

Historical Dutch Spelling Normalization

Annotation Guidelines

This annotation guideline describes the rules for correcting silver data or annotating historical Dutch 19th-century texts. These rules only apply to changing the spelling of the text, not rephrasing the sentence, changing the grammar, etc. The spelling corrections take the form of instructions to the Alpino parser. The five most important types of annotating placeholders Alpino uses are: @alt, @alt_mwu, @mwu, @phantom, and @postag. For the full documentation, see document: Alpino User Guide¹ bracketed input section.

1. When two words are merged into one, combine the words and place them after the @alt_mwu placeholder (right side); the old spelled words are added after (left side): [@alt daarom daar om]
2. When one word should be split into multiple words, apply @alt to the most important word, and insert the other words with @phantom: [@phantom aan] [@phantom te] [@alt spreken aantespreken]
3. When the silver spelling correction offers multiple possibilities, correct them to a single possibility:
de aralia's en palmen van [@alt -de-den den] corridor
de aralia's en palmen van [@alt de den] corridor
4. When correcting multiple possibilities or uncertain about a word or phrase, use the Historical Dutch dictionary² to check for the correct instance.
5. Leave Hyphenated words unchanged, (e.g. twee-en-negentig, op-zichzelf). But do correct single words which are written in old Dutch, (e.g. Noord-Duitschland → Noord-Duitsland, kampioen-athleet → kampioen-atleet)
6. Normalize adjectives ending in -en, for example: gespierden → gespierde
7. Normalize grammatical cases, for example: uwen, uwer, uwe → uw; onzen → onze
8. Keep archaic verb conjugation of words if the spelling is correct, for example: gij kwaamt, gij zijt, hij zeide, als -ie
9. Correct words which have an accent circonflexe, for example: broêr → broer; weêr → weer; elkaâr → elkaar
10. Correct French written words to Dutch alternatives, if possible. For example: billardballen → biljartballen; million → miljoen

¹Alpino User Guidelines

²Woordenboek der Nederlandsche Taal (WNT)