

# Orthographic Variations in Bavarian

## Annotation Guidelines

### Background

In this project we investigate orthographic variations in dialects. We want to identify the relationship between words in High German and the corresponding words in Bavarian (e.g., München ↔ Minga). We are particularly interested in differences that can be attributed to dialect-specific word variations.

### Task

In this annotation project we identify different dialect-specific variations (regional pronunciations and expressions) of German terms. The data comes from Wikipedia articles that are written in Bavarian.

We give you a table (see Figure 1) with highly frequent Standard German words (any POS) and 10 possible orthographic variations of the same concept in Bavarian. You can find the description of the different columns in Table 1. You are asked to indicate if the Bavarian candidate is an orthographic variation of the Standard German word. The answer to this question can be:

- **Yes (1)**: the candidate is an exact dialectal variation of the Standard German word, without inflectional modifications.
- **Inflected (2)**: the candidate represents a morphologically inflected variant of the word in Standard German. Nouns differ in gender,<sup>1</sup> number, and case; verbs in tense, person, number, mood (e.g., indicative, subjunctive) and voice (active/passive); and adjectives in gender, number, case, and degrees of comparison (superlative, comparative, diminutive). Please, consider only inflectional morphology variations.
- **No (3)**: the candidate has nothing to do with the word in Standard German.

Your task is to enter **1** in the “label”-column to indicate Yes, **2** for Inflected, and **3** for No. In Figure 1 we show an example of ten annotations for the word *Gespräch*. If you’re unsure whether a word is a Bavarian variant, and if the provided context examples don’t help, please try Googling its meaning or looking it up on Wikipedia. If you are still unsure, please state that the word is *not* a Bavarian variant. For this annotation task precision is more important than recall.

If multiple labels are applicable for a word, choose the lower label. For example, if a Bavarian word can be seen both as a strict dialect variant (**1**) and as an inflected variant (**2**) of the provided German lemma, it should be annotated with the lower label (**1**).

We provide the part-of-speech tag for each German lemma and up to three usage examples for each Bavarian candidate word (see Figure 1), which are meant to assist in understanding the word’s meaning, but should not be considered as strict references for annotation purposes.

---

<sup>1</sup>In Bavarian, nouns that end with the suffix *-in* are not inflections but derivations and should therefore be annotated with **3**. This applies only if the suffix changes the grammatical gender. For example, *Feansäämoderatorin* is a Bavarian derivation of the German word *Fernsehmoderator*.

id	de	pos	bar	label	example sentence	link
24	Gespräch	NOUN (100.0 %)			-2267-8 Ich bin Abenteuerer und nicht Dichter: Aus <b>Gesprächen</b> mit Kurt Hofmann Sämtliche Gedichte Derr Herr Nor wieder auftreten und etwa bei der Wiedergabe von <b>Gesprächen</b> geschrieben werden müssen, wurde auch ihre Schrei	<a href="#">link</a>
			Gesprächen	3	rtmann: Ich bin Abenteuerer und nicht Dichter: aus <b>Gesprächen</b> mit Kurt Hofmann (2001) ISBN 3-85002-465-2 Hans-C Midm Weihbischof wui ea a klearends <b>Gspräch</b> gflahtd hom, und dea soi zwengs da Suspendiarung	<a href="#">link</a>
			Gspräch	1	i 2007 «Wir lebten in einer Frivolitätsepoche», A <b>Gspräch</b> mit'm Philosophn Peter Sloterdijk iba de Finanzma	<a href="#">link</a>
			Onspruch	3	Se häd an Sebastian Blattner zu an <b>Gspräch</b> mid Ian Valobtn, an Anton Rambold, ibaredn soin.	<a href="#">link</a>
			Obsproch	3	emba 1918 hod de prowisorische Nazjonalvasommlung <b>Onspruch</b> auf de Gebietshoheit üwas gschlossene Siedlungsge	<a href="#">link</a>
				3	Da Artiki 5 siagt in <b>Obsproch</b> mit de Regiarunga vo de NATO-Mitgliedsstootn de W	<a href="#">link</a>
			Gspräche	2	Ijk Das Philosophische Quartett im ZDF Artike und <b>Gspräche</b> Peter Sloterdijk: "Weltsucht - Anmerkungen zum Dr Bereits seit 1920 gob 's eanstzunehmend <b>Gspräche</b> , gegenseitige Besuche zwischn Politikean, da Gene	<a href="#">link</a>
			Assproch	3	Da junge Dichda Johann Peter Eckermann hod lange <b>Gspräche</b> midm Goethe dokumentiad und so an Haffa	<a href="#">link</a>
			Gsprächn	2	Lemsweish D fonetische <b>Assproch</b> vo /x/ u /n/ is ned deitle: In Fall vo /x/ setzd moant ma owa min Dialekt meara des wos a bstimmte <b>Assproch</b> hod u duach Isoglossn definiat wiad, z. B.	<a href="#">link</a>
			Gsprech	1	Nutzung vo Preposiziona, fonetische Untaschid u d <b>Assproch</b> ; da vejat Afsotz bhandlt fonetische Ainghalt, d	<a href="#">link</a>

Fig. 1: Example for the noun *Gespräch* in Standard German and ten lexically similar candidate words. The annotation task is to determine whether these are **valid orthographic variations** in a strict sense (i.e., possible translations of the German term), **inflections of variations**, or **different words** altogether.

Column	Explanation
id	The example ID.
de	The word in Standard German.
pos	The automatically assigned part-of-speech of the word in Standard German.
bar	A lexically similar Bavarian candidate word.
label	<ul style="list-style-type: none"> <li>Enter 1 to indicate <b>yes</b>, the Bavarian candidate word is an exact translation of the Standard German word.</li> <li>Enter 2 to indicate <b>inflected</b>, if the Bavarian candidate word is a morphologically inflected variant of the Standard German word.</li> <li>Enter 3 to indicate <b>no</b>, i.e., the candidate word does not correspond to a variant of the German word.</li> </ul>
example sentences	One to three text snippets containing the Bavarian word (highlighted).
link	URL linking to the Bavarian Wikipedia page from which the example sentence was extracted.

Tab. 1: Explanation of the different columns used in the annotation table.

## Annotation Details

A word is considered a correct dialectal variant if it corresponds to a regional pronunciation or expression. Paraphrases, synonyms, as well as terms in Standard German or English should not be marked. Variants that differ in word type such as, e.g., “boarisch” (Bavarian adjective) for “Bayern” (Standard German noun), should be marked as **no**. For example, the following changes in inflectional morphology for nouns should be marked as **inflected**:

Inflection Type	Standard German	Bav. Variant: <i>yes</i>	Bav. Variant: <i>inflected</i>
<b>Number</b> (Singular, Plural) should be the same	Stadt	Schtatt, Sdadd	Schtädt, Schdedd
	Sprachen	Sprochn	Sproch
<b>Gender</b> (grammatical) should be the same.	Bürgermeister	Burgermäister, Birgermoischtr, Bürgermäister	Burgermaischteri
	Stadtpräsident	Schtadtpräsident, Stadtpresidänt	Stadtpresidänti, Stadtpresidäntin
<b>Case</b> (Nominative, Genitive, Accusative, Dative) should be the same.	Kinder	Kinda	Kinda’n

Please also consider the following guidelines for determining when the two words should be treated as unrelated (and marked as **no**).

Guideline	Standard German	Bav. Variant: <i>yes</i>	Bav. Variant: <i>no</i>
<b>Completeness and accuracy:</b> (partial) words should not be missing or added (e.g., in the case of compound words).	Zeitung	Zittung	Daageszittig
	Latein	Làtiinisch, Lätinisch	-

The following conditions should not be judged strictly. This means that variations are permitted as long as they do not violate any of the rules stated above.

Guideline	Standard German	Bav. Variant: <i>yes</i>	Bav. Variant: <i>inflected</i>
<b>Capitalization:</b> Variations are allowed.	Liebe	liebi	–
	Südfränkisch	sidfrankisch, sidfränkisch, Sidfränkisch	Siidfrenggische
<b>Hyphenation:</b> Variations are allowed.	Tages-Anzeiger	Dagesazeiger	–
	Santihansdoor	Santihanns-Door	–
	Westberlin	Wescht-Berlin	–