**STUDY #2 LOOKING FOR CORRELATIONS BETWEEN FORMS AND SPEAKERS’ CHARACTERISTICS**

In the previous study we had been looking at speakers’ characteristics that may help us predict whether they would be more prone to overabundance (i.e. using more than one form, for one cell; as the English past tense of the verb *to burn* can be either *burnt*  or *burned*). The results showerd that overabundance is something that was happening at both individual- and collective-levels and that the only determining factor that would make a speaker use more than one form was how much they had contributed to the corpus (length); i.e. the longer they spoke for, the higher the chances for them to use more than one form of the paradigm.

Now we want to see if there is any correlation between the use of specific forms and speakers characteristics. The final goal is to prove whether there is any regularity between the use of specific forms and speakers characteristics, or whether overabundance is something that just happens because of the context in which they live (where there is not a normative form and people are highly multilingual, at least, at a passive level). In other words, whether variation between forms is purely sociolinguistically neutral, or whether it is conditioned by some of speakers’ characteristics (and if so, which).

Using the available data for the nouns ‘person’ and ‘year’ we want to run two parallel analyses (with the available data) to see if they provide similar results. We have picked the most overabundant cells, where we have also a significant amount of speakers using more than one form.

For ‘year’, the main cells with more than two speakers using multiple forms would be also the gen pl (19 speakers out of 42 speakers using it more than once) and the gen sg (7 speakers out of 26 speakers using it more than once). The loc sg would also be over the threshold (5 speakers out of 13 "recidivists").

As for ‘person’, the only cell that had more than two speakers using multiple forms was the gen pl (13 speakers out of 19 speakers using it more than once).

Now getting into finer detail, I shall rescue some of the initial exploratory analyses and point some potential directions in which we could analyse these data.

# 1. The noun ‘year’

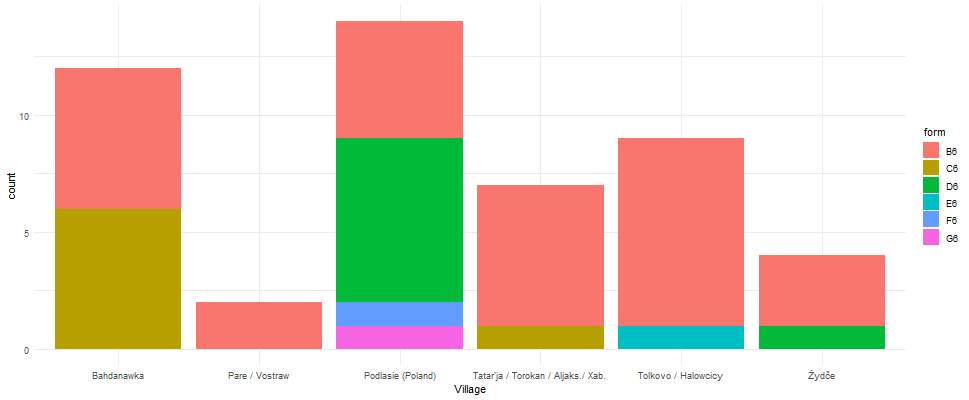
## 1.1. Genitive singular:

From a first impression, the factor that seemed to be more relevant for the distribution of forms was the origin of speakers. Some forms are concentrated in some villages, though we shall focus on the main three B6 (ˡɦoda), C6 (ˡɦodu), D6 (ˡɾoku).

A graph with different colored squares

Description automatically generated

The form D6 is the one closest to Polish an Ukrainian and C6 would be correspond to the Belarusian form (the Belarusian Academy admits both B6 and C6). B6 corresponds to the Russian form, but, as I said, it is also used in Belarusian. We therefore hypothesise that villages closer to Poland (or in Poland) or Ukraine will approach more to D6 (and less prominently C6), whereas in more Russified areas (especially, in Belarus) we would find more B6.



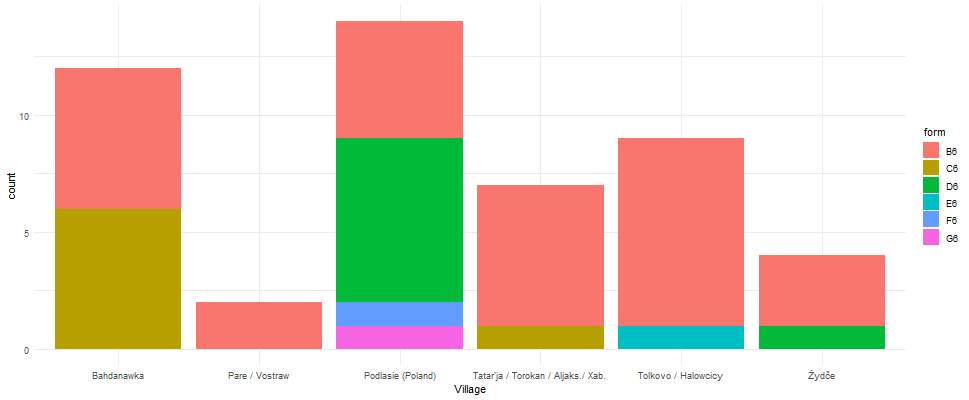
## 1.2. Genitive plural

The three most prominent forms are B7 (*ɦod*), C7 (*ɦoˡdov|ɦoˡdɪv* |… ) and G7 (*lit | ljet | lʲɪt | …*).

A graph with different colored squares

Description automatically generated

B7 is an innovation of West Polesian; C7 follows the model of Belarusian; and G7 follows the model of Russian and Polish. This one is a bit harder to predict the directionality, because G7 could be either an influence of Polish or Russian; yet we would expect less C7 outside of Belarus, and more or less B7 depending on how affected they are by other languages.



## 1.3. Locative singular

The three most commonly used forms are C11 (*ɦoˡdu | ɦaˡdu*) and F11 (*ɾoˈkovɪ*), and much less frequently B11 (*ˡɦodu| ˡɦadu*)and D11 (*ɦod*).

C11 corresponds to the Russian form. F11 would correspond to the dat.sg of that same noun in Ukrainian, but West Polesian is unique on using it as a loc.sg, so it can be considered an innovation of WP (admitting that the stem is the same as in Polish and Ukrainian). There are no languages matching the B11 and D11 form, so they are also and innovation (though the stem is the same as in Belarusian and Russian). Thus, we predict that we will hardly find any C11 in Podlasie (Poland) and probably also no B11 and D11. Conversely, F11 being quite a unique form, it may be less conditioned by the village.

A graph with different colored squares

Description automatically generated

# 2. The noun ‘person’

## 2.1. Genitive plural

The three most frequent forms are B7 (*ʧoloˈvjek |ʧɪloˈvjek |...*), E7 (*luˈdʲɪj|luˈdɪ |luˈdʲej*), G7(*duʃ*). G7 is very archaic and it only remains in WP (other than Bulgarian/Macedonian).

A graph with different colored bars

Description automatically generated

The Belarusian Academy admits both B7 and E7, with a preference for E7, which can also be applied to Russian. Ukrainian and Polish only admit E7. Therefore, we would expect G7 to be less affected by the location of the speakers, and B7 to be rare in Podlasie (Poland) (and Ukraine, if we had the data… ☹).

A group of colored squares

Description automatically generated