

PhonD2

Annotationsleitfaden

Alfred Lameli & Alexander Werth

Bearbeitet von Valeria Bunkov und Samantha Link

Marburg, Passau 2020

Forschungszentrum
Deutscher Sprachatlas



UNIVERSITÄT
PASSAU

Funded by



Deutsche
Forschungsgemeinschaft
German Research Foundation

Transkriptions- und Klassifikationsleitfaden für die PhonD2-Daten

1. Transkriptionskonventionen für die Erstellung von dialektalen Transkriptionen (Notation: X-SAMPA)

Ziel unserer phonologischen Transkription ist das Festhalten von Merkmalen, die auf der dialektalen Ebene kontrastiv sind.

1.1 Zu transkribierende Merkmale

1.1.1 Konsonanten

1.1.1.1 Glottisverschluss

Der Glottal Stop wird nur inlautend notiert (z.B. als Konsonantenersatz: bE?@ ‚Bette‘). Anlautend wird er nicht festgehalten.

1.1.1.2 Quantität

In einigen Dialekten spielt die Konsonantenlänge eine Rolle (z.B.: Ap: ‚Affe‘). („:“ in X-SAMPA-Notation).

1.1.1.3 Koronalisierung

[ʃ] und [ç] werden z.B. in westmitteldeutschen Dialekten als [ç] realisiert, wie bspw. in *Tisch* [tɪç], *Fische* [fɪçə] (Herrgen 1986, Féry 2017).

1.1.1.4 Retroflexes r

Retroflexe r-Laute treten gehäuft im (mittel-)hessischen und südmärkischen Raum auf (Lauf 1994).

1.1.2 Vokale

1.1.2.1 Quantität

Zwischen Kurz- und Langvokalen wird unterschieden (z.B. /a/ vs. /a:/).

1.1.2.2 Qualität

Geschlossene Vokale: [i], [u], [y], [o], [ø], [e]; Offene Vokale: [ɪ] [ʊ] [Y], [ɔ], [œ] [ɛ]. Hinzu kommen Vokale, die für einige Dialekte charakteristisch sind (s. literarische Umschrift/ X-SAMPA-Übersicht).

1.1.2.3 Nasalierung

Nasalität verweist auf phonologisch relevante Reduktionsphänomene (z.B. schwäb. „ba_~“ ‚Bahn‘) und wird daher transkribiert.

1.1.2.4 Mittelfränkische Tonakzente

Die Tonakzente sind bedeutungsdifferenzierend (Schmidt 1986, Werth 2011) (z.B. moselfränk. dOU_f ‚Taube‘ vs. dOU_/f ‚Taufe‘) und werden in mittelfränkischen Transkriptionen notiert.

1.2 Nicht zu transkribierende Merkmale

1.2.1 Konsonanten

1.2.1.1 Aspiration: Im Deutschen gilt die Aspiration als nicht-distinktiv (Fuhrhop/Peters 2013: 42).

1.2.1.2 Pharyngalisierung

1.2.1.3 Velarisierung

1.2.2 Vokale

1.2.2.1 Zentralisierung

1.2.2.2 Labialisierung

2. Klassifikationsmodelle für die phonologische Repräsentation

2.1 CV-Strukturen

Clements/Keyser (1983: 8ff.) / Wiese (1988: 62) / Hall (1992: 41ff.): V – Kurzvokal, VC – Langvokal u. Diphthong, C – Konsonant

2.2 Silbentyp

strong = Silbe mit Hauptakzent; weak = Silben, die keinen Akzent tragen

2.3 Silbenschema

P = Plosiv, F = Frikativ, V = Kurzvokal, V: = Langvokal, VV = Diphthong, N = Nasal, L = Liquid, G = Gleitlaut, A = Affrikate

2.4 Sonorität (angelehnt an Clements/Keyser 1983)

1 = Kurz- und Langvokal u. Diphthong, 2 = Gleitlaut, 3 = Liquid, 4 = Nasal, 5 = Frikativ, 6 = Plosiv

2.5 Phonemstatus von Affrikaten (C oder CC): für eine Variante entscheiden, Möglichkeit geben, das entsprechend umzukodieren.

2.5.1 Biphonematisch: Vgl. Menzerath (1954), Kohler (1995)

2.5.2. Monophonematisch: Vgl. Hall (1992, 2011), Wiese (1991) Argument: Sonoritätshierarchie, z.B. „des Rats“. Für das PhonD2-Korpus wird ein C für eine Affrikate angesetzt(monophonematischer Phonemstatus).

3. Klassifikationsmodelle für die morphologische Repräsentation (Wurzel 1990)

3.1 Schema: Stamm, Präfix, Zirkumfix, Infix, Fugenelement, Suffix, Umlaut, Ablaut, Diminutiv, Partikel, Suppletion.

3.2 Klassifikation der Morpheme

3.2.1 Nach Funktion Kernmorphem/Stamm (z. B. {Haus}), Flexionsmorpheme (z. B. {-e}, {-st}, {-t}), Derivationsmorpheme (z. B. {-lich}, {-heit}), Partikelmorpheme schließen Präpositionen, Konjunktionen, Adverbien, Interjektionen und trennbare Verben (Partikelverben) ein, Suppletivformen bilden eine eigenständige Morphemklasse, z. B. Kopulaverben (z. B. sein – bin – werden), unregelmäßige Komparative (z. B. viel – mehr – am meisten)

3.2.2 Klitika

An dieser Stelle ist die Erstellung einheitlicher Richtlinien wichtig, d. h. welche Formen können überhaupt klitisieren (agglutinierende vs. fusionierende Klitisierung)?

3.2.2.1 Trennschärfe zwischen Klitika, Fusionierungen, Verschmelzungen und Verschleifungen.

Festlegung: ausschließlich fusionierende Klitisierungen (am, in, ans, ins, ...).

Klitisierungen werden als ein Morphem behandelt (z.B. am statt an+dem).

3.2.3 Zweifelsfälle

3.2.3.1 Präterito-Präsentien: Zugriff auf mittelhochdeutsche Grammatik (z. B. Paul 2007)

3.3 Wortarten

3.3.1 Klassifikation nach Hentschel/Weydt (2013) & Duden-Grammatik (2016) (s. a. Lameli/Werth 2017)

3.3.1.1 Wortklassen: Lexeme und Grammeme; Substantiv, Verb, Adjektiv, Numerale, Adverb (Lexeme); Pronomen, Partikel, Artikel, Präposition, Konjunktion, Interjektion (Grammeme)

3.3.1.2 POS nach Schiller et al. (1999: 6ff.) (ggf. eigene Ergänzungen)

4. Kontrolldurchgang auf Wortebene

4.1 Korrekturtabellen

4.1.1 Zu kontrollierende Punkte

4.1.1.1 Annotation (Ziel: Vereinheitlichung der transkribierten Wörter pro Dialektraum)

4.1.1.2 Akzentsetzung (haupt- vs. nebenbetont)

4.1.1.3 Silbifizierung

Da der Silbifizierer keine Informationen zu phonologischen Wörtern hat, sind folgende Wörter besonders anfällig für Fehler (je nach Dialekt können auch noch andere betroffen sein): *Augenblick, austrinken, bestellt, eingebrochen, eingeschlafen, geblieben, gebracht, gebrannt, gedroschen, gestohlen, gestorben, hingefallen, verstehe, versteht*

- Funktionsprinzip des Silbifizierers: Es wird geprüft, ob ein Wort bereits Silbenpunkte enthält. Falls ja, wird der Input unverändert ausgegeben. Falls nein, werden Silbifizierung und Akzentsetzung (in X-SAMPA "Anführungszeichen") durchgeführt. D. h., dass Akzentsetzung und Silbifizierung immer gemeinsam in einem Wort kontrolliert werden müssen.

5. (Teil-)Korpus der freien Gespräche

5.1 Normorthographische Übersetzung (Standarddeutsche Verschriftlichung)

5.1.1 Übersetzungskonventionen (s. Handout zur morph. Annotation)

5.1.2 Zu transkribierende Wörter

5.1.2.1 Es werden ausschließlich Lexeme (Inhaltswörter) transkribiert. Zu den Lexemen gehören die Wortarten Substantiv, Verb, Adjektiv, Adverb, Numerale (Vgl. Lameli/Werth 2017). Die Lexeme werden in einem vorherigen Schritt in einem automatischen Prozess in eckige Klammern gesetzt, um kenntlich zu machen, welche Wörter transkribiert werden sollen.

5.1.3 Nicht zu transkribierende Wörter

5.1.3.1 Grammeme (Funktionswörter) werden nicht transkribiert. Zu den Funktionswörtern gehören die Wortarten Pronomen, Präposition, Artikel, Interjektion und Konjunktion bzw. Junktor (Vgl. Lameli/Werth 2017).

5.1.4 Umgang mit Klitisierungen

5.1.4.1 Klitisierte Formen werden – anders als für den PAD – sowohl im orthographischen als auch im dialektalen Transkript getrennt. Ein Beispiel: bair. ham6, std. [haben] wir; SAMPA [h a m] wir

Handouts zur phonologischen und morphologischen Annotation der PhonD2-Daten

1. Phonologie

1.1 PAD (Wenkersätze)

1.1.1 Datenbestand

PAD-Daten (PAD = Phonetischer Atlas von Deutschland). Sie bestehen aus Aufnahmen von Wenkersätzen (<https://regionalsprache.de/wenkerbogen.aspx#WS40>), die von 1956 bis 1990 an 186 Orten im gesamtdeutschen Gebiet erstellt wurden. Für diese Aufnahmen wurden bereits orthographische Transkripte für den Dialekt und das Standarddeutsche erstellt. Im Anschluss wurden für jede Aufnahme phonologische Transkripte in SAMPA (https://www.lfsag.unito.it/ipa/index_en.html) erstellt. Dabei wurde nach Dialekträumen vorgegangen. Zu jedem Dialektraum wurden Listen von zu erwartenden lautlichen Phänomenen erstellt, mit der Absicht dialektale Phänomene von Eigenheiten der Gewährspersonen zu unterscheiden.

1.1.2 Zur Genauigkeit der Transkription: Die Erstellung der Transkriptionen beruht ausschließlich auf phonologischen (auf dialektaler Ebene distinktive Merkmale) Kriterien; feinphonetische (z. B. Vokalqualitäten, Behauchung) Merkmale wurden hierbei nicht berücksichtigt.

1.1.3 Vorgehen: Die Erstellung der Transkriptionen erfolgte in Form von Textdateien (txt.); die dazugehörigen Tonaufnahmen wurden hierbei mit Hilfe von Praat (<https://www.fon.hum.uva.nl/praat/>) oder Audacity (<https://www.audacityteam.org/>) abgespielt.

1.2 Freie Gespräche

1.2.1 Datenbestand

Dieser Datenbestand setzt sich aus Tonaufnahmen von freien Gesprächen zusammen; diese stammen von denselben Gewährspersonen, zu denen die PAD-Abfragen (s. oben) vorliegen.

1.2.2 Orthographisches Standard-Transkript

1. Die Übersetzung wurde in einer Word-Datei angefertigt und nach der Sigle des Ortes (Tonaufnahme) benannt (z. B. I-481-5-6). Im Falle der Nennung der Sigle des Tonbandes (z. B. "I-5156"), wurde diese nicht ausgeschrieben, sondern als Ziffern-Zeichen-Kombination erfasst; für den weiteren Annotationsschritt aber nicht weiter berücksichtigt.

2. Der Interviewer wurde fett markiert, um eine Unterscheidung zwischen Gewährsperson u. Interviewer kenntlich zu machen. Falls es eine dritte oder vierte Person in der Tonaufnahme gab, wurde diese entsprechend kenntlich gemacht (z. B. Sprecher 1= S1, Sprecher 2= S2 usw. kursiv o. ä.); solche Passagen wurden nicht bei der Annotation nicht weiter berücksichtigt.

3. Die Übersetzung erfolgte im Fließtext, d. h. Absätze hinsichtlich der Passagen zwischen Gewährsperson und Interviewer/in mussten nicht zwangsläufig gesetzt werden.

4. Gesagtes wurde wortgenau erfasst (z. B. Interviewer, Selbstverbesserungen, Wiederholungen, Satzabbrüche, Interjektionen, etc.).

5. Der Satzbau sollte nicht verändert werden; Kasus durfte an die Standardsprache angepasst werden.

6. Zeitmarken wurde im Falle von Unklarheiten verwendet (z. B. [00:37]); diese wurden vor der Fertigstellung nochmals überprüft und die entsprechenden Stellen vollständig ausgeschrieben: Dabei wurden solche Stellen rein lautlich in erfasst und mit Klammern kenntlich gemacht.
7. Die Groß- und Kleinschreibung sowie die Interpunktion wurden beachtet.
8. Zahlen wurden ausgeschrieben (z. B. 1917 – Neunzehnhundertsiebzehn/ neunzehnhundertsiebzehn).
9. Es wurden keine Fußnoten z. B. für Anmerkungen verwendet.
10. Die Zeitmarken (z. B. [7:26]) durften im fertigen orthographischen Transkript nicht verbleiben; so wurden diese Stellen nach Gehör in Großbuchstaben aufgeschrieben (z. B. LOEHE EINGERT).
11. Klitisierungen wurden in den freien Gesprächen nicht gekennzeichnet und getrennt ausgeschrieben (z. B. bair. hama – Std. haben wir).
12. Diminutive wurden mit ^Dächlein^ gekennzeichnet (z. B. ^Schäfchen^).
13. Die Markierung nicht kognater Wörter wurde berücksichtigt, in dem ein sog. Wörterbuch am Ende des Dokuments erstellt wurde (z. B. gefestelt – gefeiert, Grombiera, wörtl. Grundbirne = Kartoffel). Da die Qualität des Webmausoutput darunter leiden würde, wenn wir hier gefeiert und Kartoffel. Solche Wörter wurden mit fortlaufenden Indexnummern mit spitzen Klammern versehen, die ohne Leerzeichen direkt an das Wort angehängt wurden, z. B. gefestelt< 1 > und Grundbirne< 2 >. Diese Indexe werden dann in einem "Wörterbuch" in der letzten Zeile der docx-Datei aufgeschlüsselt, die exakt folgendes Format haben muss und keine Leerzeichen enthalten darf:

<{1:gefeiert,2:Kartoffel,3:Wort3,4:Wort4}>

Ein Beispiel für ein orthographische Transkript:

I-56-2

Wenn Ihnen da irgendwie auch ein paar schöne Begebenheiten einfallen, dass Sie hier auch mal Einheimische getroffen haben oder was auch immer auf dem Schiff passiert ist. Aber die schönsten Begebenheiten werde ich Euch mit Sicherheit nicht erzählen. Vor zwei Jahren waren wir wieder oben in Griechenland und haben uns da einige klassische Ziele gesetzt. [7:26] wir haben uns zwar am Anfang geärgert.

1.2.3 Dialektales SAMPA-Transkript

Zu den freien Gesprächen wurden SAMPA-Transkripte erstellt, allerdings wurden hierbei nur Inhaltswörter/Lexeme (Nomen, Verben, Adjektive, Adverbien und Numerale) berücksichtigt. Diese wurden auf Grundlage des STTS automatisch ermittelt. Daraus entstanden sog. Vorlagen, in denen sämtliche zu annotierenden Wörter mit [eckigen] [Klammern] markiert waren. Die Vorlagen wurden nach der maschinellen Erstellung manuell auf die Klammersetzung (Inhaltswörter) überprüft. Ein Beispiel aus einer Vorlage:

Können *Sie* *mal* *erzählen,* *wo* *äh* *woher* *sich* *der* *Haushalt* *rekrutiert?* *Woher* *die* *Geldmittel* *kommen?* Ja , der [Haushalt] , der [besteht] aus [folgenden] [Sachen] : der [besteht] [also] [mal] aus der [Grundsteuer] A und B und [Gewerbesteuer] . [Dann] aus den [staatlichen] [Mitteln] , das [ist] die [Schlüsselzuweisung] .

1. Der Text des Interviewers (und anderer weiterer Sprecher) ist *mit* *Sternchen* *gekennzeichnet* und wurde vollständig ignoriert.

2. Zu transkribieren waren lediglich die Wörter in den [eckigen] [Klammern]. Die Klammern durften dabei nicht entfernt werden.
4. Falls während der Bearbeitung der Vorlage überzählige oder fehlende Wörter oder noch unmarkierte nicht kognate Wörter bemerkt wurden, so wurden solche Fälle sowohl in der bearbeiteten SAMPA-Vorlage als auch in der Originalvorlage korrigiert. Hier ein Beispiel für eine fertig ausgefüllte Vorlage:

Also im [vI ndr] oder im [fri:6dIN] , wenn s der [SnE:6] [vEg] [lSt] [n6] [faN6] wir [a:] mit [mi:Sgrats6] und [da] [wUt] mit den [ki:6] [hOm] wir da da [ge:gt] da [hOm] wir da eine [e:] [gmaxt] aus äh [SvatsdQ:6r6] und [ha] die [al tso:g@] in ein [xSd9l] und äh [hAd] [so:] eine [dQ6r6e:] [gmaxt] , weil sich [fe1:l] [baUr6] [hEnt@] [fel@r] keine [alsee:k] [hEnd] [lalSd6] [kEn6] .

1.2.4 WebMAUS-Textgrid

Auf Basis des orthographischen Standardtranskripts (ggf. der korrigierten Version davon) wurde daraufhin mit WebMAUS (<https://www.clarin.eu/showcase/webmaus-automatic-segmentation-and-labelling-speech-signals-over-web>) ein Textgrid erstellt. Da WebMAUS-Basic (<https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic>) keine Aufnahmen mit mehr als 3000 Wörtern verarbeiten kann, wurden die Aufnahmen über WebMAUS intern zu teilen und wieder zusammenzufügen (Chunking):

1. <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/Pipeline>
2. "Pipeline name (required)" die Option "G2P-CHUNKER-MAUS"
3. Mit drag&drop die geschnittene Tondatei und eine gleichnamige txt-Datei hochladen und auf "Run web-service" klicken. Die Erstellung des Textgrids dauert u. U. jedoch länger (im ersten Testlauf über 45 min). Wenn man möchte kann man vor dem Start unter "Expert Options" angeben, dass man nach der Fertigstellung per Mail benachrichtigt wird.
4. Im erstellten Textgrid müssen anschließend die tiers/Zeilen umbenannt werden:
 - ORT-MAU wird zu ORT-Standard
 - KAN-MAU wird zu SAMPA-Standard
 - MAU wird zu Segment
 - TRN wird gelöscht

2 Morphologie

2.1 Datenbestand

Zu jeder annotierten Datei (Wenkersätze, Funktions- und Inhaltswörter) und jedem annotierten freien Gespräch (= Inhaltswörter) wurde jeweils eine Morphemtabelle als Textdatei erstellt, in der die Morpheme jedes Wortes manuell eingeteilt wurden.

2.2 Schema und Vorgehen

Es gibt insgesamt elf verschiedene „Zellen“ zur Einteilung der Morpheme (vergleichbar mit einer Excel-Tabelle); gezählt wird hierbei ab der Position „Stamm“; das Hashtag (#) steht für eine leere Position, d. h. unbesetzte Felder bekommen einen Hashtag. Für das anschließende Einlesen ins Korpus war bei der Bearbeitung demnach wichtig, dass die Hashtags bei leeren bzw. unbesetzten Stellen dennoch erhalten wurden.

Schema:

1 2 3 4 5 6 7 8 9 10 11
 Standard | Stamm | Präfix | Zirkumfix | Infix | Fugenelement | Suffix | Umlaut | Ablaut | Diminutiv | Partikelmorphem | Suppletivform

Beispiele:

Stamm	Apfelbaum apf@l, baUm # # # # # # # # # #
Präfix	entkommen kOm Ent # # # @n # # # # #
Zirkumfix	gebaut baU g@ + # # t # # # # #
Infix	bereinigen raIn b@ + lg # @n # # # # #
Fugenelemente	Augenblick aUg@, blIk # # # n # # # # # # Freundschaftspreis frOInt, prals # # # s Saft # # # # #
Suffix	Schafe Sa:f # # # # e # # # # #
Umlaut	gäbe ge:b # # # # @ gE:b # # # #
Ablaut	schwamm SvIm # # # # # # Svam # # #
Diminutiv	Häuschen haUs # # # # # # hOIs # C@n # #
Partikelmorphem	morgen # # # # # # # # # mOrg@n #
Suppletivform	besser gu:t # # # # # # # # # bEs6

2.3 Einteilung von Wörtern nach Wortarten

Wörter werden in Wortarten eingeteilt. Das gilt auch in der Einteilung für Morpheme. So stehen im Stamm überwiegend Substantive (z. B. Apfel, Haus, Auto), Verbstämme (z. B. bau-, lieb-) und Adjektive (in ihrer Grundform, z. B. schön). Die folgenden Beispiele zeigen, welche Wortarten in welche Morphemkategorie eingeteilt wurden:

- **Stamm:** Substantive, Adjektive, Verben, Artikel, Pronomen (außer Personalpronomen)
- **Präfix:** (= Vorsilbe, z.B. *ge-* oder *ent-*), Hilfestellung:
https://dewiki.de/Lexikon/Liste_deutscher_Pr%C3%A4fixe
- **Achtung:** Bitte keine Verwechslung mit Partikelverben (z. B. *hingehen*, *fortfahren*; *hin-* und *fort-* sind in diesen Fällen Partikeln)
- **Zirkumfix:** bes. bei Partizipien (z.B. *ge+bau+t*) sind Zirkumfixe involviert. In solchen Fällen kommt ein Plus (+) in die Zirkumfixspalte.
- **Achtung:** In manchen Dialekten gibt es in Partizipformen zwar ein Präfix, allerdings kein Suffix (z. B. std. *gestohlen* – dia. *geSto:l*) oder einen Suffix, aber kein Präfix (z. B. std. *gestohlen* – dia. *Sto:l:n*). In solchen Fällen bleibt die Zirkumfixspalte leer (d. h. der Hashtag bleibt erhalten).
- **Infix:** Infixe kommen i. d. R. nur sehr selten vor und sind in der Fachliteratur umstritten (vgl. *-ig* in *bereinigen*, Infix oder Adjektiv?)
- **Fugenelement:** Im Standarddeutschen gibt es mehrere Fugenelemente (z. B. *-n-*, *-e-*, *-s-*, *-es-*; Bsp.: *Augenblick*, *Sonnenaufgang*, *Wirtshaus*). Aber: In einigen Fällen sind es keine Fugenelemente, sondern Suffixe (Kasus, z. B. *Bär+en+hunger*)
- **Suffix:** Ein Suffix (= Nachsilbe) liegt vor, wenn ein Wortstamm ein agglutinierendes Morphem aufweist (z.B. *bau+en*, *schön+er*, *Gesund+heit*)
- Verben, Adjektive, Substantive, Hilfestellung:
https://de.wikipedia.org/wiki/Liste_deutscher_Suffixe
- **Umlaut:** nahezu jede Wortart
- **Ablaut:** v. a. Verben
- **Diminutiv:** v. a. Substantive
- **Partikel:** Verbzusätze (z.B. *fortfliegen*, *hinhinmüssen*, *anfangen*), Adverbien (z.B. *früher*, *also*), Konjunktionen (*und*, *dass*), Präpositionen (*auf*, *unter*, *hinter*), Interjektionen (*aua*), Partikel (*ja*, *nein*, *nicht*); Partikelspalte = „Restkategorie“

- **Suppletion:** Bestimmte Verben (z.B. Verbstamm: *sein*, flektierte Verben „bin“, „bist“, etc.: Suppletion) und Adjektive (z.B. gut – besser), Personalpronomen (ich, du, er/sie/Sie, ...). Bei Unsicherheiten der Wortartzugehörigkeit wurde sich am Duden (<https://www.duden.de/woerterbuch?s=&scope=all>) und/oder am DWDS (<https://www.dwds.de/d/woerterbuecher>) orientiert.

2.4 Vorgehen

```
morphologie-dialect-l-62-1_2-3-4.txt - Editor
Datei Bearbeiten Format Ansicht Hilfe
früher | fROY6 | # | # | # | # | # | # | # | # | # | #
war | vat | # | # | # | # | # | # | # | # | # | #
Astfeld | asfElt | # | # | # | # | # | # | # | # | # | #
da | da | # | # | # | # | # | # | # | # | # | #
mussten | mOstn | # | # | # | # | # | # | # | # | # | #
Hüttenleute | hUdnLOY@ | # | # | # | # | # | # | # | # | # | #
Landarbeiter | lanta:ba: | # | # | # | # | # | # | # | # | # | #
noch | nO | # | # | # | # | # | # | # | # | # | #
bisschen | bIdn | # | # | # | # | # | # | # | # | # | #
Brennholz | bREnhOlt | # | # | # | # | # | # | # | # | # | #
holen | ha:l n | # | # | # | # | # | # | # | # | # | #
```

Ausschnitt einer unbearbeiteten Morphemtabelle

- Je nach Dialektraum weisen Verben verschiedene Stämme auf (z. B. Auxiliärverb *sein*). Das obige Beispiel stammt aus dem ostfälischen Gebiet (Großraum: Niederdeutsch). Im Falle von *sein* müsste man die Grundform in einem ostfäl. Wörterbuch nachschlagen; diese Vorarbeit wurde bereits geleistet (zumindest für die frequentesten Wörter):

Sein	Sin, wesen, ween Prät. Was, wörs, was, wören Part. Perf. Ewesen, eween
------	--

Ausschnitt der Lemmata-Tabelle für das Ostfälische. Beispiel: *sein*

```
*morphologie-dialect-l-62-1_2-3-4.txt - Editor
Datei Bearbeiten Format Ansicht Hilfe
früher | # | # | # | # | # | # | # | # | # | # | fROY6 | #
war | sIn | # | # | # | # | # | # | # | # | # | # | vat
Astfeld | asfElt | # | # | # | # | # | # | # | # | # | # | #
da | # | # | # | # | # | # | # | # | # | # | da | #
mussten | mOs | # | # | # | # | # | # | # | # | # | # | #
Hüttenleute | hUd, lOY@ | # | # | # | # | # | # | # | # | # | # | #
Landarbeiter | lant, a:ba: | # | # | # | # | # | # | # | # | # | # | #
noch | # | # | # | # | # | # | # | # | # | # | nO | #
bisschen | bIdn | # | # | # | # | # | # | # | # | # | # | #
Brennholz | bREn, hOlt | # | # | # | # | # | # | # | # | # | # | #
holen | ha:l | # | # | # | # | # | # | # | # | # | # | #
```

Ausschnitt von einer bearbeiteten Morphemtabelle

2.5 Sonstiges

- Zwei oder mehr Stämme wurden durch ein Komma getrennt (z. B. Apfelbaum: apf@l, baUm)
- Flektierte Formen und Suppletivformen werden auf den Infinitivstamm zurückgeführt:
 - z. B. musst | mYs | # | # | # | # | # | # | # | # | # | # | #
 - besser | gu:t | # | # | # | # | # | # | # | # | # | # | # | bEs6
 - Bäume | baUm | # | # | # | # | # | # | # | # | # | # | # | #

2.6 Sonderfälle

- Ortsnamen, Eigennamen und Familiennamen wurden nicht klassifiziert und blieben im Stamm umklassifiziert stehen
- Berufsbezeichnungen wurden klassifiziert: Zimmermann → Zimmer, Mann (beides in der Stamm-Position)
- Endungen wie *-(a)teur*, *-ation* in Fremd- und Lehnwörtern fungieren wie das deutsche *er*-Suffix
- Weitere Fälle sind in der Dokumentation bzw. im Leitfaden (s. oben) gelistet

2.7. Korrekturen

- Falls es Korrekturen gab, die vorgenommen werden mussten (etwa nachträgliche Markierung nicht kognater Wörter, z. B. std. *Kartoffeln* – dia. *Grumberren*), so wurden solche Fälle in einem separaten Dokument gesammelt. Da für die Nachkorrektur die Korpuszeilen benötigt wurden, mussten diese der aktuellsten Korpusversion entnommen und entsprechend dokumentiert werden.

Literatur

- Clements, George N. / Keyser, Samuel J. (1983): *CV Phonology: A Generative Theory of the Syllable*. Cambridge, MA: MIT.
- Duden-Grammatik (2016) = Duden, die Grammatik. Unentbehrlich für richtiges Deutsch. In: Wöllstein, Angelika (Hrsg.). 9., vollst., überarb. u. akt. Aufl. Bd. 4. Berlin: Dudenverlag.
- Féry, Caroline (2017): Die allophonischen Frikative in der Standardsprache und in den hessischen Dialekten. In: Konopka, Marek / Angelika Wöllstein (Hgg.), *Grammatische Variation. Empirische Zugänge und theoretische Modellierung*. Berlin u. a.: De Gruyter, 181–202.
- Fuhrhop, Nanna / Jörg Peters (2013): *Einführung in die Phonologie und Graphematik*. Stuttgart: Metzler.
- Hall, Tracy Alan (1992): *Syllable Structure and Syllable-Related Processes in German*. Berlin / New York: Niemeyer.
- Hall, Tracy Alan (2011): *Phonologie. Eine Einführung*. 2., überarb. Aufl. Berlin, New York: De Gruyter.
- Hentschel, Elke / Harald Weydt (2013): *Handbuch der deutschen Grammatik*. 4., vollst. überarb. Aufl. Berlin, Boston: Walter de Gruyter.
- Herrgen, Joachim (1986): *Koronalisierung und Hyperkorrektur: das palatale Allophon des /ch/-Phonems und seine Variation im Westmitteldeutschen*. Layoutgetreues Digitalisat der Ausg.: Stuttgart: Steiner. Universität Marburg, Forschungszentrum Deutscher Sprachatlas. Erschienen als: *Mainzer Studien zur Sprach- und Volksforschung*; Bd. 9. Zugl.: Mainz, Univ., Diss.
- Kohler, Klaus J. (1995): *Einführung in die Phonetik des Deutschen*. 2. Aufl. Berlin: Erich Schmidt.
- Lameli, Alfred / Alexander Werth (2017): *Komplexität und Indexikalität – Zum funktionalen Gehalt phonotaktischer Wortstrukturen im Deutschen*. In: Hennig, Mathilde (Hrsg.), *Linguistische Komplexität – ein Phantom?* Tübingen: Stauffenburg, 73–96.
- Menzerath, Paul (1954): *Die Architektonik des deutschen Wortschatzes*. Stuttgart: Dümmler (Phonetische Studien 3).
- Paul, Hermann (2007): *Mittelhochdeutsche Grammatik*. 25. Aufl. neu bearb. v. Thomas Klein, Hans-Joachim Solms und Klaus-Peter Wegera. In: Klein, Thomas / Ingo Reiffenstein (Hgg.), *begründet von Wilhelm Braune, Sammlung kurzer Grammatiken germanischer Dialekte*. Tübingen: Niemeyer.
- Schiller, Anne / Simone Teufel / Christine Stöckert (1999): *Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset)*. Universität Tübingen. <http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf> (Stand: 12.10.2021).
- Schmidt, Jürgen Erich (1986): *Die mittelfränkischen Tonakzente: (rheinische Akzentuierung)*. *Mainzer Studien zur Sprach- und Volksforschung*, Bd. 8. Stuttgart: Steiner.
- Werth, Alexander (2007): *Perzeptionsphonologische Studien zu den mittelfränkischen Tonakzenten*. In: *Zeitschrift für Dialektologie und Linguistik* 74, 292–316.

- Wiese, Richard (1988): Silbische und lexikalische Phonologie. Studien zum Chinesischen und Deutschen. Tübingen: Niemeyer.
- Wiese, Richard (1991): Was ist extrasilbisch im Deutschen und warum? In: Zeitschrift für Sprachwissenschaft 10, 3–36.
- Wurzel, Wolfgang Ulrich (1990): The Mechanism of Inflection: Lexical Representation, Rules and Irregularities. In: Dressler, Wolfgang U. / Hans C. Luschützky / Oskar E. Pfeiffer / John R. Rennison (Hgg.), Contemporary Morphology. Berlin, New York: De Gruyter, 203–216.