, J., Fasoglio, D., Trimbos, B. 2010, Eindtermen hao..wo. Europees Referentiekader Talen. Nationaal Expertisecentrum leerplanontwikkeling (SLO). Web. http://www.erk.nl/docent/streefniveaus/havo/

[5] Boersma, P., Weenink, D. 2014. Praat: Doing phonetics by computer. Computer programme version 5.3.62. http://www.praat.org/.

[6] Capelli, C. A., Nakagawa, N., Madden, C. M. 1990. How children understand sarcasm: The role of context and intonation. *Child Development* 61(6), 1824-1841.

[7] Cheang, H. S., Pell, M. D. 2008. The sound of sarcasm. *Speech Communication* 50(5), 366-381.

[8] Cheang, H. S., Pell, M. D. 2011. Recognizing sarcasm without language: A cross-linguistic study of English and Cantonese. *Pragmatics & Cognition* 19(2), 203-223.

[9] Chen, A., De Jong, D. (to appear). Prosodic expression of sarcasm in L2 English. *Spoken discourse and prosody in L2* (ed) Marina Chini. Series Materiali Linguistici. Milan: Franco Angeli.[10] Haiman, J. 1998. *Talk is Cheap: Sarcasm, Alienation, and the Evolution of Language* Oxford: Oxford University Press.

[11] Kreuz, R. J., Glucksberg, S. 1989. How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of Experimental Psychology: General* 118(4), 374-386.

[12] Missaglia, F. 2007. Prosodic training for adult Italian learners of German: The Contrastive Method. In: Trouvain, J. & Gut, U. (eds.), *Non-Native Prosody: Phonetic Description and Teaching Practice*. Berlin: Mouton de Gruyter, 237-252.

[13] Pincus, N. 2014. Teaching intonation with explicit instruction and audio-visual techniques. *Proc. 7th INTED.* Valencia, 3428-3435.

[14] Veenker, T. J. G. 2014. The Zep Experiment Control Application. Computer programme version 1.8. Beexy Behavioral Experiment Software. http://www.beexy.org/zep/.

[15] Wilcox, R. 2012. *Introduction to robust estimation and hypothesis testing*. London: Academic Press.

[16] Zhang, Z. & Yuan, K.-H. 2013. Robust Cronbach's alpha. R package version 0.2.6. http://www.inside-r.org/packages/cran/coefficientalpha/docs/cronbach/.