US English attitudinal prosody performances in L1 and L2 speakers

Albert Rilliard [1], *Donna Erickson* [2], *Takaaki Shochi* [3], *João Antônio de Moraes* [4]

[1] LIMSI-CNRS, France, [2] Kanazawa Medical University & Sophia University, Japan
[3] CLLE-ERSSàB UMR 5263, France [4] Laboratório de Fonética Acústica, FL/UFRJ/CNPq, Brazil
rilliard@limsi.fr, ericksondonna2000@gmail.com, shochi38@gmail.com, jamoraes3@gmail.com

Abstract

Expressive behavior linked to paralinguistic meanings finds grounds in codes proposed as universals, as well as in culture-specific conventions. This study observes performances in such kinds of attitudinal prosody for USA English, produced by L1 and L2 speakers. The results show that the observed variance is linked to individual competence, to the linguistic context, and to the cultural background of the speakers. They also show that the code used to express a given speech act, code learned in the L1 language by L2 speakers of English, may be used in their L2 language. For some of these expressions, L2 speakers received higher scores than L1 speakers, suggesting that expressions conventionalized in a foreign language, are adequately fulfilling not-conventionalized expressions in the L1 culture.

Index Terms: prosody, attitude, cross-cultural, first and second language

1. Introduction

Prosodic performances help a speaker to express a position about her/his speech and also about the addressee [1].They help to actualize speech acts (e.g. to convey doubt or authority) in a non-verbal way, thus allowing a smoother interaction – it is easier to retract from a position that has not been verbalized. Such attitudinal prosody has been theorized in pragmatic terms as intermediate between emotional expressions and illocutions [2,3]. Differing from emotions, they are intentionally produced and controlled by the speaker; unlike e.g. wh-questions, attitudinal prosody may not have a one-to-one relationship with the meaning of the intended speech act. Attitudinal prosody is rather part of a contextualized communication process, and speakers may use many tools, including prosody, to reach their goals.,

Most studies of attitudinal prosody focus on performance in a given language, either proposing inventories of the existing clichés observed in a language [4,5,6,7], or studying the various means used to express a given speech act in a given language (e.g. politeness in [8,9]). Some studies are interested in the cross-cultural comparison of attitudinal prosody perception by listeners of various linguistic origins [10, 11]. To that aim, they take samples of social affects from a given language's inventory to present stimuli to listeners from various languages, studying the biases induced by cultural backgrounds in the reception of prosodic attitudes. In all these works, attitudes are obtained and labeled according to the cultural specificities of a given language; that is they are interested in e.g. the performance of Catalan politeness, of French irony, etc.

But, as advocated by Wierzbicka [12,13,14] for emotions, the label of an attitude in a given language (e.g. French *ironie*, USA English *irony* or Brazilian *ironia*) does not necessarily cover the exact same concept: these three ironies are not produced in similar social context, or with a similar communicative goal. It is interesting to study these conceptual differences, but we'll here concentrate on how prosodic performances may vary across languages in identical situations – trying to freeze such conceptual mismatches (i.e., what would have been the prosodic performances of Brazilian Portuguese speakers expressing the speech act corresponding to French "*ironie*"?) The aim of this ongoing research is precisely to record speakers from various linguistic and cultural origins, and to compare their prosodic performances in the same situations of communication. Speakers are not asked to produce a sentence with e.g. irony or authority, but rather to behave in situations conducive of these attitudes, as exemplified below:

- *Authority*: Speaker A is a custom agent; speaker B is a traveler. B is in front of A, requesting permission to enter the country; A needs to impose his authority; the scene is at a custom counter at the airport.
- *Irony*: A & B are friends, same age; A is going to Boston to see an important baseball game, and B, who is living in Boston calls A. Unfortunately, the weather in Boston is rainy and B says its wonderful; the scene is at an airport.

The interaction contexts as well as the communication goals of speakers are the same, whatever the language of the speakers. Speakers may be L1 or L2 speakers of the studied language. Several speakers from the same linguistic origin, and of both genders, are recorded to enable a comparison of the cross-speaker, cross-gender, and cross-cultural influence on the prosodic performances. Among the hypotheses of the work are the codes that have been proposed in the literature as universal of prosodic expressivity – namely the frequency code [15], and the effort and production codes [16]. These codes would predict for example the use of higher pitch for expressions where the speaker is in an inferior position, and of lower pitch in the case of dominant expressions. Another research aim is to observe the possible influence of the linguistic and cultural backgrounds on the performances of speakers, for expressions corresponding to complex social settings. All interaction contexts do not necessarily correspond to a prototypical prosodic attitude in each culture, but this is precisely part of what is under investigation here. Will speakers use their L1 prosodic typology speaking in L2? If a situation is (or nor) conventionalized in the speaker's L1, will it help (or restrain) the performances of her/his productions, in L1 and in L2?

This paper focuses on the results obtained for USA English for L1 speakers, compared to L2 speakers whose L1 language is either Japanese (L2JP) or French (L2FR). The aims of the study include observations of prosodic performances across languages and cultural backgrounds, in comparable communication situations, and examination of the perception of these prosodic changes by L1 listeners of these languages. We will here review the evaluation by L1 listeners of USA English of the performances of L1 and L2 speakers in expressing speech acts that correspond to sixteen interaction situations.

10.21437/SpeechProsody.2014-167

2. Corpus

Sixteen communication situations have been designed, where the recorded speakers interact with an experimenter (L1 speaker of the target language) in order to elicit the corresponding attitudes. These attitudes are performed on two target sentences ("A banana" and "Mary was dancing") that are produced in the end of a small dialogue picturing a situation where these sentences may be produced with the intended attitudes. A complete description of the recording setting may be found in [17]. The sixteen situations correspond roughly to expressions described by the following labels: admiration (ADMI), arrogance (ARRO), authority (AUTH), contempt (CONT), doubt (DOUB), irony (IRON), irritation (IRRI), neutral declarative sentence (DECL), neutral question (QUES), obviousness (OBVI), politeness (POLI), seduction (SEDU), sincerity (SINC), surprise (SURP), uncertainty (UNCE), "Walking On Eggs" (WOEG), often referred to as "walking on eggshells" or "walking on thin ice."

Labels "sincerity" and "walking on eggs" correspond to situations typical of the Japanese society, where the speaker is supposed to behave in a specific way because of the hierarchical relationship with the interlocutor and the intended speech act. For sincerity, the speaker asserts the sincerity of her/his utterance to a higher-level interlocutor. The WOEG situation is close to the Japanese concept of *kyoshuku*, defined by [18] as "*corresponding to a mixture of suffering ashamedness and embarrassment, which comes from the speaker's consciousness of the fact his/her utterance of request imposes a burden to the hearer*" (p. 34). Expressions labeled "irony" and "seduction" correspond to attitude types frequently used in the U.S.A., where two speakers have mutually positive, friendly attitudes. Irony may often be used to help lighten an otherwise negative situation. Other types of "negative" irony, intending to hurt the other person, are not addressed in this study. The attitude of "seduction" might best be defined here as where the speaker interacts with her/his interlocutor with the intention of being attractive, fascinating, inviting to the listener. This type of attitude may be akin to that frequently seen in Hollywood movies, but without necessarily being sexually provocative.

Among the speakers recorded for this study on USA English, eight (5 females and 3 males) are L1 speakers; six (3 females and 3 males) are L2 speakers with Japanese (Tokyo variety) as their L1; five (3 males and 2 females) are L2 speakers with French (standard French variety) as their L1. Most speakers are university students. To select L2 speakers – and to ensure a basic level in the target language-selected L2 speakers all spent one year studying in the USA. All speakers were audio-visually recorded performing the two sentences in the 16 situations. The video were hand segmented, resulting in 256+192+160=608 stimuli.

3. Performance evaluation

3.1. Experimental paradigm

These audio-visual performances were presented to listeners, whose L1 is USA English, who had to judge the performances on a raw 1 to 9 scale. Three different groups of listeners evaluated each of the three groups of speakers (L1, L2 Japanese, L2 French). Subjects were first presented with the 16 situations, explained showing the pictures used to elicit the 16 attitudes for "banana." The stimuli were then presented in isolation, by speaker (in order to give subjects a better idea of the expressive habit of this speaker) for both sentences and were randomized. Prior to the presentation of a stimulus, the name of the targeted attitude was displayed, then the stimuli was presented (only once), and listeners had 10 seconds to use the keyboard to give a performance score; then the next stimuli was presented. In cases of no answer, a 0 score is given to the stimulus (it represents less than 0.1% of the number of answers) – these 0 scores were latter removed from the data. A test session lasts typically 30 to 45 minutes. To move to the next speaker, subjects were instructed to press the "enter key."

3.2. Subjects

17 subjects (7 females, mean age 25) evaluated the L1 speakers' performance; 16 subjects (6 females, mean age 21) evaluated the L2 speakers with Japanese L1; 35 subjects (26 females, mean age 24) evaluated the L2 speakers with French L1. The evaluations for all three groups were conducted in Midwest U.S.A., (the prevailing dialect is Midwestern English). The first two evaluations were conducted in Ohio and South Dakota, the third group, in New Mexico.

3.3. Performance judgments analysis

The performance judgments received by each stimulus were analyzed using a mixed-effects model [19] based on the lme4 library [20] of the R software [21]. The performance scores were normalized using a z-score for each subject to avoid individual differences in the use of the answer scale. These performance measures were fitted as a function of the subjects and the speakers – as random effects; the latter being nested into the group of linguistic origin factor, and the linguistic group factor interacts with the targeted attitude and the sentence (the last three being fixed factors).

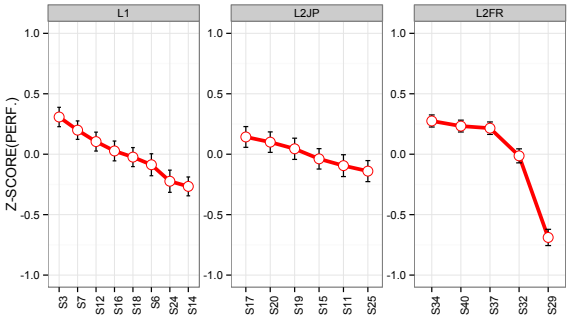

Figure 1: *Mean performances (expressed in z-scores) obtained by individual speakers, nested in the linguistic groups (L1 speakers, L2 Japanese, L2 French); the error bars represent the confidence intervals at 5%.*

Note that, because each of the three groups of listeners evaluate one of the three groups of speakers, the raw performances between speakers' group cannot be directly compared – thus the use of z-scores. But z-scores, since they are standardized, also prevent us from comparing the average performances of two groups of speaker. The present analysis focuses on the differences inside each group of speakers, and compares their relative performances in the 16 situations.

An ANOVA based on the mixed-effects model presented above shows that the linguistic group and the subject factors

did not have a significant main effect (this result is obvious, after the preceding comment on z-scores), but the linguistic group shows a significant interaction with the attitude ($\chi^2_{(30)}$=156.2, p<2.2e-16), with the sentence ($\chi^2_{(2)}$=22.1, p=1.6e-5), and the triple interaction was also significant ($\chi^2_{(30)}$=80.9, p=1.4e-6). The main effects of targeted attitudes and support sentences are also significant (respectively: $\chi^2_{(15)}$=1105.5, p<2.2e-16; $\chi^2_{(1)}$=7.3, p=0.007). The speaker factor also has a significant effect on the performances (effect tested using the method presented in [19:242]; comparing this model with another model, it is identical except it did not have the speaker factor – $\chi^2_{(1)}$=940.1, p<2.2e-16).

3.3.1. *Speakers performances*

The mean performances reached by each speaker, in each linguistic group, are displayed in figure 1. There are important variations across speakers. These variations are mostly linear in the L1 and L2 Japanese groups; one speaker (S29) departs from the others in the L2 French group. Performance variations are more restricted (the group of speakers is more homogeneous) in the L2 Japanese groups, compared to the two other groups. The following analyses consider performances averaged across all speakers of each of the three groups.

3.3.2. *Effect of sentences*

Targeted expressions were produced on two sentences ("A banana" and "Mary was dancing"), chosen for their syntactic simplicity and absence of affective meaning. These individual sentences did have a small – yet significant – effect (cf. Figure 2) on the performances. This effect is different in the case of the L1 group of speakers, as compared to the two L2 groups: L1 speakers received higher scores for their performances on the "Mary" sentence than on the "banana" one, while the reverse is observed for both groups of L2 speakers.

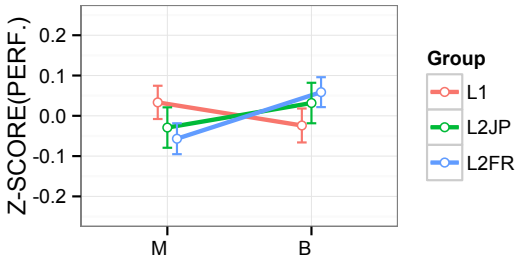

Figure 2: *Mean performances (in z-scores) for both sentences ("Banana" and "Mary was dancing"), for each linguistic group (L1, L2 Japanese, L2 French); error bars represent the confidence intervals at 5%.*

3.3.3. *Effect of attitudes*

Expressing each of the 16 individual attitudes is not achieved by speakers with the same accuracy. The significant main effect of attitude on the performance is depicted in figure 3, and shows important differences (of about one standard deviation) between the best-elicited expression (the situations of surprise) and the worst one (the situation of irony).

The tendency observed for the speakers of all linguistic groups for the two best (surprise and doubt) and the two worst (irony and seduction) expressions holds for the L1 speakers, while it seems culture-specific competences or constraints induce varying patterns of performances for the two L2 groups of speakers.

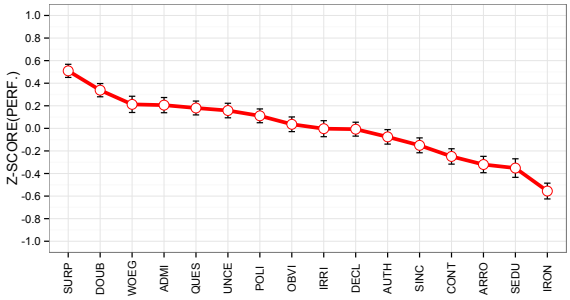

Figure 3: *Mean performances (in z-scores) for each attitude (see text for labels), all linguistic groups averaged; the error bars represent the confidence intervals at 5%. Attitudes are sorted in descending level of performances.*

These cultural-specific patterns, as well as some cross-cultural similarities, are displayed in figure 4, where attitudes are sorted in descending order of performances for the L1-speaker group. The expression of surprise received the highest performance scores whatever the linguistic origin of speakers; the expression of doubt is also ranked amongst the best performances for all speaker groups (if only the third best for the L2 Japanese speakers).

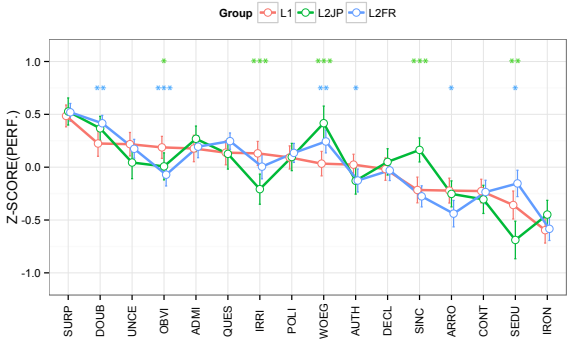

Figure 4: *Mean performances (in z-scores) for each attitude (see text for labels), and for each linguistic group (L1, L2 Japanese, L2 French); the error bars represent the confidence intervals at 5%. Attitudes are sorted in descending level of performances for the L1 speaker group. Stars above attitudes indicate significant differences between the L1-speaker group and L2 Japanese group (top line of stars) or L2 French group (bottom line of stars).*

More differences are observed amongst the performances of the lowest ranked attitudes. The expression of irony seems to be difficult to elicit for speakers of all linguistic origins, but it is not necessarily the most problematic one: L2 Japanese speakers received their lowest performances scores for their behavior in the situation of seduction. The expression of seduction received also the second lowest scores for L1 speakers, while it received only the fifth lowest score for L2 French speakers (close to the mean performance level of this group). Other important differences are observed between the L1 speaker group and the two L2 speaker groups.

In order to focus on the most important differences between language groups, T-tests were run to test significant

differences between performances for a given attitude of the L1-speaker group and each L2-speaker groups. Significant differences are indicated on figure 4 by stars above the significant differences observed for a given attitude (*: $p<0.05$; **: $p<0.01$; ***: $p<0.001$).

The L2 Japanese speakers received significantly different scores from the L1 group for five interaction situations. For irritation, obviousness, and seduction, Japanese speakers received significantly lower scores than L1 speakers; conversely, Japanese speakers outperformed L1 speakers in the cases of sincerity and WOEG expressions.

The L2 French speakers received significantly different scores from the L1 group for six interaction situations. For arrogance, authority, and obviousness, they received lower scores than the L1 speakers, while for the expressions of doubt, seduction, and WOEG, L2 French speakers outperformed the L1 speakers.

4. Discussion & conclusions

Speakers, depending on their personalities among other things, performed differently. The interaction situations were prepared to ease the expression of the targeted attitude in a laboratory environment – but still, in the laboratory as in real life, individuals behave differently, and express themselves with more or less change in voice and face; thus the observed variation in the performances. The focus of this paper is not on individual differences. In order to have a better idea of performance differences at the linguistic and cultural levels, several speakers have been recorded in each language group. It would of course be of a great interest to enter the details of these differences and study the influence of individual characters on expressivity (cf. [22, 23, 24] on this topic). Along this line, it is interesting to note that some speakers received low mean scores for their performances, but may have received the best score for their elicitation of a given situation. It is typically the case for L1 speaker S24, whose mean performance scores are rather low, but who was rated amongst the best for expressing surprise.

Also, the variations induced in performances by the sentences may be related to the type of elicitation contexts used for each of the sentences. L1 speakers tended to perform better for the "Mary was dancing" sentence, which is introduced by a written dialogue, while the L2 speakers received higher scores in the "banana" situations, which are presented using iconic presentations based on images. The language level, and the linguistic complexity may be an explanation for the observed differences. Another explanation of this observation (not necessarily exclusive of the first one) could be linked with the greater linguistic complexity of the "Mary was dancing" sentence, compared to "Banana".

Most of the explained variance is linked with the varying performances of the 16 attitudes, and with the differences between linguistic groups in the performance of individual attitudes. A first remark about these differences is that they concern only a minority of the communication contexts: in most cases, L1 and L2 speakers do perform comparably and this result is in itself interesting. This supports literature reporting few differences found cross-culturally e.g. in [25]. The differences observed between L1 and L2 speakers can be explained by different factors.

- A first kind of differences is the one where L1 speakers outperform one (and only one) group of L2 speakers in performing a social attitude. It is the case for irritation with L2 Japanese speakers, and for arrogance and authority with L2 French speakers. These situations are conventionalized in the considered cultures. Whether they are based on different codes may be answered by comparing the productions of both L1 and L2 groups of speaker.
- In the situation leading to expressions of obviousness (a propositional attitude, see e.g., [26]), both groups of L2 speakers are outperformed by L1 speakers. Although the expression of obviousness may be conventionalized in each of these cultures, propositional attitudes address the linguistic content of the utterances and thus are subjected to more variations in their pitch contours [26]. One can suppose that the L2 speakers did not catch the typical prosodic patterns used by L1 speakers.
- The situation demanding speaker to have a seductive behavior shows an interesting partition, with L2 Japanese speakers receiving lower score than the L1 speakers, while L2 French speakers received higher scores. In this case, the social interaction is not conventionalized in the Japanese society, while it is – if under different conventions in the French culture. The lack of – or presence of – a conventionalization seems to be a critical point in the measured performance of speakers.
- In the case of the situation leading to the expression of sincerity and WOEG, the L2 Japanese speakers outperformed the L1 speakers. These expressions correspond to common situations in the Japanese culture, while they are not conventionalized in the culture of the USA, where the society is less based on hierarchical relations. This lack of conventionalization in the L1 culture may also, as in the previous case, be a key point in the observed differences in performance.
- The L2 French outperformed L1 speakers in the WOEG situation to a lesser degree, although it is not a conventional situation in the French culture. A possible explanation may lie in the more hierarchical nature of social relationships in France, as compared to the USA, that may play role for an unconventional expression, but not for conventional ones such as arrogance or authority.

A question remains. In the cases where L2 speakers outperformed L1 speakers, (perhaps because the L2 culture conventionalization "trained" the L2 speaker how to behave in these situations), why were these performances evaluated so highly by L1 speakers of USA English who did not know the conventions. Why did they give high ratings to L2 French speakers' seduction and L2 Japanese speakers' WOEG and sincerity? Could it be possible that the judges do have stereotypic representations of peoples from these cultures, and thus relate Japanese behaviors more easily to politeness expressions, and French to seductive behavior? It could be possible to test for such stereotypes by running performance judgments of the same expressions with native listeners of the two L2 cultures less biased by such stereotypes.

5. Acknowledgements

This work was supported by the following grants: ANR PADE, JSPS Grants A #25240026 and A #23320087, and PEPS IDEX MAVOIX. The authors warmly thanks M. Kondo & S. Detey from Waseda University for their help; they also thank all the speakers and listeners for their participation.

6. References

[1] Moraes, J. A., "The pitch accents in Brazilian Portuguese: Analysis by synthesis", in Proceedings of Speech Prosody 2008, Campinas, 389–397, 2008.

[2] Wichmann, A., "The attitudinal effects of prosody, and how they relate to emotion", in Proceedings of the ISCA Workshop on Speech and Emotion, Newcastle, 143–148, 2000.

[3] Wichmann, A. "Attitudinal intonation and the inferential process", In Speech Prosody 2002, 11-16, 2002.

[4] Martins-Baltar, M., "De l'énoncé à l'énonciation: une approche des fonctions intonatives", Paris: Didier, 1977.

[5] Fujisaki, H. & Hirose, K., "Analysis and perception of intonation expressing paralinguistic information in spoken Japanese", Proceedings of the ESCA Workshop on Prosody, 254-257, Lund, Sweden, 1993.

[6] Morlec, Y., Bailly, G. & Aubergé, V., "Generating prosodic attitudes in French: Data, model and evaluation", Speech Communication, 33(4):357–371, 2001.

[7] Gu, W., Zhang, T. & Fujisaki, H., "Prosodic Analysis and Perception of Mandarin Utterances Conveying Attitudes", Proceedings of Interspeech, Firenze, Italy, 1069-1072, 2011.

[8] Wichmann, A., "The intonation of Please-requests: a corpus-based study", Journal of Pragmatics, 36: 1521–1549, 2004.

[9] Nadeu, M. & Prieto, P., "Pitch range, gestural information, and perceived politeness in Catalan", Journal of Pragmatics, 43(3): 841-854, 2011.

[10] Scherer, K. R., Brosch, T., "Culture-Specific Appraisal Biases Contribute to Emotion Dispositions", European Journal of Personality, 23: 265–288, 2009.

[11] Shochi, T., Rilliard, A., Aubergé, V. & Erickson, D., "Intercultural perception of English, French and Japanese social affective prosody", in S. Hancil [Ed.], The role of prosody in affective speech, Linguistic Insights 97, Bern: Peter Lang, AG, Bern, 31-59, 2009.

[12] Wierzbicka, A., "A semantic metalanguage for a cross-cultural comparison of speech acts and speech genres", Language in Society 14(4): 491-513, 1985.

[13] Wierzbicka, A., "Defining Emotion Concepts", Cognitive Science 16:539-581,1992.

[14] Wierzbicka, A., "Empirical Universals of Language as a Basis for the Study of Other Human Universals and as a Tool for Exploring Cross-Cultural Differences", Ethos 33(2):256–291, 2005.

[15] Ohala, J.J., "An ethological perspective on common cross-language utilization of F0 of voice", Phonetica, 41:1-16, 1984.

[16] Gussenhoven, C., "The Phonology of Tone and Intonation", Cambridge: Cambridge University Press, 2004.

[17] Rilliard, A., Erickson, D., Shochi, T., Moraes, J., "A. Social face to face communication – American English attitudinal prosody", in Proceedings of Interspeech, Lyon, 1648-1652, 2013.

[18] Sadanobu, T., "A natural history of Japanese pressed voice", Journal of the Phonetic Society of Japan 8(1): 29-44, 2004.

[19] Baayen, R. H., Davidson, D. J., Bates, D. M., "Mixed-effects modeling with crossed random effects for subjects and items", Journal of Memory and Language, 59: 390–412, 2008.

[20] Bates, D., Maechler, M., Bolker, B. & Walker, S., "lme4: Linear mixed-effects models using Eigen and S4". R package version 1.0-5. 2013. http://CRAN.R-project.org/package=lme4

[21] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

[22] Scherer, K. R., "Personality markers in speech", in K.R. Scherer & H. Giles [Eds.]. Social markers in speech. Cambridge: Cambridge University Press, 147-209, 1979.

[23] Sadanobu, T., "Nihon shakai, Nozoki Chara-kuri - Kaotsuki, Karadatsuki, kotobatsuki-", Sanseido Publisher, 2011.

[24] Sadanobu, T. & Luo, M., "Bumpo, Para-gengojouhou, Character ni motozuku Nihongo meishisei bunsetsu no tougoteki kijyutsu", Journal CAJLE (ISSN 1481-5168), 12: 77-95, 2011.

[25] Rilliard, A.; Erickson, D.; Moraes, J. A.; Shochi, T., "Cross-Cultural Perception of some Japanese Expressions of Politeness and Impoliteness", in F. Baider, G. Cislariu [Eds.] Linguistic approaches to emotions in context. Amsterdam: John Benjamins, 251-276, 2014.

[26] Moraes, J. A., Rilliard, A., "Illocution, Attitudes and Prosody", In T. Raso et al. [Eds.], Spoken Corpora and Linguistic Studies, Amsterdam: John Benjamins, to appear.