The following analysis

**Basic Statistical description**

* Basic description of the data sets. I suggest two tables one with the demographic information so

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | | N | % |
| Gender and age | |  |  |
|  | Women < 75 |  |  |
|  | Women 75+ |  |  |
|  | Men |  |  |
| Area | |  |  |
|  | East |  |  |
|  | West |  |  |
|  | Poland |  |  |
| Interview Length etc | |  |  |
|  |  |  |  |

* A second one that deals with the response for the different cells

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Cell | No of speakers using cell in conversation | Maximum no of usages of cell by an individual speaker | No of speaker using cell more than once | No of forms produced by speakers | No of Speakers using multiple forms | Maximum number of forms used by an individual speaker |
| Person | | | | | | |
| Nominative Plural |  |  |  |  |  |  |
| Nominative Singular |  |  |  |  |  |  |
| Genitive Plural |  |  |  |  |  |  |
| Genitive Singular |  |  |  |  |  |  |
| Etc.. |  |  |  |  |  |  |
| Year | | | | | | |
| Nominative Plural |  |  |  |  |  |  |
| Nominative Singular |  |  |  |  |  |  |
| Genitive Plural |  |  |  |  |  |  |
| Genitive Singular |  |  |  |  |  |  |
| Etc.. |  |  |  |  |  |  |

These tables will give a reader a quick overview of the data. The second tables allows them also to know:

1. Why you have chosen to use only specific cells for your more complex analyses and not others. For instance, if there are very few times any speaker user a particular cell in the conversation then that cell need not be analysed later
2. That the elements that affect variation are inter-speaker parts of language yet it is clear that some speaker do use multiple forms for the same cell.

**Intra Speaker variation**

What I propose is to look across all cells and see if an individual has ever produced multi-forms for any cell. It would then be possible to do a straightforward test to see whether these multi-forming speakers differ by age-gender, area or number of forms etc. This could all be done using chi-squared but we could also try using logistic regression. This might be tied in with close qualitative analysis of the way speakers use multiple forms in a single cell i.e. looking at context to see if any pattern that might explain why one form and not another is used. This will tell you whether any of the basic

**Inter Speaker Variation**

Most of the explanatory variables are between speaker. i.e. a can only be male or female during the conversation and cannot swap genders half way through, they are speaking in a specific area etc. Thus most of the information is on between person variation. I am going to suggest a present or absent of each form for each person and then do multi-level logistic regression. This will allow us to find out if variation in form is related to gender, area, number of times used. These would have to be done for each cell separately.