

Information on the spreadsheets including the data

Data_1 and **Data_2** list all occurrences of the words ‘bishop’ and ‘child’ in the codices, respectively, as well as their derivatives. These two spreadsheets are structured in the same way.

Column A: Source (i.e., the individual codices from which the data comes; information on the codices is available here: <http://omagyarkorpusz.nytud.hu/en-codices.html>).

Column B: Date of source. Note that since all the codices come from a period covering hardly more than a century and represent a variety of dialects, the dates do not necessarily correspond to a diachronic sequence.

Column C: The exact locus of the word form within the source as given in the Old Hungarian Corpus. For the sake of identifiability, we have kept all the abbreviations as they are in the OHC though these reflect Hungarian names even in the English version (e.g. *Zsolt* refers to the Book of Psalms in Bible translations etc.).

Column D: The word forms; in some cases, due to erroneous data entry or some other reason, the word form is not coextensive with the sequence the OHC recognizes as one form; in such cases the form given by the OHC is given in square brackets (e.g. the definite article is frequently written together with the following word and the two are not consistently separated in the OHC). @@ indicates line break.

Column E: Morphological analysis of the word forms; where potentially ambiguous, the most probable analysis is given; the analysis was disambiguated in those cases only where both of the following conditions apply: (i) there is a category difference between affixes in competing analyses (Type A vs B, see Appendix) and (ii) labiality is crucially involved. Since many of the forms are analysed in the OHC, we keep the same category labels in these two spreadsheets as those used in the OHC. The full list of these labels can be found here: http://omagyarkorpusz.nytud.hu/morf_eng.html.

Column F: The labial (L) or non-labial (l) quality of the vowel in the first syllable of the word stem (labial only in the case of ‘bishop’; in ‘child’ the first vowel is never labialized)

Column G: The labial (L) or non-labial (l) quality of the vowel in the second syllable of the word stem

Column H: The labial (L) or non-labial (l) quality of the vowel in the suffix immediately following the word stem; when no suffix follows, i.e. the word is in the nominative singular, cells in this column are empty

Column I: The Type of the first suffix following the word stem (A vs B, for explanation see Appendix); when no suffix follows, i.e. the word is in the nominative singular, cells in this column are empty

Columns J–O: If more than one suffix follows the stem, the same information as in columns G and H is given in these cells. This is only for the sake of completeness; from the point of view of regressive labialization these suffixes are irrelevant.

Data_3 and **Data_4** list all occurrences of word forms affixed with the allative and the multiplicative suffix in the codices, respectively. These two spreadsheets are structured in the same way.

Column A: Source (i.e., the individual codices from which the data comes).

Column B: Date of source. Note that since all the codices come from a period covering hardly more than a century and represent a variety of dialects, the dates do not necessarily correspond to a diachronic sequence.

Column C: The exact locus of the word form within the source as given in the Old Hungarian Corpus. For the sake of identifiability, we have kept all the abbreviations as they are in the OHC though these reflect Hungarian names even in the English version (e.g. *Zsolt* refers to the Book of Psalms in Bible translations etc.).

Column D: The word forms; in some cases, due to erroneous data entry or some other reason, the word form is not coextensive with the sequence the OHC recognizes as one form; in such cases the form given by the OHC is given in square brackets (e.g. the definite article is frequently written together with the following word and the two are not consistently separated in the OHC). @@ indicates line break.

Column E: The quality of the vowel in the syllable immediately preceding the suffix in question (allative in *Data_3* and multiplicative in *Data_4*).

B=Back (only in *Data_4*)

L=front labial

I=front non-labial

Column F: The quality of the vowel in the suffix in question (allative in *Data_3* and multiplicative in *Data_4*).

B=Back (only in *Data_4*)

L=front labial

I=front non-labial

Column G: An approximate gloss for the stem

Data_5 includes all the remaining data we refer to throughout the paper; these are presented here for illustration rather than any kind of quantitative inquiry. Sources are indicated from which we took the data.

Column A: Gloss

Column B: Transcription

Column C: Date, written form, name of original source; if no source is indicated at this point, the given form was taken from the online etymological dictionary of Hungarian (uesz.nytud.hu)

Column D: Source reference. These are of three kinds:

(i) a link to the relevant entry in one of two dictionaries, either the online etymological dictionary of Hungarian (uesz.nytud.hu) or the online Uralic etymological dictionary (uralonet.nytud.hu);

(ii) a query expression to find a particular form in the Old Hungarian Corpus, in each case beginning with the sequence “[W FOCUS”]; these expressions can be entered into the query editing window directly in the search interface of the corpus (<http://omagyarkorpusz.nytud.hu/en-search.html>);

(iii) a link to a published source that lists the forms in question with further source references; this is used when the ultimate source references are only available in printed form.

All empty cells in Column D inherit the value of the cell above them transitively.