Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Speech prosody and possible misunderstandings in intercultural talk – A study of listener behaviour in Standard Vietnamese and German dialogues

Kieu-Phuong Ha, Samuel Ebner, Martine Grice

Institute for Linguistics, Department of Phonetics, University of Cologne
{hak;samuel.ebner;martine.grice@uni-koeln.de}

Abstract

A perception experiment testing the interpretation of backchannels by Vietnamese native listeners (Ha 2012) indicates that the pitch reflecting affective meanings may not be derived from the Frequency Code as proposed for a large number of languages (Ohala 1983, Gussenhoven 2004). In the current study we investigate the prosodic patterns of backchannels used by Vietnamese and German speakers in map task dialogues. The analysis focuses on the rechecking phase of the map task. Our findings show that Standard Vietnamese backchannels are produced consistently with a falling/level pitch contour. For German, although there is variation in the form of backchannels, they are predominantly produced with a rising contour. The study points out potential misunderstandings that may occur in intercultural talk in general, and in Vietnamese-German dyads in particular.

Index Terms: intonation, backchannels, interpersonal communication, intercultural talk.

1. Introduction

Intonation not only serves linguistic functions, such as indicating whether an utterance is a question or a statement, but also conveys interpersonal meanings such as whether someone is being polite or is an interested listener. The mapping between intonation – more specifically pitch modulation – and the functions it expresses is not always the same across languages and cultures.

Perception studies have shown that listeners with different language backgrounds can have distinct perception strategies and that different speech communities may associate the same pitch contours with meaning in diverging ways [1], [2], [3]. A recent study on Vietnamese backchannels [4] has shown that Standard Vietnamese might be an exception to the intonational universals as derived from the Frequency Code [2], [5] which makes reference to the relationship between the size of the larynx and the vibration of the vocal folds across speakers. This code is claimed to explain the association of high pitch with questions and affective meanings such as friendliness, politeness and submissiveness as well as the association of low pitch with confidence and dominance [1], [2], [5].

The current study compares Vietnamese and German, two languages and cultures that differ in many ways. German is an intonation language with stress-timed rhythm that makes use of pitch, among other parameters, to convey meaning at utterance or discourse level [6]. Vietnamese is, on the other hand, a syllable-timed language with a complex lexical tone system, in which pitch is essentially used to convey word meaning [7].

As far as we know, there have been no studies that directly compare these two languages in terms of culture-specific discourse strategies. Our aim in this paper is to investigate how native speakers of these two languages use pitch modulation when interacting with a conversation partner. More specifically, we are concerned with how they give feedback whilst engaged in task-oriented dialogues. In section 2 we briefly report on a perception test conducted for Vietnamese [4], addressing how Vietnamese listeners interpret the affective meaning of pitch modulation on backchannels. This part serves as background for our production experiment involving Vietnamese and German map tasks, which we present in section 3. In section 4, we conclude and discuss the findings in the context of potential misunderstandings in intercultural talk.

2. Background

Backchannels are commonly short utterances (e.g. *yes*, *right* or *exactly*) or even utterances with non-lexical items (e.g. *uh huh* or *mhm*) with a distinctive prosody and which can be accompanied by facial expressions or specific gaze behaviour. The primary function of backchannels is to signal the listener's attention. In addition, they have been found to signal affiliation or agreement with the content of the talk at hand [8], [9], [10]. Recent studies on the prosody of backchannels in other languages such as British English show that a falling tone plus a rapid speech tempo can convey non-supportiveness, such as when the listener attempts to end the speaker's turn/topic [11]. In American English a high pitch on backchannels can show that the listener is interested and encourages the speaker to say more about the current topic [12].

A previous study on the perception of affective meanings of pitch modulation in Vietnamese backchannels [4] found that level or falling pitch on backchannels is interpreted by native listeners of Standard Vietnamese as significantly more polite than rising pitch. In this study two backchannel tokens excised from two contexts signalling attention and agreement were manipulated for the last 50% of the word. The manipulation created four contours with endings at four pitch levels, the difference between each level being 30 Hz, as illustrated in Figure 1.

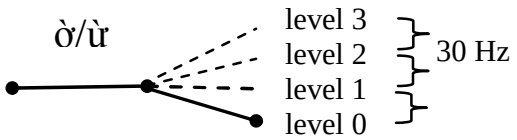

Figure 1: *Manipulated pitch levels on ờ/ừ.*

 doi: 10.21437/SpeechProsody.2016-164

In a semantic scaling task, 24 subjects judged the hearer's politeness and 25 subjects judged the hearer's dominance on a 5-point scale (higher values indicate greater level of politeness and dominance).

The results for politeness showed that across both contexts level/falling contours were perceived as more polite than rising contours, see Figure 2. For dominance, as illustrated in Figure 3, there is a tendency for level/falling pitch to be perceived as less dominant than rising pitch, although there were considerable differences across the two backchannel types (signalling attention vs. agreement).

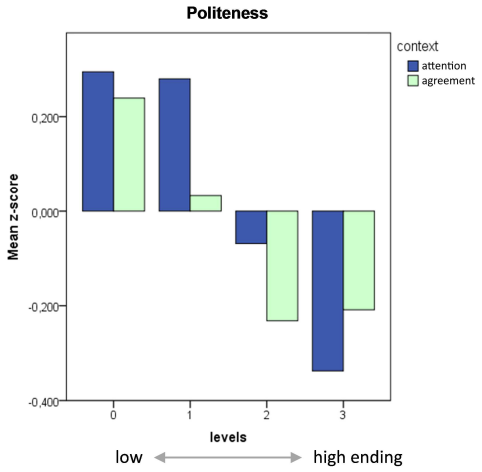

Figure 2: *Mean z-scores for 24 subjects rating politeness in backchannels paying attention and signalling agreement.*

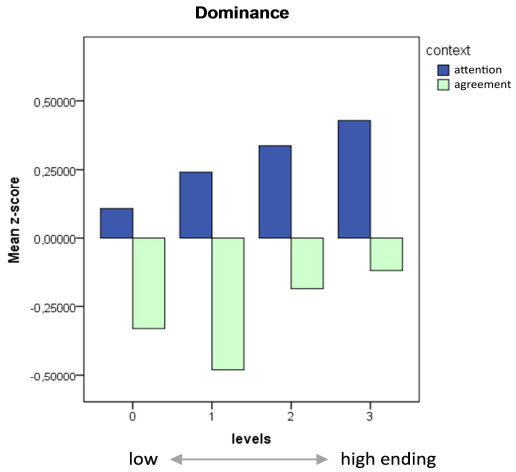

Figure 3: *Mean z-scores for 25 subjects rating dominance in backchannels paying attention and signalling agreement.*

These results suggest that there is a difference in the way pitch signals affective meanings in this language as compared to cross-linguistic tendencies derived from the Frequency Code, according to which a higher pitch is associated with a higher perceived degree of politeness and a lower pitch is associated with a higher perceived degree of dominance [1], [2], [5] [cf. 13].

3. Production experiment

Studies on German show that backchannels in the language are also realised with a number of intonational patterns. These include falling and falling-rising pitch as expressions of acknowledgment, and confirmation or agreement, respectively [14], [15], [16]. In this section we analyse task-oriented dialogues to investigate how speakers of Vietnamese and German signal that they are listening so that the interlocutor can continue to talk.

3.1. Participants and map task

We recorded a corpus of map task dialogues involving native speakers of Vietnamese and German. Since the number of male speakers was not comparable across the two languages, in this study we only focus on the data of female speakers. Participants were 10 Vietnamese female speakers (aged 20 to 35) and 10 German female speakers (aged 22 to 26) in Hanoi and in Cologne. All of them were recorded in their native languages and without eye contact. We used the map task design from the HCRC Map Task Corpus [17]. Two participants have slightly different maps with 11 or 12 landmarks. Some of the landmarks are located differently. While only one map has a route marked on it, the other has only the starting point of the route. The task is for one participant to instruct the other to reproduce the route on the second map without either participant seeing the other's map. This is achieved by discussing the route and landmarks. Towards the end of the task both participants recheck the route, making sure that the reproduced route corresponds to the original one. We refer to this as the rechecking phase.

Two sets of maps were used with each dyad, so that each participant was recorded once giving instructions and once drawing the route. In total, our corpus consisted of 20 dialogues, 10 Vietnamese and 10 German. The basic structure of every dialogue contains the following moves: instructing, checking, query, explaining and aligning [cf. 18]. The aligning move involves the rechecking phase, the part of the task we focus on here, specifically on the backchannels produced by the instructor whilst listening to the follower describing the route that she has drawn. This ensures a consistent comparison of backchannels of one and the same category across the languages.

3.2. Data and methods

Excerpts (1) and (2) represent Standard Vietnamese and German backchannels from the rechecking phase analysed in this study. For the annotation of non-lexical tokens, we use *mm* for monosyllabic tokens and *mm mm* for disyllabic ones. The acoustic features of disyllabic tokens are provided below.

(1) A: *Đi thẳng tuốt về hướng cần tây ở ngay dưới.*
I'm going straight on in the direction of the celery, right at the bottom.

B: *Mm.*

A: *Lượn qua đầu của cần tây đi vào giữa cần tây và cánh cam.*

I'm going around the top of the celery and between the celery and the lady bug.

B: *Ờ.*

Yes.

(2) A: *Ich gehe über den Storch.*
I'm going over the stork.

B: *Mm mm.*

A: *Links am Storch vorbei.*

Past the stork on the left.

B: *Mm mm.*

A: *An der Seite, wo nicht die Schrift ist.*

On the side where there is no text.

In general speakers of both languages used a number of words as backchannels in the rechecking phases: In German they were *okay*, *richtig* 'right', *genau* 'correct', *ja genau* 'yes, right', and the non-lexical items *mm*, *mm mm*. In Vietnamese the backchannels are *chuẩn* 'correct', *rồi* 'got it!', *đúng rồi* 'right', *ôkê* 'okay', *ừ/ờ* 'yes', and the non-lexical items *mm*, *mm mm*. In this analysis we focus on the prosody of the three backchannels which occur most frequently in both languages in the corpus. They are *ja*, *ừ/ờ* (German and Vietnamese 'yes'), and in both languages, the monosyllabic *mm* and the disyllabic *mm mm*. The latter non-lexical items are particularly relevant for our analysis as they can be seen as the carrier of intonation without any lexical tone in Vietnamese. Tokens that occur non-finally (e.g. turn-initially) were excluded, the same applies to tokens functioning as responses to tag or yes/no questions, e.g. to questions seeking confirmation. In total we analysed 310 backchannels. Table 1 provides the frequency of these tokens across the two languages.

Table 1: *Frequency of the backchannels investigated.*

Token / Language	mm	mm mm	ja / ừ / ờ	Total
Vietnamese	134	3	33	170
German	24	57	59	140
Total	158	60	92	310

The target tokens were annotated using conversation-analytical methods [19] to identify the pragmatic context of backchannels on the one hand, and auditory as well as acoustic methods on the other. Disyllabic tokens in German were identified by a clearly discernable voiced glottal fricative [ɦ] between the two syllables. In cases where it was not clear whether the syllable was disyllabic, we took into account the speech pressure waveforms and intensity contours. Disyllablic tokens in Vietnamese were more straightforward to identify, with a glottal stop marking the onset of each of the two syllables.

F0 values and duration of the tokens were extracted for further analysis. The F0 contour was analysed by calculating the intervals (in semitones relative to the respective utterance mean) between the F0 of the time point at 10% (A) and the time point at 90% (B) into the tokens. This measurement was conducted in order to minimise the effect of microprosody and glottalization at the beginning and the end of the syllable. If the value of the interval between B and A was positive, this was an indication of a rising contour, if it was negative or zero, it indicated that the contour was falling or level. Note that across the annotated tokens, there are no major inflection points between A and B that change the form (rising, falling or level) of the contour. The F0 contours were extracted in Praat [20], corrected using a customized version of mausmooth [21], and plotted in Praat and R [22].

3.3. Results

In general we found that Vietnamese non-lexical backchannels are either falling or have a level contour, while the German equivalents are predominantly rising. For the lexical token meaning 'yes' in both languages, the picture is somewhat different. The *ja* tokens in German can either rise or fall, falls being often accompanied by glottalization. For Vietnamese *ờ* and *ừ*, the contours are consistently level or falling, similar to the citation form realisation of the lexical falling tone of the words.

We examined the distribution of the F0 rises and falls by calculating the intervals between the F0 of the time point at 10% (A) and the time point at 90% (B) into the tokens. The distributions of the calculated intervals are plotted in Figure 4 for lexical (i.e. 'yes') vs. non-lexical tokens across the two languages investigated. The boxplot shows the quartiles of the data (whiskers represent 1.5x inner quartile range, and notches ±1.58x IQR/sqrt(n)). This figure shows an asymmetry in the distribution of the pitch contours across the two languages: The values for Vietnamese are predominantly below the zero level, indicating that Vietnamese backchannels have a falling or level F0 contour. There is only one outliner for a non-lexical *mm* the value of which displays a clear further distance to the zero line.

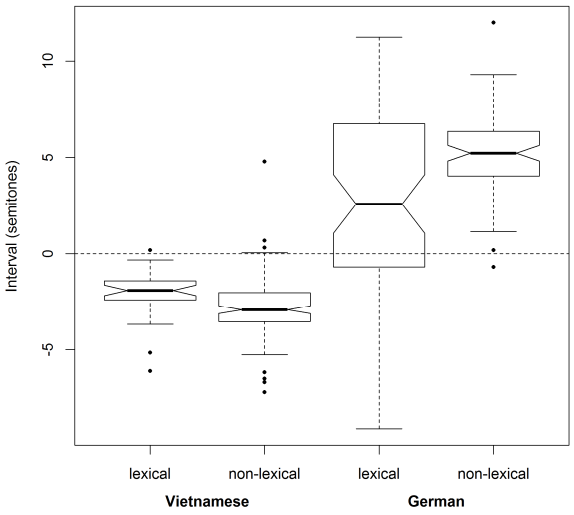

Figure 4: *Distribution of F_0 intervals in semitones between the time points of 10% and 90% into the tokens (negative values = falling, positive values = rising).*

By contrast, the values for German are more variable, especially in the case of the lexical *ja* tokens indicating that these tokens can have a rising or a falling contour. The clearest difference lies between the non-lexical backchannels of the two languages. In Vietnamese they fall or have a level contour. Since *mm*s have no lexical tone, this finding provides evidence that we are dealing with the falling or level intonational contours. By contrast, German *mm*s rise. While disyllabic *mm mm* tokens in the corpus appear to occur frequently in German (n=57), they are rare in Vietnamese (n=3).

In terms of duration, there is a tendency for German that higher pitch corresponds to longer duration. This tendency is clearer for the backchannels meaning 'yes', which are plotted in Figure 5 with overall trendlines calculated by linear regression for the respective languages (not accounting for the

factor speaker). While the duration for *ja* ranges from around 100 to 500ms, the duration for *ờ/ừ* ranges between 200 and around 350ms. The falling *ja* tokens are the shortest on average and the rising *ja* tokens the longest. The Vietnamese falling/level *ờ/ừ* tokens are midway between the two.

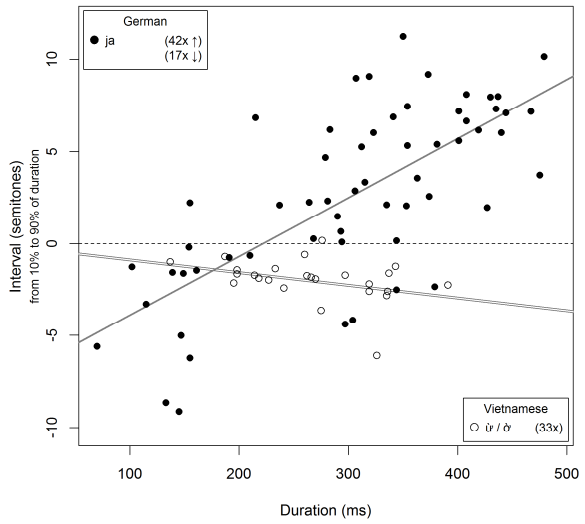

Figure 5: *Distribution of F_0 intervals in semitones between the time points of 10% and 90% of the tokens* ja, ờ/ừ, *plotted against token duration*

Although the majority of the *ja* tokens have a rising contour, a considerable number have a level or falling contour. Our sequential analysis, i.e. the analysis of the conversational context [cf. 19] of the *ja* utterances in German, shows that while the rising *ja* tokens predominantly indicate "Go on! I'm listening" (n=42), the falling ones (n=17) indicate other pragmatic meanings such as "Done!" or "Done, you can go to the next landmark!". The falling/level pitch contour is often accompanied by non-modal voice such as glottalization, which occurs at the end of the word. A closer look at these tokens shows that they also appear to have a function in structuring discourse. Across the board, by using these tokens the listener conveys some 'closure' notion, i.e. her agreement with the content of the talk, and at the same time closing (a sequence of or a move in) the conversation. The existence of these two intonationally distinguished categories within the totality of the German *ja* tokens explains the great variance in the F0 contours found.

4. Conclusion and discussion

Our production study revealed crucial differences between the backchannel intonation patterns of German and Vietnamese listeners during the rechecking phase of the map task. While Vietnamese "Go on!" signals are consistently level or falling, German equivalents are predominantly rising. We also found that the intonation patterns in German are more flexible than those in Vietnamese, most probably due to the functional load of pitch in conveying word meanings in the latter. The German *ja* tokens appear to have a rising as well as a falling contour. The former contour signals attention or "Go on!", while the latter contour appears to have a closing confirmation function (such as "Done!") occurring toward the end of a move or at the end of a conversation. In terms of duration, there is a tendency for German "Go on!" backchannels (rising contour) to be longer on average than the Vietnamese ones, whereas the German "Done!" backchannels (falling contour) tend to be shorter.

Considering the results from the previous study investigating how native listeners perceive the affective meanings of backchannels in Standard Vietnamese, the current study pinpoints candidates for possible misunderstandings in intercultural talk, particularly in Vietnamese-German dyads (i.e. one of the speakers being non-native in the other language). In Standard Vietnamese, a rising pitch might be interpreted as impolite. For German natives, the level/falling pitch might lead to irritation [cf. 23], or might give the impression of impoliteness, e.g. it could be interpreted as indicating disinterest or even an attempt to end the current speaker's turn, as evident in the function of the falling *ja* backchannels.

It seems the case that speakers of Standard Vietnamese associate the rising pitch pattern with emphasis and interpret it as displaying too much energy or being too emphatic, resulting in a higher degree of perceived dominance and a lower degree of perceived politeness. An explanation for this tendency might be culturally motivated. A final high pitch can be linguistically interpreted as a question. When addressing others, in particular people who are not familiar or in a higher position, questioning is generally not considered to be acceptable in Vietnamese. Furthermore, the form of a falling or non-rising contour in Vietnamese backchannels, which is perceived as more polite, might be not arbitrary, but rather iconic [cf. 24]. There could be a relationship between the intonational form of a fall, the meaning of politeness and the lowering of the eyes, head and body gestures that have been documented as a sign of respect in Asian cultures cf. [25], [26].

The current study provides information for learners of Vietnamese, in particular for learners with German as a native language. Backchannels in Standard Vietnamese are commonly produced with a low level or falling pitch. A rise is not unnatural but it can be perceived as impolite. Misinterpretation during communication with native speakers of Vietnamese are thus likely if German speakers transfer the prosodic patterns from their first language to Vietnamese speech.

Further investigation is needed to ascertain the potential misunderstandings in Vietnamese-German dyads mentioned above. Follow-up study will look at the backchannels across the two languages occurring during the real tasks (i.e. during the instructing, checking, query and explaining moves) from the same corpus, on the one hand and at the perception of native listeners of German in terms of how they interpret the affective meanings when the intonation of the signals are modified towards the Vietnamese patterns, on the other.

5. Acknowledgements

We would like to thank our participants for taking part in the experiments, Trần Thị Minh for her help in recruiting subjects for the perception test, and Nguyễn Xuân Hằng for her help with the data annotation. This work is funded by the German Research Foundation on the Project "Tone and Intonation in Vietnamese" and by the University of Cologne on the project "Prosody and Conversational Interaction: A study of Vietnamese and German dialogues".

6. References

[1] C. Gussenhoven, "Intonation and interpretations. Phonetics and phonology", in B. Bel and I. Marlin, (eds.): *Proceedings of Speech Prosody, Aix-en-Provence*, pp. 45–57, 2002.

[2] C. Gussenhoven, *The phonology of tone and intonation*. Cambridge: CUP, 2004.

[3] A. Chen, *Universal and language-specific perception of paralinguistic intonational meaning*. Utrecht: LOT, 2005.

[4] K. P. Ha, *Prosody in Vietnamese – Intonational Form and Function of Short Utterances in Conversation*. Asia-Pacific Linguistics 002 (SEAMLES 001). PhD thesis. Canberra: The Australian National University, 2012.

[5] J. J. Ohala, "Cross-language use of pitch: an ethological view", in *Phonetica*, vol. 40, pp. 1–18, 1983

[6] M. Grice, S. Baumann and R. Benzmüller, "German Intonation in Autosegmental-Metrical Phonology", in S.-A. Jun (Ed.) *Prosodic typology. The phonology of intonation and phrasing*. OUP, pp. 55–83, 2005.

[7] D. H. Nguyen, *Vietnamese*. Amsterdam/Philadelphia: John Benjamins, 1997.

[8] S. Duncan, "On the structure of speaker-auditor interaction during speaking turns", in *Language in Society*, vol. 3, pp. 161–180, 1974.

[9] F. E. Müller, "Affiliating and disaffiliating with continuers: prosodic aspects of recipiency", in E. Couper-Kuhlen and M. Selting (eds.), *Prosody in Conversation*. Interactional Studies. Studies in Interactional Sociolinguistics 12. Cambridge and New York: CUP, pp. 131–176, 1996.

[10] R. Gardner, „Rezipientenpartikeln in der englischen Konversation: Mm, Mm Hm (Uh huh) und Yeah", in *Essener Linguistische Skripte – elektronisch*, vol. 3, no. 1, pp. 15–29, 2003.

[11] A. Wichmann, *Intonation in Text and Discourse*. Pearson Education (Longman), 2000.

[12] N. Ward, "Pragmatic Functions of Prosodic Features in Non-Lexical Utterances", in *Speech Prosody*, pp. 325–328, 2004.

[13] E. Uldall, "Dimensions of meaning in intonation", in D. Abercrombie, D. B. Fry, P. A. C. McCarthy, N. C. Scott, J. L. Trim (eds.), *In Honour of Daniel Jones: Papers Contributed on the Occasion of his Eightieth Birthday*, London: Longman, pp. 271-279, 1964.

[14] K. Ehlich, *Interjektionen*. Tübingen: Niemeyer (Linguistische Arbeiten 111), 1986.

[15] J. E. Schmidt, "Bausteine der Intonation?", in *Germanistische Linguistik*, vol. 157/158, pp. 9–32, 2001.

[16] M. Müller, *Prosodie von Hörersignalen in deutschen Telefongesprächen*. Unpublished Master thesis, Universität zu Köln, 2011.

[17] A. Anderson, M. Bader, E. Bard, E, Boyle, G. M. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. S. Thompson, and R. Weinert, "The HCRC Map Task Corpus", in *Language and Speech*, vol. 34, pp. 351–366, 1991.

[18] J. Kowtko, S. Isard, G. M. Doherty, "Conversational Games within Dialogues", in *Proceedings of the ESPRIT Workshop on Discourse Coherence*, University of Edinburgh, pp. 1–12, 1993.

[19] H. Sacks, E. A. Schegloff, and G. Jefferson, "A Simplest Systematics for the Organization of Turn-Taking for Conversation", in *Language*, vol. 50, no. 4, pp. 696–735, 1974.

[20] P. Boersma and D. Weenink, "Praat: doing phonetics by computer". Retrieved from http://www.praat.org, 2015.

[21] F. Cangemi, "Mausmooth". Retrievable online at http://phonetik.phil-fak.uni-koeln.de/fcangemi.html, 2015.

[22] R Core Team, "R: A Language and Environment for Statistical Computing". *R Foundation for Statistical Computing*, Vienna, Austria. Available online at: https://www.R-project.org, 2015.

[23] T. Stocksmeier, S. Kopp and D. Gibbon, "Synthesis of prosodic attitudinal variants in German backchannel *ja*", in *Interspeech*, 2007.

[24] D. Bolinger, "The Inherent Iconism of Intonation", in J. Haiman (ed.), *Iconicity in Syntax*, Amsterdam and Philadelphia: John Benjamins, pp. 97–108, 1985.

[25] B. Ingersoll-Dayton and A. Saengtienchai, "Aspect for the elderly in Asia: Stability and change", in *International Journal of Aging and Human Development*, vol 48(2), pp. 113–130, 1999.

[26] K. Sung, "Elder respect: Exploration of ideals and forms in East Asia", in *Journal of Aging Studies*, vol 15, pp. 13–26, 2001.