International Linguistics Research; Vol. 7, No. 2; 2024
ISSN 2576-2974 E-ISSN 2576-2982
https://doi.org/10.30560/ilr.v7n2p13

Gender- and Age-related Constraints in Japanese Politeness Marking: A Statistical Analysis of the Tokyo Story Characters' Usage of Keigo

Albéric Derible[1]

[1] Komaba Organization for Educational Excellences, University of Tokyo, Tokyo, Japan

Correspondence: Albéric Derible, KOMEX, Komaba 3-8-1(18-1121), Meguro-ku, Tokyo-to, 153-8902, Japan. Tel: 81-3-5454-4826. E-mail: aderible@gmail.com

Received: May 15, 2024; Accepted: June 5, 2024; Published: June 7, 2024

Abstract

This quantitative study explores the relationship between three sociological factors (age, gender and possible in-law relationship) and the use of Japanese honorifics (*keigo*). Statistical analyses are conducted on the occurrence of Japanese politeness markers found in the script of the movie *Tokyo Story* directed by Yasujiro Ozu in 1953. Results notably confirm the highly significant correlation between age, gender and the use of *keigo*, women and younger speakers using consistently more politeness markers than men and older speakers, even though Japanese polite speech does not seem to reflect the order of magnitude derived from the age difference numeric value but only the acknowledgement of one interactant's seniority.

The study also delves into the distinction between the mere presence of *keigo* in discourse and its intensity of use in actual speech (i.e. number of politeness markers), revealing complex relationships. For example, same-sex dyads exhibit different patterns of politeness, with men-to-men politeness associated more with the former and women-to-women politeness more with the latter. The interplay between age and gender in *keigo* use seems to defy systematic rules, highlighting the importance of considering unique gender- and age-based pairings, in other words: distinct categories of address styles that are both interrelated and individual.

Keywords: Japanese honorifics, Politeness markers, Statistics, Sociological factors

1. Introduction

It is a commonly held belief that women's speech is more polite than men's. According to Ide, this is especially true for Japanese women, who use honorifics a lot more frequently than men, and sometimes even excessively (Ide, 1982: 378). Sociolinguists such as Holmes (1995) and Suzuki (2007) have in part denounced this this very popular belief, shared by many other linguists and folks alike, by narrowing the interactional and contextual settings within which this linguistic practice operates. Holmes argues for example that it is only valid if you consider very specific conversational topics or speech acts, such as compliments, as men and women have different views on what it means to be polite, and therefore develop different communication strategies. My own research takes a more quantitative approach to this subject and is more in line with Suzuki's work in that respect. Our work focuses less however on the underlying strategies displayed by speakers and more on the statistical significance of the occurrence of politeness markers in their speech.

The present study aims to determine the role played by gender in Japanese speakers' use of *keigo* (Japanese polite language), especially in relation to their respective age. Linguists have named a number of other social or situational features such as age, education, status, origin, and formality, as factors contributing to the use of Japanese honorifics (Fukuda and Asato, 2003, Hori, 1986). They characterize the vertical and horizontal distances (i.e. power relationship and intimacy level) that separate the speaker from the hearer, and consequently with what kind of *keigo*, and to what degree, they must address one another. The present study focuses on three variables: gender, age and in-law relationship. It aims to assess their respective and combined effect, if any, on politeness marking within the Japanese family. Ultimately, the study shall reveal what are the inter- and intragender differences in terms of politeness, and how the age difference between two family members, or the fact that they are related by blood or by marriage, can account for these differences.

To that end, statistical analyses will be conducted on a corpus based on the screenplay (脚本, *kyakuhon*) of the movie *Tokyo Story*, directed by Yasujiro Ozu in 1953, where the talk of three generations of Japanese family members account for more than 95% of the dialogue.

Once the theoretical framework underpinning this study had been established in section two, the reasons for choosing this corpus will be briefly stated in section three. This third section will also show what kind of Japanese politeness markers have been selected as tokens for the statistical analyses that will follow in section four, alongside speakers' individual features used in designing variables for said analyses. The data thus gathered will be presented and discussed in section five. This section will be divided into subsections covering the three stages in which our statistical analysis was conducted. Section 6 will conclude this article. It will present a brief summary of the research findings and outline possible leads for future research on the subject.

2. Theoretical Framework

2.1 Conflicting Approaches

The "Politeness Theory" is a framework originally laid out by Brown and Levinson (1960, 1987) and later implemented by Leech (2014). It explains how people manage and negotiate "face-threatening acts" in communication while still maintaining social harmony and politeness. Brown and Levinson propose several politeness strategies used in interaction to mitigate potential face threatening acts, along with a number of underlying key concepts, such as "social distance", "power" and "rank of imposition" (1987: 15). They are designed by the authors to be used as tools during analysis. They will not however be presented or discussed any further in this paper, since arguing about their respective merits would far exceed the scope of the current study. A significant number of non-Western authors – many of them Japanese – have objected to the ideology underlying this framework. We shall briefly address their claims to better present our own epistemological premises. The crux of the matter lies in the understanding of the relationship between the particular concept of politeness and language taken as a whole. Ide for example calls for cultural aspects to be considered when evaluation strategies are displayed by the speaker. She differentiates between linguistic and behavior strategies adopted by the speaker and therefore argues that, in Brown and Levinson's model, confusion sometimes arises between propositional contents and speech acts (1989: 239). She puts forward the concept of "discernment" (弁え, *wakimae* in Japanese) as both a cognitive and social process that involves perceiving and understanding the relevant social and cultural factors in a given situation (1999: 448). In short, speakers no longer chose freely what strategy best suits their communication goals, but are – at least in part – guided by their own "discernment". This concept is widely acknowledged to be part of the Japanese speaker's psyche and consequently very often cited in academic papers on Japanese politeness by authors such as Matsumoto (1988: 407) and Okamoto (1999:51), to name but a few. It is also closely related to the concept of "social deixis", a term that is very often associated with the works of Japanese linguists such as Ide (1989), Minegishi Cook (1999:88) and Okamoto (2011: 3675). It presents language as a way to index or convey social relationships, roles and hierarchies between individuals in a given context.

Depending on which of the two theoretical concepts presented above the linguist will base their analysis, linguistic expressions can therefore be understood as either *deriving from*, or *signaling* social variables such as relative status, familiarity, or power dynamics. To our knowledge, no consensus has yet been reached in answer to the question: is politeness merely a tool, or is it akin to grammar, in a sense that it more or less dictates how speakers not only should behave, but also express themselves? Okamoto, for one, seems to lean more toward the second interpretation: "while politeness in languages like English is mostly thought to concern speaker's volitional strategies, the use of honorifics and other formal expressions in Japanese is often said to be governed by rigorous situation-based rules, or conventions" (1999: 51). Except that, when conducting research on sentence-final forms and honorifics, she warns that "these linguistic forms cannot be adequately accounted for by directly relating linguistic forms to certain social variables" (1997: 807). We obviously do not presume to possess the necessary wisdom to adjudicate the matter once and for all, nor would we want to. We will however state that we have adopted a prescriptive approach to this question. That is why, when interpreting the results brought forth by statistical analysis, we will assume that there is causality between social variables and linguistic expressions. Said variables will therefore be hereafter referred to as "factors" constraining the use of Japanese honorifics. This working hypothesis is less motivated by our understanding of language and politeness as being comprised of grammar rules and social norms, rather than by the objective we had in mind when we first considered doing quantitative research on Japanese politeness marking.

As a foreign language instructor, and a student of Japanese, we have very often deplored that Japanese grammar reference books focused more on the *how* to express oneself politely – by resorting to the right *keigo* markers and striking the appropriate language register – than the *when* or the *why*. Shimamori (2001: 304) explains for example that "addressee honorifics" basically differ from "referent honorifics" in that their use characterizes the presence of either a "horizontal" or a "vertical" distance between speakers. But she says very little as to what social factors determine these two types of distances. The present study aims in part to help teachers of Japanese as a foreign language and their students, who would by definition be unaware of the implicit factors governing Japanese

interpersonal relationships, to assess such distances, especially since the very phrasing used in Japanese grammar or textbooks can be easily misinterpreted by students. When Japanese instructors present referent honorifics as conveying either "respect" or "humility", when respectively talking about *sonkeigo* (尊敬語) and *kenjogo* (謙譲語), they indirectly grant emotional content to linguistic forms. In so doing, they perhaps unwittingly adopt what we would call a *sentimentalist* approach to this question, which might lead their students to believe that if they did not feel such emotions toward their addressee, then they might not have to resort to either "respectful" or "humble" languages.

For the remainder of this article therefore, we will assume that Japanese honorifics do in fact characterize a horizontal and/or a vertical distance (i.e. "social distance" and "power" in the Politeness Theory), and that these distances must be evaluated by the speaker on a case-by-case basis in respect to a number of factors that will be detailed in the next section. We will not however rule conclusively on the question of whether or not such distances rigorously dictate the use of honorifics, as the concept of "discernment" would suggest they do, merely reflect a specific "social deixis", or could even be shortened in some instances by a particular communication strategy put in place by the speaker, as implied by the "volitional model" presented at the beginning of this section. Readers will decide for themselves how much room for maneuver is left to the Japanese speaker depending on the theoretical framework they choose to support.

2.2 Factors Determining Politeness Marking in Japanese

To the best of our knowledge, no consensus has yet been reached in regard to the nature, and number, of possible determining factors for Japanese politeness marking. Along with gender and age most often cited, education and region of origin have also been mentioned (Fukuda and Asato, 2004: 9). Ide adds that these factors can be of a social or a psychological nature and that they "are numerous and intertwined". She identifies the major ones as being "social position, […] power, […] age, and […] formality" (1982: 366). For Okamoto, if "interpersonal distance is usually characterized in terms of a status difference and/or the degree of intimacy, or a *uchi-soto* (in-group/out-group) distinction" she also allows for other, more contextual, factors to be considered such as: "formality of the setting, type of genre, means of communication, topic" (1999: 53). As mentioned in the introduction, the present study will only consider three factors: age, gender and in-law relationship. Several reasons presided over this choice. First, age and gender were also the key variables selected by Suzuki (2007: 80-81), to whom the current study owes a great deal in terms of methodology (i.e. gender- and age-based pairings presented in subsection 4.3). Second, these factors can be objectively and categorically evaluated, which could not be said about "intimacy" or "formality" for instance. And third, because, as sociological variables pertaining to individual speakers, age and gender constitute, in our opinion, some of the fundamental features that Japanese speakers consider when evaluating – using "discernment" and/or following their own personal communication objectives – the horizontal and the vertical distances that separate them from their addressee and that will, in turn, guide them toward particular honorifics. In the words of Brown and Levinson, "power" and "distance" can be seen as the two axes that define interpersonal relationships, but the only way to position interactants in relation to each other along these axes is to assess and compare their respective age, gender, etc.

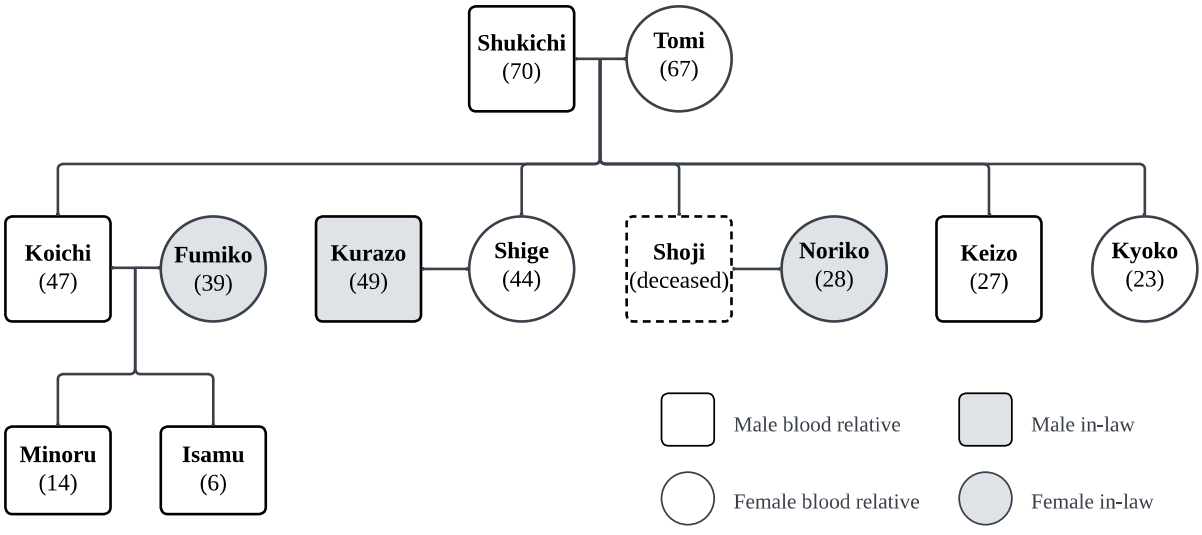
Note, however, that gender is not universally perceived to be as significant a factor as age. Japanese women's speech has long been a topic of discussion in linguistic literature. Ide, examining what makes Japanese women's speech more polite, cites Peng (1981) and states that "the first notable characteristic of Japanese women's speech is the frequent use of honorifics as was verified in three surveys on sex differences in language use" (Ide: 1982, 378). They are moreover more prone to "hypercorrection", which leads them to an increase in use of honorifics (Ide: 1982, 378). Women also are less likely than men to resort to bad language and swear words (Ide: 1982, 381). Okamoto and Shibamoto trace back the origin of Japanese women's speech to the *nyobo kotoba* (女房詞), a cant that was originally used by Japanese court ladies during the Muromachi era and later came to thought of as a general women's language. Together with *genbun itchi* (言文一致), a nineteenth and early-twentieth century literary movement, they greatly contributed to the epitomized image of the *ryosai kenbo* (良妻賢母), "good wife, wise mother", that still set standards for today's Japanese women's speech.

Nonetheless, a workshop we personally conducted in 2021 on the subject of Japanese polite speech elicited several remarks in contradiction with this view from a number of Japanese women in attendance. They disputed then the fact that gender should still be considered a factor when determining the use of Japanese honorifics. Their position was in fact not unlike Mizutani and Mizutani's who have remarked, on the subject of change in Japanese honorifics, that the "gender differences in language usage have been minimized" (1987: 1-2). Hori even argues that communicative strategies employed by women have less to do with their gender *per se* than the role they play in Japanese society, meaning the "social networks" and the "social motivations" behind their language choices (1986: 374). By Hori's understanding, "if the same role-relations and responsibilities were given [to men and women],

language use in Japanese society would not greatly differ between sexes" (1986: 385). Roles still assigned, however, to men and women, and their corresponding use of honorifics, "in the model conversations in the textbooks largely conform to traditional gender norms" remark Okamoto and Siegal in their survey conducted on seven popular textbooks used in Japanese language classrooms in the United States (2003: 51). Given that its corpus is scripted speech by a writer, the present study will not be able to address the diversity and change in gender roles and relations in contemporary Japan, but it shall at least give us an idea of what role gender might have played in the 1950's in regard to Japanese polite speech.

2.3 Field of Study

For Leech, politeness is both a linguistic and a social/cultural phenomenon and can therefore be analyzed within two subdomains called "pragmalinguistics" and "sociopragmatics". The first is "oriented to linguistic realizations of politeness, and the [second] oriented to the social or cultural determinants of politeness" (2014: 13). Since the current study aims at evaluating the relevance of sociological factors in regard to Japanese politeness marking, it will concentrate on sociopragmatics. As it will however not address all possible contextual factors (e.g. intimacy, formality, etc.), and only consider the selected few as *constraining* linguistic forms, it will not rule on what Leech (1983: 11) and later Suzuki (2007: 106-107) refer to as the "appropriateness" of politeness marking in a given utterance. Judging on appropriateness involves positioning a given utterance on a "relative politeness" bipolar scale that ranges from "not polite enough" to "too polite", the accepted and expected politeness marking, in accordance with a given situation, being situated partway between these two extremes. That is also the reason why no analysis will be made on the "illocutionary functions" played by politeness markers (Suzuki: 2007: 79). In a volitional model, politeness can be interpreted as a tool used, for example, to implement speakers' collaborative or conflictive strategies in regard to their addressee. This would require a case-by-case analysis of every speech act expressed in the corpus, and thus far exceed the scope set for the current study. In short, context will only be deemed relevant as long as pertaining to the three sociological factors presented in section 2.2 and propositional content will be disregarded in favor of individual Japanese politeness markers traditionally associated with *keigo*. In that sense, the data selected for the study will strictly deal with what Leech used to call "absolute politeness"



3. Corpus

3.1 Movie Plot and Main Characters

Tokyo Story (東京物語, *Tokyo Monogatari*) was directed by Yasujiro Ozu and released in 1953. The story follows an aging couple, Shukichi and Tomi Hirayama, who travel from their hometown of Onomichi to Tokyo to visit their adult children. However, their children are too busy with their own lives to spend much time with their parents. The couple is eventually forced to return home, but their journey is marked by a series of emotional encounters

and poignant moments. The movie explores themes of intergenerational conflict, family dynamics, and the changing values of postwar Japan. Names and ages of each family member are shown in figure 1 below. All family members have professions associated with middle or upper-middle class standing. The eldest son, Koichi, is a physician who runs a small clinic in Tokyo's suburbs. His wife, Fumiko, is a stay-at-home mother. The Hirayamas' eldest daughter, Shige, runs a hairdressing salon and her husband, Kurazo, is a white-collar worker (サラリーマン, Japanese *salary-man*), and so are their youngest son, Keizo, and their widowed daughter-in-law, Noriko, who was married to their middle son, Shoji. He went missing in action during the Pacific War and is now presumed dead. Shukichi and Tomi live with their youngest and unmarried daughter, Kyoko, who is a primary school teacher.

3.2 Choice of Corpus

Several practical reasons presided over the choice of this corpus. First, the screenplay was readily available for a modest price (990 yen, about $10) which meant that the dialogue of that movie did not need to be painstakingly transcribed prior to study. Second, the script was written by the director and his long-time collaborator Kogo Noda who included, at the first appearance of each character in the plot, his or her name, age and occupation. Having thus exact numbers to draw upon enabled us to establish a precise age difference between any two protagonists engaged in conversation. Occupation was also at first thought to be a useful indicator of the disparity in status that might exist between characters but was later disregarded as irrelevant. Firstly, because all movie characters more or less belong to the same social class, and secondly, because, as mentioned earlier, the vast majority of interactions are confined to a household belonging to one or another member of the Hirayama family, that is to say in an informal environment that supersedes traditional status-related hierarchy. Third, as shown in the table 1 below, about half of the lines written in the script were delivered by or to a family member related by marriage (i.e. not a blood relative per se). Screenplays of other movies, and even plays, were also considered for this study, but *Tokyo Story* was eventually selected because of this specific feature. Our reasoning was that it might yield some new and interesting insights on Japanese politeness marking since interactions between Japanese in-laws had, to our knowledge, never been the subject of a quantitative study. Lastly, and perhaps more importantly, as shown in table 1 below, the corpus had also the additional merit of presenting a fairly balanced characters' male per female ratio.

This corpus is not however without flaws. We will presently address three of them. If the script is a little bit more than 1150-line long, it only covers the speech of a dozen individuals or so, which is understandably not ideal for a quantitative study. It is also very dated: it shows at best how Japanese people *used* to speak to members of their family seventy years ago. Critics could also point out that a movie script hardly qualifies as authentic discourse (i.e. attested language practice), since it is always "fabricated", as one might say, by a screenplay writer. In short, it would be hazardous to draw conclusions regarding present Japanese politeness marking on that basis alone. We acknowledge that our chosen corpus is not a perfect and accurate rendition of actual, or even past, Japanese speech. As it is however designed to foster our understanding of linguistic politeness, we do not see the authors' one-sided creating process, or the fact that they were Japanese men in their late forties and fifties, as necessarily biased or misleading. In our opinion, it is still compatible with the prescriptive approach generally associated with discourses given on the subject of politeness, where representations and ideologies do in fact guide or control language use, at least in part. For example, according to Okamoto, "recognizing linguistic ideologies in the use of honorifics can also account for their non-reciproqual use as well as the variations in their non-reciproqual use" because "the complexity of indexical processes in which the relationship between social context and forms of speaking is construed through the filter of one's belief about language use" (1997: 814).

Ideology might not then be a determining factor of the same prominence as age or gender, but it clearly influences how much weight these factors will carry in a given situation.

This choice was therefore in line with the epistemological positioning adopted for this study (section 2). And finally, even if this corpus is admittedly just an example of how people should have use *keigo* seventy years ago, its analysis has nevertheless borne novel and surprising results that, we feel, could still give some indication on current verbal manifestations of Japanese politeness.

Table 1. Number of Utterances, gender of speakers and hearers, in-law relationship

	Utterances		
	Male hearer	Female hearer	Total
Male speaker	229	306	535
Female speaker	315	308	623
Total	544	614	1158

4. Token Selection

4.1 Keigo *Markers*

For this study, only Japanese politeness markers of purely morpho-syntactic and lexical natures where taken into consideration, that is to say: verbs, adverbs, adjectives and affixes, together with forms of address or other formulaic expressions that are commonly listed in Japanese grammar books as belonging to expressions of verbal politeness. Table 2 below lists examples of markers selected in our corpus. Each marker has been assigned a value ranking from 1 to 3 depending on the degree of politeness it conveys, in accordance with the work of the two Japanese grammarians used in this study: Shimamori (2001) and Kuwae (1999). Japanese language is comprised of three registers: a "neutral register", reserved for family members and intimates, a "polite register" and "deferent register" (Shimamori, 2001: 305). Addressee honorifics account for the last two registers and were assigned the corresponding grades 1 and 2. As for referent honorifics, their assigned grades were based on their inherent syntactic complexity. Passive voice and other, more conventional, honorific constructions (e.g. *o-V ni naru*, *o-V nasaru*, *o-V suru*, etc.) and lexical items (e.g. *mieru*, 見える, *itadaku*, 頂く, formulaic expressions standing respectively for "to come" and "to receive") were therefore assigned the default grade 1, eventually raised to 2 or 3 when adjustments needed to be made to account for indirection (e.g. *V-(sa)sete itadaku*, (さ)せて頂く, literally: "to humbly be let to do something") and/or a more prestigious register.

This grade system was meant to account for word tokens that carried an exceptionally higher reverential value, or when they encompassed more than one category of *keigo*. For instance, the honorific title *sama* (さま) was assigned a grade of 2 whereas the more common suffix *san* (さん) was only assigned the default grade of 1. Another example is the morphologic construction *go-annaishimasu* (ご案内します) that can be translated in English as "Follow me" or "I will lead the way". It is a compound composed of the prefix *go*, that makes the intended action (i.e. guiding) belong to the *sonkeigo* category, and the verbal ending *masu* which belongs to *teineigo*. This word token was therefore assigned a grade of 2. It is indeed well-known that different categories of *keigo* are very often combined in speech and that a given utterance can bear several such markers (Okamoto, 1997: 811). This grading device was indeed meant to reproduced as faithfully as possible the degree of reverence adopted by the individual speaker. It was implemented as consistently as possible during data collection and, given the length of the corpus and reasonable allowance made for error, we are confident that our grading choices could not be as questionable as to impair the validity of the ensuing statistical analysis. Not to mention that, as will be exposed in the following section, this device also converted elegantly into a quantitative variable.

Table 2. Examples of keigo markers by categories

Word token	Keigo marker syntactic category	Grade
o-bento "lunch box"	prefix *o*	1
o-machi desu "Wait, please."	prefix *o* + verb *desu*	2 (1+1)
o-kaa-san "Mother"	prefix *o* + title *san*	2 (1+1)
o-kaa-sama "Mother"	prefix *o* + title *sama*	3 (1+2)
go-annaishimasu "Follow me"	prefix *go* + verb ending *masu*	2 (1+1)
o-hima itadakimasu "I'll leave you alone."	prefix *o* + verb *itadaku* + verb ending *masu*	3 (1+1+1)
Kochira "Me" or "I"	pronoun *kochira*	1

For a more comprehensive, yet non-exhaustive, list of markers, see Appendix A.

4.2. Other Politeness Markers

Note that variables chosen for the statistical analysis did not reflect the three categories of *keigo* commonly described in Japanese grammar and reference books. *Keigo* is actually made of *teineigo* (丁寧語), *jokengo* (謙譲語) and *sonkeigo* (尊敬語) which respectively translate in English as: polite, humble and respectful languages. The former category covers "addressee honorifics" which are used regardless of the people involved in the

interaction, whereas the latter two refer to "referent honorifics" which are used toward the person (i.e. the speaker, the hearer or a third person associated to one or the other) one is talking about. In Brown and Levinson's model, "addressee honorifics" characterize the interactants' relative positions on the horizontal axis (i.e. the socio-affective distance that separate them: "Distance"), and "referent honorifics" their position on the vertical axis (i.e. "Power"). As already stated in section 2 however, we only chose to consider what markers strictly fell into the range of "absolute politeness", regardless of the politeness strategies adopted by speakers. Differentiating between the three categories of *keigo* would have required a more qualitative approach and a case-by-case analysis of all occurrences of politeness, taken in the socio-pragmatic sense of the word. Since the choice made by the speaker of one or another category of *keigo* is necessarily based on context, it is clearly contingent to his or her politeness strategies. *Why is the speaker humbling him- or herself? How is it necessary for him or her to praise the person he or she is talking to?* etc. are examples of questions that would have had to be answered to be consistent in terms of methodology. It should also be noted that, in effect, referent honorifics are almost always used conjointly with addressee honorifics, since the vertical distance characterized by the first kind of honorifics necessarily implies that a horizontal distance be marked by the second kind of honorifics as well (Shimamori, 2001: 320). After all, acknowledged hierarchy seldom coincides with intimacy. For the same reasons, we did not select as markers "camaraderie building devices", such as jokes or playful banter (Leech, 2014: 165) that, too, can contribute to shortening the distance separating the speaker form the hearer. We also overlooked expressions that are exclusively syntax-based (i.e. not purely morphological or lexical), such as negative questions or any other "indirect" phrasings that are universally associated with politeness (Manno, 2002: 14). Similarly, analyzing propositional contents, speech acts, "illocutionary force indicating devices" (Leech, 2014: 58) and so forth would have far exceeded the scope we set for this study (Note 1).

4.3 Variables Selection and Speakers' Pairings

4.3.1 Variables Selection for Stage 1

To account for interpersonal dynamics that govern the use of *keigo* within the Japanese family, three characterizing social features were considered: age, gender and possible in-law relationship. The last feature logically provided us with a binary variable that we dubbed **inlaw**. The gender feature (i.e. male vs. female) was coupled with an interactional component (i.e. speaker vs. hearer) to produce four binary variables: male speaker (hereafter referred to as **sm** for "speaker is male"), female speaker (**sf**), male hearer (**hm**) and female hearer (**hf**)[2]. The age feature constituted two other variables: one binary variable to account for the fact that a given speaker could either be younger or older than the hearer: **sy** for "speaker is younger" than the hearer, and a continuous variable **agediff** that was the numeric value of the age of the speaker subtracted by the age of the hearer. The last variable showed therefore a negative number whenever **sy** equaled 1. The mere use of **keigo** was also differentiated from its intensity – deriving from the grading system explained in the section above – and respectively translated into a binary variable and a continuous variable named **keigo** and **markers**. In our corpus, the variable markers ranged between 0 and 8. It equaled zero whenever **keigo** equaled zero, but had potentially no limit because it was the sum of all the values attributed to *keigo* markers in a given utterance.

4.3.2 Variables Selection for Stage 2 and 3

Since *keigo*, depending on its category, can either convey respect, humbleness, or politeness, all Japanese speakers, male or female, young or old, resort to it at some point in their interactions with others. And what is more, honorifics can even be present in the speech of both interactants. This might be caused in some cases by the formality of the exchange, since, according to O'Neill, formality often "causes a reciprocal display of respect language" (O'Neill, 1981: 3). More generally, Ide explains that "reciprocity is observed when rules come to conflict". By rules, the author means the three following factors, ranked in order of decreasing importance: "dominance in social position, power, and age" (1982: 369). That is why, for example, a doctor and his older patient might both use honorifics when talking to each other. If gender is understood as providing some speakers with a higher social position (i.e. males), then the comparative use of *keigo* between genders and age groups can only be drawn *in regard to* the age and gender of the speaker and the age and gender of the hearer. That is why, in the second stage of our analysis, the variables listed above that pertained to individual speakers (**sm**, **sf**, **hm** and **hf**) (Note 2) were further combined to generate four other binary variables: **smhf**, **smhm**, **sfhf** and **sfhm**. These accounted for different types of utterance, each characterized by a specific type of gender-based speaker-hearer pairing. For example, **smhf** referred to a line spoken by a male to a female addressee. These pairings also composed the four possible categories of a nominal variable: **genderdyad**. By introducing these new sets of predictor variables, we were able to confirm and further analyze some the findings brought out during the first stage of our study. Stage 3 necessitated similarly that we further subdivided these four dyads into the following eight: **smhfo**, **smhfy**, **smhmo**, **smhmy**, **sfhfo**, **sfhfyo**, **sfhmo** and **sfhmy**. The letters "o" and "y" tagged to the name of the

gender-based variables stood for the comparative adjective “older” or “younger” which referred then to the second member of the pairing. In this later model, **smhfo** indicated for instance a line told by a male speaker to an older female hearer. These last pairings composed the eight possible categories of a nominal variable: **agegenderdyad**.

As a rule, whenever the context of the movie could not provide a distinct interlocutor for a given speaker, when characters spoke to their family as group for instance, the hearer consistently corresponded, for the purpose of this study, to the eldest person in attendance. Our assumption was that he or she embodied by default the most “distinguished”, in Japanese: *erai* (偉い), character present at the scene. This decision was made in accordance with our working hypothesis which posited age as the primary social feature to consider when analyzing the use of *keigo*, and with what Brown and Levinson referred to as “situational ‘formality’” caused by the presence of an audience (1987: 16).

5. Data and Analysis

Data collection and ensuing statistical analysis were conducted throughout the study using the open-source software R. Student’s T tests, Pearson correlation coefficients and logistic models were used for binary outcomes, while single and linear regressions were used to assess correlations involving one or several continuous variables. As a programming language, R also enables its users to generate a wide variety of graphics, such as scatter and bar plots found in this study.

5.1 Stage 1

5.1.1 Data

Throughout all three stages of our analysis, *keigo* use (**keigo**) and the number of *keigo* markers (**markers**) will be systematically presented in regard to sample units, or utterances, which is to say: lines spoken by movie characters as written in the script. The number of words per occurrence of *keigo*, or per *keigo* marker, within a given utterance was not selected as an accurate benchmark for presenting data for two reasons. First, because polite language in Japanese almost always results, due the various formulaic expressions, affixes and verbal compounds mentioned in section 4, in the lengthening and/or syntactic complexification of utterances. Note that Japanese shows in that respect a very common trend, seeing that, to ensure that no one loses face, polite phrasing is “superimposed on the raw form of the statement” (i.e. pure and simple transmission of information) and results in an additional cognitive cost for both speaker and hearer (Manno, 2002: 9). Second, because only 76 out of the 1158 utterances (about 6%) of the corpus went longer than a line on a .doc file. This mode of display was therefore deemed sufficiently representative of the characters’ use of *keigo* not to warrant any further adjustment.

Table 3. Binary predictor variables for stage 1, number of utterances, number of *keigo* use and average *keigo* markers’ value per utterance.

Predictor variables	Utterances	*Keigo*	Markers	Markers per utterance
sy	553	323	680	1.23
so	582	168	307	0.53
sm	535	177	326	0.61
sf	623	321	670	1.08
hm	544	289	557	1.02
hf	614	209	437	0.71
inlaw	725	308	585	0.81

At first glance, data shown in the far-right column in Table 3 above seem to match our expectations regarding *keigo* use in regard to speakers’ age and gender. Indeed, the speech of younger Japanese shows more than twice as many *keigo* markers as older Japanese. Similarly, female speakers use almost twice as many markers as male speakers and, in reverse, utterances directed at female hearers bare 30% less markers than utterances delivered to male hearers. At this point however, it seems hazardous to comment on the number of markers per utterance that can be attributed to the variable **inlaw**.

5.1.2 Analysis

Figures in table 4 below show that all predictor variables but **inlaw** are strongly correlated to the dependent variable **keigo**, thus confirming observations made in the subsection above. Given the correlation coefficients of the predictor variables for which statistical significance was established, we can moreover hypothesize that age, as a sociological feature, has a slightly stronger impact than gender on the speakers’ use of *keigo*, since **sy** shows a 0.30

correlation coefficient and **sm**, **sf**, **hm** and **hf** only a 0.19 correlation coefficient. Bearing in mind that coefficients whose magnitude ranges between 0.3 and 0.5 indicate a rather low correlation, it is worth noting that their positive or negative value are also in accordance with our working hypothesis. Note that histograms of residuals and intercorrelation testing results for all simple and multiple linear models conducted in the study are available in appendixes B and C.

Table 4. Student's T test results and Pearson correlation coefficients for binary predictor variables, Stage 1 (dependent variable: **keigo**)

	sy	**sm**	**sf**	**hm**	**hf**	**inlaw**
p	2.20E-16	2.63E-10	2.63E-10	5.86E-11	5.86E-11	0.6422
r	0.30	-0.19	0.19	0.19	-0.19	-0.01

Table 5 below shows the results of the linear model analysis conducted with the same five binary predictor variables and another variable (**agediff**) to account for the dependent variable **markers**. Note that even if, in this linear model, these last two variables are continuous, the correlation between **markers** and all predictor variables but **inlaw** is just as significant as with **keigo**. And since all Japanese speakers resort to *keigo* at some point in their interaction with others, **markers**, as a continuous variable, gives a more faithful and precise account of *keigo* usage. That is why, for the remainder of the article, this variable will be preferred to **keigo** whenever possible.

Table 5. Linear model results, Stage 1 (dependent variable: **markers**)

	sy	**sm**	**sf**	**hm**	**hf**	**agediff**	**inlaw**
p	2.00E-16	1.24E-09	1.24E-09	6.01E-05	6.01E-05	2.00E-16	0.074
Estimate	0.702	-0.466	0.466	0.309	-0.309	-0.014	-0.142

For instance, the *Estimate* value for **sy** indicates that, all else being equal, the speech of a younger Japanese will feature 0.702 more *keigo* markers than an older Japanese'. In other terms, a younger speaker will use in average three more *keigo* markers every two utterances. This makes for a considerable difference with the speech of an older Japanese since the average number of markers per utterance ranges between 0.53 and 1.23 (See table 3.). The output given for **agediff** however means that a speaker will use 0.014 more *keigo* markers for each year that he or she is younger than the hearer (i.e. *Estimate* value for **agediff** is negative). This means that the age difference has to equal around 70 years for the speech of a younger Japanese to feature on average one more *keigo* marker. We hypothesize that such a disparity is to be imputed to the fact that Japanese speakers assess the age difference separating them from their hearers in absolute terms. The question of *seniority* therefore, rather than the age difference numeric value, seems to be the prime factor to consider when evaluating the use of *keigo*, as shown by the almost horizontal regression line in figure 2 below. For the remainder of the article, we will therefore no longer consider **agediff** but only **sy** as the predictor variable referring to age.

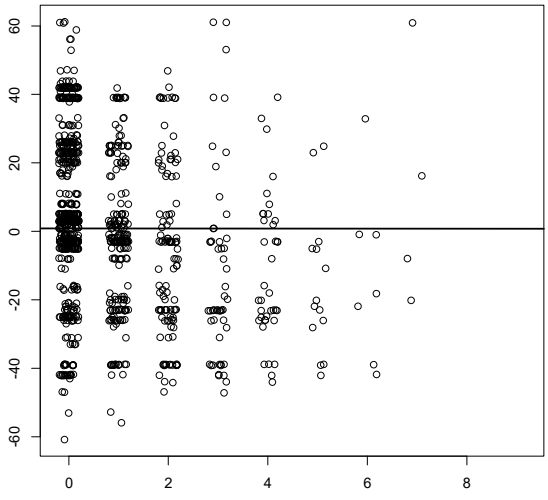


Figure 2. X axis: number of *keigo* markers per utterance. Y axis: age difference between speaker and hearer

On the other hand, the five other binary variables show *Estimate* values in accordance with remarks made earlier in this study. All else being equal for example, the speech of a Japanese male will feature 0.47 less *keigo* markers than a female.

As expected, there is also no statistically significant correlation to be made between the variables **inlaw** and **markers** (p = 0.07). As this was confirmed by analysis conducted in stages 2 and 3, results concerning **inlaw** will be no longer discussed in the following subsections. In truth, this undeniable lack of correlation between **inlaw** and our dependent variables, be they binary or continuous (i.e. **keigo** and **markers**) was one of the most surprising findings of the current study. It contradicted our intuitive assumption that in-law relationship would trigger more polite forms of address and therefore induce, in the lesser of the two protagonists' speech at least, a more frequent use of *keigo*. By "the lesser of the two protagonists" we mean the speaker who, by age and/or gender, is positioned lower on the social scale than his or her hearer: either because the speaker is a female addressing a male and/or is younger than the hearer. We will not delve here on how these particular results affect our working hypothesis. Variables for which statistical significance cannot be demonstrated will be addressed collectively at the end of this section. Note however that, if some languages do allocate specific honorifics for in-laws to address each other (Note 3), this does not seem to be the case for Japanese. *Keigo* usage might therefore depend on the individual's personal rapport with his or her in-laws, whether they are seen as "regular" family members, or as people somewhat closer than acquaintances but not intimate friends either.

To further investigate the gender-age dynamic at play between Japanese speakers and hearers, we will now present the results of four multiple linear regression models, each model accounting for one of the four possible gender-based pairings, or dyads. Predictor variables were consequently considered in sets of three, one referring to age (**sy**), one to the gender of the speaker (**sm**, **sf**) and one to the gender of the hearer (**hm**, **hf**). The results are shown in Table 6 below.

Table 6. Multiple Linear models' results, Stage 1 (dependent variable: **markers**)

	sy	**sm**	**hm**	**sy**	**sm**	**hf**
p	5.52E-12	4.10E-04	2.40E-01	5.52E-12	4.10E-04	2.40E-01
Estimate	0.584	-0.281	0.095	0.584	-0.281	-0.095

	sy	**sf**	**hm**	**sy**	**sf**	**hf**
p	5.52E-12	4.10E-04	2.40E-01	5.52E-12	4.10E-04	2.40E-01
Estimate	0.584	0.281	-0.095	0.584	0.281	0.095

However, Table 6 above consistently showed hearer gender variables (**hm**, **hf**) that were not statistically significant enough (p-values = 0.24) to make these models additive. To determine precisely the effect of predictor variables that did appear to be statistically significant – seniority (**sy**) and speaker gender (**sm**, **sf**) – first independently from each other, and later their joint effect, on the dependent variable (**markers**), another, more parsimonious, model was designed. This new multiple linear model excluded hearer gender variables. Its results appear in Table 7 below. Intercorrelation tests results for this linear model appear in Appendix C.

Table 7. Parsimonious Multiple Linear models' results, Stage 1 (dependent variable: **markers**)

	sy	**sm**	**sy**	**sf**
p	1.07E-14	4.32E-04	1.07E-14	4.32E-04
Estimate	0.618	-0.279	0.618	0.279

All p-values being well under the 5% threshold, the joint effect of seniority and speaker gender on the average number of *keigo* markers in a given utterance could be easily calculated. As expected, a younger male speaker will use in average about 0.34 marker (0.618 minus 0.279), while a younger female speaker nearly 1 marker, that is to say three times as much. These results are moreover consistent with Suzuki's findings in regard to the average scores attained for absolute politeness by the four age-gender groups considered in his study: JHAF >JYAF>JYAM>JHAM (i.e. Japanese Higher Age group Female, Japanese Younger Age group Female, etc.) (2007: 140-141) (Note 4).

What, then, of hearer gender variables (**hm**, **hf**)? Considered individually, they did not seem to be correlated to the dependent variable **markers**. We felt however intuitively that Japanese speakers did take into account the gender

of their addressee when deciding on their use of *keigo*. We surmised that, to better understand the inter- and intra-gender dynamics at play, these variables needed to be coupled with speaker gender variables into the four possible types of gender-based speaker-hearer pairings. That was the object of the second stage of our research.

5.2 Stage 2

5.2.1 Data

Tables 8 and 9 below show data collected on stage 2.

Table 8. Number of Utterances, *Keigo* use and number of Markers for each gender-based dyad

	Utterances	*Keigo*	Markers
smhf	306	62	133
smhm	229	114	193
sfhf	308	147	326
sfhm	315	174	355

Table 9. Ratio of *Keigo* use and number of Markers per Utterance and average number of Markers per use of *Keigo* for each gender-based dyad

	Keigo/U	Markers/U	Markers/*Keigo*
smhf	0.20	0.43	2.15
smhm	0.50	0.84	1.69
sfhf	0.48	1.06	2.22
sfhm	0.55	1.13	2.04

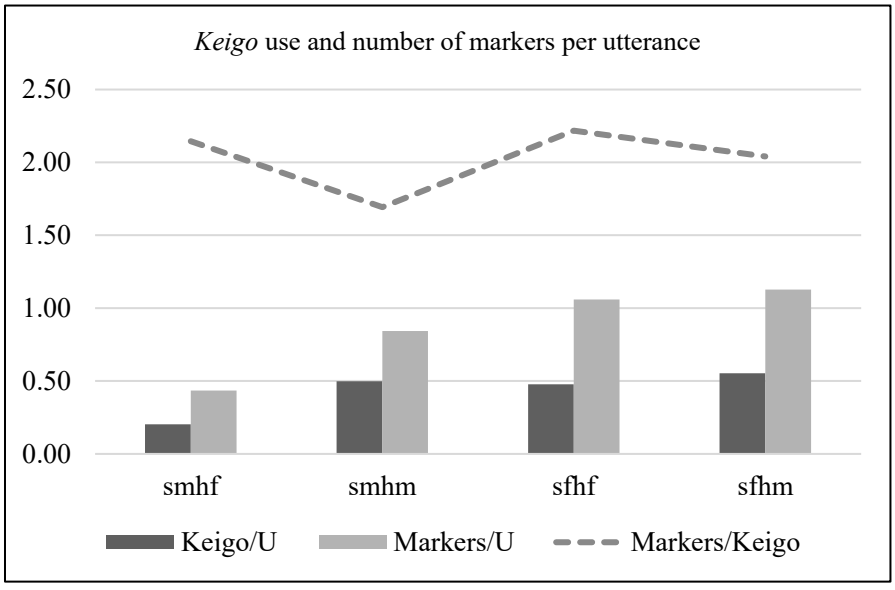


Figure 3. Ratio of *Keigo* use and average number of Markers per Utterance for each gender-based dyad

As shown in Table 9 and figure 3, the highest disparity in keigo usage is found between mixed-gender dyads: **smhf** and **sfhm**. The fact that these two types of pairings occupy opposite ends of the spectrum in terms of keigo is in accordance with our working hypothesis and the linguistic literature on the subject. Indeed, Japanese female speech addressed to a male (**sfhm**) shows Keigo/U and Markers/U values that are more than twice superior to Japanese male speech addressed to a female (**smhf**). In same-gender dyads however (**smhm** and **sfhf**), it is worth noting that the difference seems to lie, not in the likelihood to resort to keigo per se (Keigo/U shows very similar values), but rather in the degree in which keigo use will be expressed, that is: the average number of markers per use of keigo. If confirmed by statistical tests in the following sub-section, that would indicate a very specific male strategy in regard to *keigo*. When speaking to another male (possibly older than they are), Japanese male speakers would be just as likely as Japanese female speakers to acknowledge their addressee's dominance by resorting to *keigo*,

but never to the extent demonstrated by Japanese female speakers in terms of intensity (i.e. the number of *keigo* markers used for a given *keigo* marked utterance). We are of the opinion that this consistent, yet diminutive, use of *keigo* amongst Japanese male speakers could be interpreted as a verbal token of submission that does not entirely debase its user in the eye of his addressee.

We are aware, however, that this discrepancy might also be imputed to other factors not considered in our model, such as certain categories of addressee. In Hori's work for example, while "men choose forms higher in honorifics in certain categories, such as 'neighbor' and 'sibling'", "women exceed men in the choice of honorifics only in regard to 'spouse' and 'friend'" (1986: 382). Within our framework however, categories of hearers notwithstanding, only the age factor can give us a more detailed view on this issue.

5.2.2 Analysis

Table 10 below shows the results of the linear model analysis conducted to account for the dependent variable **markers** with six binary predictor variables, including variables corresponding to the four gender-based dyads that have been discussed above, and the nominal variable **genderdyad** (Note 5).

Table 10. Linear model results, Stage 2 (dependent variable: **markers**)

	sy	**inlaw**	**smhf**	**smhm**	**sfhf**	**sfhm**	**genderdyad**
p	2.00E-16	0.074	2.44E-11	0.824	0.0371	2.52E-06	2.34E-11
Estimate	-0.014	-0.142	-0.578	-0.023	0.994	0.406	NA

Table 11 below shows the results of the logit model analysis conducted to account for the dependent variable **keigo** with the same binary predictor variables.

Table 11. Logit model results, Stage 2 (dependent variable: **keigo**)

	sy	**inlaw**	**smhf**	**smhm**	**sfhf**	**sfhm**
p	2.00E-16	0.642	2.00E-16	0.014	0.051	3.39E-07
Estimate	1.241	-0.057	-1.417	0.363	0.261	1.241

Setting aside **inlaw**, already known to be non-statistically significant in this study, these results confirm what was hypothesized regarding mixed-gender dyads, since **smhf** and **sfhm** both show P-values, in the linear as well as in the logit model, that are well under the 5% error margin usually allowed for these tests. Perhaps more surprising are the P-values obtained for same-gender dyads: **smhm** and **sfhf**. The former is above the 5% threshold in the linear model but below it in the logit model, whereas we find the exact opposite results for the latter. In our opinion however, P-values results shown in tables 10 and 11 regarding these same-gender dyads need not necessarily be interpreted as contradictory. We posit that these inconsistent results are to be imputed to the nature of the dependent variable. If, as we hypothesized earlier, Japanese male-to-male speech can indeed be best characterized by the question of resorting to *keigo* or not, rather than its intensity (i.e. number of *keigo* markers per utterance), then it stands to reason that only the *binary* dependent variable **keigo** would yield statistically significant results, as it did in the logit model. Similarly, we suggest that the same reasoning be applied to Japanese female-to-female speech but in reverse. Since, as stated in section 2, Japanese females are a lot more likely to resort to *keigo* than their male counterparts, we should not expect the *binary* dependent variable **keigo** to help us differentiate their speech in terms of politeness. On the contrary, only the *continuous* variable **markers** can provide us with an appropriate basis for comparison. In any case, whether or not we are correct in our interpretation, it should be noted that the nominal variable **genderdyad**, of which **smhm** is only one of the possible values, is undoubtedly very strongly correlated to **markers** (P value = 2.34E-11) in the linear model.

A word of caution regarding estimate values shown in tables 10 and 11: if they are, for the sake of clarity, recorded here together in the same tables, they cannot be simply compared two by two as each column pertains to a simple linear regression analysis conducted separately for each individual predictor variable. In fact, except for **sy** and **inlaw**, the other four binary variables refer to gender-based combinations that are mutually exclusive. *Estimate* values shown in table 10 only indicate the average number of markers per utterance for all utterances positively defined by a given binary predictor variable, in comparison to the average number of markers per utterance for all utterances where the same variable does not apply. For instance, **smhf** shows an *Estimate* value of -0.578 in the

linear model but that only means that, when addressing a woman a Japanese male speaker will use, in average, one less *keigo* marker every two utterances than a Japanese male speaker addressing another man (**smhm**), a Japanese female speaker addressing another woman (**sfhf**) or a Japanese female speaker addressing a Japanese man (**sfhm**), which is to say: all the other gender-based dyads considered as a whole… Perhaps more instructive, especially in light of our working hypothesis, is the fact that *Estimate* values can either be positive or negative. *Estimate* values for **smhf** (-0.578) and **sfhm** (0.406), for instance, would indicate that Japanese women are expected to use, regardless of their respective age, a lot more *keigo* when addressing men, than men are when addressing women.

As for *Estimate* values recorded in Table 11, it should be pointed out that, being generated by a logit model, they can only be used to calculate the corresponding odd's ratio for each variable. They cannot therefore be matched to estimate values in table 9 for comparison.

To assess the effect of age disparity between speaker and hearer within same and mixed gender pairings, multiple linear models that also comprised **sy** were then considered. Table 12 below however shows inconclusive P-values for **smhm** and **sfhm**, meaning that the increasing effect of hearer seniority could only be accurately measured for only two gender-based dyads (**smhf** and **sfhf**). As indeed the last two variables made for additive models, estimate values for **sy** (0.58805 and 0.70983) could be added to estimate values for **smhf** (-0.36109) and **sfhf** (0.21485). An older Japanese female will therefore be, on average, addressed with four times more *keigo* markers by another female speaker than by a man. If, when the addressee is an older woman, speaker's gender can undoubtedly have, as expected, an increasing or a decreasing effect on *keigo* usage, the absence of statistically significant results for **smhm** and **sfhm** however makes it impossible to measure, with any certainty, the joint effect of hearer seniority and speaker gender when a man is being addressed, probably because of the ambivalent nature of Japanese male-to-male speech in regard to *keigo*.

Table 12. Multiple linear model results, Stage 2 (dependent variable: **markers**)

	sy	**smhf**	**sy**	**smhm**
p	4.75E-13	7.17E-05	2.00E-16	9.43E-01
Estimate	0.58807	-0.36109	0.702267	-0.006877

	sy	**sfhf**	**sy**	**sfhm**
p	2.00E-16	1.18E-02	1.35E-15	1.81E-01
Estimate	0.70983	0.21485	0.66103	0.12214

5.3 Stage 3

5.3.1 Data

The third and last stage of the current study aimed to account for hearer's or speaker's seniority conjointly with the four existing variables derived from gender-based dyads. As stated in section 4.3.2, eight new binary predictor variables were consequently designed as followed: **smhfy**, **smhmy**, **sfhmy**, **sfhfy**, **sfhmy**, **smhfo**, **smhmo**, **sfhmo** and **sfhfo**. These new variables were implemented to foster our understanding of borderline cases, that is to say, Japanese speakers' pairings where gender and age could possibly have opposite effects. Analyses conducted in stages 1 and 2 have shown that in mixed gender dyads, females were statistically more likely than males to resort to *keigo* when addressing a man, that they were conversely less likely to be shown linguistic deference by men, and that these trends could even be amplified by the seniority of their male counterpart; begging the questions: what of pairings involving an older female and a younger male: **smhfo** and **sfhmy**? Which of age and gender is going to prove a stronger factor in regard to keigo usage? Having experienced on a personal level the second of the two possible conflicting situations – when interacting some 20 years ago with middle-aged Japanese women as a young male speaker – I intuitively felt that age was, mutatis mutandis, the most defining factor of the two, unless the age difference between a female speaker and a male hearer amounted to at least 20 or 30 years, in other words, if speaker and hearer were born more than a generation apart. As a male university professor, I was indeed on more than one occasion spoken to in very colloquial terms by elderly female staff members (cafeteria staff, office workers, etc.) that were more than 25 years my senior and addressed then with very little if no *keigo*.

Tables 13 and figure 4 below show data collected on stage 3.

Table 13. Number of Utterances, Keigo use and number of Markers for each gender- and age-based dyad

	Utterances	*Keigo*	Markers
smhfy	246	37	72
smhfo	59	24	60
smhmy	105	39	61
smhmo	108	72	126
sfhfy	166	60	109
sfhfo	138	86	196
sfhmy	65	32	65
sfhmo	248	141	298

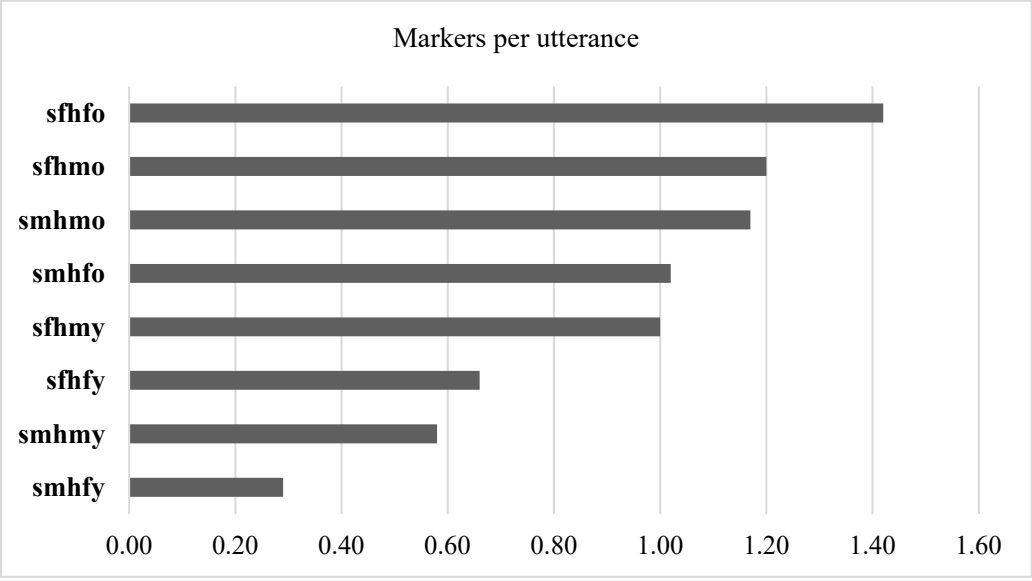


Figure 4. Average number of Markers per Utterance for each gender- and age-based dyad

Note that, for increased visibility, utterances and their corresponding number of *keigo* markers appear in figure 4 in descending order. This manner of display was chosen to better illustrate the following remarks.

At first glance, it would seem to corroborate findings made during stages 1 and 2. First because it nicely breaks down all eight types of utterances by seniority. The top half, those showing the highest average number of keigo markers, all involve an older addressee, and the bottom half, with the lowest number of keigo markers, all involve conversely a younger addressee. Second because, those two groups can further be divided into two sub-groups: the top two being consistently defined by their female speaker; confirming thus the higher social status enjoyed by older and/or male Japanese speakers as far as linguistic deference is concerned. It is worth noting however that the gender of the addressee does not seem to bear any particular impact on how utterances are ranked in terms of keigo use within these four sub-groups. We would have expected for instance **sfhmo** to be atop **sfhfo**, in accordance with our findings regarding male addressees. A possible explanation for this might be found in Smoreda and Licoppe's work on the duration of phone calls and the receiver's gender. Authors warn of a "receiver effect" that causes conversation to be lengthened when a woman receives a call from a man, but which does not apply when a woman calls a man. Here too, perhaps must we "consider not only gender differences in 'interpersonal styles' of interaction, but also gender-situated identities" (2000: 246). Just as with phone call duration, it is possible that speaker-hearer status might too be asymmetrical in regard to Japanese honorifics. Along with age and gender, an "addressee effect" (e.g. who is talking to whom?) might also be a determining factor that should not be treated individually but only as part of a whole, more interaction-based dynamic. The fact that, not **sfhmo** but **sfhfo** generates the highest use of honorifics might then only be the result of a particular combination off all three factors.

5.3.2 Analysis

Table 14 below show Chi-square test results and Pearson correlation coefficients for the eight predictor variables derived from age- and gender-based dyads in regard to the binary dependent variable **keigo**. Table 15 shows the

results for the linear model analysis conducted with same predictor variables but this time in regard to the continuous dependent variable **markers**. In both tables, predictor variables were assigned columns following the same order as in table 14 and figure 4 to facilitate analysis.

Table 14. Chi-square test results and Pearson's correlation coefficients for all binary predictor variables, Stage 3 (dependent variable: **keigo**)

	sfhfo	**sfhmo**	**smhmo**	**smhfo**	**sfhmy**	**sfhfy**	**smhmy**	**smhfy**	**agegenderdyad**
p	1.05E-06	6.71E-07	1.83E-07	0.711	0.297	5.37E-02	2.03E-01	2.20E-16	2.20E-16
r	0.143	0.146	0.153	-0.011	0.031	-0.057	-0.037	-0.293	NA

Table 15. Linear model results, Stage 3 (dependent variable: **markers**)

	sfhfo	**sfhmo**	**smhmo**	**smhfo**	**sfhmy**	**sfhfy**	**smhmy**	**smhfy**	**agegenderdayd**
p	7.56E-08	3.40E-06	1.07E-02	0.346	0.376	3.07E-02	2.21E-02	9.76E-15	2.20E-16
estimate	0.636	0.435	0.338	0.165	0.148	-0.238	-0.307	-0.720	NA

If the correlation between the nominal variable **agegenderdyad** and both **keigo** and **markers** is undoubtedly statistically significant (P value = 2.20E-16), thus proving the strength of our working hypothesis, a more enlightening perspective would be to consider all eight binary predictor variables as marking out eight points along what we would call a statistical significance continuum (See figure 5 below). This continuum would present predictor variables baring the strongest correlation to dependent variables at its opposite ends, and predictor variables baring no correlation at the center.

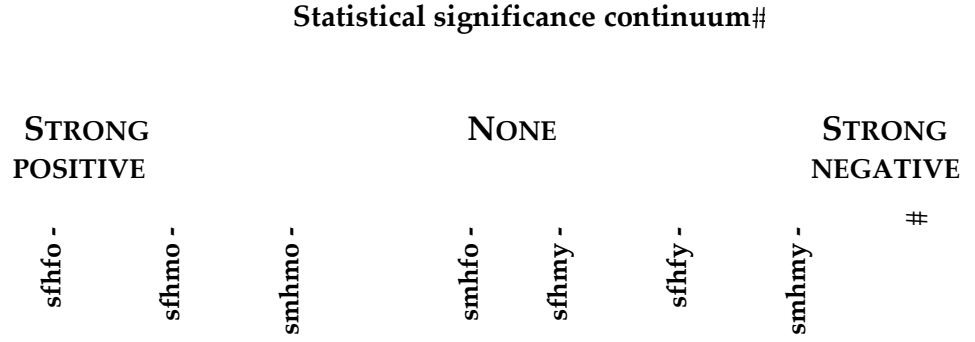


Figure 5. Statistical significance continuum for each gender- and age-based dyad

As shown in tables 14 and 15, on the left of said continuum, **sfhfo**, **sfhmo** and **smhmo** would stand out as variables which are both strongly and positively correlated to **keigo** and **markers**: their corresponding P-values are well under or very near the 1% mark in both tables and their Pearson correlation coefficients and Estimate values are positive, albeit decreasing the closer we get to the center of the spectrum. On the right, **smhfy**, **smhmy** and **sfhfy** could be seen as their mirror opposites, since their Pearson correlation coefficients and Estimate values make them more and more strongly, but negatively, correlated to **keigo** and **markers** as their P-values get closer and closer to 0. Except that we would then need to disregard P-values given for **sfhfy** and **smhmy** in table 14, as they are both well above the 5% threshold. Even if we were to adopt here a similar reasoning as the one chosen during stage 2 in regard to same-gender dyads, that would only explain why **sfhfy** is not correlated to **keigo**.

For the time being, we can only impute these surprising results to the "addressee effect" outlined in the preceding subsection and surmise that Japanese male-to-male speech operates in regard to keigo usage differently, depending on who, between the addresser and the addressee, holds seniority. Conflicting P-values for **smhmy** between tables 14 and 15 would tend to suggest that, when a Japanese male addresses a younger male, his speech is less likely to bear keigo markers than when addressing an older male, but that his resorting to the polite language keigo itself will not be affected one way or another by these particular conversation settings, since **smhmy** is only correlated to **markers**, not **keigo**. Note however that the correlation between **smhmo** and both **keigo** and **markers** is undeniable. A Japanese man is therefore a lot more likely to resort to *keigo*, and to use more *keigo* markers in his speech when addressing an older man. It is worth noting moreover that Pearson correlation coefficients, while

displaying a decreasing trend compatible with our analysis, only range between +0.15 and -0.3, which makes for very weak correlations between **keigo** and the eight predictor variables, regardless of P-values associated with them.

What is perhaps more remarkable is what tables 14 and 15 tell us about mixed-gender pairings involving a younger male and an older female. P values for **smhfo** and **sfhmy** are consistently over 25%, which place them at the center of our continuum. In accordance with our working hypothesis, we posit that this absence of correlation does not make these results inconclusive, but rather attests to the fact that these particular pairings are not governed by any set rules in regard to *keigo* usage. In our opinion, it is, as matters stand in the current study, impossible to know which of age and gender is going to prove a stronger factor in regard to *keigo* usage, because polite language is then left to the individual's choice. In linguistic terms, interpersonal dynamics would fall in both instances into the realm of idiolectal variation. This would mean that an older woman addressing a younger man (**sfhmy**) could therefore choose to rank his gender above her age, so to speak, and consequently make use of *keigo*, or, on the contrary, rank her age above his gender and forgo with *keigo*. Similarly, a younger man could decide, when addressing an older woman (**smhfo**), that his gender gives him a higher status than her and choose not to resort to *keigo*, regardless of her age. He could also want to pay respect to an elderly lady and adopt the opposite strategy. Note that, depending on the strategies preferred by the younger male and the older woman, *keigo* could possibly be used by both, one, or none of the interactants.

5.4 Gender and age constraints in regard to *keigo* markers

Figure 6 below compiles findings brought forth during stages 1 to 3 in accordance with the theoretical framework adopted for the study. The sequence in which age- and gender-based dyads are arranged along the continuum in figure 5 is reinterpreted and represented here as a decision flowchart based on interactants' respective sociological attributes.

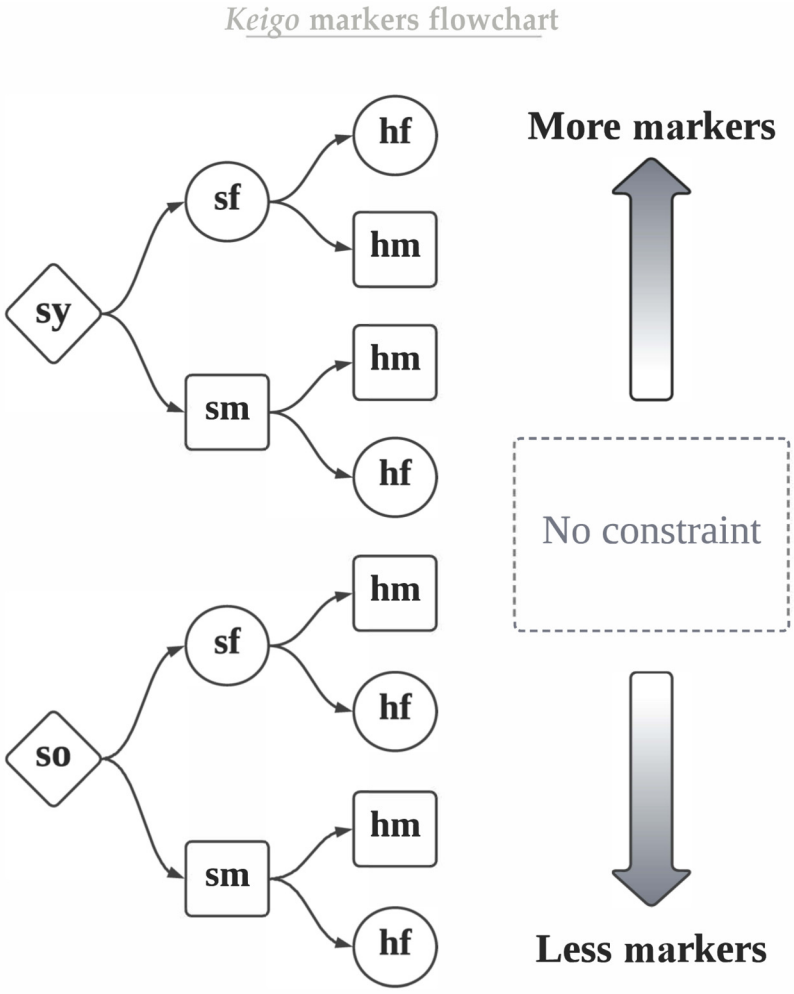


Figure 6. *Keigo* markers flowchart based on gender and age constraints

Items on the left of the flowchart shall therefore no longer be seen as predictor variables but rather interpreted as sociological factors governing *keigo* marking in Japanese speech. Since seniority was the obvious polarizing variable in the continuum, this model posits age (**sy**, **so**) as the first factor to be considered. The four sub-groups identified for age- and gender-based dyads in figure 4 (section 5.3.1) supplied us accordingly with the speaker's gender (**sf**, **sm**) and then the hearer's gender (**hf**, **hm**) as the next factors to be considered. Note that the speaker's age and gender are systematically confronted to the hearer's at each step of the decision-making process.

In other words, statistical analysis conducted throughout the study established correlations, or lack of thereof, between variables derived from gender- and age-based dyads, that are reinterpreted in the model as creating potential linguistic constraints. The rank of imposition indicated on the schematic's right side is presented as guiding the individual speaker's decision process outcome. We hasten to say however that this model is more to be construed as a linguist's artifact, than a faithful rendition of a Japanese speaker's decision process. Its purpose is to shed some light on *keigo* marking if understood as resulting from age and inter- or intra-gender dynamics that are specific to the Japanese speaking community. We intended for it to be mostly used as a tool by Japanese language instructors and their students. We consequently do not claim for it to hold any scientific value in the field of neurolinguistics for example.

6. Conclusion

The current study has yielded a number of promising, and sometimes even surprising results that will be briefly reiterated here before addressing some of its more concerning limitations, along with suggestions of possible leads for future research to remedy said limitations. Among its principal findings are undoubtedly the highly significant correlation between age, gender and the use of *keigo*. Simply put, women and younger speakers do tend to use more *keigo* markers than men and older speakers. Even if, given the existing literature, such results could have been easily expected, what we did not consider when devising our working hypothesis however, was how statistical analysis would foster our understanding of age, or rather, our understanding of how the age difference between speaker and hearer operates in regard to *keigo*. As a sociological factor indeed, its numeric value has proven of little importance compared to the single criterion of seniority. Japanese polite speech, it seems, does not reflect the order of magnitude derived from the age difference but only the acknowledgement of one interactant's superiority. In that respect, age is therefore tantamount to gender as it should only be analyzed in absolute and binary terms (Note 6). The former has moreover proven to hold more of a secondary importance than the latter. Results brought forth during stage 3 have indeed pointed out age as the main factor to be considered in regard to *keigo* use. Note that, personal communications with several Japanese native speakers and Japanese instructors would likewise suggest that this trend has probably heightened since the 1950's (i.e. when our corpus was originally written). Current Japanese speakers do not in fact consider gender as a relevant factor in regard to *keigo* use. If that be the case, it would in essence be very similar to in-law relationship since this variable (**inlaw**) has not, contrary to our assumptions, proven to be correlated to either **keigo** or **markers** either.

Differentiating between the mere use of *keigo* in discourse (**keigo**) and its intensity of use in actual speech (**markers**) has likewise brought forth mixed results that need not be altogether disregarded as inconclusive. For example, analyses with same-sex dyads have shown that men-to-men politeness (**smhm**) has more to do with the first variable and women-to-women politeness (**sfhf**) more with the second. As for cases where we might have expected age and gender to work conjointly toward a heightened (**sfhmo**), or on the contrary a lessened (**smhfy**), use of *keigo*, even though it has not always been possible to measure the respective weight of each variables in an additive model; or cases where idiolectal variation was the only possible interpretation (**smhfo**, **sfhmy**), they too had a welcome, albeit unforeseen, outcome. They indirectly contributed to our understanding of *keigo*. In essence, and contrary to our linguist's intuition, Japanese polite speech does not seem to methodically obey rules set by gender and age, taken as superseding sociological factors. Combined together with the "addressee effect", they define what we would call "categories" of styles of address: each gender- and age-based pairing being unique and its interpersonal dynamics entailing specific requirements or lack of thereof in regard to keigo. It therefore must be considered both as a whole and on its own. In that sense, figures 5 and 6 were only designed to help us visualize how all eight possible pairings fair in this complex system of address but do not claim to give a perfect and holistic rendition of Japanese age and inter- and intra-gender dynamics.

Finally, even if each consecutive stage of the current study has provided more precise variables to allow for a more refined analysis of the results yielded in a previous stage, we must acknowledge that this did little to address the numerous flaws already outlined for our chosen corpus in section 3.2. This corpus is regrettably very limited in its scope. It indeed only pertains to scripted polite speech within a single, and what is more, fictional Japanese family that lived in the 1950's. Future studies should therefore adopt more extensive corpuses, ideally made of attested utterances spoken in non-family settings, such as those that could be acquired from conversations between friends,

co-workers, costumers and store clerks, etc. that should also have the added advantage of being contemporary examples of present Japanese speech and therefore provide a more realistic and current portrayal of *keigo* usage. Said corpuses could, for example, be drawn from non-scripted TV dramas or interviews and even give useful insights on potential regional variations. The greater number and diversity of speakers thus involved would moreover strengthen the correlations that could then be drawn from statistical analyses made from such longer and wider datasets. Variables pertaining to dialectal and sociolectal variations could then be considered as well to further our understanding of Japanese politeness marking. As for our own personal research, we shall at the very least endeavor to make use of the same predictor variables but widen our existing corpus by adding scripts of other movies made by the same director, or even better, scripts of more recent movies that also depict day-to-day interactions between family members. It could then be possible, for example, to revisit the matter of in-law relationship and its possible connection with age in regard to *keigo* use. Indeed, when Okamoto remarks on the minimal use of addressee honorifics, used conjointly with plain forms, in a conversation between family members, she argues that "age difference is not generally regarded as constituting a vertical distance (i.e. hierarchy) in families in contemporary Japan, although this does not seem to apply to in-law relationships (2011: 3680)". In that respect, we feel that our research would greatly benefit from the works of such directors as Hirokazu Koreeda that, just like Yasujiro Ozu in his time, focuses on human relationships and family dynamics. What is more, his feature films, such as *Nobody Knows* (誰も知らない, *Daremo shiranai*) released in 2003 and *Our Little Sister* (海街 diary, *Umimachi Diary*) released in 2015, often depict very complex in-law and/or step-family interactions in post-2000 Japanese society. Koreeda's works would therefore provide us with the means to revisit the in-law relationship hypothesis which, for lack of conclusive results, we were forced to abandon in the present study. We are convinced that this diachronic approach (ranging from the 1950s to the 2000s) could also yield some very surprising and interesting results, notably in regard to gender and its impact on Japanese politeness marking.

Acknowledgments

I would like to express my gratitude to my two research assistants: Nozomi Kamisasa for transcribing the Japanese script in Roman alphabet, and Hoorossadat Zafarmand for collecting data from the Romanized text. Their hard work and dedication have been essential in creating a workable corpus for this study.

References

Brown, P., & Stephen C. L. (1987). *Politeness: Some Universals in Language Usage*. Cambridge, U.K.: Cambridge University Press. https://doi.org/10.1017/CBO9780511813085

Dixon, R. M. W. (1972). *The Dyirbal language of North Queensland*. Cambridge, England: Cambridge University. https://doi.org/10.1017/CBO9781139084987

Fukuda, A., & Asato, N. (2004). Universal politeness theory: application to the use of Japanese honorifics. *Journal of Pragmatics 36* (pp. 1991-2002). https://doi.org/10.1016/j.pragma.2003.11.006

Holmes, J. (1995). *Women, Men and Politeness*. Edinburg Gate, England: Pearson Education Limited.

Hori, M. (1986). A Sociolinguistic Analysis of the Japanese Honorifics. *Journal of Pragmatics 10* (pp. 373-386). https://doi.org/10.1016/0378-2166(86)90007-X

Ide, S. (1982). Japanese Sociolinguistics Politeness and Women's Language. *Lingua 57* (pp. 357-385). https://doi.org/10.1016/0024-3841(82)90009-2

Ide, S. (1989). Formal forms and discernment: two neglected aspects of universals of linguistic politeness. *Multilingua 8(2-3)* (pp. 223-248). https://doi.org/10.1515/mult.1989.8.2-3.223

Kogo, N., Yasujiro O. (1953). *Tokyo Monogatari* (Tokyo Story). Tokyo: Romando Scenario Library. Kindle Edition.

Kuwae, K. (1999). *Manuel de japonais, volume 2*. Paris : Langues & mondes, l'asiathèque.

Leech, G. (1983). *Principles of pragmatics*. London: Longman.

Leech, G. (2014). *The Pragmatics of Politeness*. New York, U.S.: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195341386.001.0001

Manno, G. (2002). La politesse et l'indirection : un essai de synthèse. *Langage et société 100* (pp. 5-47). https://doi.org/10.3917/ls.100.0005

Matsumoto, Y. (1988). Reexamination of the Universality of Face: Politeness Phenomena in Japanese. *Journal of Pragmatics 12* (pp. 403-426). https://doi.org/10.1016/0378-2166(88)90003-3

Mizutani, O., & Mizutani, N. (1987). *How to Be Polite in Japanese*. Tokyo: The Japan Times.

Minegishi Cook, H. (1999). Situational Meanings of Japanese Social Deixis : The Mixed Use of the Masu and Plain Forms. *Journal of Linguistic Anthropology 8(1)* (pp. 87-110). https://doi.org/10.1525/jlin.1998.8.1.87

Okamoto, S. (1997). Social context, linguistic ideology, and indexal expressions in Japanese. *Journal of Pragmatics 28* (pp. 795-817). https://doi.org/10.1016/S0378-2166(97)81491-9

Okamoto, S. (1999). Situated Politeness: Manipulating Honorifics and non-Honorifics Expressions in Japanese Conversations. *Pragmatics 9(1)* (pp. 51-74). https://doi.org/10.1075/prag.9.1.05oka

Okamoto, S., & Siegal, M. (2003). Toward Reconceptualizing the Teaching and Learning of Gendered Speech Styles in Japanese as a Foreign Language. *Japanese Language and Literature 37(1)* (pp. 49-66). https://doi.org/10.2307/3594875

Okamoto, S., & Shibamoto-Smith, J. S. (2004). *Japanese Language, Gender and Ideology: Cultural Models for Real People*. New York, U.S.: Oxford University Press. https://doi.org/10.1093/oso/9780195166170.001.0001

Okamoto, S. (2011). The use and interpretation of addressee honorifics and plain forms in Japanese: Diversity, multiplicity, and ambiguity. *Journal of Pragmatics 43* (pp. 3673-3688). https://doi.org/10.1016/j.pragma.2011.06.012

O'Neill, P. G. (1981). *A Programmed Course On Respect Language In Modern Japanese*. Guildford, Great Britain: Hodder and Stoughton.

Peng, F. et al. (1981). Male/female differences in Japanese. Tokyo: The East-West Sign Language Association.

Shimamori, R. (2001). *Grammaire japonaise systématique, Volume II, Les expressions verbales et les expressions de la politesse*. Paris: Jean Maisonneuve.

Siegal, M., & Okamoto, S. (2003). Toward Reconceptualizing the Teaching and Learning of Gendered Speech Styles in Japanese as a Foreign Language. *Japanese Language and Literature 37* (pp. 49-66). https://doi.org/10.2307/3594875

Smoreda, Z., & Licoppe, C. (2000). Gender-Specific Use of the Domestic Telephone. *Social Psychology Quarterly 63(3)* (pp. 238-252). https://doi.org/10.2307/2695871

Suzuki, T. (2007). *A Pragmatic Approach to the Generation and Gender Gap in Japanese Politeness Strategies*. Tokyo, Japan: Hituzi Syobo Publishing.

Notes

Note . Suzuki found and elegant, albeit more qualitative, solution to the challenge posed by the multifacetedness of politeness marking. His research involved independent raters who were asked to judge both "the degree of politeness (i.e. absolute politeness) not just by looking at the use of honorifics [but also] the illocutionary force of an utterance created by its 'propositional content' and 'lexicogrammatical strategies' [and] the degree of appropriateness (i.e. relative politeness) of the responses in relation to the situation" (2007: 126).

Note 2. Variables deriving from sociological factors, as well as the corresponding speakers' pairings obtained from them, appear in bold typeface throughout the article.

Note 3. See for example a type of bystander honorifics, dubbed by Dixon the "mother-in-law language" of Dyirbal; an Australian Aboriginal language that provides a set of lexical terms to be substituted in the presence of certain "taboo" relatives such as a parent-in-law of the opposite sex; and, by the symmetry rule, a child-in-law of the opposite sex (Dixon, 1972: 32).

Note 4. Parallels between Suzuki's findings and our own could not be drawn for stages 2 and 3 as his research did not systematically consider differences in terms of age and gender between speaker and hearer.

Note 5. In this study, P values for nominal variables **genderdyad** and **agegenderdyad**, that could respectively take four and eight values, were generated in R using the "drop1" function associated with linear regression models. The function allows users to assess the significance of individual terms by sequentially dropping one term at a time from the model and comparing the reduced model to the full model.

Note 6. The term "binary" strictly refers here to the predictor variable **gender**. It does not aim to exclude or misrepresent the experiences of individuals whose gender identity does not align with either "male" or "female" categories.

Appendix A

List of word tokens, lexical and morphosyntactic categories of keigo markers and corresponding grades

Table A1. Word tokens, lexical and morphosyntactic categories of keigo markers and corresponding grades (non-exhaustive)

Word token	Affixes, adjectives, nouns, verbs and verb endings	Grade
mairimasu	*mairu* + *masu*	2
o-bento	*o*	1
o-hayou	*o*	1
gozaimasu	*gozaru* + *masu*	2
o-cha, o-tsukudani, o-hiru	*o*	1
o-furo, o-senbei, o-niku	*o*	1
o-machi desu	*o* + *desu*	1 + 1
o-kage-sama de	*o* + *sama*	3
o-tanoshii	*o*	1
go-yukkuri	*go*	1
o-kaa-san	*o* + *san*	2
o-kaa-sama	*o* + *sama*	3
yoroshiku douzou	*yoroshii*	1
o-kaeri	*o*	1
o-genki	*o*	1
o-hisashiburi	*o*	1
o-yasumi ni naru	*o* + *naru*	1
o-kaeri nasai	*o* + *nasai*	2
o-yasumi nasai	*o* + *nasai*	2
irasshaimase	*irassharu* + *mase*	1 + 1
itterasshai	*irassharu*	1
o-sewa-sama	*o* + *sama*	3
o-saki ni	*o*	1
go-kuroo-sama	*o* + *sama*	3
o-tsukare-sama	*o* + *sama*	3
o-daiji ni	*o*	1
go-kuroo-sama	*o* + *sama*	3
o-tsukare-sama	*o* + *sama*	3
go-chisoo-sama	*go* + *sama*	3
o-sewa ni narimashita	*o* + *narimasu*	1 + 1
o-sewa-sama de	*o* + *sama*	3
o-hitori	*o*	1
o-rusuban	*o*	1
sumimasen	*masen*	1
doumo sumimasen	*doumo* + *masu*	2
o-negai	*o*	1
o-yoroshii	*o* + *yoroshii*	2
o-warui	*o*	1
yoroshii, yoshii	*yoroshii*	1
itadaku	*itadaku*	1
go-chisoo	*go*	1
go-meitei	*go*	1
go-men kudasai	*go* + *kudasaru*	2
kudasatte	*kudasaru*	1
go-aisatsu suru	*go* + *suru*	1
go-kigen	*go*	1
o-tsutome	*o*	1
itashimasu	*itasu* + *masu*	1 + 1
o-kawari ni narimasen	*o* + *narimasu*	1 + 1

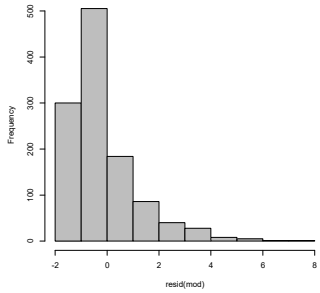
o-baachama	*o* + *chama*	2
o-baa-chan	*o* + *chan*	2
go-annai suru	*o*	1
go-issho ni	*go*	1
o-yasumi ni naru	*o*	1
o-yasumi !	*o*	1
o-yasumi nasai	*o* + *nasaru*	1 + 1
o-yasumi	*o*	1
tsukare nasatta	*nasaru*	1
o-machi dousan	*o* + *san*	2
o-machi doosama	*o* + *sama*	3
ukagaimasu	*ukagau* + *masu*	1 + 1
o-negai itashimasu	*o* + *itasu* + *masu*	3 + 1
o-negai shimasu	*o* + *shimasu*	1 + 1
itasu	*itasu*	2
donata	*donata*	1
go-kuroo-san	*go* + *san*	2
o-tomo suru	*o* + *suru*	1
kanja-san	*san*	1
go-you	*go*	1
oboete rasshai	*rasshai*	1
go-yakkai ni naru	*go*	1
o-isha-san	*o* + *san*	2
o-shigoto	*o*	1
o-taku	*o*	1
o-kirei	*o*	1
o-isogashii	*o*	1
o-hayai	*o*	1
o-hima itadatku	*o* + *itadaku*	1 + 1
o-matase suru	*o* + *saseru*	2
o-mukai ni agarimasu	*o* + *agarimasu*	1 + 1
o-tazune kudasai mashita	*o* + *kudasai* +*masu*	2 + 1
go-joukyou	*go*	1
watakushi	*watakushi*	1
kochira	*kochira*	1
sochira	*sochira*	1
o-atsumari negaimasu	*o* + *ganau* + *masu*	1 + 1
o-nenne	*o*	1
o-kamai dekimasen	*o* + *masen*	1 + 1
o-toshirori	*o*	1
o-niai ni naru	*o* + *naru*	1
o-tsukare ni naru	*o* + *naru*	1
go-ran nasai	*goran*	1
orareru	passive voice for *iru*	1
yamete oru	*oru*	1
o-hazukashii	*o*	1
o-shaku suru	*o*	1
gomen nasai	*nasai*	1
o-kozukai	*o*	1
o-wakare	*o*	1
o-fukuro	*o*	1
o-shoukou	*o*	1
o-yome ni iku	*o*	1
o-yaku ni tachimasen	*o* + *masen*	1 + 1
ossharu	*ossharu*	1
futtote orareta	*oru* + passive voice	1 + 1

o-miokuri suru	*o + suru*	1
o-katami hosshii	*o*	1
watakushi	*watakushi*	1
washira	*washira*	1
o-rei	*o*	1
o-sabishii	*o*	1

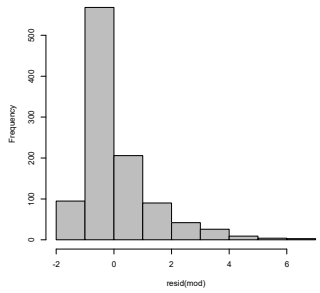
Appendix B

Histograms of residuals for linear models, Stages 1, 2 and 3

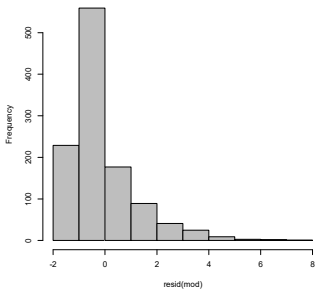
Table B1. Simple linear models, Stage 1



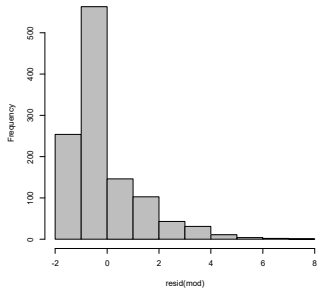
Variables: markers, sm



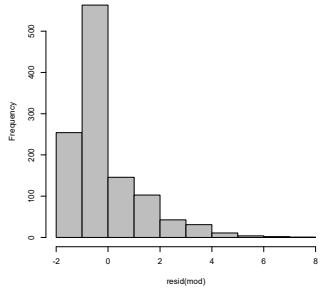
Variables: markers, agediff



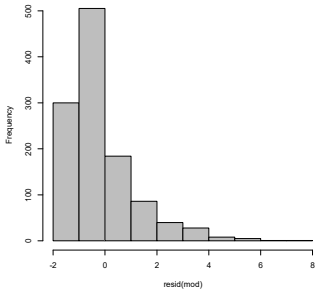
Variables: markers, sy



Variables: markers, hm

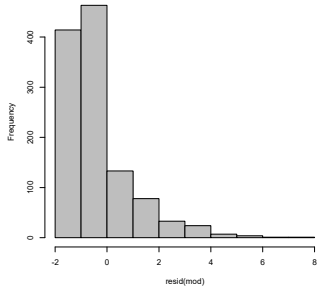


Variables: markers, hf

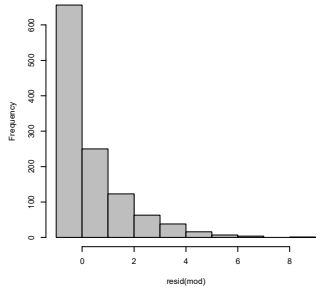


Variables: markers, sf

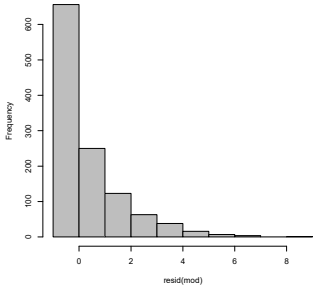
Table B2. Simple linear models, Stage 2



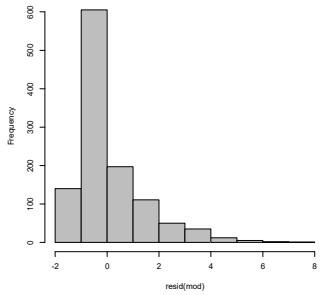
Variables: markers, smhf



Variables: markers, smhm

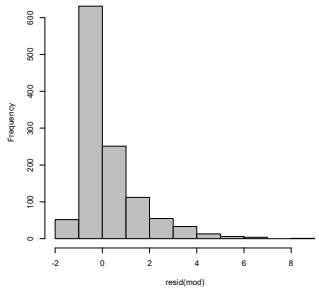


Variables: markers, sfhf

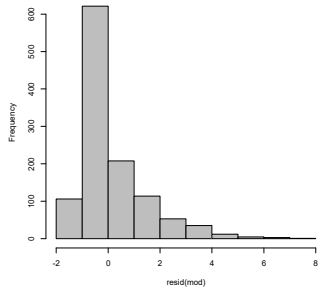


Variables: markers, sfhm

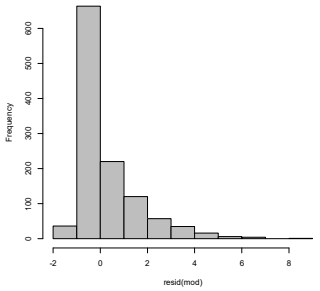
Table B3. Simple linear models, Stage 3



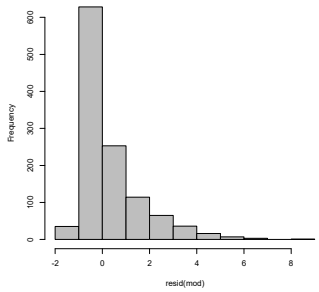
Variables: markers, sfhfo



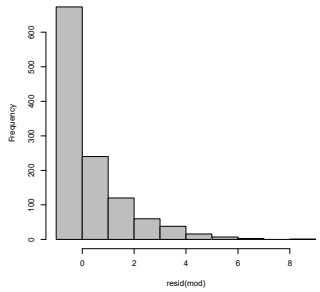
Variables: markers, sfhmo



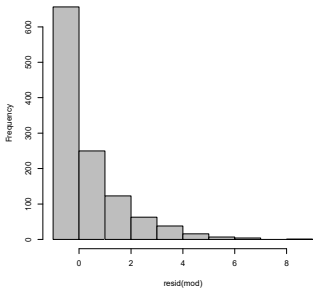
Variables: markers, smhmo



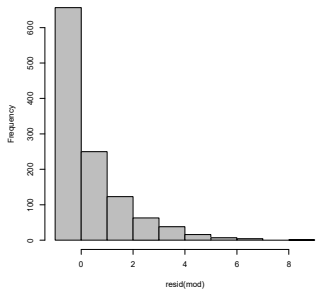
Variables: markers, smhfo



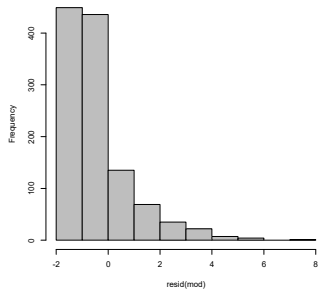
Variables: markers, sfhmy



Variables: markers, sfhfy



Variables: markers, smhmy



Variables: markers, smhfy

Appendix C

Intercorrelation testing for linear models, Stages 1 and 2

Table C1. Multiple linear models, Stage 1

Call: lm(formula = markers ~ sy + sm + sy * sm, data = tnm9)	Call: lm(formula = markers ~ sy + sf + sy * sf, data = tnm9)
Residuals: Min 1Q Median 3Q Max -1.2798 -0.7532 -0.3789 0.6211 7.7202	Residuals: Min 1Q Median 3Q Max -1.2798 -0.7532 -0.3789 0.6211 7.7202
Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 0.75325 0.08328 9.044 < 2e-16 *** sy 0.52655 0.10530 5.001 6.62e-07 *** sm -0.37433 0.10724 -3.490 0.000501 *** sy:sm 0.20831 0.15889 1.311 0.190118 --- Signif. codes: 0 ‚Äò***‚Äô 0.001 ‚Äò**‚Äô 0.01 ‚Äò*‚Äô 0.05 ‚Äò.‚Äô 0.1 ‚Äò ‚Äô 1	Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 0.37892 0.06756 5.608 2.57e-08 *** sy 0.73486 0.11899 6.176 9.17e-10 *** sf 0.37433 0.10724 3.490 0.000501 *** sy:sf -0.20831 0.15889 -1.311 0.190118 --- Signif. codes: 0 ‚Äò***‚Äô 0.001 ‚Äò**‚Äô 0.01 ‚Äò*‚Äô 0.05 ‚Äò.‚Äô 0.1 ‚Äò ‚Äô 1
Residual standard error: 1.266 on 1131 degrees of freedom (23 observations deleted due to missingness) Multiple R-squared: 0.08231,Adjusted R-squared: 0.07988 F-statistic: 33.81 on 3 and 1131 DF, p-value: < 2.2e-16	Residual standard error: 1.266 on 1131 degrees of freedom (23 observations deleted due to missingness) Multiple R-squared: 0.08231,Adjusted R-squared: 0.07988 F-statistic: 33.81 on 3 and 1131 DF, p-value: < 2.2e-16
Variables: sy, sm	Variables: sy, sf

Table C2. Multiple linear models, Stage 2

Call: lm(formula = markers ~ sy + smhf + sy * smhf, data = tnm9)	Call: lm(formula = markers ~ sy + sfhf + sy * sfhf, data = tnm9)
Residuals: Min 1Q Median 3Q Max -1.2551 -0.6994 -0.2927 0.3006 7.7449	Residuals: Min 1Q Median 3Q Max -1.4203 -0.6566 -0.4760 0.5240 7.8337
Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 0.69940 0.06898 10.139 < 2e-16 *** sy 0.55566 0.08942 6.214 7.24e-10 *** smhf -0.40672 0.10611 -3.833 0.000133 *** sy:smhf 0.16861 0.20395 0.827 0.408571	Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 0.47596 0.06227 7.644 4.49e-14 *** sy 0.69030 0.08812 7.834 1.08e-14 *** sfhf 0.18066 0.11660 1.550 0.122 sy:sfhf 0.07336 0.17079 0.430 0.668

--- Signif. codes: 0 ‚Äò***‚Äô‚Äô 0.001 ‚Äò**‚Äô‚Äô 0.01 ‚Äò*‚Äô‚Äô 0.05 ‚Äò.‚Äô‚Äô 0.1 ‚Äò ‚Äô‚Äô 1 Residual standard error: 1.264 on 1131 degrees of freedom (23 observations deleted due to missingness) Multiple R-squared: 0.08421,Adjusted R-squared: 0.08178 F-statistic: 34.67 on 3 and 1131 DF, p-value: < 2.2e-16	--- Signif. codes: 0 ‚Äò***‚Äô‚Äô 0.001 ‚Äò**‚Äô‚Äô 0.01 ‚Äò*‚Äô‚Äô 0.05 ‚Äò.‚Äô‚Äô 0.1 ‚Äò ‚Äô‚Äô 1 Residual standard error: 1.27 on 1131 degrees of freedom (23 observations deleted due to missingness) Multiple R-squared: 0.07614,Adjusted R-squared: 0.07369 F-statistic: 31.07 on 3 and 1131 DF, p-value: < 2.2e-16
Variables: sy, smhf	Variables: sy, sfhf