Random Effects in the Third Survey of the Okazaki Survey on Honorifics

Kenjiro Matsuda
Kobe Shoin Women's University

1 Introduction

Do interviewers play any significant role in deciding the quality or performance of the respondents in questionnaire-based surveys? Recent research has shown that individuals are important elements in accounting for language variation (Johnson 2009), regardless of whether the data are from natural speech interviews or questionnaire-based surveys (Maekawa 2017). Nevertheless, the statistical status of the interviewer effect has not been examined in variationist sociolinguistics. Under the generalized linear mixed-effect model (GLMM), both these effects can be entered into the model as random factors to test their statistical significance. This study aims to examine the status of several random effects using data from the third survey of the Okazaki Survey on Honorifics (OSH) conducted in 2008.

2 Problem: Effect of the individual and interviewer on linguistic variation

The status of individual speakers in the speech community has been an issue of contention for variationist sociolinguists (Guy 1980, Romain 1982, Wolfram and Thomas 2002). Despite various proposals, researchers are far from a consensus on how to assess and incorporate the individual's effects in building the variable grammar of the speech community. One major obstacle in tackling of this problem is the absence of a proper statistical method to exactly estimate the individual effect using sociolinguistic data. Simply coding the individual as an independent factor has been one solution, as it automatically groups individuals according to their linguistic performances (Rousseau and Sankoff 1978), but this has not become a common practice in the field.

The introduction and the rapid diffusion of GLMM (also known as *hierarchical model* or *multilevel model* in other fields; Johnson 2009) brought a revolutionary change to the situation[1]. The most innovative feature of the GLMM is that, by introducing the random factor into the regression model, it can accommodate data wherein each observation is not independent of the other. This is typical in familiar situations such as where multiple observations are made by the same speaker, or where the data are collected by multiple observers. In the former case, observations from the same individual are not independent from each other, and they should exhibit some mutual similarity. In the latter case, the data collected by each observer are not independent from each other (as they should show similarities). GLMM easily accommodates both cases by postulating a *random factor*, which accounts for the variation due to the individual speaker/observer, in addition to the *fixed factors*. A variable is set to a random factor when, for example, it has too many levels, not all the levels are covered in the dataset, or the effects of those levels is not the main interest of the research. In contrast, factors whose levels are all exhausted in the data and the effects of those levels are among the main interests of the research are set as fixed factors. The general formula of the GLMM takes the following form:

(1) $y = X\beta + Zu + \varepsilon$,

where y is a vector of outcomes, X is a matrix of predictor (fixed) variables, β is a matrix of fixed-effects regression coefficients, Z is a matrix of random factors, u is a matrix of their effects, and ε is a residual. The random effects can be a slope or intercept of the regression model, or both.

The introduction of GLMM into variationist sociolinguistics has opened a new possibility to explore the

[*] Thanks to Kikuo Maekawa for his comments on the earlier version of this paper. This study was made possible by the 2019 NINJAL Collaborative Research Project "The Analysis of the Questionnaire and the Visiting Record for the First, Second, and Third Okazaki Survey on Honorifics," and JSPS KAKENHI Grant #16H03420.

[1] See Matsuda (2018) for a detailed history of the method's introduction into variationist sociolinguistics.

conundrum mentioned above: it is now possible for linguists to incorporate into the regression model the effects of the individual speaker and interviewer on linguistic performance to account for their variation. GLMM has already become a default analytical method for variational linguistics, and in Japanese linguistics, following Hibiya (2012), Maekawa (2017) used GLMM with a Bayesian estimation to examine the individual's effect using a language standardization survey in Tsuruoka, a large-scale real-time language survey project conducted by National Institute of Japanese Language and Linguistics (NINJAL 1954, 1974, 2007). He demonstrated convincingly that the model incorporating the respondent as a random factor shows best fit to the data.

Maekawa's (2017) analysis, however, left one possibility unexamined: the effect of the survey interviewer. Since the questionnaire-based survey in Tsuruoka involved each researcher interviewing multiple respondents, we have reason to believe that results from the same interviewer are not independent from each other. That is, we can expect that a better model would have, aside from fixed effects, a random effect for the respondent and another random effect for the interviewer who interviewed those respondents.

This study attempts to examine the possibility that a proper model of the variation requires an interviewer random factor in addition to the respondent random factor, using the dataset from the OSH. The OSH is another large-scale real-time language survey project conducted by NINJAL, similar to the Tsuruoka project but with honorifics as its main focus instead of language standardization.

3 Data and method

The OSH was conducted by NINJAL with the purpose of capturing the use of honorifics and the speaker's consciousness about their use in Okazaki city, Japan (NINJAL 1957, 1983, Abe 2010, Nishio et al. 2010). The OSH has been conducted thrice in 1957, 1972, and 2008 (OSH I, II, and III), and its data come from a trend study based on the random sample of residents in the area, and a panel study that follows up with the speakers sampled in the preceding surveys. The entire dataset is available on the NINJAL website (https://www2.ninjal.ac.jp/ longitudinal/okazaki.html). Figure 1 shows the size of each sample for OSH I to III:

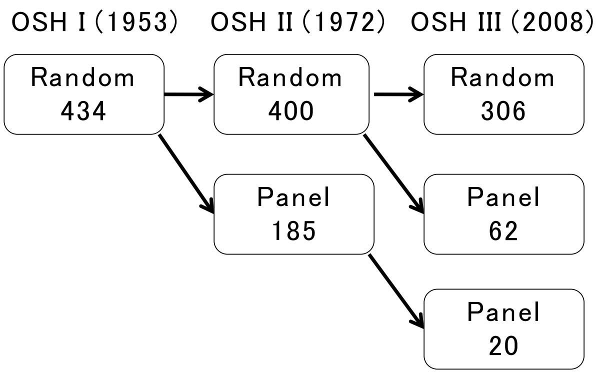

Figure 1: OSH Samples

The main components of all three OSH surveys are the linguistic and social life sections. The former has a series of 18 questions on the use of honorifics in specific situations, an honorifics recognition test, and a section asking about opinions on the actual use of honorifics. This section is conducted through face-to-face interviews with the respondents. Each question asks what the respondents would say to a specific interlocutor in a specific situation. Most of the questions also have a picture aid to facilitate the interview. All the responses were written down for OSH I and II on the spot, and they were also recorded on an integrated chip recorder for OSH III. The responses were given an honorifics-level rating from one to three (with one being the highest), so that each response for each question has exactly one honorific level. We use 11 of these questions for the current study. Figure 2 shows the distribution of each question by the honorifics level with which the respondents answered the question, with the error bar indicating the standard deviation.

Of the 28 interviewers in OSH III, 21 were professional researchers and seven were graduate students of linguistics. Each interviewer was given a list of respondents, whom they visited with variable success rates. Table 1 gives the relevant statistics concerning the number of interviews they conducted, and Figure 3 illustrates the distribution of the number of interviews each interviewer conducted[2]. For the present study, we use the trend sample part of OSH III, and so the data structure is most simplified.

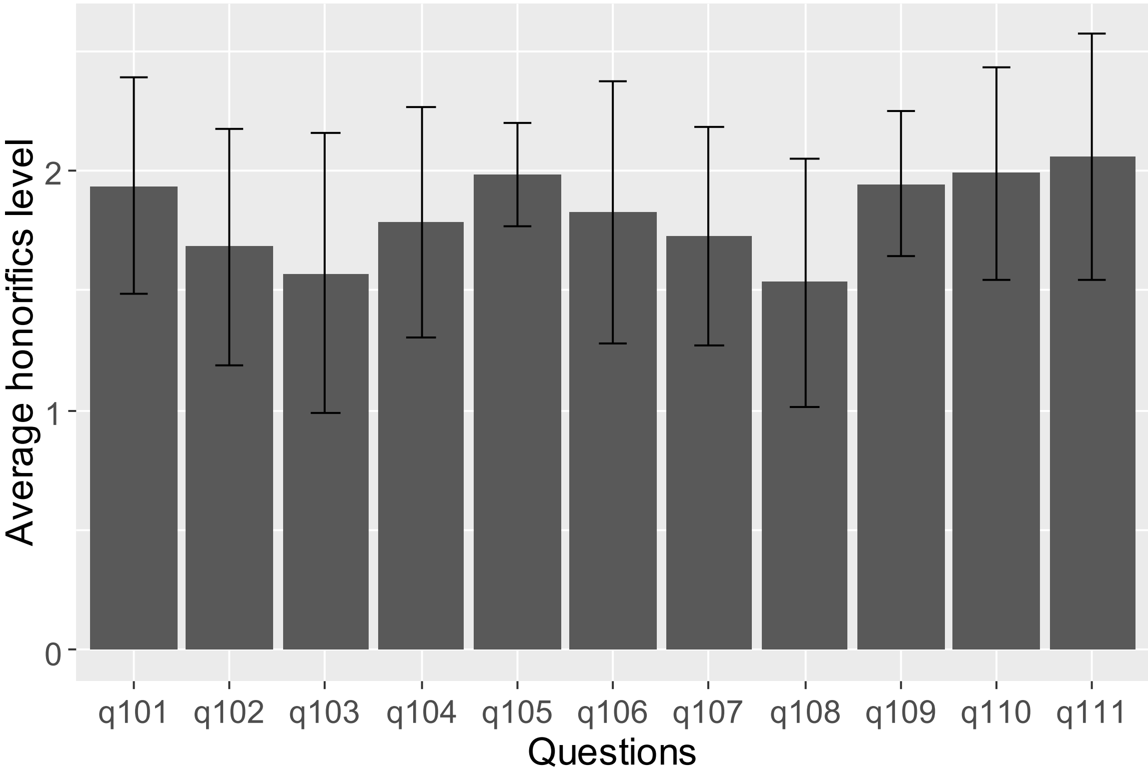

Figure 2: Distribution of the 11 questions by the average honorifics level

Table 1: Interviewer statistics

Total number of interviews	306
Mean	10.89
Median	8.50
Minimum	2.00
Maximum	25.00

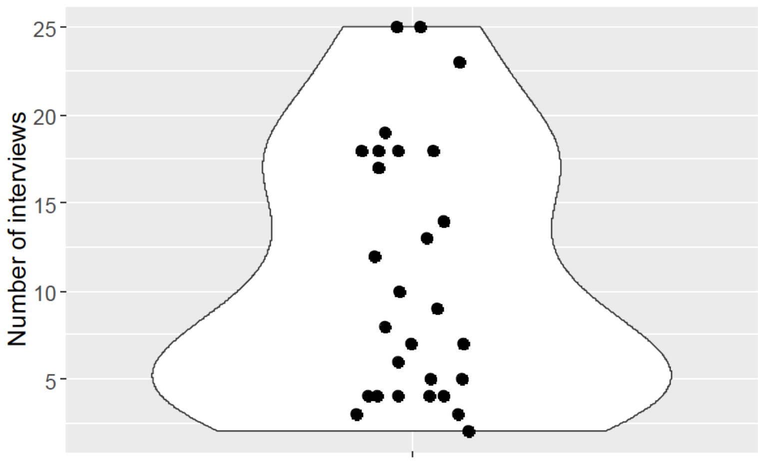

Figure 3: Distribution of the number of interviews by each interviewer

[2] The OSH database does not provide the interviewer's name for each respondent. They were retrieved from the original questionnaire by the author.

We can easily imagine that these 28 interviewers had different interview skills, and those differences could affect the respondents' responses, including the levels of honorific expressions they use in answering the questions. Such an expectation seems plausible when we look at Figure 4, which plots each respondent's average honorifics level for the 11 questions by the interviewer. With the boxplot for each interviewer showing a sizable variation, we have reasons to include the variable as one of the random variables.

From the OSH database, we also use the respondents' sex and the birth year (age) as social variables. The distribution of the sample by age and sex is described in Table 2. In view of the distribution of the honorifics level by age in Figure 5, the variable is divided into three categories, 10s-20s, 30s-40s and 50s-70s. Figure 6 plots the distribution of the average honorifics level for all the questions by sex and age. There are more females in the lower half of the scatterplot, suggesting that they are more polite in answering the questions than males.

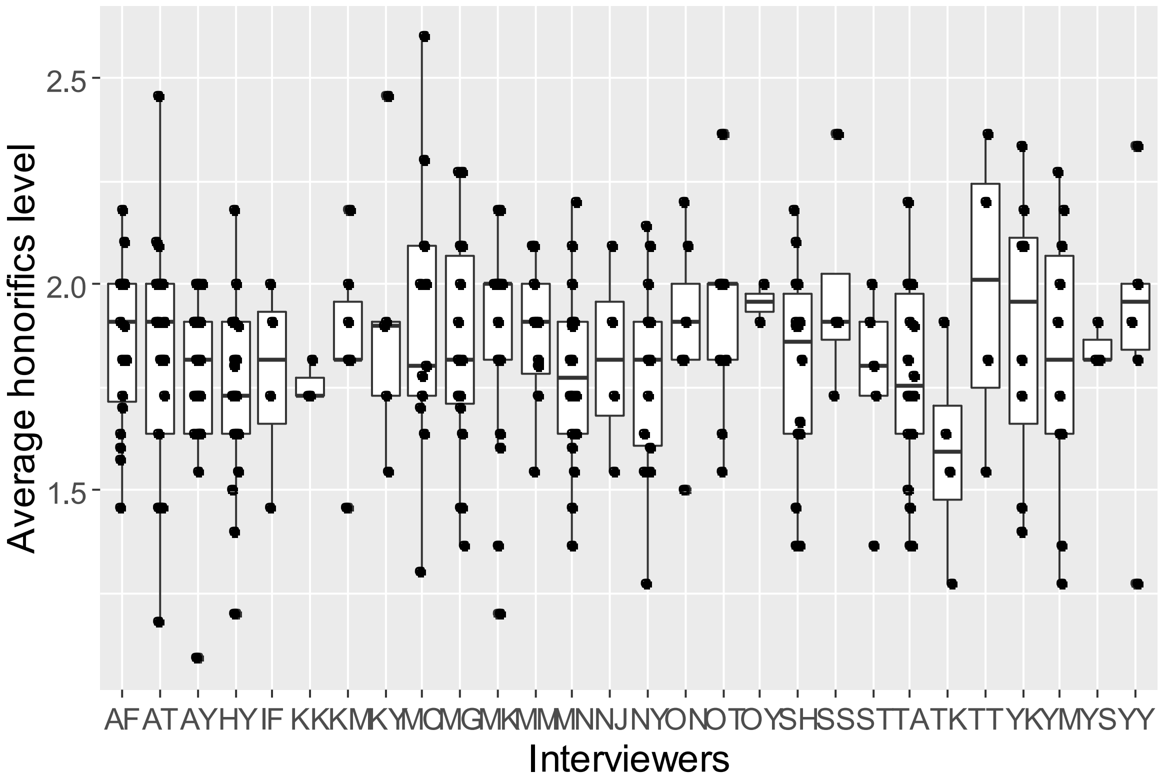

Figure 4: Distribution of the respondents' honorifics level by the interviewer

Table 2: Distribution of the sample by sex and age

	10s	20s	30s	40s	50s	60s	70s	Total
Male	8	23	31	26	22	31	19	160
Female	7	16	24	28	29	23	19	146
Total	15	39	55	54	51	54	38	306

From the OSH III trend study dataset, a total of 305 cases (excluding one case for which we could not obtain the interviewer's name) are used for statistical analysis. The sex and the age of the speaker are entered as the fixed variable into the model. For the dependent variable, the level of honorifics for each question is used, as it is deemed as a continuous variable. The lme4 package (version 1.1.21) of R (version 3.6.1 for Windows 64bit) is used for the statistical analysis, with an option to use the maximum likelihood in place of the restricted maximum likelihood estimation method.

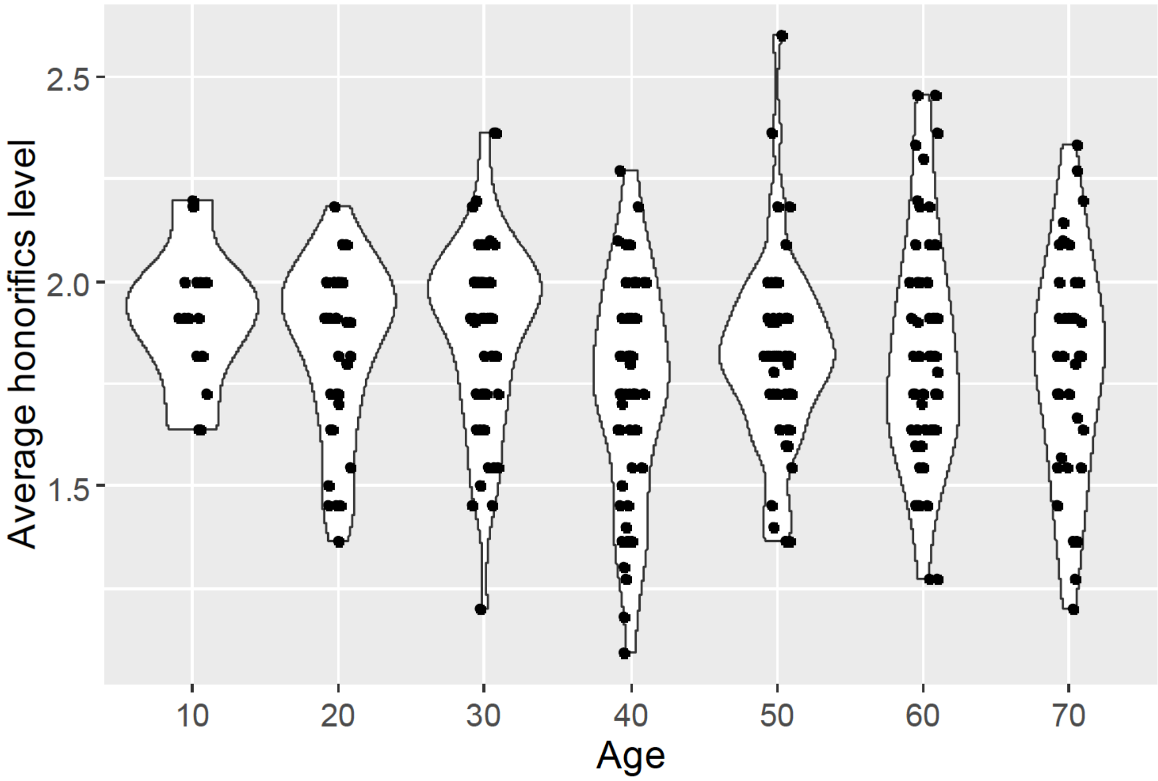

Figure 5: Distribution of the average honorifics level for the 11 questions by respondents' age

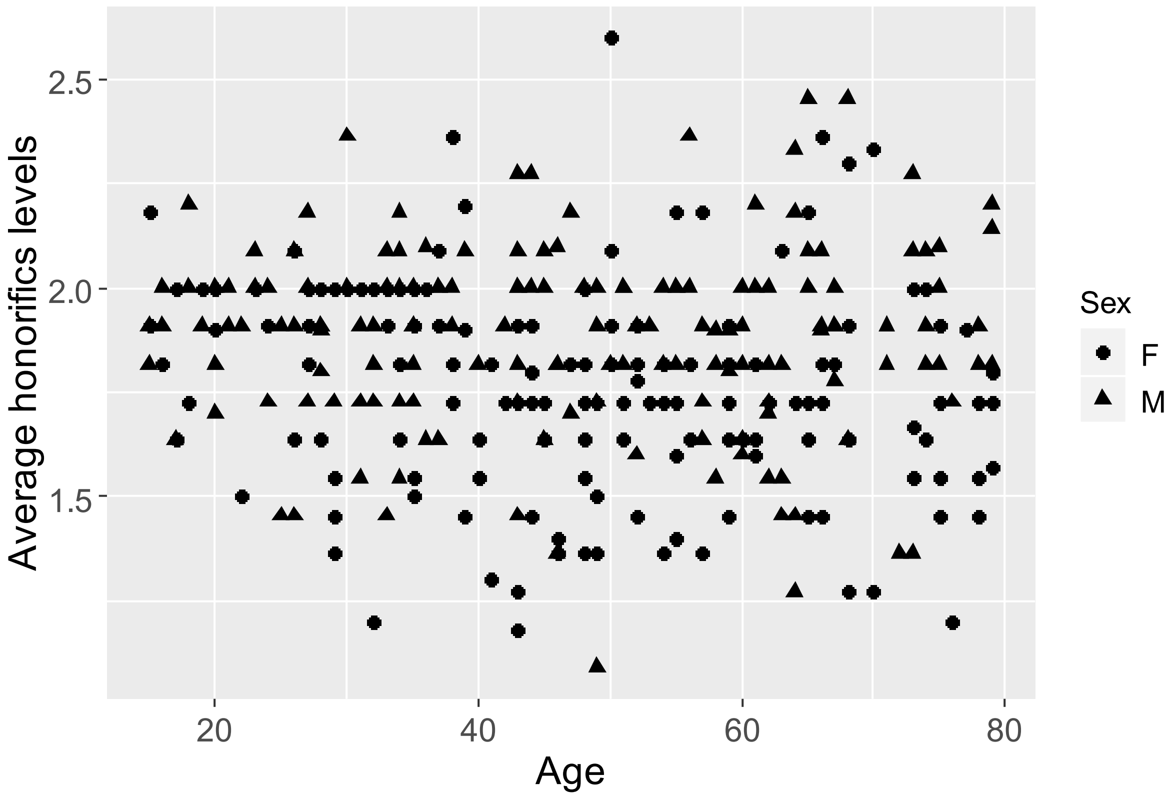

Figure 6: Distribution of average honorifics level for the 11 questions by respondents' sex and age

4 Result

A series of analyses using the step() package selected model (2) as the best one:

(2) Honorifics level for each question for each respondent = Sex + Age + Sex × Age + Respondent + Question

Table 3 summarizes the detailed statistics of model (2).

Table 3: Result of the linear mixed-effect model

Akaike information criterion	3,954.6
Bayesian information criterion	4,015.6
Log likelihood	-1,967.3
Deviance	3,934.6

RANDOM EFFECTS:

Groups	Variance	Std. Dev.
Respondent	0.0391630739	0.1978966
Interviewer	0.0000002684	0.0005181
Questions	0.0281569691	0.1678004

FIXED EFFECTS:[3]

	Estimate	Std. Error	df	*t*	*p*
Intercept	1.889120	0.058691	19.016462	32.188	<2e-16 ***
Sex [Female]	-0.035477	0.045320	300.925985	-0.783	0.4343
Age [30-49]	-0.015866	0.045056	301.151391	-0.352	0.7250
Age [50-]	-0.006491	0.044593	302.271917	-0.146	0.8844
Sex [Female]:Age[30-49]	-0.134350	0.064520	301.241242	-2.082	0.0382 *
Sex [Female]:Age [50-]	-0.145694	0.067308	303.614488	-2.165	0.0312 *

*** 0.001 ** 0.01 * 0.05

As expected from the relevant plots in the previous section, the interaction term for sex and age was selected as a significant fixed effect. With 1 being the highest honorific level and 3 the lowest, the females seem more polite than the males, and the age effect seems rather small. Turning to the interaction term, we see that the age effect is clear in the age group of 30s and up for females.

The most notable point of the result, however, is that among the random factors (intercepts, in this case), the respondent and question turned out to be significant. That is, the interviewer effect was deemed as non-significant, as confirmed by its extremely small variance ($2.684e^{-7}$). The result clearly shows, in agreement with Maekawa (2017), that the respondent is a significant factor in accounting for linguistic variation, but this is not the case with the interviewer.

5 Discussion

While the significant status of the respondent factor may not be surprising given the individual variability of individual respondents, the nonsignificant status of the interviewer calls for an explanation. At this point, we can conceive the following three hypotheses. First, the use of the questionnaire should help to make the quality of the interviews equal. In natural speech interviews, interviewers guide the flow of the interviews by asking various questions to elicit as much spontaneous speech as possible. Accordingly, one might expect that the interviewer's skill plays an important role in deciding the quality of the interviews. In contrast, it is usual that every effort is taken to minimize the differences in the interview styles in questionnaire-based surveys, with

[3] Note that because the lme package uses the sum contrast parameterization, the male level of sex is set to zero, and so is the 10-20s level of age and Sex [Female]:Age [30-49].

procedures and wording being fixed in detail for all speakers. Naturally, then, there would be little differences in the interview quality due to the interviewer.

Second, the sheer number of interviewers could be responsible for the small differences due to the interviewer. With 28 members, OSH III had the largest survey team of all OSH projects (Table 4). This could have cancelled out the skill differences and, consequently, the differences in the quality of the interviews.

Third, the makeup of the interviewers could be a factor. If we deem an interviewer as a "specialist" if his/her main subject of study is descriptive linguistics/dialectology/quantitative sociolinguistics or he/she at least has experience in working as an interviewer in a similar survey, then OSH III had the highest rate of specialists of all the OSH projects (Table 4)[4]. If 75% of the interviewers were specialists in linguistic interviews, it would be expected that the quality of the interviews would be more or less equal. At this stage, we can only suggest these three possibilities as the causes for the absence of the interviewer effect.

Table 4: Breakdown of the interviewers by their specialties

	Specialist	Non-specialist	Total
OSH I	6 (38%)	10 (63%)	16 (100%)
OSH II	7 (58%)	5 (42%)	12 (100%)
OSH III	21 (75%)	7 (25%)	28 (100%)
Total	34	22	56

OSH = Okazaki Survey on Honorifics.

Note that all these hypotheses can be tested by analyzing the data from OSH I and II in a similar way as in the current study. According to the first hypothesis, the interviewer effect would be absent in OSH I and II as well, but the second the third hypotheses would make predictions in the opposite direction.

6 Conclusion

We demonstrated that at least for OSH III data, of the three random factors, the respondent and question factors are significant, and the interviewer effect was found to be negligible. As for the fixed effects, the interaction term for sex and age of the respondent turned out to be significant in the model.

Our study also leaves several issues to be tackled in future studies. First, we need to analyze OSH I and II data and see how the random effects behave there. If the interviewer effect turned out to be significant in both these surveys, then the most plausible explanation would be the specialty of the interviewer as noted above. In contrast, if the factor is found to be nonsignificant in the two surveys as well, we can conclude that in questionnaire-based surveys such as the OSH, the interviewer effect plays no significant role.

As another issue, there is a problem of a proper statistical estimation method. The GLMM method used in our study is based on the maximum likelihood estimation, but statisticians are not in agreement regarding the exact treatment of the likelihood value. This is one reason why Maekawa (2017) adopted the Bayesian method. The Bayesian method is now more accessible than before thanks to popular software such as BUGS and Stan with easy-to-use interfaces, and they should bring new perspectives to our analysis on the variation in OSH datasets.

References

Abe, Takahito. (2010) *Keigo-to Keigo Ishiki: Aichi-ken Okazaki-shi-niokeru Dai-sanji Chosa.* Report of the Grant-in-Aid for Scientific Research (A) "Socio-Psychological Survey of Honorifics: The Third Survey in Okazaki, Aichi Prefecture" (No.19202014). Vol. II.

Guy, Gregory R. (1980) Variation in the group and the individual: The case of final stop deletion. In William Labov (ed.) *Locating Language in Time and Space,* 1-36. New York: Academic Press.

Hibiya, Junko (2012). The velar nasal in Japanese revisited. Paper presented at NWAV-Asia Pacific 2 at the National Institute for Japanese Language and Linguistics, August 3, 2012.

Johnson, Daniel Ezra. (2009) "Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule

[4] The percentage values are rounded up at the first decimal, and they may not add up to 100%.

analysis". *Language and Linguistics Compass* 3/1, 359-383
Maekawa, Kikuo. (2017) "Tsuruoka-shi kyootsuugoka choosa deeta-no kakuritsuronteki saikentoo". *Proceedings of Language Resources Workshop* 2, 164-181.
Matsuda, Kenjiro (2012) What happened to the honorifics in a local Japanese dialect in 55 years: A report from the Okazaki Survey on Honorifics, *University of Pennsylvania Working Papers in Linguistics* 18(2), Article 7. Available at: https://repository.upenn.edu/pwpl/vol18/iss2/7 [Accessed on December 27, 2019].
Matsuda, Kenjiro. (2018) Hen'i riron-niokeru ippanka senkeemoderu-no doonyuu. *Keiryoo Kokugogaku* 31(6), 402-416.
National Language Research Institute. (1954) *Chiiki Shakai-no Gengo Seekatsu: Tsuruoka-shi-ni-okeru Jittai Choosa*. Tokyo: Shuei Shuppan.
National Institute for Japanese Language and Linguistics. (1957) *Keigo-to Keigo Ishiki.* Tokyo: Shuei Shuppan.
National Institute for Japanese Language and Linguistics. (1974) *Chiiki Shakai-no Gengo Seekatsu: Tsuruoka-shi-ni-okeru 20-nen-mae-to-no Hikaku.* Tokyo: Shuei Shuppan.
National Institute for Japanese Language and Linguistics. (1983) *Keigo-to Keigo Ishiki: Okazaki-ni okeru 20-nen mae-to-no ikaku.* Tokyo: Shuei Shuppan.
National Institute for Japanese Language and Linguistics. (2007) *Chiiki Shakai-no Gengo Seekatsu: Tsuruoka-shi-ni-okeru 20-nen-kankaku 3kai-no Keezoku Choosa.* Tokyo: NINJAL.
Nishio, Junji, Kayoko Tsuji, and Megumi Kukita. (2010) *Kigo-to Keigo Ishiki: Aichi-ken Okazaki-shi-niokeru Dai-sanji Chosa.* Report of the Grant-in-Aid for Scientific Research (A) "Socio-Psychological Survey of Honorifics: The Third Survey in Okazaki, Aichi Prefecture" (No.19202014) Vol. IV.
Romaine, Suzanne. (1982) "What is a speech community?". In Suzanne Romaine (ed.) *Sociolinguistic Variation in Speech Communities,* 13-24. London: Edward Arnold.
Rousseau, Pascal and David Sankoff. (1978) "A solution to the problem of grouping speakers". David Sankoff (ed.) *Linguistic Variation: Models and Methods,* pp. 97–117. New York: Academic Press.
Wolfram, Walt and Erik R. Thomas. (2002) *The Development of African American English.* Oxford: Blackwell.