J.Vilkevic (Fall 2017)

The usage of singular and plural "you" by native speakers of Mandarin Chinese and Lithuanian

The data and the goal of the project

The data that I am using are the answers to a discourse completion task. Discourse completion task is a type of questionnaire, where the respondents read about different situations and then write what they would say. The respondents of the questionnaires used for this project were native speakers of Mandarin and Lithuanian. These questionnaires focused on requests, so in each situation the respondent had to ask for something. For example, in one of the situations, respondents had to write how they would ask a fellow student to move a bit, so that they could sit down and eat in the cafeteria. I am using my data in Mandarin, that will also be used to compare linguistic politeness in Mandarin and Chinese in my thesis, as well as my professor's Lithuanian data, which she used to compare Lithuanian and English politeness. There are 12 situations in total and for this project I used 30 answers in each situation for both languages.

At first, I wanted to focus on all of the forms of address but I decided not to do that, because of two reasons. First of all, I would have to deal with some problems with Mandarin and Voyant, because in my experience it is a bit complicated to use Voyant while looking for Chinese names/nicknames etc., because of how it sometimes treats characters of those names as separate words. And more importantly, even though it would be an interesting thing to look into, it would not really tell me anything valuable about my data when it comes to politeness. In the end, I decided to focus on the usage of the singular "you" and plural "you" (which is often used to signify respect or distance between the speaker and the addressee). Also, even though I had access to the data in Lithuanian, Mandarin Chinese and English, I decided not to use English, because you cannot really tell what kind of "you" is used. In order to make the data more manageable, I put each set of answers of both languages in separate files and named them LT1 – LT12 and M1 – M12.

The goal of this project is to see if there is any difference in the usage of the singular and plural "you" in the requests made by native speakers of Lithuanian and Mandarin Chinese. For example, if the speakers of Mandarin use the respectful "you" more often in certain situations than the speakers of Lithuanian.

Calculating the instances of the singular and plural forms second person address

I decided to start with working on Mandarin because it seemed like the easier one. One thing that I should mention about this data is that some of the answers had multiple instances of "you" in them. This is obviously a problem because if, for example, one person used singular "you" five times in his answer and another three people used one plural "you" each, the data would show, that singular "you" was a more popular choice, which would not be true. I removed them, by putting the text into Word and searching for the instances of repeated usage of "you" in the answers, and then deleting them. There are probably multiple more sophisticated and quicker ways to do it, but I wanted an easy and "foolproof" way to do it and this seemed to work quite well. The data included with the project has all of the unnecessary pronouns deleted. All I had to do after that is to upload the first file into Voyant and search for 你* and 您* (searching without * would lose some of the pronouns, for example in words like 您好 (*hello*)). Then I repeated it with all of the other 11 files. I created a table in Excel of how many instances of the searched words I found in each situation.

Working with the Lithuanian data was a more complicated task because people do not always use the pronoun and the pronoun can also be reflected in the ending of the verb. For example, by saying *rašyk*, the speaker orders one person to write without having to use the singular pronoun. Meanwhile, by saying *rašykite*, the speaker orders the same thing to multiple addressees or to one addressee whom he/she would address by using the plural "you". I looked at the data for quite a while and decided on the best way to treat this data in order to find out how the addressee is addressed in every answer (that it is visible in) and getting the most accurate numbers. Firstly, I searched in Voyant for all of the cases of tav* (tu/singular you), jūs*, jum* (jūs/plural you). Then, I removed these rows with the help of Open Refine. If I did not do it and then looked at the verbs, my numbers would be wrong, because some of the answers had both the pronoun and the verb with the ending that reflects the addresse. Some of the answers that had the pronouns also had multiple verbs with those endings, so it was a good way to get rid of them too. I did it by selecting the „text filter", looking for those words and then removing the rows that contained them. Then I used Voyant again and looked for all of the instances of *uok,*ėk,*ali,*um,*yk,*kš,*ysi,*eik,*iepk,*elk,*iepsi (all singular, I also at first had *unk and *ėsi, but then saw that I did not need them) and *ite,*ėte, *ėt, *it (all plural). Unfortunately, this did not always work with all of the endings. For example in the first situation, it did work with *uok, but when I searched for *ali, I got an error message. Later, I

tried it again and it worked fine, I am not sure what that depends on because I did everything the same way. Sometimes when it still did not work, I would look at the text, search for the whole words with that ending in Voyant and then also check in Word if my numbers were right. This problem with Voyant really made this part more time consuming and tedious than I expected.

The analysis and visualization

The numbers that I got were:

Mandarin

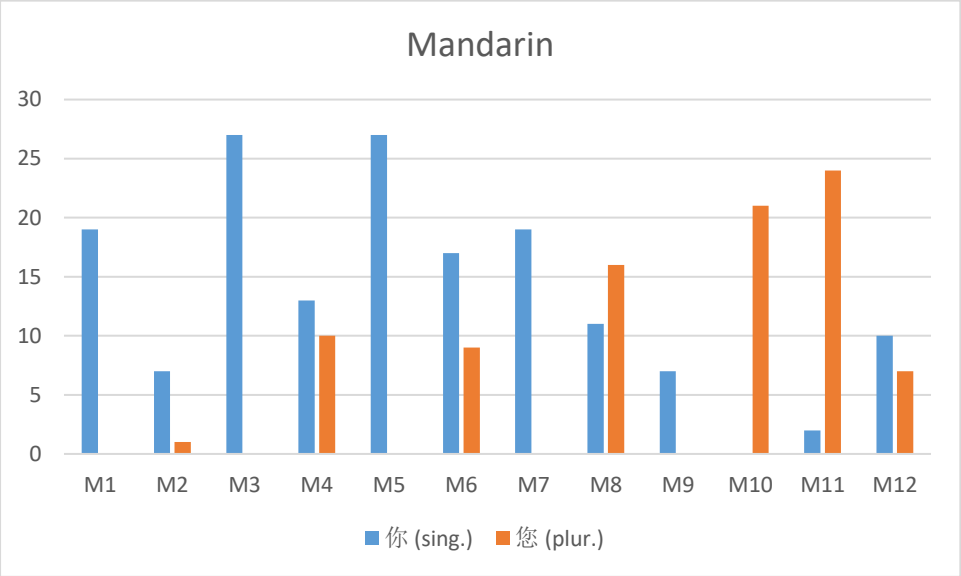
Column1	你 (sing.)	您 (plur.)
M1	19	0
M2	7	1
M3	27	0
M4	13	10
M5	27	0
M6	17	9
M7	19	0
M8	11	16
M9	7	0
M10	0	21
M11	2	24
M12	10	7

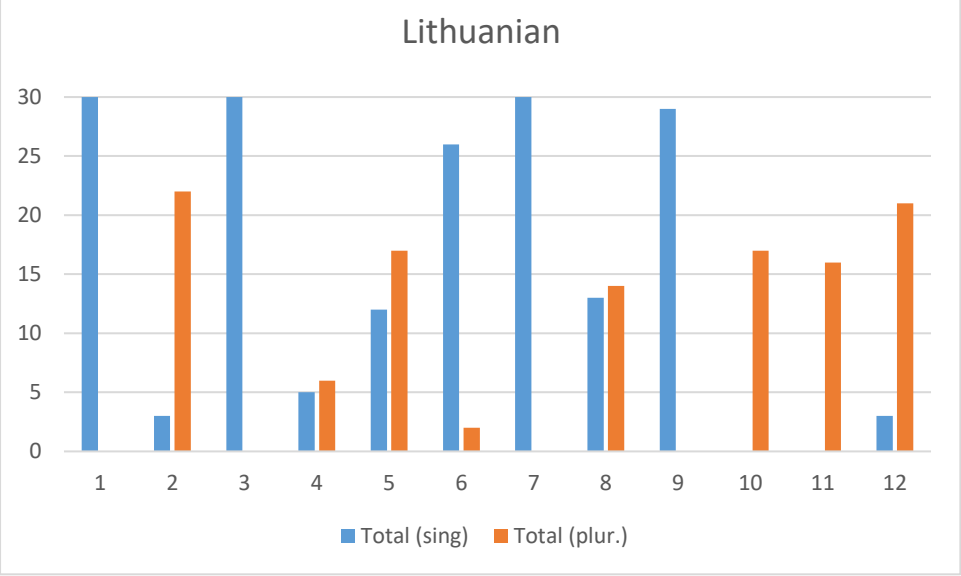
Lithuanian

Column1	Tu (sing.)	Jūs (plur)	Verb (sing.)	Verb (plur.)	Total (s	Total (p	r.)
LT1	4	0	26	0	30	0	
LT2	0	1	3	21	3	22	
LT3	17	0	13	0	30	0	
LT4	0	1	5	5	5	6	
LT5	2	10	10	7	12	17	
LT6	1	1	25	1	26	2	
LT7	1	0	29	0	30	0	
LT8	11	13	2	1	13	14	
LT9	0	0	29	0	29	0	
LT10	0	11	0	6	0	17	
LT11	0	0	0	16	0	16	
LT12	0	4	3	17	3	21	

The reason why the total of singular + plural is not always 30 is that in some instances there is no indication of how the addressee is addressed and in some instances I could tell it only from the context and not the things that I was looking into.

Here are some charts that I have made so that I could easier see if there are any tendencies:





Since the numbers are a lot higher in Lithuanian, there is no point in comparing how many instances of usage of singular and plural forms were found in each situation. What I want to look at is what forms Lithuanian and Mandarin speakers tend to use in which situations and if there are any noteworthy instances.

Looking at the first situation, we can see, that in both languages the speakers used only the singular form. That is not surprising, since in the situation, they were required to ask for

something from a friend. The same result is achieved in the third (again, asking for something from a friend), seventh (mother asking her son to go to the store) and ninth (older sibling asking his sister to pick up the phone/open the door) situations. Meanwhile in the tenth situation the speakers of both languages decided to use the plural form. This situation required them to ask their professor to extend the deadline for their essay, so it is not surprising, that they chose to be extremely respectful.

In the second situation, the speakers of Mandarin used the singular form a lot more often, while the Lithuanian preferred the plural one. In this situation the professor had to ask a student to close a window. I expected the speakers of both languages to use overwhelmingly the plural form.

In the fourth situation the speakers of both languages used both singular and plural forms in quite similar frequencies, but the speakers of Mandarin preferred the singular form and those who spoke Lithuanian preferred the plural form. In this situation respondents had to ask a stranger to move a bit, so that they could sit.

In the fifth situation Mandarin speakers used overwhelmingly the singular form, while the Lithuanian speakers used both but preferred the plural form. In this situation the professor had to again ask for something from their student. This is very interesting, because it seems like Mandarin speakers when speaking as a professor to their student decided to use the singular form and the Lithuanian speakers did the opposite in both situations.

In the sixth situation, where the speakers had to confront a neighbor who listened to music too loudly and ask them to turn it down, the speakers of both languages preferred the singular form, but Lithuanian speakers used it a lot more often.

In the eighth situation the usage of both forms was frequent with the preference of the plural form. In this situation the speakers had to ask for something from their uncle and in the past they did something that would affect their chances of getting what they want negatively.

In the eleventh situation both Lithuanians and the Chinese preferred the plural form, but the latter also used the singular form a couple of times. In this situation they had to ask a person, who works in the university, for a piece of paper.

In the last situation they had to ask for a referral from a doctor that they knew quite well, the Lithuanian speakers preferred the plural form and the Mandarin speakers used both, but the singular more often.

The conclusions

Overall, Mandarin speakers preferred the singular form in eight situations and plural form in other three, while the Lithuanian speakers preferred the singular form only in five situations, and plural form in other seven. While in order to really find out why the numbers are the way that they are a more extensive research is needed, this project showed me some overall tendencies. Before completing this project, I expected the Mandarin speakers to use the plural (the more "respectful" and "distant") form more often and the Lithuanian speakers to use the singular (the more "friendly" and "casual") form more often, so the results really surprised me. This is especially true for the requests made by professors to their students. It also surprised me, that there were similarities in quite a few of the situations.

After completing the counting part of the project, I would definitely not use Voyant for this again, because it is not really reliable when it comes to this task. I would, however, use it for other things, like making wordclouds. I will use OpenRefine in the future, it is definitely a very useful tool.

When it comes to the further research, I will take these results into account when analyzing the data for my thesis. I am not sure if I will use similar tools for future analysis. The most valuable thing that I got by making this project is a chance to really familiarize myself with the data that I will have to work with in the future, which is incredibly useful.