


## Annotating German in Austria: A Case-study of manual annotation in and for digital variationist linguistics

Markus Pluschkovits <markus\_dot\_pluschkovits\_at\_univie\_dot\_ac\_dot\_at>, University of Vienna 

### Abstract

The following is a case study of manual annotation as used in a large-scale variationist linguist project focusing of spoken varieties of German in Austria. It showcases the technical architecture used for annotation, which unifies different granularities of tokenization through the corpus in a distinct entity designed for annotation. It furthermore lays out the hierarchical stand-off annotation used in the project Deutsch in Österreich (“German in Austria”) and demonstrates how a semantic model for the hierarchical organization of annotations can bring transparency to the annotation process while also providing a sounder epistemological basis than monodimensional annotation.

## 1. Introduction

The process of annotation has been referred to as one of the oldest and most ubiquitous research practices there is (cf. [Lordick et al. 2016, 187]), and yet, especially in linguistics, specific annotations and their annotation guidelines are seldomly published or discussed, and are often unavailable (cf. [Ide 2017, 2]). Moreover, with some exceptions (e.g. [Ide and Pustejovsky 2017], [Consten and Loll 2012]), the process of annotation in linguistics itself is rarely subject to critical reflection, despite being fundamental for many branches of empirical linguistics. The purpose of this case-study is to add to this reflection process of annotation in linguistics, especially from a variationist linguist perspective. Furthermore, this case-study showcases examples of a hierarchical annotation system with a meaningful structure, developed by and used in the FWF-funded<sup>[1]</sup> Special Research Program “Deutsch in Österreich” (FWF F060 “German in Austria,” henceforth DiÖ). Additionally, it aims to contribute to the discussion of the epistemological status of annotation (in the sense of classification of data, see Section 4 below). We also propose that the annotation structure used by DiÖ can be generalized for a variety of research interests.

The paper is structured as follows: Section 2 gives context to the DiÖ research project, which is the basis of this case-study. Section 3 discusses automatic and manual annotation in linguistics generally and the DiÖ project specifically. It lays out the workflow and provides examples, such as the annotation of tag question markers and the phonetic variation of <-ig> endings. These examples are embedded in a discussion of the structure of annotations. Section 4 broadens the scope beyond this specific case-study, moving from the particularities of annotation in the DiÖ-project to annotation as an (often under-theorized) problem field in the (digital) humanities in general. The paper closes with a conclusion in Section 5, aiming to synthesize the insights for a broader audience beyond linguistics.

## 2. Context

This section gives a brief context to both the theoretical, linguistic assumptions behind the DiÖ project in Section 2.1 and methodological and technical aspects in Section 2.2.

### 2.1 Theoretical Assumptions

DiÖ is a variationist linguistic research program investigating the variation, contact and perception of different varieties of German in Austria. The variety of research goals of DiÖ necessitates heterogeneous data sources (ranging from historical census data on multilingualism to recent spoken data) and methods (ranging from traditional sociolinguistic interviews and online questionnaires to computer-aided language production experiments). The following will focus on the research on the variation of spoken German in Austria and the development of a collaborative online research platform (for more general information on the DiÖ project, see e.g. [Budín et al. 2019] or [DiÖ 2022]).

In order to elaborate on the annotation strategies used, a brief overview of the research goals as well as the data used in the project, both from a formal and a semantic point of view is necessary. Additionally, we give a short contextualization of the variationist linguistic research context. Within DiÖ, inter- and intraindividual variation of different varieties of German are a focus of research. The basic assumption underlying this field is that there is not a single, coherent German variety, but rather different varieties of the language, with different speakers having access to different subsets of German (they have, so to say, access to a repertoire of different ways of speaking – see e.g. [Koppensteiner and Lenz 2021]). These different varieties are made up and signified by certain variants of a linguistic variable (e.g. different pronunciations of one and the same word, or different terms for the same concept). While some speakers are assumed to have their language use structured as one variety with different poles, other speakers are assumed to have two (or more) distinct and separated varieties. Usually, these poles of one (or two distinct) variety/ies are associated with specific extra-linguistic factors, such as the communicative situation. In a specific communicative situation, certain variants indicate either the usage of e.g. high-prestige varieties (usually corresponding to standard language) or a more localized variety (usually dialectal). The research goal – which can be roughly summarized as the question of “Who speaks with whom in what way in which situation?” – is therefore the organization of these varieties within the speakers, how they relate to the language system itself and to extralinguistic factors, such as the age of the speaker, their dialect region, or the appropriateness for specific situations, and how these varieties are changing.

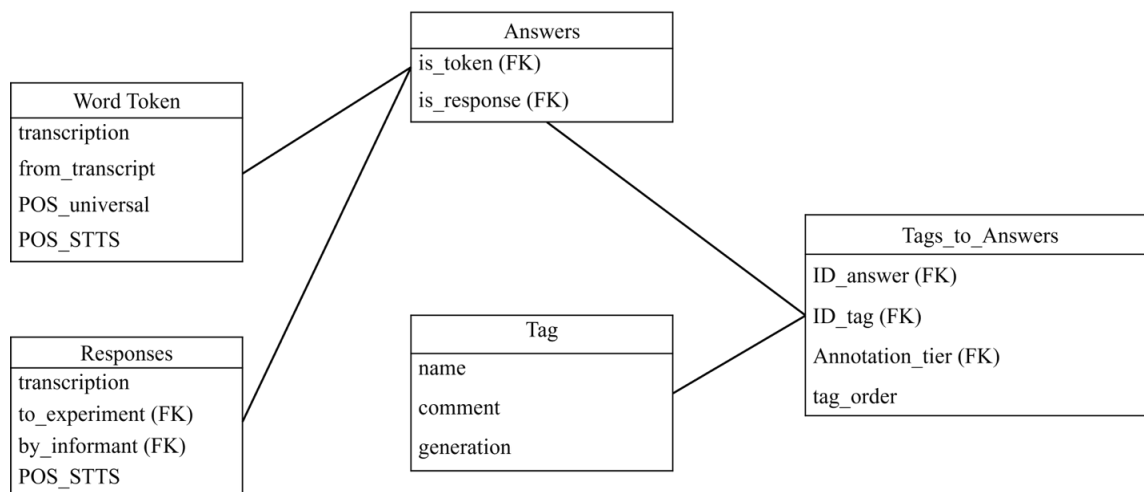
### 2.2 Methodological and Technical Aspects

To reflect these complex, interlinking factors, we collected language data in a variety of ways: we used fairly traditional, sociolinguistic research methods, such as structured interviews or conversations among peers in absence of a researcher. Additionally, more recent methods from the field of digital humanities were used as well, such as computer-assisted language production experiments. These language production experiments were used to evoke variants usually occurring too rarely in conversations to make reliable claims about their distribution, e.g. relative clauses or complementizer agreements (for more on language production

experiments for corpus building, see e.g. [Bülöw and Breuer 2019], [Fingerhuth and Breuer 2022] or [Lenz et al. 2019]). We conducted these language production experiments, whereby a stimulus is shown to the participant and their verbal response is recorded, using OpenSesame, a program to build and conduct experiments for social sciences (see [Mathöt et al. 2012]). Additional reading- and translation tasks completed the corpus material. Overall, just for a subcorpus of the research project, more than 500 hours of spoken language were recorded.

The language data was then integrated into the cooperative online research platform, which encompasses all different project parts and their respective research data. The research platform is based on a PostgreSQL database with several custom-made frontend interfaces for data entry, based on previous work by [Breuer 2021]. The audio was segmented and transcribed (partially using software developed by DiÖ, see [Breuer et al. 2022]). We tokenized the transcribed data in two separate ways: for the more open settings, such as the structured interview and the conversation among peers, the data was tokenized on the word level. The more structured settings, such as the translation tasks, language production experiments and reading lists were tokenized on the level of (half-) sentences. These (half-)sentences correspond to responses of our informants (for more on language production experiments see Section 3.2). In order to be able to produce coherent, but also flexible tagging, a further entity (named “answers”) was created, which unifies the different levels of tokenization. One, or several word-tokens can be subsumed as an answer, which is the actual entity in the database that is being tagged. This allows for the coherent tagging and easier data extraction of tagged material, as the source of everything being tagged (whether it is a half-sentence stemming from a controlled setting or individual word tokens coming from a transcript from an open setting) is a centralized entity. Figure 1 provides a schematized and simplified model of the data structure.

7



**Figure 1.** A schematized and simplified model of the data structure used for annotation. Every box represents an entity in the data model, rows in a box represent attributes of the datasets, and the lines represent relations between entities. “FK” stands for foreign key and indicates that a dataset of this entity is related to a dataset of a different entity. Word tokens and responses can be made to answers, which are the entity that tags are given to. As one tag can be given to multiple answers, and one answer may contain multiple tags (i.e., an m:n relation), there needs to be a table purely to connect these two entities.

As can be seen in Figure 1, annotations are not directly given to either word token or responses, but instead to the beforementioned centralized entity answers. We discuss this model again in Section 3.2.

8

### 3. Annotation in (Variationist) Linguistics and DiÖ

After having briefly established the research context and some basic technological background of DiÖ, this next section will briefly touch on annotations in linguistics in general before showcasing the annotation system(s) and logic deployed by DiÖ.

9

#### 3.1 (Semi-)Automatic Annotations in Linguistics and DiÖ

Annotation in linguistics, which can be very broadly defined as the act of associating “descriptive or analytic notations with language data” [Ide 2017, 2], usually falls into one of two general categories: there is, on the one hand, (semi-)automatic annotation, and manual annotation on the other. The (semi-)automatic approach is considered first. This kind of annotation is usually done with large corpora and is carried out using natural language processing tools, such as the python library spaCy ([Honnibal et al. 2020]) or the NLTK (“Natural Language Tool Kit”, [Bird et al. 2009]). These kind of annotations are usually – but not always – fairly one-dimensional. A prime example of these kinds of annotation processes is POS (“part of speech”) tagging, which usually associates a token with a label clarifying their part of speech function within a sentence. Table 1 showcases an example from the DiÖ-corpus for an utterance tagged for part of speech according to two different models – spaCy’s universal tagset and the Stuttgart-Tübing-Tagset, a tagset modelled for German ([Schiller et al. 1999]):

10

<b>0029:</b>	<b>für</b>	<b>mich</b>	<b>ist</b>	<b>das</b>	<b>schon</b>	<b>normal</b>	<b>dann</b>	<b>in</b>	<b>dem</b>	<b>Fall</b>	<b>.</b>
POS universal	ADP	PRON	AUX	PRON	ADV	ADJ	ADV	ADP	DET	NOUN	PUNCT
	adposition	pronoun	auxiliary	pronoun	adverb	adjective	adverb	adposition	determiner	noun	punctuation
POS German	APPR	PPER	VAFIN	PDS	ADV	ADJD	ADV	APPR	ART	NN	\$.
	preposition; circumposition left	non- reflexive personal pronoun	finite verb, auxiliary	substituting demonstrative pronoun	adverb	adjective, adverbial or predicative	adverb	preposition; circumposition left	definite or indefinite article	regular noun	sentence- final punctuation mark
English translation:	for	me	is	that	pretty	normal	then	in	this	case	.
	"for me, that is pretty normal then in such a case."										

**Table 1.** Example of an utterance from the DiÖ corpus, tagged for part of speech with two different models. Explanation of the labels added by the author.

As can be seen, every word (and punctuation) is associated with a specific label, which clarifies their part of speech function within the utterance. Within DiÖ, two different sets of labels are used – a “universal” (i.e. not language specific) part of speech tagging, and part of speech tagging specifically for German as based on the Stuttgart-Tübingen-Tagset ([Schiller et al. 1999]). Tagging was done automatically using SpaCy ([Honnibal et al. 2020]).

11

The annotations used here are very typical for these kinds of NLP-based automatic annotations. The labels appear straightforward, usually unambiguous and one-dimensional, and the process of assigning them to the individual tokens is opaque to most researchers. Different levels of granularity are observable: The tags designed for the German language provide a more granular approach to the classification. We can additionally observe a misassignment of a tag: the “ist” is wrongfully labelled as auxiliary or respectively finite auxiliary, when it is a full verb in this case (just as in the English translation).

12

Nevertheless, automatic annotation is tremendously important in linguistics in general and in DiÖ as well: it allows researchers to tag vast material of data at a very low cost and with great accuracy. This makes the data easier to search and analyze. There are, on the other hand, definite caveats to automatic annotation: the processes and parameters of assigning specific labels to individual tokens is usually not entirely transparent to the researchers. This results in two distinct problems: on the one hand, the relationship between the labels and the linguistic theory used to justify them is obscured, which can be problematic when researchers explicitly work within a specific theoretical framework (cf. [Newman and Cox 2020, 43]). Even more so, the “theoretical assumptions, complex social factors, and linguistic intuitions,” which are important factors in building an annotation scheme, are irrecoverably lost in automatic annotation and not “identified and clearly reflected,” as [de Marneffe and Potts 2017, 412] argue (see also Section 4). While e.g. part of speech tags are highly canonized and widely used, they are still not theory-neutral and carry specific assumptions about language. The comparatively low cost of automatic annotation and the omission of a human annotator results in an annotation process and result that appears highly objective. When such “assumptions underlying a classification scheme” become “effectively invisible and thus no longer subject to challenge or rethinking,” this invisibility becomes “dangerous and should be avoided,” as e.g. Sperberg-McQueen argues [Sperberg-McQueen 2016, 378].

13

Thankfully, linguists, programmers and DH-professionals working on the intersections of those fields are aware of the limitations of automatic annotations. Usually, success rates of individual NLP tools are indicated in the tool’s documentation, and manual corrections of the automatic annotations are often part of the workflow. There are furthermore several works dealing excessively with automatic and semi-automatic annotations, e.g. [Ide and Pustejovsky 2017].

14

For DiÖ, automatic annotation played a minor role. The corpus was part of speech tagged with the SpaCy library, both with the universal language tags and the Stuttgart-Tübingen tagset. Additionally, we performed lemmatization using SpaCy. The core of the annotation work done was tagged manually. The different status of the tags is even indicated in the data model itself: whereas the automatic annotations are saved as attributes to the individual tokens, the manual tags are distinct entities in the database, and are given to answers (which consist of either tokens on the word level or on the level of half-sentences, see Section 2.2), meaning that they are recorded in their own table. We lemmatized and part of speech tagged the corpus to ensure better searchability, but it is not a basis to answer the overarching research questions of the project.

15

### 3.2 Manual Annotation in DiÖ

These overarching research questions require a deeper and more complex analysis of the language data than automatic annotations can provide. As e.g. [Smith et al. 2008, 164] note, these automatic annotations are very often “inadequate” for many research questions because of their simplicity. This means that researchers generally need to classify their data further. Even though it is very often not labelled as such, such a classification of data according to the needs of a linguist is nothing else than the further annotation of linguistic data. [Kendall 2008, 336] for example showcases this problem perfectly, explaining how there is no unified method of achieving further classification of linguistic data (i.e. annotation), while also not referring to this process as annotation (instead using the terms “variable tabulating” and “coding”):

16

Variable tabulating – the extraction and coding of different realizations of the same linguistic variable – is the central methodology within quantitative variationist sociolinguistics and the primary means for undertaking this filtration process (cf. Labov 1966; Wolfram 1993, 2006). While variationists may agree in principle on how they do their work, there are no agreed upon standards for how they move from a real-world speech event to a set of quantified data for a particular linguistic feature.

17

While it appears to be the rule rather than the exception that a linguistic research question working with corpora requires researchers to further (manually) annotate their data in some way (cf. [Smith et al. 2008, 165]), research in annotation tends to focus on automated annotation (cf. [Newman and Cox 2020, 26]). Coverage of manual annotation has been described as “rather patchy,” with [Smith et al. 2008, 167] noting how underrepresented the process of manual annotation is in introductory textbooks. The field of corpus linguistics appears to therefore stand before the problem that one of its fundamental methods – manual annotation of

18

language data – is chronically underrepresented and understudied.

This is a further reason to lay out the manual annotation process of DiÖ. In contrast to the automatic annotations discussed in Section 3.1, the manual annotation process functions quite differently, both on the level of the data model of the cooperative online research database and on the level of tag design for individual annotations. Picking up the discussion in Section 2.2, the data model is explained in more depth before turning to the general semantics of the tags used, which is then further illustrated using examples of individual tags and the tagging process from research from DiÖ.

19

Figure 1 above lays out a (simplified) database diagram of how the manual annotations are handled within DiÖ. As mentioned in Section 2.2, annotations are not directly given to either individual word-tokens or half sentences, but rather to a different entity in the data model, which are termed “answers”. These answers may consist of a response from a language production experiment<sup>[2]</sup>, or one or more word tokens from a transcript of a free speech setting. Thereby, we assign the annotations to answers that contain the relevant tokens or responses. This has the benefit that these answers act as entities that unify the heterogenous data sources (word tokens from transcripts, responses from language production experiments). It furthermore allows for annotating several word tokens as one answer, which is immensely helpful for the annotation of linguistic phenomena beyond the individual word level, such as syntactic structures like the passive. Additionally, it leaves the primary language data intact.

20

Another important entity within the database scheme is that of the tags themselves: these are, in contrast to the part of speech tags (see Section 3.1), not saved as a varchar-attribute to the individual word tokens<sup>[3]</sup> but rather as distinct data sets within the database. This annotation method can therefore be considered as a form of stand-off annotation (i.e., an annotation not saved directly in the text itself, such as e.g. TEI-coding, where the annotations are set directly in the text). The tagging process itself is thus conceived as assigning an m:n relation between the tags and the answers. There are again benefits to this: it allows for proper documentation of the meaning of individual tags as attributes of an entity within the database, it means that constraints for the combination of tags to ensure well-formed annotations can easily be set, and it furthermore eliminates the issue of mis- or heterogenous spellings of one tag. Moreover, this allows to easily change tags – if researchers become unsatisfied with the definition or name of a tag, only the entry in the table specifying the tags needs to be changed instead of manipulating every instance of the tag, as would be the case with inline annotation.

21

Consider, for example, a researcher being interested in a specific morphological phenomenon (e.g. variation in comparative and superlative forms of adjectives, such as <dunkler>/<dunklicher> “darker” as comparative of <dunkel> “dark”<sup>[4]</sup>). They will comb through the corpus and search for adjectives (utilizing the POS-tags for the transcripts tokenized on word level). Whenever they find a token that is an adjective in the comparative, they will create an answer consisting of that token and tag it. They can then go through the responses of the language production experiments concerned with comparative forms of adjectives, and create an answer consisting of the relevant response and tag it as well. “Tagging,” in this case, means that an answer is recorded in the database, which consists of one or several token or a response from a language production experiment. That answer is then linked with the relevant tags, which are datasets in their own right. The table labelled “Tags\_to\_Answers” in Figure 1 represents this. In order to adapt a tag, it is not necessary to change it in every tagged answer – instead, only the dataset that represents the tag needs to be changed because the answers refer to that dataset instead of having the tag itself recorded in it.

22

In Figure 1, we can additionally observe that tags are associated with one or more annotation tiers. Whenever an answer is being tagged, it first needs to be specified on which annotation tier the researcher wants to tag. This limits the number of available tags to the researcher to the relevant one for the phenomenon being tagged, and avoids ambiguities and redundant tags, as tags can be assigned to a tier specifying their use-case (e.g. a tag for number can be used to both specify singular / plural for phenomena concerning nouns or verbs, depending on which tier it is used). The main entities characterizing the annotation within DiÖ are therefore the answers (i.e., a homogenized version of the language material being tagged), the individual tags (i.e. the linguistic labels carrying a specific, analytic meaning) and their relationship to each other (i.e. the annotation of language data with tags).

23

A further attribute of individual tags deserves more detail: that of the generation. DiÖ employs a hierarchical tagging system for manual annotations (in contrast to the flat/one-dimensional model used for automatic annotations, where any notion of hierarchy is, at best, implicit). The individual tags are organized as tag families in a parent-child relationship on different generations, wherein certain tags (the “children”) may only follow certain other tags (their “parent”). This organization follows a specific semantic model, as illustrated by Figure 2:

24

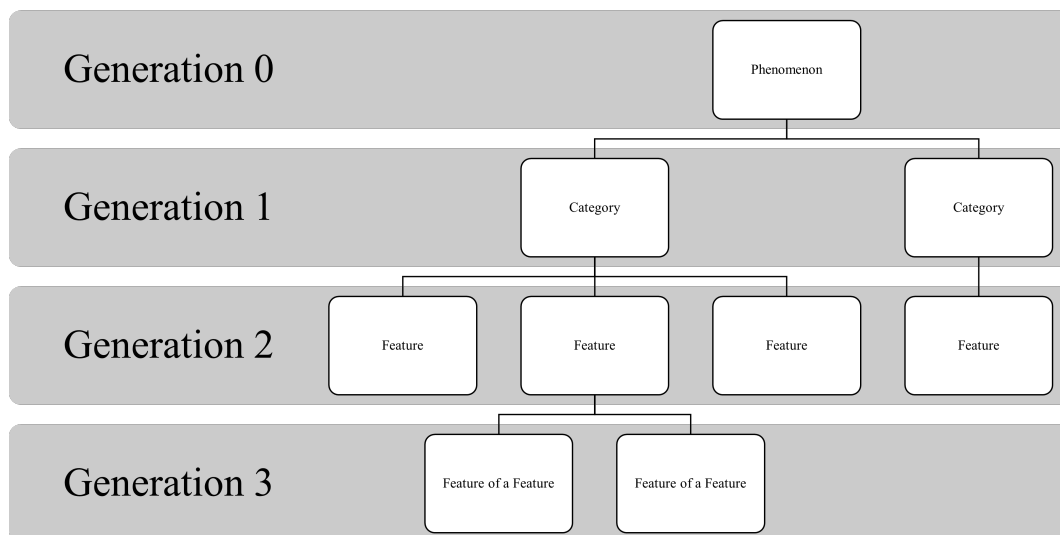


Figure 2. Schema of the semantic model used for annotation design in DiÖ

As can be seen, the basic parent tag on generation 0 on any annotation tier will be a tag characterizing the linguistic phenomena being annotated – mostly, this will be a single tag per tier (relative clauses, where the variation is studied on two levels – that of the noun (phrase) in the matrix clause and the referent in the relative clause – for example have two tags on the generation 0, one for each element). Like the stem of a tree, this tag creates the basis from which the other tags stem. Tags on the first generation generally denote categories that characterize the phenomenon being studied. These category tags compliment each other, and usually, every category tag is applicable to a specific instance of the phenomenon being annotated (e.g. the tags on the first generation for the phenomenon “variation of /au/,” which are “realization of /au/,” “sound preceding,” “sound following,” and “etymological root,” must always apply to any instance of the phenomenon, as any word containing /au/ has a specific realization of the sound, a sound preceding and following it<sup>[5]</sup>, and an etymological root). Whereas the tags on the first generation denote categories of the linguistic phenomenon and complement each other, tags on the second generation denote features of the categories, and usually exclude each other. To stick with the example of the “variation of /au/,” the second generation tags following e.g. the tag “etymological root” may be a tag for Middle High German /ou/ or Middle High German /û/, but it cannot be both – these tags describing the Middle High German vowel of the word exclude each other. The same goes for “sound following” or “sound preceding” – here, only one of the possible children of that tagset apply for every specific instance of the phenomenon. As such, these tags usually exclude each other. The third generation in a tagset usually specifies features further, and offers what could be considered features of features, categorizing the phenomenon with a higher granularity. An example would be a further specification of the grammatical voice of a verb – while usually, the features (second generation tags) of a category “grammatical voice” (generation one tag) would be “active” or “passive,” the kind of passive used can be further specified: a dative passive, a processual passive or a statal passive. These would be features specifying other features further, which puts them on the third generation.

25

While technically, this hierarchical system can expand, offering as many successive generations as needed, most of the tagsets for a specific phenomenon do not exceed the first three generations and follow the semantics of the generational structure closely. In order to devise a tagset for a specific linguistic phenomenon, the researchers interested in that phenomenon set up a draft for a specific annotation system following the logic laid out above. This draft is then brought forward to the researchers which manage all annotations within the cooperative online research platform. The draft tagset is then checked for consistency and applicability as well as redundancy with existing annotations, and is then brought back to the researchers for review of the changes made. Once the researchers agree on a first draft, we implement the tags and their hierarchies into the database. Usually, after some preliminary annotations being made, the tagset will be expanded or changed to encompass finer distinctions or additional categories wherever needed. This follows the idea of an iterative, cyclical model of the creation of tagsets and annotation schemes (cf. e.g. [Zinsmeister 2015, 87]).

26

The tagging itself can be done via various interfaces of the database: for annotations on responses to language production experiments, the interface (see Figure 3) allows for transcription and organization of the utterances to responses, which are then saved as answer, and can then be annotated in the same interface. The researchers select the annotation tier from a drop-down menu and adds the relevant tags. The tags and their children are displayed as nested, reflecting the parent-child-structure in order to make the user interface clearer (see e.g. [Johnson 2014, 29–36] for the role of visual structure in user interface design). Additionally, color feedback indicates the hierarchy, with darker blue tones for deeper layers. For convenience, pre-set tag groups can be saved and applied to answers, which allows for faster tagging. During the saving process, the annotations are checked if they adhere to well-formedness (i.e. no tag in a hierarchy missing).

27

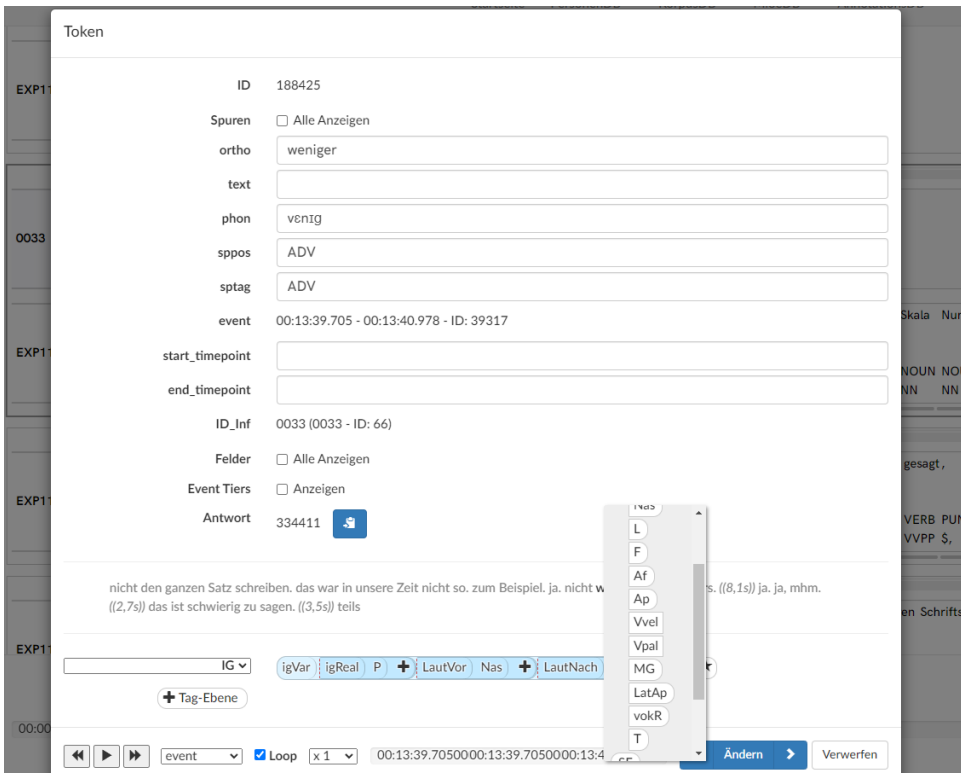
# 0025 - 8 - Vorlesen der Aufgabenstellung (L02\_ig)

Aufgabe: ID: 504 - Aufgabenset: L02\_ig - Variante: 8 - Stimulus Dialekt: Nein - Evokationsziel Dialekt: Nein - Beschreibung: Vorlesen der Aufgabenstellung  
 Aufgabenstellung: einig

**Figure 3.** Interface used for transcription and annotation of closed settings. Here, an item from a reading task (einig, “united”) is tagged on two different tiers, once for variation of <-ig> and once for variation of /ei/.

Transcription and annotation of the open settings is separated in two different interfaces.<sup>[6]</sup> The annotation interface initially displays the transcript in tokenized form (see Figure 4) and allows to search transcript with pure text, wildcard or RegEx search. Individual tokens, token spans consisting of several continuous tokens from a speaker, and groups of discontinuous tokens can be subsumed as answers and then annotated in this interface. The annotation interface for the open settings itself (see Figure 5) is visually similar to the closed settings, adhering to the principle of consistency in user interface design (cf. [Shneiderman and Plaisant 2004, 74]). As before, in this interface, annotators choose the annotation tier via a dropdown and then select either the tags themselves or a pre-set of a certain tag combination.

**Figure 4.** Interface for transcripts of open settings. Tokens are displayed, with their part of speech tags toggled on. The token <weniger> (“fewer”) was searched, with current search result highlighted in blue. A blue border around a token signals that this token is already tagged.



**Figure 5.** Annotating a token from an open setting. The token <weniger> (“fewer”) being annotated for the variation of <-ig>, with the drop-down menu selecting the feature tag on generation two following the generation one tag LautNach (“preceding sound”).

The usual process of annotating is very often non-linear in DiÖ. Whereas we designed the language production experiments to evoke specific linguistic constructions, the respective phenomena may occur anywhere in the open settings (e.g. a set of language production experiments was explicitly designed to evoke responses of adjectives in the comparative and superlative forms, but these forms may appear anywhere in the open settings). In order to make manual annotation efficient, very often student assistants were tasked with finding instances of the linguistic phenomenon within the transcripts of the open settings and tag them only with the generation 0 tag of that tier. An additional interface (see Figure 6) then allows to retrieve all instances of a tag (e.g. a generation 0 tag) on a specific tier, and further annotate the answers. As a result, the annotation process frequently involved only giving bare annotations to make hard-to-find phenomena easily searchable (mostly the tag on generation 0), and then tag them with additional annotations via this interface. Since this interface allows for the search of any tag on any tier, it is also regularly used to verify tagged instances (e.g. make sure that the tag only covers relevant instances, and is neither too broad nor too narrow).

#	R	Inf.	Sätze - Ortho / lu	Sätze - IPA	Ausgewählte Tokens - lu / Ortho / Phon	Tagebenen
1	1	0206	99	noʎnonoɔɛnʔɛsɪk*		IG: igVar, igReal, P, LautVor, Af, LautNach, MG
2	1	0206	66	segsonseɕʔɛɪɕ		IG: igVar, igReal, F, LautVor, Af, LautNach, MG
3	1	0206	44	fieonfɛɪʔɛɪk*		IG: igVar, igReal, P, LautVor, Af, LautNach, MG
4	1	0206	20	ʔɛvanʔɛɪk*		IG: igVar, igReal, P, LautVor, Af, LautNach, MG
5	1	0216	99	noʎnonoɔɛnʔɛsɪk*		IG: igVar, igReal, P, LautVor, Af, LautNach, MG
6	1	0216	vierundvierzig	fɪɛɔndfɪɛʔɛɪk*		IG: igVar, igReal, P, LautVor, Af, LautNach, MG
7	1	0216	zwanzig	ʔɛvanʔɛɪk*		IG: igVar, igReal, P, LautVor, Af, LautNach, MG
8	1	0216	66	segsonseɕʔɛɪɕ*		IG: igVar, igReal, P, LautVor, Af, LautNach, MG
9	1	0216	20	ʔɛvanʔɛɪk*		IG: igVar, igReal, P, LautVor, Af, LautNach, MG
10	1	0216	44	fieonfɛɪʔɛɪk*		IG: igVar, igReal, P, LautVor, Af, LautNach, MG
11	1	0029	99	noʎnonoɔɛnʔɛsɪk*		IG: igVar, igReal, P, LautVor, Af, LautNach, MG
12	1	0029	20	ʔɛvanʔɛɪk*		IG: igVar, igReal, P, LautVor, Af, LautNach, MG

**Figure 6.** Answer with the tag “Affricates” on the annotation tier for variation of <-ig>. The results seen here are numbers, accompanied by their transcription in the international phonetic alphabet. The right-hand column shows the tags. All of these annotations can be further tagged and modified from this interface, see also Figure 7 for a more detailed view.

All the tagged material can be exported from the database in an Excel-readable format. The export is conducted based on annotation tiers, and includes the relevant answers, their source (free or closed settings), metadata such as age and gender of the speaker, and other details. Furthermore, the individual tags for every answer are separated using different punctuation depending on their generation, which makes it easier to organize them into individual columns. Any annotations on a different tier for the same answer are also reproduced in this way, in order to make interaction between different linguistic phenomena easily observable.

Before turning to specific examples of tagsets used within DiÖ, a brief note on the technical advantages of setting up annotations this way is in order (advantages of the semantics of the tagging system are discussed in Section 4). There are a variety of benefits to using a relational database structure for annotations: the stand-off nature of the annotations ensures that the metalinguistic data, such as annotations, is clearly separated from the language data, making it easily recoverable. Additionally, it becomes easy to change the description, definition or names of individual tags without combing through the annotated material. The primary data itself can easily be modified as well (e.g. in the case of spelling mistakes, or incomplete transcripts) without any additional workload of making sure that the annotations still refer to the correct answers. Furthermore, the annotations become part of the database that also hosts the corpus, in line with suggestions made by e.g. [Smith et al. 2008, 175]. By including the annotations within the database structure, they are easily recoverable and accessible to other researchers once the corpus is published (which DiÖ plans to do by the end of the project runtime).

In order to provide a more tangible view on the annotation logic of DiÖ, the following presents two examples of research on two different language aspects: the variation of the pronunciation of unstressed <-ig> on the phonetic level, and the variation of tag questions on the pragmatic level.

The first example of a tagset as used in DiÖ is concerned with the variation of unstressed <-ig>, especially in more standard-speaking contexts, and the data tagged with it resulted in the publication [Lanwermyer et al. 2019]<sup>[7]</sup>. It serves as an example because of its illustrative simplicity. Unstressed <-ig> may be pronounced either as a plosive sound (usually [ǰ], [k] or [kʰ]) or as a fricative sound ([ç]), where pronunciation norms clearly prefer plosive realization for spoken German in Austria. Results, however, showed that there is variation in the pronunciation of <-ig> in spoken standard, and it is dependent on complex linguistic factors such as part of speech and phonetic contexts.

The tagset used to arrive at the conclusions of [Lanwermyer et al. 2019] is reproduced in a condensed version Table 2, with an additional example of one of the tagged items in the database frontend (Figure 7).

Generation 0	Generation 1	Generation 2	Generation 3
igVar "variation of <-ig>"			
	igReal "realization of <-ig>"		
		P "plosive"	
		F "fricative"	
		? "unclear"	
	LautVor "preceding sound"		
		P "plosive"	
		F "fricative"	
		N "nasal"	
		VVel "velar vowel"	
			lang "long"
			kurz "short"
		MG "morpheme border"	
		(...)	
	LautNach "following sound"		
		P "plosive"	
		F "fricative"	
		N "nasal"	
		VVel "velar vowel"	
			lang "long"
			kurz "short"
		MG "morpheme border"	
		(...)	

**Table 2.** Tagset for the annotation of variation of <-ig>. Some phonetic contexts have been omitted for readability, indicated by (...).



As can be seen, this tagset is fairly simplistic, and follows the semantic logic stringently: following a generation 0 tag detailing the linguistic phenomenon (which coincides with the annotation tier used), the phenomenon is annotated for three different categories, which are necessary to describe the variation encountered in the phenomenon. These are, on the one hand, the concrete realization of the <-ig> graphemes, and the phonetic context (i.e. the sound following and preceding <-ig>) on the other. These are the linguistic categories that need to be annotated (other variables, such as speaker age, origin, or setting, are not tagged explicitly, but are added when results are exported based on the records in the database), and they complement each other in the description of the phenomenon. In contrast, the features of these categories – recorded on the generation 2 tags – exclude each other, and are specific for each occurrence of unstressed <-ig>. Both of these categories and their relevant features refer to the same linguistic system – that of the pronunciation of the specific instance of the phenomenon.

Antwort bearbeiten: ID 24920 ×

Antwort 24920

Ortho / lu

vierundvierzig

IPA

fɪːɛrʊnfɛt͡sɪkʰ

IG ▾ igVar igReal P + LautNach MG + + + ★  
 + Tag-Ebene

⏪ ⏩ ⏮ ⏭ ⏯ ⏸ Aufgabe ▾  Loop x 1 ▾ 00:01:46.918000 00:01:46.918000 00:01:50.592000 Speichern Schließen

**Figure 7.** Annotation of the word <vierundvierzig> (“forty-four”) from a reading task according to variation of <-ig>. The phonetic realization was as a plosive, the sound preceding was tagged as affricate, and the sound following was a morpheme border (i.e. no sound followed <-ig>).

A more complex tagging system, which also follows the semantic model laid out earlier in this Section, is the annotation of tag questions. Tag questions are words, phrases and particles that suggest an interrogative meaning even if no actual answer is needed (a famous example would be the “isn’t it?” phrase, often associated with varieties of British English). German (in Austria) offers a variety of such tag questions, which can be heterogenous in both grammatical form and meaning. The tagset<sup>[8]</sup> is laid out in Table 3, again with omissions for readability, with a tagged example from the DiÖ corpus provided in Figure 8.

Generation 0	Generation 1	Generation 2
QuestTag "tag questions"		
	QTForm "grammatical form of the tag question"	
		QTAffPart "affirmative particle"
		QTNegPart "negative particle"
		QTKonj "conjunction"
		(...)
	Int "intonation"	
		rise "rising intonation"
		level "level intonation"
		fall "falling intonation"
	PosSatz "position in sentence"	
		initial
		mid
		final
	PosTurn "position in turn"	
		initial
		mid
		final
	RespType "type of (possible) response"	
		0int "none intended"
		polar "polar answer"
		pause "a pause follows"
		resp "a response is given"
		unexresp "unexpected response"
	QTFunk "assumed pragmatic function"	
		unsidek "declaring insecurity"
		bestford "wants affirmation"
		evimark "marks information as given"
		diskgli "structures discourse"
		konsak "signals consent"
		sonsfunk "miscellaneous"

**Table 3.** Tagset used to classify tag questions. Some feature tags for grammatical form have been omitted for readability.

Once again, the generation 0 tag specifies the linguistic phenomenon being tagged, coinciding with the annotation tier. The six categories describing the phenomenon are the grammatical form (QTForm), the intonation (Int), its position in the sentence (PosSatz) and in the utterance (PosTurn), as well as the intended kind of response (RespType) and its possible function (QTFunk). In contrast to the example taken from phonetics, not all of these categories refer to the same linguistic system – we can observe a grammatical category (QTForm), a category on the phonetic level (Int), a syntactical category (PosSatz) and a discourse category (PosTurn) as well as two pragmatic categories (RespType and QTFunk). The categories also differ in how much interpretation they demand of the annotators: whereas syntactical and grammatical categorization with the relevant features requires a negligible amount of interpretation, other features, such as the intended type of response or the possible function require a significant effort of interpretation from the annotator. We can expect lower inter-annotator agreement in these categories.

As for the features of the categories laid out in Table 3, it is clear that some of the features form parts of a necessarily closed group (such as medial, final and initial for position in turn and in sentence, or rise, level and fall for intonation), as these are relatively conventionalized grouping systems in linguistics, and they characterize a more formal level. The category of function, on the other hand, which includes features like “demanding confirmation” or “declaring insecurity,” is much more open-ended, and the features of this category only showcase one possibility of dividing up pragmatic functions of question tags, and – as indicated by a categorization of “miscellaneous” – this might not yet be the best possible way of doing it. The next iteration of this tagging system might see different features for this category.

Antwort bearbeiten: ID 304499

×

Antwort 304499

Ortho / lu

Aufgabe machen, das, wie die Zeit war, **nicht**? wenn die die Wiesen abgemäht sein gewesen, dann

lu / Ortho / Phon

Aufgabe machen, das, wie die Zeit war, **nit** ? wonn de de wiesan oogmaht sein gwejn, donn

Phon / Ortho / lu

Aufgabe machen, das, wie die Zeit war, **nicht**? wenn die die Wiesen o:gma:d sein gewesen, dann

QUESTAGS ▾

QuestTag

QTForm QTNegPart + Int rise + PosSatz final + PosTurn medial +

RespType Oint + QTFunk diskgli +

+  
+  
★

+ Tag-Ebene

⏪ ⏩ ⏴ ⏵

Satz ▾  Loop x 1 ▾ 00:10:50.979000 00:10:50.979000 00:10:52.886000

Speichern Schließen

**Figure 8.** Tagged example of <nicht> (“not”) as a tag question. The complete example, taken from an open setting, translates to “[...] doing homework, ah, how the time was, [was it] not? When the lawn was mowed, then [...]”

The example tagged in Figure 8 shows that the <nicht> (“not,” bolded in the interface) was classified as a negative particle with rising intonation. It stands at the final position of the sentence, but not of the speaker’s turn – thereby, the annotator assumed that there was no response intended. The function of this tag question was classified as structuring the discourse, but the additional ??-tag (a universal generation 0 tag for every tier) suggests that this annotation still needs to be reviewed.

39

As can be seen, the annotation model used by DiÖ is versatile enough to accommodate coherent tagging on all linguistic levels. However, the semantic model of DiÖ annotations can be further generalized to be employed not just for linguistics by clarifying the epistemological status of annotations in the following Section 4.

40

## 4. Annotations, Classification and Epistemology

After having explained the technical aspects of the DiÖ annotation system and provided examples of the semantic logic behind the hierarchical structure, this section aims to briefly problematize annotation as a (linguistic) research method based on the approach to annotation within the DiÖ project. Furthermore, we want to showcase the benefits of working with the multi-dimensional method of annotation, and how it can be generalized for other fields of study.

41

First and foremost, two aspects, which have so far only been implicitly assumed, need to be laid out. First, it is imperative to conceive of the process of annotation as something that is necessarily interpretative, and not purely objective. Annotation is always, “at least in some degree, the product of the human mind’s understanding of the text,” as [Leech 1997, 2] has stated. Secondly, annotation is essentially a question of the classification of data. This necessarily has the implication that, insofar as the classification of whatever phenomena can be considered production of knowledge, annotation is research (see also e.g. [Breuer and Seltmann 2018, 146]) – and a kind of research that is very often undertheorized (cf. [Lordick et al. 2016, 188]). These two factors feed into each other: annotation, as categorization, cannot be purely objective, as whatever classification scheme that is used is necessarily imperfect and incomplete (cf. [Sperberg-McQueen 2016, 389]), meaning that researchers must select foci which they deem worth categorizing as well as possible. As [Sperberg-McQueen 2016, 378] demonstrates, “perfect classification would require, and a perfect classification scheme would exhibit, perfect knowledge of the object,” meaning that if there theoretically was a phenomenon which researchers would be able to categorize fully and perfectly, this phenomenon would no longer be worth researching, as there would be nothing left to learn about it.

42

Annotation is therefore very often a necessary part of the production of knowledge in the humanities without standing on stable theoretical footing. Ideally, researchers are aware of these pitfalls, and adapt their method of data classification (i.e. annotation) accordingly. Very often, however, we can observe that annotations end up in a circular logic. [Consten and Loll 2012] provide a very enlightening study of such cases, demonstrating how “hypotheses of a theory are (partly) tested with derived (annotated) data that have been generated in the framework of the same theory” ([Consten and Loll 2012, 712]). This, simply put, means that without critically reflecting the process of annotation, researchers fall into the danger of subscribing to unfalsifiable hypotheses, as all the data on which they test their hypothesis is annotated in such a way that it makes falsification impossible (i.e. by annotating the data in such a way as the theory would suggest). To return to the example of tag questions and their annotation provided in Section 3.2, this danger becomes imminently clear: if the researchers assume, based on whatever theory, that there are X different kinds of tag questions, and annotate them accordingly, it would be surprising if the results of the annotation

43

process used to test the theory differed in any way from the theory used to annotate them.

This fact is among the reasons of why within DiÖ, we annotate the corpus with several dimensions instead of just classifying data as a monodimensional variants of a variable. In such a way, the annotations are easily recoverable, and the groupings of linguistic phenomena to a variant becomes transparent, and thereby an easier subject to review. Because of a meaningful, hierarchical structure of the annotations, based on categories and features that describe the phenomenon, the researchers need to make explicit what they consider constitutive of a certain variant. Thereby, problematic cases where the data does not fall neatly into pre-described variants can still be annotated according to their categories and features and thereby possibly suggest new, unfound regularities. Additionally, such a structure helps the researchers in their design of an annotation scheme, as there is no need to start from scratch for every linguistic phenomenon worth investigating.

44

Finally, since classification – of whatever sort – can be conceived of “as identifying locations in an n-dimensional space,” where “[e]ach dimension is associated with an axis, and the set of possible values along any one axis is sometimes referred to as an array” [Sperberg-McQueen 2016, 380]), there is definite value in making this annotation space more complex than simply one axis with a singular array. Instead, this semantic model makes the annotation space as complex as the object at hand demands while also giving meaning to not just the values in an array, but to the axis itself. To return to the example of tag questions provided in Section 3.2, instead of locating the data points as array of variants on the axis of a variable “tag questions,”<sup>[9]</sup> the DiÖ annotation logic provides a much richer annotation space with several, transparent axes that describe the data being tagged with a finer granularity. In such a way, this annotation logic can be beneficial not just for corpus linguistics, but for research in humanities in general, as it gives a meaningful and transparent structuring to the annotation space.

45

## 5. Conclusion

This paper has laid out the annotation processes conducted by DiÖ. By narrowing the scope from annotation in general to manual annotation, we set the focus on manual annotation as a research practice on the example of corpus and variationist linguistics. From the digital perspective, it was laid out how stand-off annotation can be beneficial for the annotation of linguistic corpora, allowing proper documentation of tags seamlessly, make it painless to edit already existing tags, showing a clear distinction between the annotations and the data being annotated, and being able to set constraints that only allow well-formed annotations according to the annotation manual. Then, the semantic layer of the DiÖ annotation process was laid out. The benefits of a multi-dimensional, hierarchical structure for tags, where generations correspond to either categories, features of a category, or features of a feature were shown. Two examples from contemporary variationist linguistics, the pronunciation of unstressed <-ig> in standard varieties of German in Austria, and the variation of tag questions, exemplified this. Finally, we argued that annotation in general can be understood as a form of – necessarily subjective and incomplete – categorization. The hierarchical, multi-dimensional model of annotations used in DiÖ provides a richer, more transparent annotation space than e.g. variant tabulation could. For this reason, the model used in DiÖ may be considered a better fit for many annotation projects in humanities research than simple, uni-dimensional annotation.

46

As has been argued that annotation, even though it is an elementary research practice in the humanities, is often severely undertheorized (cf. [Lordick et al. 2016, 188]), we hope that the present paper furthers works with a focus on annotation, especially how annotation as a practice is situated within different research fields. Additionally, it should be closer investigated on what epistemological footings annotation can stand within a certain discipline. Moreover, it would be worth questioning, possibly based on concrete examples, in how far annotation as categorization can have ontological implications for the humanities, insofar as the categorization achieved by annotation in the humanities can be considered a natural kind or “merely” a representational convention (cf. [Magnus 2018, 1427]).

47

We hope that this case study on annotation in variationist linguistics can help further the discourse on the topic of annotation in the humanities and help to stifle further discussions on this important topic.

48

## Notes

[1] The FWF (“Fonds zur Förderung der wissenschaftlichen Forschung,” “Austrian Science Fund”) is Austria’s largest funding body.

[2] These language production experiments usually consist of a visual stimulus (a picture or video), accompanied by an audio cue that tries to provoke either a specific morphological form (e.g. pictures of dogs in several sizes, accompanied by a voiceover providing a template such as “This dog is big. The other dog is...” to provoke comparative and superlative forms) or syntactic structures (e.g. a video of someone getting a tooth pulled, accompanied by the spoken question “What is happening to the person?” to provoke passive constructions) or other linguistic phenomena. For more detailed information on language production experiments see [Bülow and Breuer 2019], [Fingerhuth and Breuer 2022] or [Lenz et al. 2019].

[3] Responses are not part of speech tagged as their smallest unit is not the word level.

[4] See [Korecky-Kröll 2020] for further information on variation of comparative forms in German in Austria

[5] This may also be “none,” as even if the sound is at the beginning or end of a word, syllable or utterance, this position constitutes a specific phonetic context.

[6] For the transcription interface, which was also developed in-house, see [Breuer et al. 2022].

[7] The tagset itself was designed by Johanna Fanta-Jende and Manuela Lanwermyer with the help of Melanie Seltmann.

[8] This tagset was developed by Stefanie Edler, Katharina Korecky-Kröll, Markus Pluschkovits and Anja Wittibschlager.

[9] Even this would not be an unproblematic effort, as the relation of variants to a linguistic variable is not necessarily unproblematic outside of phonetics, see. e.g. [Lüdeling 2017].

## Works Cited

- Bird et al. 2009** Bird, Steven, Ewan Klein, Edward Loper. (2009) *Natural language processing with Python: analyzing text with the natural language toolkit*. Sebastopol, CA: O’Reilly Media, Inc.
- Breuer 2021** Breuer, Ludwig Maximilian. (2021) *“Wienerisch” vertikal: Theorie und Methoden zur stadtsprachlichen syntaktischen Variation am Beispiel einer empirischen Untersuchung in Wien*. Dissertation: University of Vienna.
- Breuer and Seltmann 2018** Breuer, Ludwig Maximilian and Melanie Seltmann. (2018) “Sprachdaten(banken) – Aufbereitung und Visualisierung am Beispiel von SyHD und DiÖ” in Ingo Börner, Wolfgang Straub, Christian Zolles (ed.) *Germanistik digital: Digital Humanities in der Sprach- und Literaturwissenschaft*. Wien: facultas, pp. 135–152.
- Breuer et al. 2022** Breuer, Ludwig Maximilian, Arnold Graf, Tahel Singer, Markus Pluschkovits. (2022) “Transcribe: a Web-Based Linguistic Transcription Tool,” *Working Papers in Corpus Linguistics and Digital Technologies: Analyses and Methodology* 7, pp. 8–24. Available at: <https://doi.org/10.14232/wpcl.2022.7.1>

(accessed 3 December 2022).

- Budin et al. 2019** Budin, Gerhard, Stephan Elspaß, Alexandra N. Lenz, Stefan Michael Newerkla, Arne Ziegler. (2019) The research project "German in Austria. Variation–Contact–Perception." in Lars Bülow, Ann-Kathrin Fischer, Kristina Herbert (ed.) *Dimensionen des sprachlichen Raumes. Variation – Mehrsprachigkeit – Konzeptualisierung*. Frankfurt am Main: Peter Lang Verlag, pp. 7–35. Available at: <https://www.peterlang.com/document/1063161> (accessed 30 November 2023).
- Bülow and Breuer 2019** Bülow, Lars and Ludwig M. Breuer. (2019) "Experimental approaches in the realm of language variation – How Language Production Tests can help us to better understand language variation". In Lars Bülow, Ann-Kathrin Fischer, Kristina Herbert (ed.) *Dimensionen des sprachlichen Raums. Variation – Mehrsprachigkeit – Konzeptualisierung*. Frankfurt am Main: Peter Lang Verlag, pp. 251–269. Available at: <https://www.peterlang.com/document/1063161> (accessed 30 November 2023).
- Consten and Loll 2012** Consten, Manfred and Annegret Loll. (2012) "Circularity effects in corpus studies – why annotations sometimes go round in circles," *Language Science* 34(6), pp. 702–714. Available at: <https://doi.org/10.1016/j.langsci.2012.04.010> (accessed: 3 December 2022).
- DiÖ 2022** DiÖ. (2022) *SFB German in Austria*. Available at <https://www.dioe.at/en/> (accessed 3 December 2022).
- Fingerhuth and Breuer 2022** Fingerhuth, Matthias, Ludwig Maximilian Breuer. (2022) "Language production experiments as tools for corpus construction: A contrastive study of complementizer agreement," *Corpus Linguistics and Linguistic Theory* 18(2), pp. 237–262. Available at: <https://doi.org/10.1515/cllt-2019-0075> (accessed: 27 September 2022).
- Honnibal et al. 2020** Honnibal, Matthew, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. (2020) *spaCy: Industrial-strength Natural Language Processing in Python*. Available at: <https://doi.org/10.5281/zenodo.1212303> (accessed: 30 November 2023).
- Ide 2017** Ide, Nancy. (2017) "Introduction: The Handbook of Linguistic Annotation" in Ide, Nancy, Pustejovsky, James (ed.) *Handbook of Linguistic Annotation*. Dordrecht: Springer, pp. 1–18.
- Ide and Pustejovsky 2017** Ide, Nancy and James Pustejovsky (ed). (2017) *Handbook of Linguistic Annotation*. Dordrecht: Springer.
- Johnson 2014** Johnson, Jeff. (2014) *Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Guidelines*. Burlington, MA: Morgan Kaufmann.
- Kendall 2008** Kendall, Tyler. (2008) "On the History and Future of Sociolinguistic Data." *Language and Linguistics Compass* 2(2), pp. 332–351. Available at: <https://doi.org/10.1111/j.1749-818X.2008.00051.x> (accessed 5 October 2022).
- Koppensteiner and Lenz 2021** Koppensteiner, Wolfgang and Alexandra N. Lenz. (2021) "Standard(s) aus der Perspektive von 'Nicht-LinguistInnen' in Österreich" in Markus Hundt, Toke Hoffmeister, Saskia Naths (ed.) *Laien, Wissen, Sprache. Theoretische, methodische und domänenspezifische Perspektiven*. Berlin: De Gruyter, pp. 391–416. Available at: <https://www.degruyter.com/document/doi/10.1515/9783110731958/html> (accessed 30 November 2023).
- Korecky-Kröll 2020** Korecky-Kröll, Katharina. (2020) "Morphological dynamics of German adjective gradation in rural regions of Austria" *Zeitschrift für Dialektologie und Linguistik* 87(1), pp. 25–65. Available at: <https://doi.org/10.25162/zdl-2020-0002> (accessed 30 November 2023).
- Lanwermer et al. 2019** Lanwermer, Manuela, Johanna Fanta-Jende, Alexandra N. Lenz, Katharina Korecky-Kröll. (2019) "Competing norms of standard pronunciation. Phonetic analyses of the <-ig>-variation in Austria," *Dialectologia et Geolinguistica* 27(1), pp. 123–141. Available at: <https://doi.org/10.1515/dialect-2019-0008> (accessed 3 December 2022).
- Leech 1997** Leech, Geoffrey. (1997) "Introducing Corpus Annotation," R. Garside, Geoffrey Leech, Anthony Mark McEnry (ed.) *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London: Longman, pp. 1–18.
- Lenz et al. 2019** Lenz, Alexandra N., Ludwig Maximilian Breuer, Matthias Fingerhuth, Anja Wittibschlager, Melanie Seltmann. (2019) "Exploring syntactic variation by means of 'Language Production Experiments' – Methods from and analyses on German in Austria," *Journal of Linguistic Geography* 7(2), pp. 63–81. Available at: <https://doi.org/10.1017/jlg.2019.7> (accessed: 3 December 2022).
- Lordick et al. 2016** Lordick, Harald, Rainer Becker, Michael Bender, Luise Borek, Canan Hastik, Thomas Kollatz, Beata Mache, Andrea Rapp, Ruth Reiche, Nils-Oliver Walkowski. (2016) "Digitale Annotation in der geisteswissenschaftlichen Praxis," *Bibliothek - Forschung und Praxis* 40(2), pp. 186–199. Available at: <https://doi.org/10.1515/bfp-2016-0042> (accessed: 27 September 2022).
- Lüdeling 2017** Lüdeling, Anke. (2017) "Variationistische Korpusstudien" in Marek Konopka, Angelika Wöllstein (ed.) *Grammatische Variation: Empirische Zugänge und theoretische Modellierung*. Berlin: de Gruyter, pp. 129–144. Available at: [doi.org/10.1515/9783110518214-009](https://doi.org/10.1515/9783110518214-009) (accessed: 6 October 2022).
- Magnus 2018** Magnus, P. (2018) "Taxonomy, ontology and natural kinds" *Synthese*, 195(4), pp. 1427–1439. Available at <https://www.jstor.org/stable/26750695> (Accessed 29 September 2022).
- de Marneffe and Potts 2017** de Marneffe, Marie-Catherine, Christopher Potts. (2017) "Developing Linguistic Theories Using Annotated Corpora" in Nancy Ide, James Pustejovsky (ed.) *Handbook of Linguistic Annotation*. Dordrecht: Springer, pp. 411–438.
- Mathôt et al. 2012** Mathôt, Sebastiaan, Daniel Schreij, Jan Theeuwes. (2012) "OpenSesame: An open-source, graphical experiment builder for the social sciences," *Behavior Research Methods* 44, pp. 314–324. Available at: <https://link.springer.com/article/10.3758/s13428-011-0168-7> (Accessed 30 November 2023)
- Newman and Cox 2020** Newman, John and Christopher Cox. (2020) "Corpus Annotation" in Stephan Gries, Magali Paquot (ed.) *A Practical Handbook of Corpus Linguistics*. Cham: Springer, pp. 25–48.
- Schiller et al. 1999** Schiller, Anne, Simone Teufel, Christine Stöckert. (1999) "Guidelines für das Tagging deutscher Textcorpora mit dem STTS (Kleines und großes Tagset)." Available at: <https://www.ims.uni-stuttgart.de/documents/ressourcen/lexika/tagsets/stts-1999.pdf> (Accessed 30 November 2023).
- Shneiderman and Plaisant 2004** Shneiderman, Ben and Catherine Plaisant. (2004) *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. 4th edn. Boston, MA: Pearson.
- Smith et al. 2008** Smith, Nicholas, Sebastian Hoffmann, Paul Rayson. (2008) "Corpus Tools and Methods, Today and Tomorrow: Incorporating Linguists' Manual Annotations," *Literary and Linguistics Computing* 23(2), pp. 163–180. Available at: <https://doi.org/10.1093/llc/fqn004> (Accessed: 26 September 2022).
- Sperberg-McQueen 2016** Sperberg-McQueen, C. M. (2016) "Classification and its Structures," Susan Schreibman, Ray Siemans, John Unsworth (ed.) *A New Companion to Digital Humanities*. 1st edn. Hoboken, NJ: John Wiley & Sons, pp. 377–393.
- Zinsmeister 2015** Zinsmeister, Heike. (2015) "Chancen und Grenzen von automatischer Annotation," *Zeitschrift für germanistische Linguistik* 43(1), pp. 84–111. Available at: <https://doi.org/10.1515/zgl-2015-0004> (accessed 3 December 2022).

