

Guidelines for the Morphological Annotation of Nouns in the Falko Learner Corpus

Anke Lüdeling^a, Julia Lukassek^a & Roodabeh Akbari^a

^aHumboldt-Universität zu Berlin

These guidelines describe the annotation of different types of nominal word formation in German. We also report on the evaluation of these guidelines in the Falko corpus.

Keywords: Nominal morphology, manual annotation, corpus linguistics

1 Research question and goal of the annotation

The complexity and the origin of nouns have been argued to vary with register: Academic registers, for example, contain more neoclassical words and it has been shown that they display a more 'nominal' style (see [Biber 2006](#), [Hennig 2020](#) among others) which results in more derivations. It has also been shown that nominal style involves a higher degree of modification and that learners of German use less (optional) modification than L1 speakers (cf. e. g. [Hirschmann 2015](#)). We are concerned with the acquisition of register-based variation in noun modification and nominal word formation here by learners of German as a Foreign Language.

The primary goal of these annotation guidelines is thus to prepare corpus data for research on modifiers within the nominal domain. Nouns are syntactic anchors for modifiers, cf. *Qualle* 'jellyfish' in (1a). However, they can also include modifiers in their internal structure in case the noun is a compound, cf. the modifier *Haus* 'house' within the compound *Haus•qualle* in (1b).¹

¹Some conventions: When we want to refer to the morphemes, we use the center dot to mark morpheme boundaries in these guidelines. We will write stems in lower case when we speak

- (1) a. Qualle, die im Haus gehalten wird
jellyfish that at home kept is
'jellyfish that is kept at home'
- b. Haus•qualle
house•jellyfish
'jellyfish that is kept at home'

The syntactic modification of nouns can be investigated using part-of-speech annotation and syntactic annotation (such as dependency parsing). In order to compare the frequency and types of modifiers in the nominal domain, we also need an annotation of the internal morphological structure of nouns. The word-internal structure can be annotated according to these guidelines.

The guidelines were developed for research questions on modifiers in the nominal domain within the Falko L1 and L2-corpus (Reznicek et al. 2012, Hirschmann et al. 2022). Falko is a freely available learner corpus containing written texts by advanced learners of German as a Foreign Language and German L1 control texts. The corpus is annotated on many layers pertaining to different levels of grammar, see Hirschmann et al. (2022) for an overview of annotation layers in Falko. However, the resulting annotation is supposed to enable different types of research questions. Furthermore, the guidelines do not include any restrictions specific to learner language, which makes them reusable for different types of corpus data. Apart from the documentation of compounds, it provides a full classification system for the word formation morphology of nouns in general.

1.1 Scope of the annotation in the Falko corpus

As stated, the guidelines can be used for any corpus. In this subsection, we briefly describe the specifics of our annotation. Falko includes a layer with part-of-speech (POS) annotations, following the Stuttgart-Tübingen Tagset.² The scope of the reported annotation in these guidelines amounts to all units that have the

about their morphological properties and mark affixes with a hyphen (such as the suffix *-ung* or the prefix *be-*). Verbs are quoted as stems instead of as infinitives when we focus on their potential in a morphologically complex word (cf. *lach* 'laugh' in *Lach•krampf* 'fit of laughter'). For these guidelines, we assume a binary model of word formation (loosely based on some form of an item-and-arrangement model). We also use the terms *head* and *non-head/modifier* here. German complex words are mostly right headed. We do not want to claim that this is necessarily how word formation works but it is a straightforward model for consistent annotation.

²STTS, see Schiller et al. (1999), <https://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/germantagsets/>

tags NN or TRUNC³ on the annotation layer ZH1gpos. This layer is based on the target hypothesis 1 (Zielhypothese 1 = ZH1). This is a layer with minimal corrections within the Falko corpus, cf. Reznicek et al. (2013) for an explanation.⁴ The ZH1 layer also functions as basis for the dependency parses and POS-annotations, which will be used in the noun modification study.

2 Morphological annotation: Annotation layers and tags

The annotation of word formation morphology is encoded on the layer **morph-anno**. We differentiate between native and non-native nouns. Native nouns receive one of the tags from Section 2.1 to Section 2.8 on this annotation layer. Non-native nouns are marked with a specific tag on the morph-anno layer and their morphological structure is specified in the layer **morph-anno-fremd**. Furthermore, we document the origin of a given noun or one of its parts on a third layer named **herkunft**. The precise course of action for non-native nouns is described in Section 2.9.

For morphologically complex nouns, we take the highest morphological process as the base for our annotation. For instance, the noun *Veränderungsmitteilung* ‘change notice’ is constructed in several steps (roughly: *Veränderung* is derived from the prefix verb *ver•änder* ‘to change’ and the suffix *-ung*, *Mitteilung* is derived from the particle verb *mit•teil* ‘to notify’ and the suffix *-ung*). The highest process is the compounding of *Veränderung* and *Mitteilung* and therefore the noun is tagged as compound.⁵

We analyze nouns within their context and in cases of lexical ambiguity, we consult this context for disambiguation. For instance, the noun *Druck* could be derived either from *drucken* ‘to print’ or *drücken* ‘to press’. Context will usually clarify whether the noun is a case of conversion (*print* reading) or the nominalisation (*press* reading).

The nouns annotated according to these guidelines are structurally transparent. We do not consider semantic transparency.

In the following sections, we describe all tags in detail.

³NN is the tag for common nouns (appellatives). Tokens with the tag TRUNC can be part of a nominal ellipsis, such as in *Haus- und Garten•türen*, lit. ‘house and garden doors’, which is expanded to ‘house doors and garden doors’. The truncated *Haus-* is given the label TRUNC in the STTS scheme.

⁴Falko is a multi-layer corpus, the original learner utterance is also available.

⁵In an additional step, we also use the semi-automatic process developed in the RUEG project (we are grateful to Martin Klotz and Nadine Zürn, see <https://korpling.german.hu-berlin.de/rueg-docs/standalone/noun-formation-german/>) to find all morphemes and the morphological structure of these nouns.

2.1 Determinative compound

We use the tag kdet for determinative compounds. They can be identified by the following criteria.

General criteria:

- The noun is composed of (at least) two freely occurring morphemes.
 - The head of the compound contributes the semantic core of the complex word.
 - The non-head determines (modifies) the head.
- (2) Haus•qualle, Quallen•käfig, Käfig•haltung, Haltungs•note,
home•jellyfish, jellyfish•cage, cage•keeping, posture•grade,
Noten•durchschnitt
grade•average

The examples in Example 2.1 are clear determinative compounds. In the following we add cases that could also receive other analyses. We decided to include them here as determinative compounds.

Specific cases:

- The noun consists of a nominalised infinitive and further VP-components functioning as non-heads
- (3) Weiter•geben, Bewusst•sein, Kuchen•essen
on•passing, conscious•being, cake•eating
- Compounds in which the head governs the non-head (or: the head is de-verbal and the non-head fills an argument role of the underlying verb, sometimes called verbal compounds, synthetic compounds, German: *Reaktionskomposita*) are labelled here as determinative compounds.
- (4) Quallen•zucht, Bürger•versther, Streit•schlichter
jellyfish•cultivation, citizen•understander, dispute•mediator
- The compound consists of a numeral and *Jahre* 'years'.⁶

⁶The abbreviated forms *Zwanziger* 'twenties' or *80er* 'eighties' are annotated as kurz.

- (5) Zwanziger•jahre
twenty•years

- If the complex noun consists of a deverbal noun (head) and a preposition (non-head) and there is no verb with the preposition as a prefix or particle, we treat the noun as a determinative compound.⁷

- (6) Vor•teil, Bei•spiel, Nach•richt, In•halt, An•walt, Aus•sicht
advantage, example, news, content, advocate, view

2.2 Copular compound

Copular compounds are a small subgroup of compounds (which occurs rarely in general and in Falko). We use the tag `kkop` for copular compounds. They can be identified by the following criteria.

- The noun consists of (at least) two freely occurring morphemes.
- The components are equal with respect to their semantic contribution.

- (7) Fürst•bischof, Spieler•trainer
prince•bishop, player•coach

2.3 Phrasal compound

Phrasal compounds consist of a lexical head and a phrase, e. g. a quantifier phrase, as non-head. Their complexity and spelling vary. Phrasal compounds are tagged `kphras`.

- (8) Trimm-dich-Pfad, Schönes-Wochenende-Ticket, 40-Stunden-Woche,
keep_fit-you-trail, nice-weekend-ticket, 40-hours-week,
(ein) ‘Ich schaue heute besser nicht in den Spiegel’-Tag, 1-Euro-Jobber
(a) ‘Today, I would rather not look in the mirror’-day, 1-Euro-worker

⁷This may be a controversial decision. Our reasoning here is that the verbs *vorteilen*, *beispielen*, *nachrichten*, *inhalten*, *anwalten* and *aussichten* are not attested and cannot, therefore, function as a base for deverbal nominalisation. The cases here are typically highly lexicalized.

2.4 Derivations

For derivations, we use the tag der. The following criterion identifies derivations:

- The noun consists of at least one freely occurring morpheme and at least one discernible affix.

- (9) Halt•ung, Weck•er,
keep+affix_{abstract-object}, wake+affix_{agent},
Möglich•keit, Versuch•ung
possible+affix_{abstract-object}, tempt+affix_{abstract-object}
keeping, alarm clock, possibility, temptation

2.5 Conversion

If a stem form occurs in the nominal domain and some other morphosyntactic category, but no suffix signals the derivation, we assign the tag kon for conversion. The direction of the conversion is irrelevant, i. e. all nouns that participate in a conversion relation as source or result are assigned to this class.⁸ The following criterion is used for identification:

- The noun occurs as a stem in a different morphosyntactic category than noun.

- (10) Verlauf, Grün, Besuch, Französisch, Bett, Salz
course_N, green_N, visit_N, French_N, bed_N, salt_N

2.6 Transposition

Syntactic reclassifications of inflected forms are tagged trans for transposition. The following criteria are used for identification:

- The noun corresponds formally to a verbal infinitive.

⁸We decided on this very broad understanding of conversion so that we can later see whether and how morphological networks are formed. We expect that many cases will be overlooked during annotation. Therefore, all simplex nouns will be cross-checked for potential conversion relations after annotation. In case they occur in a different morphosyntactic class, we revise the annotation and assign the tag kon.

- (11) (das) Sitzen, (zum) Heulen, Leben, Unternehmen
(the) sit_N, (to_the) cry_N, live_N, undertake_N
sitting, crying, life, enterprise
- The noun corresponds formally to a inflected adjective
- (12) (der) Grüne, (die) Jugendlichen
(the) green_N, (the) young_N
(the) green / (the) member of the green party, (the) adolescents
- The noun corresponds formally to an adjectivised participle
- (13) (ein) Verrückter, Studierende
(an) insane_N, studying_N
(an) insane person, students

2.7 Other nominalisations

Some nominalisations cannot be assigned to any of the aforementioned categories, but they are related to stems in other categories. These nominalisations are tagged as nom. They can be identified by the following criteria:

- The noun is formed by the nominalising suffixes *-t*, *-st* or *-e* from a verbal or adjectival base.⁹
- (14) Fahrt, Aufgabe, Suche, Höhe, Stärke, Emanze
journey, task, search, height, strength, women's libber
- The noun is derived from a verb by changing the stem vowel or the stem form (often called implicit derivation).
- (15) Abschluss, Bezug, Zugang, Zukunft, Herkunft, Begierde,
graduation, relation, access, future, origin, desire,
Behörde
authorities

⁹Strictly speaking, this would be derivation. The suffixes are not productive (except perhaps *-e* in some cases) and very often not even transparent.

- The noun consists of a verbal stem affixed by the circumfix *Ge...(-e)* or a morpheme variant of a verbal stem.¹⁰

(16) Gesetz, Gesuch, Gespräch, Gedanke, Geschichte, Geheue,
law, request, conversation, thought, history, honking,
Gelaber
blabbering

- The noun resembles an infinitive, but different from actual transpositions it has masculine grammatical gender.

(17) (der) Schaden, Nutzen, Laden, Schrecken, Rahmen
(the) damage, benefit, shop, horror, frame

2.8 Simplex

We use the tag sim for nouns without (synchronically discernible) internal structure.¹¹ The following criteria identify simplex nouns:

- The noun is a freely occurring morpheme.
- The noun does not have any internal structure.

(18) Baum, Hund, Hand, Wolke, Mut
tree, dog, hand, cloud, courage

Plural nouns with no singular form (pluralia tantum) like *Eltern* ‘parents’, *Leute* ‘people’ are treated as simplex nouns. The fact that they have an obligatorily plural form is not documented.

2.9 Non-native nouns

Non-native nouns of neoclassical, English, or other origin are annotated on three layers.

- **morph-anno:** Non-native nouns are tagged as nnat, except for cases where they follow a native word formation rule on the highest morphological level. For such cases, see the tags in Section 2. Examples are:

¹⁰We acknowledge that the circumfix *Ge...-e* allows for new word formations and is productive in some readings. In order to keep the annotation simple and systematic, we nevertheless treat even newly or ad hoc formed cases of *Ge...-e* nouns as nom.

¹¹For the Falko annotation, we use this tag generously. If relatedness to a stem in a different category is not visible, we annotate sim even if there might be some historical relationship.

- determinative compounds: *Radiofeature* ‘radio feature’, *Lektüreliste* ‘reading list’
- derivations: *Jobber* ‘worker’, *Regierung* ‘government’
- transposition: *Theoretisieren* ‘theorising’

<

- **morph-anno-fremd**: This layer is used for all non-native simplex nouns or nouns that follow non-native word formation rules on the highest morphological level. Two tags are used in this layer:

- com: The noun has a discernible internal structure following non-native or neo-classical word formation rules or it is an English compound. Central for this class is the visibility of systematic word formation processes. Examples:

- (19) *Demokratie* ‘democracy’ (reason: formations like *Autokratie* ‘autocracy’, *Bürokratie* ‘bureaucracy’ and others prove that the word formation process is systematic)
- (20) *Theorie* ‘theory’ (reason: there are derivations like *theoretisieren* ‘to theorize’ and *theoretisch* ‘theoretic’ without the affix *-ie*),
- (21) *Automatismus* ‘automatism’ (reason: formations like *Feminismus* ‘feminism’, *Chauvinismus* ‘chauvinism’, *Optimismus* ‘optimism’ and others prove that the word formation process with *-ismus* is systematic)
- (22) *Innovation* ‘innovation’ (reason: formations like *Religion* ‘religion’ and *Faszination* ‘fascination’ prove that the word formation process with *-ion* is systematic.)
- (23) *Teamwork*, *Deadline* (reason: these words are not formed productively in German but are borrowed as complex words. Nevertheless, they have an internal structure.)

The most frequent suffixes in this class are:¹² *-ament/-ement*, *-ant/-ent*, *-anz/-enz*, *-age*, *-ar/-är*, *-arium*, *-at*, *-aille*, *-ade*, *-asmus/-ismus*, *-ee*, *-esse*, *-elle*, *-ette*, *-(er)ie*, *-eur*, *-ier*, *-iere*, *-ik*, *-iker*, *-ine*, *-(at/-t/-x)ion*, *-ist*, *-(i)tät*, *-(at/it)or*, *-ose*, *-ur*, *-iv*, *-al/ell*, *-abel/-ibel*, *-ive*, *-logie*,

¹²Our list is basically an extended version of the list provided in Meibauer et al. (2002: p. 55)

-loge, -kratie, -krat, -grafie, -graf, -sophie, -soph, -nomie, -nom, -esk, -oid, -os/-ös;

The most frequent prefixes are: a-/an-, anti-, de-/des-, dis-, ex-, hyper-, in-/il-/im-/ir-, inter-, ko-/kon-/kol-/kor-, kom-, non-, para-, prä-, post-, pro-, re-, super-, trans-, ultra-

- sim: The noun is non-native and has no morphological structure that would be discernible for a German speaker (e. g. *Profil* ‘profile’, *Fake* ‘fake’, *Niveau* ‘level’).¹³
- **herkunft**: All nouns that include a non-native component are annotated with one of the following tags:
 - neokl: Nouns of Latin or Greek origin that may follow German word formation rules and are pronounced according to German pronunciation rules.
 - engl: Nouns of English origin with a pronunciation resembling English pronunciation rules.
 - fremd: Nouns of any other foreign origin.

For complex nouns, it suffices for the assignment of a tag on this level if one of the components is non-native. If a compound is formed from two non-native words with different origins, the tag is chosen according to the head noun. For instance, the compound *Kriminalitätsniveau* ‘criminality level’ receives the tag fremd, as the head is of French origin.

All native nouns are tagged as nat on this annotation layer.

2.10 Other conventions

- **Abbreviations**: All formations that involve a process of abbreviation are tagged as kurz on the morph-anno layer. (e. g. *B.*; *PKW*, *Uni*, *Bafög*, *Sechziger*, *80er*, *§*). On the **herkunft** layer, they are all annotated with nat.
- **Nominalised complex VPs**: Formations like *Stellung•nahme* ‘statement’ or *Bezug•nahme* ‘reference’, whose derivation is disputable, as they are related to complex lexicalised VPs, are tagged as nom on the **morph-anno** layer.

¹³This is of course a simplification, as non-native nouns of Greek, Latin, English or any other origin can be structurally complex according to the original language’s word formation rules. However, this is irrelevant for the annotation intended in our guideline, as we want to represent the linguistic knowledge of a German speaking person.

Guidelines for the Morphological Annotation of Nouns

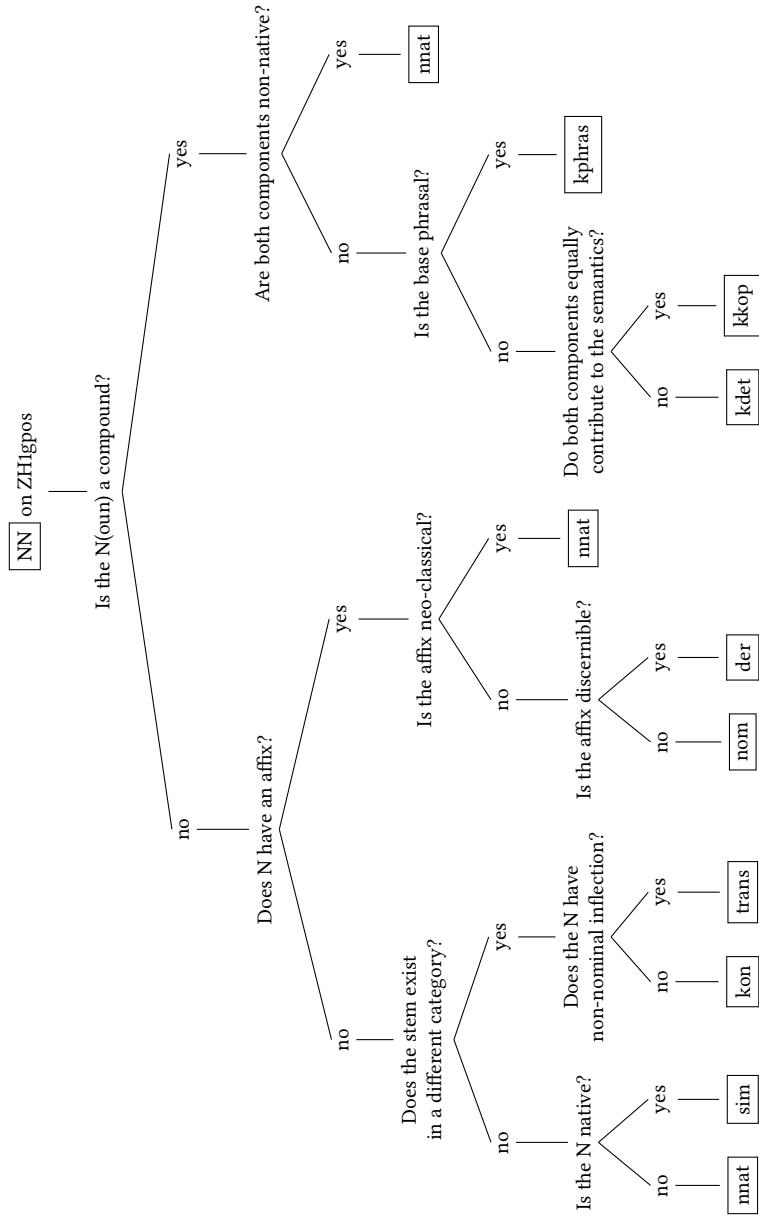
- If a formation based on a derivation or transposition and a compound cannot unequivocally be traced back to one order of processes, we use the tag `kdet` (e. g. *Hoch•schul•berechtigte* ‘somebody who is qualified to attend university’).
- In some rare cases ambiguities cannot be resolved contextually. We use the tag `amb` on the **morph-anno** layer, if no disambiguation is possible.
- If the meaning of a noun is unknown to the annotators, it is tagged as `unknown` on the **morph-anno** layer.
- For nouns that are tagged as TRUNC on ZH1gpos, we use the noun including its elided part as a base for the morphological annotation. For instance, in *Ehe- und Hausfrau* ‘housewife’ (literally ‘marriage- and house•woman’) we annotate `kdet` for *Ehe-*, as *Ehefrau* is the base for the annotation.

Table 1: Annotation layers and their tags

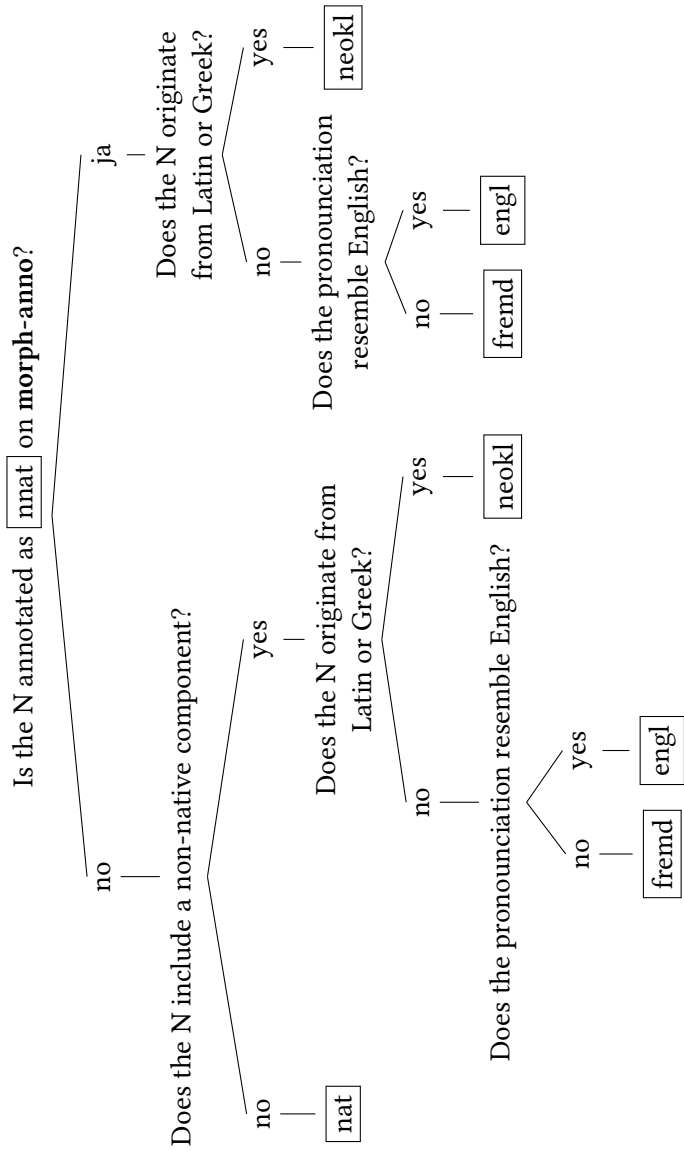
Layer	Tags
morph-anno	kdet, kkop, kphras, der, kon, trans, nom, sim, nnat, kurz, amb, unknown
morph-anno-fremd	sim, com
herkunft	nat, neokl, engl, fremd

2.11 Decision trees

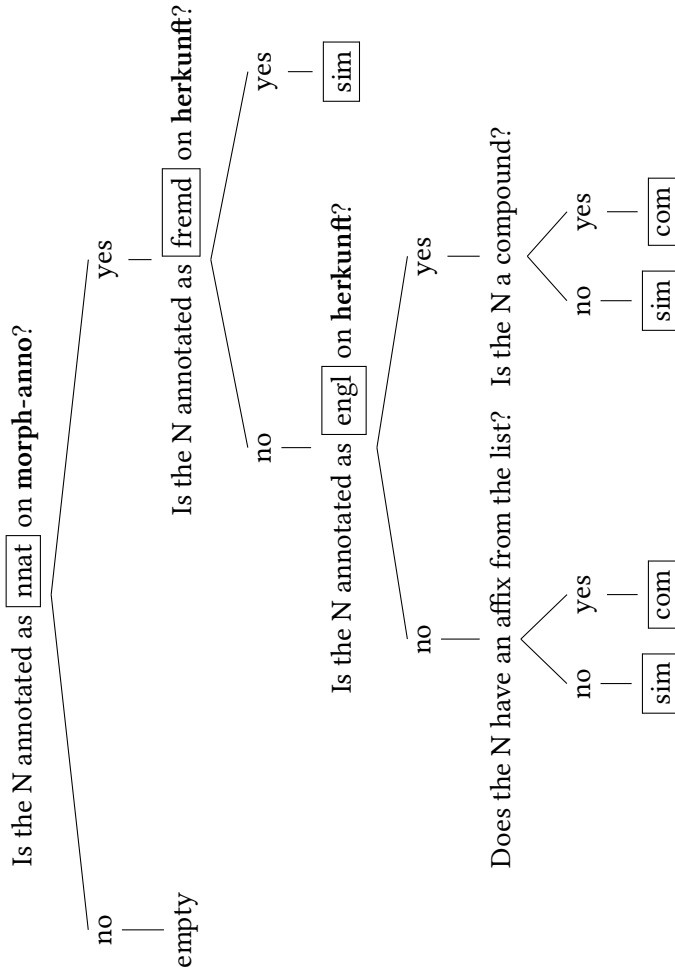
2.11.1 Layer *morph-anno*



2.11.2 Layer *herkunft*



2.11.3 Layer *morph-anno-fremd*



2.12 Evaluation of the guidelines

The guidelines were developed in several iterations. We wrote a version of the guidelines and then proceeded to test it on a small sub-corpus taken from the Falko corpus. Usually, two or three parallel annotations were done in each loop. Each test loop included a semi-automatic comparison of the assigned tags for each token. This resulted in a list of mismatches we discussed in our meetings and accounted for in the next version of the guidelines.

The corpus texts were annotated in spreadsheets, where each line corresponded to one token. The texts were automatically preprocessed such that for every token that was annotated as NN on ZH1gpos a dummy tag was added on the **herkunft** layer. This preprocessing step ensured that all nouns would be identified and annotated. Furthermore, we excluded any other entries than the tags in Table 1 in order to avoid typing errors.

Prior to the actual annotation, we asked three annotators that were not involved in the development of these guidelines to do a test run. They were instructed by one of the authors and received the annotation guidelines together with five texts that were preprocessed in the described manner. The annotators were supposed to annotate the texts accordingly. Based on these annotations, we computed a Fleiss' kappa (Artstein & Poesio 2008) for three raters on each annotation layer. The results are presented in Table 2.

Table 2: Inter-annotator agreement for each layer measured by Fleiss' kappa

Layer	Fleiss' kappa
morph-anno	0.81
morph-anno-fremd	0.57
herkunft	0.85

Our annotation guidelines reached almost perfect agreement on the **morph-anno** and **herkunft** layers. For the **morph-anno-fremd** layer, the agreement was moderate. The weaker agreement on morph-anno-fremd can be explained by the fact that on this layer annotations are generally only necessary for a small amount of nouns, namely those that were tagged `nmat` on morph-anno. Due to the low overall number of nouns with this tag in the five annotated texts, each mismatch weighed into the agreement measures stronger than it did on the other layers.

After the guidelines were tested and we reached sufficient agreement even with annotators that were not involved in the development of the guidelines, we annotated both essay sub-corpora of the Falko corpus (L1 and L2). In order to provide additional quality control, 10 % of the texts were annotated by two annotators. All identified mismatches were checked and corrected. Furthermore, we conducted a series of automatic corrections. First, we checked a list of noun tokens¹⁴ for which we expected possible mismatches. The list is available in the appendix (Section 3). If the annotation of one of these nouns differed from the one provided in the list, it was corrected automatically. Importantly, the words included in the list are not exceptions from our annotation guidelines, but nouns that occurred with a moderate frequency and were difficult to annotate. Second, we verified the consistency of annotations by searching for instances of the same noun and comparing the annotations. If they were different, we examined the case and corrected it, if necessary. Third, we verified whether all nouns were annotated.

Funding information

The research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1412, 416591334.

Acknowledgements

The authors are grateful to Sophia Döring, Pia Linscheidt, and Sarah Schneider for the test annotation and valuable feedback. We are also grateful for the very helpful comments by two anonymous reviewers.

References

- Artstein, Ron & Massimo Poesio. 2008. Inter-coder agreement for computational linguistics. *Computational Linguistics* 34(4). 555–596.
- Biber, Douglas. 2006. *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.
- Hennig, Mathilde. 2020. *Nominalstil. Möglichkeiten, Grenzen, Perspektiven*. Tübingen: Narr.

¹⁴As our annotation files did not include noun lemmatisation, we had to add all inflected variants of the nouns in our list.

- Hirschmann, Hagen. 2015. *Modifikatoren im Deutschen. Ihre Klassifizierung und varietätenspezifische Verwendung*, vol. 89 (Studien zur deutschen Grammatik). Tübingen: Stauffenburg.
- Hirschmann, Hagen, Anke Lüdeling, Anna Shadrova, Dominique Bobeck, Martin Klotz, Roodabeh Akbari, Sarah Schneider & Shujun Wan. 2022. Falko. Eine Familie vielseitig annotierter Lernerkorpora des Deutschen als Fremdsprache. *Zeitschrift Korpora Deutsch als Fremdsprache (KorDaF)*. 139–148. DOI: <https://doi.org/10.48694/kordaf.3552>.
- Meibauer, Jörg, Ulrike Demske, Jochen Geilfuß-Wolfgang, Jürgen Pafel, Karl Heinz Ramers, Monika Rothweiler & Markus Steinbach. 2002. *Einführung in die germanistische Linguistik*. Stuttgart: Metzler.
- Reznicek, Marc, Anke Lüdeling & Hirschmann. 2013. Competing Target Hypotheses in the Falko Corpus: A Flexible Multi-Layer Corpus Architecture. In *Automatic treatment and analysis of learner corpus data*, 101–123. Amsterdam: John Benjamins.
- Reznicek, Marc, Maik Walter, Karin Schmidt, Anke Lüdeling, Hagen Hirschmann, Cedric Krummes & Torsten Andreas. 2012. *Das Falko-Handbuch: Korpusaufbau und Annotationen*. Tech. rep. Institut für deutsche Sprache und Linguistik, Humboldt-Universität zu Berlin. https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko/FalkoHandbuchV2/at_download/file.
- Schiller, Anne, Simone Teufel, Christine Stöckert & Christine Thielen. 1999. *Guidelines für das Tagging deutscher Textcorpora mit STTS*. Universität Stuttgart, Institut für maschinelle Sprachverarbeitung; Universität Tübingen, Seminar für Sprachwissenschaft.

Appendix

Table 3: Nouns checked and corrected automatically after annotation

noun	morph-anno	morph-anno-fremd	herkunft
§	kurz		nat
%	kurz		nat
€	kurz		nat
1-Euro-Jobber	kphras		engl

Table 3: Nouns checked and corrected automatically after annotation

noun	morph-anno	morph-anno-fremd	herkunft
40-Stunden-Woche	kphras		nat
60ern	kurz		nat
70er-	kurz		nat
70er”	kurz		nat
70ern	kurz		nat
8-Stunden-Tag	kphras		nat
80ern	kurz		nat
Abertausende	kdet		nat
Absichten	nom		nat
Acht	kon		nat
Akt	nmat	sim	neokl
Alkohol	nmat	sim	fremd
Alter	kon		nat
Analyse	nmat	com	neokl
Anbetracht	kon		nat
Anreiz	kon		nat
Anschein	kon		nat
Anspruch	nom		nat
Ansprüche	nom		nat
Anzahl	kdet		neokl
Arbeit	kon		nat
Art	kon		nat
Aufenthalt	nom		nat
Aufenthalte	nom		nat
Ausdruck	nom		nat
Ausdrücke	nom		nat
Auto	nmat	sim	neokl
Autokratie	nmat	com	neokl
Autos	nmat	sim	neokl
B.	kurz		nat

Guidelines for the Morphological Annotation of Nouns

Table 3: Nouns checked and corrected automatically after annotation

noun	morph-anno	morph-anno-fremd	herkunft
Bachelor	nnat	com	nat
Bachelorn	nnat	com	engl
Bachelors	nnat	com	engl
Begierde	nom		nat
Behörde	nom		nat
Behörden	nom		nat
Beispiel	kdet		nat
Bereich	nom		nat
Bereiche	nom		nat
Bereichen	nom		nat
Beruf	kon		nat
Berufe	kon		nat
Berufen	kon		nat
Beute	sim		nat
Bezirk	sim		nat
Bibel	nnat	sim	neokl
Büro	sim		nat
Cent	nnat	sim	neokl
Cents	nnat	sim	neokl
Chance	nnat	sim	fremd
Chancen	nnat	sim	fremd
Chauvinismus	nnat	com	neokl
Chef	nnat	sim	fremd
Chefs	nnat	sim	fremd
Computer	nnat	com	engl
Dealer	der		engl
Debatte	nnat	com	fremd
Debatten	nnat	com	fremd
Delikt	nnat	com	neokl
Demokratie	nnat	com	neokl

Table 3: Nouns checked and corrected automatically after annotation

noun	morph-anno	morph-anno-fremd	herkunft
Diebstahl	kdet		nat
Ding	sim		nat
Dingen	sim		nat
Dings	sim		nat
Ebene	nom		nat
Ebenen	nom		nat
Einblick	kon		nat
Emanze	nom		neokl
Emanzen	nom		neokl
Ende	nom		nat
Endes	nom		nat
Epoche	nmat	sim	neokl
Epochen	nmat	sim	neokl
Erde	nom		nat
Erfolg	kon		nat
Euro	sim		nat
Fall	kon		nat
Fälle	kon		nat
Fällen	kon		nat
Familie	nmat	sim	neokl
Familien	nmat	sim	neokl
Feminismus	nmat	com	neokl
Feministinnen	der		neokl
Fernsehen	trans		nat
Fernseher	der		nat
Film	nmat	sim	engl
Filme	nmat	sim	engl
Filmen	nmat	sim	engl
Films	nmat	sim	engl
Finanzen	nmat	com	neokl

Table 3: Nouns checked and corrected automatically after annotation

noun	morph-anno	morph-anno-fremd	herkunft
Firma	nnat	sim	fremd
Folie	nnat	sim	neokl
Folien	nnat	sim	neokl
Fortbildung	der		nat
Fortbildungen	der		nat
Fortschritt	nom		nat
Fortschritten	nom		nat
Fragestellung	kdet		nat
Fragestellungen	kdet		nat
Freude	nom		nat
Freuden	nom		nat
Gebiet	nom		nat
Gebieten	nom		nat
Gegeneinander	kdet		nat
Gegenwart	kdet		nat
Geisteswissenschaftler	kdet		nat
Grenze	nom		nat
Grenzen	nom		nat
Grund	nom		nat
Gründe	nom		nat
Gründen	nom		nat
Gruppe	sim		nat
Haschisch	nnat	sim	fremd
Heroin	nnat	sim	fremd
Ikone	nnat	sim	neokl
Individuen	nnat	com	neokl
Individuum	nnat	com	neokl
Inhalt	kdet		nat
Inhalte	kdet		nat
Inhalten	kdet		nat

Table 3: Nouns checked and corrected automatically after annotation

noun	morph-anno	morph-anno-fremd	herkunft
Inhalts	kdet		nat
Interesse	nnat	com	neokl
Internet	nnat	com	engl
Jahr	nom		nat
Jahre	nom		nat
Jahren	nom		nat
Jugend	sim		nat
Jugendliche	trans		nat
Jugendlichen	trans		nat
Jugendlicher	trans		nat
Junge	nom		nat
Jungen	nom		nat
Jungs	nom		nat
Kampf	nom		nat
Karriere	nnat	sim	fremd
Karrieren	nnat	sim	fremd
Kino	nnat	sim	neokl
Klausel	nnat	sim	neokl
Kokain	nnat	sim	fremd
Konflikt	nnat	sim	neokl
Konflikte	nnat	sim	neokl
Konflikten	nnat	sim	neokl
Konflikts	nnat	sim	neokl
Kontakt	nnat	sim	neokl
Kontakte	nnat	sim	neokl
Kontakten	nnat	sim	neokl
Kontakts	nnat	sim	neokl
Kontrolle	nnat	com	fremd
Kontrollen	nnat	com	fremd
Konzern	nnat	sim	neokl

Guidelines for the Morphological Annotation of Nouns

Table 3: Nouns checked and corrected automatically after annotation

noun	morph-anno	morph-anno-fremd	herkunft
Konzerne	nnat	sim	neokl
Konzernen	nnat	sim	neokl
Konzerns	nnat	sim	neokl
Kopf	nom		nat
Köpfe	nom		nat
Köpfen	nom		nat
Kopfes	nom		nat
Kosten	kon		nat
Krawall	sim		nat
Krawalle	sim		nat
Krawallen	sim		nat
Krawalls	nom		nat
Kriminalität	nnat	com	neokl
Kriminelle	trans		neokl
Kriminellem	trans		neokl
Kriminellen	trans		neokl
Krimineller	trans		neokl
Kriminelles	trans		neokl
Kriterien	nnat	com	neokl
Kriterium	nnat	com	neokl
Kritik	nnat	com	neokl
Küche	sim		nat
Kumpanen	nnat	sim	fremd
Kunde	sim		nat
Kunden	sim		nat
Laden	nom		nat
Läden	nom		nat
Land	kon		nat
Linie	nnat	sim	neokl
Linien	nnat	sim	neokl

Table 3: Nouns checked and corrected automatically after annotation

noun	morph-anno	morph-anno-fremd	herkunft
Literatur	nnat	com	neokl
Mafia	nnat	sim	fremd
Magister	nnat	sim	neokl
Magistern	nnat	sim	neokl
Magisters	nnat	sim	neokl
Manager	der		engl
Material	nnat	com	neokl
Menge	sim		nat
Minute	nnat	sim	neokl
Miteinander	kdet		nat
Mode	nnat	sim	neokl
Module	nnat	sim	neokl
Moment	nnat	sim	neokl
Monat	nnat	sim	neokl
Monate	nnat	sim	neokl
Monaten	nnat	sim	neokl
Monats	nnat	sim	neokl
Motiv	nnat	com	neokl
Module	nnat	sim	neokl
Naturwissenschaftler	kdet		nat
Nutzen	nom		nat
Opfer	kon		nat
Park	sim		nat
Partei	nnat	sim	neokl
Parteien	nnat	sim	neokl
Partner	nnat	com	neokl
Patrizier	der		neokl
Person	nnat	sim	neokl
Pflicht	nnat		neokl
Platz	sim		nat

Guidelines for the Morphological Annotation of Nouns

Table 3: Nouns checked and corrected automatically after annotation

noun	morph-anno	morph-anno-fremd	herkunft
Polizei	nmat	sim	neokl
Polizist	nmat	com	neokl
Polizisten	nmat	com	neokl
Praktika	nmat	com	neokl
Praktikum	nmat	com	neokl
Praxis	nmat	com	neokl
Prestige	nmat	sim	neokl
Problem	nmat	sim	neokl
Projekte	nmat	com	neokl
Punkt	kon		nat
Punkte	kon		nat
Punkten	kon		nat
Rahmen	nom		nat
Recht	kon		nat
Rechte	kon		nat
Rechten	kon		nat
Rechts	kon		nat
Regierung	der		neokl
Regierungen	der		neokl
Religion	nmat	com	neokl
Religionen	nmat	com	neokl
Risiko	nmat	sim	fremd
Rolle	nom		nat
Schaden	nom		nat
Schicksal	der		nat
Schicksale	der		nat
Schrecken	nom		nat
Schulden	nom		nat
Sekunde	nmat	sim	neokl
Seminar	nmat	com	neokl

Table 3: Nouns checked and corrected automatically after annotation

noun	morph-anno	morph-anno-fremd	herkunft
Seminare	nnat	com	neokl
Seminars	nnat	com	neokl
Sicht	kon		nat
Sichten	kon		nat
Sinn	kon		nat
Sinne	kon		nat
Sinnen	kon		nat
Sinnes	kon		nat
Sport	nnat	sim	engl
Stereotyp	nnat	com	neokl
Stress	nnat	sim	engl
Studie	nnat	com	neokl
Studien	nnat	com	neokl
Studium	nnat	com	neokl
Stunde	nom		nat
Stunden	nom		nat
Sucht	nom		nat
Summe	nnat	sim	neokl
Summen	nnat	sim	neokl
Tag	kon		nat
Tage	kon		nat
Tagen	kon		nat
Tages	kon		nat
Tags	kon		nat
Tausende	trans		nat
Text	nnat	sim	neokl
Texte	nnat	sim	neokl
Texten	nnat	sim	neokl
Textes	nnat	sim	neokl
Theorie	nnat	com	neokl

Guidelines for the Morphological Annotation of Nouns

Table 3: Nouns checked and corrected automatically after annotation

noun	morph-anno	morph-anno-fremd	herkunft
Theorien	nmat	com	neokl
Therapie	nmat	com	neokl
Therapien	nmat	com	neokl
Theorie	nmat	com	neokl
These	nmat	sim	neokl
Thesen	nmat	sim	neokl
Titel	nmat	com	neokl
Titeln	nmat	com	neokl
Titels	nmat	com	neokl
Tod	nom		nat
Tode	nom		nat
Traum	nom		nat
Träume	nom		nat
Träumen	nom		nat
Umfang	kon		nat
Vernunft	nom		nat
Waffe	sim		nat
Waffen	sim		nat
Weise	nom		nat
Weisen	nom		nat
Welle	nom		nat
Wellen	nom		nat
Wort	sim		nat
Worte	sim		nat
Worten	sim		nat
Wörter	sim		nat
Zigarette	nmat	com	neokl
Zigaretten	nmat	com	neokl
Unternehmen	trans		neokl
Einkommen	trans		nat

Table 3: Nouns checked and corrected automatically after annotation

noun	morph-anno	morph-anno-fremd	herkunft
Maschine	nmat	sim	neokl
Maschinen	nmat	sim	neokl
Armut	sim		nat
Urlaub	sim		nat
Zeitung	der		nat
Respekt	nmat	com	neokl
Angelegenheit	kdet		nat
Fach	sim		nat
Fächer	sim		nat
Fächern	sim		nat
Lust	sim		nat
Zahl	kon		nat
Aspekt	nmat	com	neokl
Aspekte	nmat	com	neokl
Aspekten	nmat	com	neokl
Englisch	kon		nat
Deutsch	kon		nat
Französisch	kon		nat
Spanisch	kon		nat
Russisch	kon		nat
Angst	sim		nat
Ängste	sim		nat
Ängsten	sim		nat
Italienisch	kon		nat
Schweden	kon		nat
Griechisch	kon		nat
Fenster	sim		nat
Haus	kon		nat
Hause	kon		nat
Häuser	kon		nat

Guidelines for the Morphological Annotation of Nouns

Table 3: Nouns checked and corrected automatically after annotation

noun	morph-anno	morph-anno-fremd	herkunft
Häusern	kon		nat
Hauses	kon		nat
Stellungnahmen	kdet		nat
Beamter	nom		nat
Beamte	nom		nat
Beamten	nom		nat
Thema	nnat	sim	neokl
Hinsichten	nom		nat
Hinsicht	nom		nat
Sklave	sim		nat
Sklaven	sim		nat
Gedanke	nom		nat
Gedanken	nom		nat
Alltag	kdet		nat
Arbeitslose	trans		nat
Arbeitsloser	trans		nat
Arbeitslosen	trans		nat
Unterricht	kon		nat
Buch	kon		nat
Bücher	kon		nat
Büchern	kon		nat
Studiums	nnat	com	neokl
Kultur	nnat	com	neokl
Kulturen	nnat	com	neokl
Projekt	nnat	com	neokl
Projekten	nnat	com	neokl
Projekts	nnat	com	neokl
Praktikums	nnat	com	neokl
Probleme	nnat	sim	neokl
Hierarchie	nnat	com	neokl

Table 3: Nouns checked and corrected automatically after annotation

noun	morph-anno	morph-anno-fremd	herkunft
Hierarchien	nnat	com	neokl
Medizin	nnat	com	neokl
Themen	nnat	sim	neokl
Argument	nnat	com	neokl
Argumente	nnat	com	neokl
Argumenten	nnat	com	neokl
Arguments	nnat	com	neokl
Natur	nnat	sim	neokl
Naturen	nnat	sim	neokl
Kollege	nnat	com	neokl
Kollegen	nnat	com	neokl
Politiker	nnat	com	neokl
Musik	nnat	com	neokl
Magazin	nnat	sim	fremd
Magazine	nnat	sim	fremd
Magazinen	nnat	sim	fremd
Magazins	nnat	sim	fremd
Million	nnat	sim	neokl
Millionen	nnat	sim	neokl