

Das ZDL-Regionalkorpus

*Ein Korpus für die Untersuchung diatopischer Variation
im Standarddeutschen*

Dr. Andreas Nolda

Berlin-Brandenburgische Akademie der Wissenschaften (BBAW)

Zentrum für digitale Lexikographie der deutschen Sprache (ZDL)

<https://andreas.nolda.org>

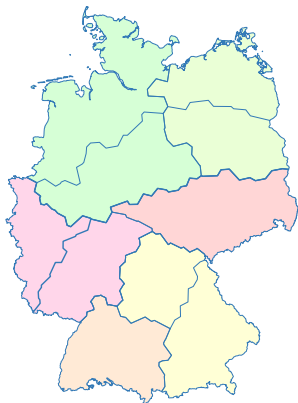
Seminar „e-Lexikographie“

Universität Potsdam

23. Februar 2023

Das ZDL-Regionalkorpus im Überblick

Das ZDL-Regionalkorpus ist ein Korpus für die Untersuchung diatopischer Variation im geschriebenen Standarddeutsch.



Zur Erstellung dieser Grafik wurde Kartenmaterial von www.regionalsprache.de verwendet.
Die Abzugsgrenzen orientieren sich an der Dialektgliederung bei Lutz (2013: 204).

Das ZDL-Regionalkorpus im Überblick

Das ZDL-Regionalkorpus ist ein Korpus für die Untersuchung diatopischer Variation im geschriebenen Standarddeutsch.

Es enthält aktuell 32,4 Mio. Artikel aus Lokal- und Regionalressorts von 26 Tageszeitungen aus Deutschland mit 9,4 Mrd. Tokens.



Zur Erstellung dieser Grafik wurde Kartenmaterial von www.regionalpraech.de verwendet.
Die Ansgrenzen orientieren sich an der Dialektgliederung bei Lenz (2012: 214).

Das ZDL-Regionalkorpus im Überblick

Das ZDL-Regionalkorpus wird gehostet am Zentrum für digitale Lexikographie der deutschen Sprache (ZDL).



Das ZDL-Regionalkorpus im Überblick

Das ZDL-Regionalkorpus wird gehostet am [Zentrum für digitale Lexikographie der deutschen Sprache \(ZDL\)](#).

Es wurde konzipiert und implementiert an der [Berlin-Brandenburgischen Akademie der Wissenschaften \(BBAW\)](#) von Andreas Nolda, Adrien Barbaresi, Alexander Geyken, Bryan Jurish, Frank Wiegand u. a.



Das ZDL-Regionalkorpus im Überblick

Das ZDL-Regionalkorpus wird gehostet am [Zentrum für digitale Lexikographie der deutschen Sprache \(ZDL\)](#).

Es wurde konzipiert und implementiert an der [Berlin-Brandenburgischen Akademie der Wissenschaften \(BBAW\)](#) von Andreas Nolda, Adrien Barbaresi, Alexander Geyken, Bryan Jurish, Frank Wiegand u. a.

Das ZDL wird getragen von der BBAW und den [Akademien in Göttingen](#), [Leipzig](#) und [Mainz](#) unter dem Dach der Union der deutschen Akademien der Wissenschaften.



Akademie der Wissenschaften
zu Göttingen



Akademie der Wissenschaften
und der Literatur | Mainz

Das ZDL-Regionalkorpus im Überblick

Das ZDL-Regionalkorpus wird gehostet am [Zentrum für digitale Lexikographie der deutschen Sprache \(ZDL\)](#).

Es wurde konzipiert und implementiert an der [Berlin-Brandenburgischen Akademie der Wissenschaften \(BBAW\)](#) von Andreas Nolda, Adrien Barbaresi, Alexander Geyken, Bryan Jurish, Frank Wiegand u. a.

Das ZDL wird getragen von der BBAW und den [Akademien in Göttingen, Leipzig und Mainz](#) unter dem Dach der Union der deutschen Akademien der Wissenschaften.

Das ZDL kooperiert mit dem [Institut für deutsche Sprache \(IDS\) in Mannheim](#) und wird finanziert durch das [Bundesministerium für Bildung und Forschung \(BMBF\)](#).



Akademie der Wissenschaften
zu Göttingen



Akademie der Wissenschaften
und der Literatur | Mainz

IDS

LEIBNIZ-INSTITUT FÜR
DEUTSCHE SPRACHE

Bundesministerium
für Bildung
und Forschung

Das ZDL-Regionalkorpus im Überblick

Das ZDL-Regionalkorpus ist Datengrundlage für die Regionalangaben im „Digitalen Wörterbuch der deutschen Sprache“ (DWDS).

Fasching



Fasching, der

Grammatik Substantiv (Maskulinum) · Genitiv Singular: **Faschings** · Nominativ Plural: **Faschinge**

Aussprache [ˈfaʃɪŋ]

Worttrennung Fa-sching

Wortbildung mit ›Fasching‹ als Erstglied: ↗ [Faschingsball](#) ... [19 weitere](#) · mit ›Fasching‹ als Letztglied: ↗ [Kinderfasching](#) · ↗ [Weiberfasching](#)

Herkunft nicht mehr durchsichtiges Kompositum aus ↗ [fasten](#) und ↗ [Schank](#)

Bedeutung weitere Informationen zum Sprachareal

eWDG und ZDL

↘ [besonders D-Südost](#) , [oft D-Mittelost](#) , [A](#) [Synonym zu \[Fastnacht\]\(#\) \(●\), \[Karneval\]\(#\)](#)

Das ZDL-Regionalkorpus im Überblick

Das ZDL-Regionalkorpus ist Datengrundlage für die Regionalangaben im „Digitalen Wörterbuch der deutschen Sprache“ (DWDS).

Das ZDL-Regionalkorpus kann von jedem [angemeldeten DWDS-Nutzer](#) für eigene Recherchen genutzt werden.

ZDL-Regionalkorpus (ab 1993)

Hinweis: Dieses Korpus ist erst nach [Anmeldung](#) recherchierbar.

Suche in: ZDL-Regionalkorpus (ab 1993)



Gliederung

- 1 Diatopische Variation im Deutschen
- 2 Design des ZDL-Regionalkorpus
- 3 User-Interface des ZDL-Regionalkorpus
- 4 Daten zur diatopischen Variation im ZDL-Regionalkorpus

Gliederung

- 1 Diatopische Variation im Deutschen
- 2 Design des ZDL-Regionalkorpus
- 3 User-Interface des ZDL-Regionalkorpus
- 4 Daten zur diatopischen Variation im ZDL-Regionalkorpus

Sprachliche Variation

Sprachliche Variation in einer Sprache lässt sich mit (*sprachlichen*) *Variablen* beschreiben.

Die Ausprägungen einer sprachlichen Variablen sind (*sprachliche*) *Varianten*.

Sprachliche Variation

Sprachliche Variation in einer Sprache lässt sich mit (*sprachlichen*) *Variablen* beschreiben.

Die Ausprägungen einer sprachlichen Variablen sind (*sprachliche*) *Varianten*.

Aussprache von <ich>

- [ɪç]
- [ɪ]
- [ɪk(ə)]
- [i]

Sprachliche Variation

Sprachliche Variation in einer Sprache lässt sich mit (*sprachlichen*) *Variablen* beschreiben.

Die Ausprägungen einer sprachlichen Variablen sind (*sprachliche*) *Varianten*.

Genus von „Joghurt“

- Maskulinum
- Neutrum

Sprachliche Variation

Sprachliche Variation in einer Sprache lässt sich mit (*sprachlichen*) *Variablen* beschreiben.

Die Ausprägungen einer sprachlichen Variablen sind (*sprachliche*) *Varianten*.

Bezeichnung für die Zeit vor Aschermittwoch

- „Fasching“
- „Karneval“
- „Fas(t)nacht“

Sprachraum

Der *Sprachraum* einer Sprache ist das geographische Gebiet, in dem die Sprache gemeinsame Erstsprache der Bevölkerung (smehrheit) ist.

Sprachraum

Der *Sprachraum* einer Sprache ist das geographische Gebiet, in dem die Sprache gemeinsame Erstsprache der Bevölkerung (smehrheit) ist.

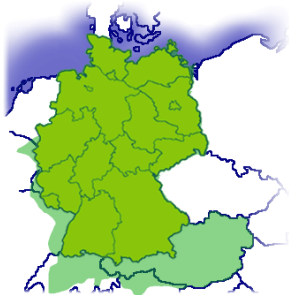
Oft hat die Sprache dort zudem den Status einer Amtssprache.

Sprachraum

Der *Sprachraum* einer Sprache ist das geographische Gebiet, in dem die Sprache gemeinsame Erstsprache der Bevölkerung (smehrheit) ist.

Oft hat die Sprache dort zudem den Status einer Amtssprache.

Der deutsche Sprachraum:



nach https://wiki-de.genealogy.net/Datei:Lokal_Dialekt_Dialekte.png
[25. Okt. 2013]

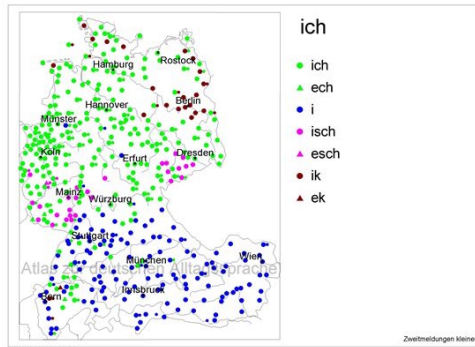
Diatopische Variation

Diatopische Variation in einer Sprache ist sprachliche Variation im Sprachraum der Sprache.

Diatopische Variation

Diatopische Variation in einer Sprache ist sprachliche Variation im Sprachraum der Sprache.

Bei diatopischen sprachlichen Variablen sind die Varianten nicht gleichförmig über den Sprachraum verteilt:

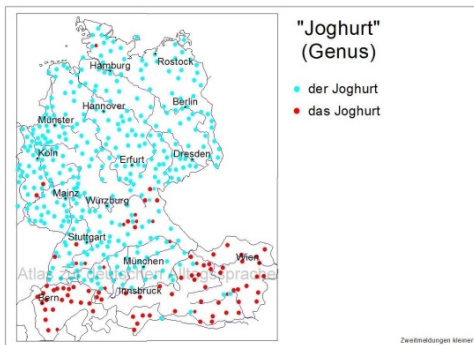


<https://www.atlas-alltagssprache.de/runde-2/f25c/> [2. Sep. 2017]

Diatopische Variation

Diatopische Variation in einer Sprache ist sprachliche Variation im Sprachraum der Sprache.

Bei diatopischen sprachlichen Variablen sind die Varianten nicht gleichförmig über den Sprachraum verteilt:

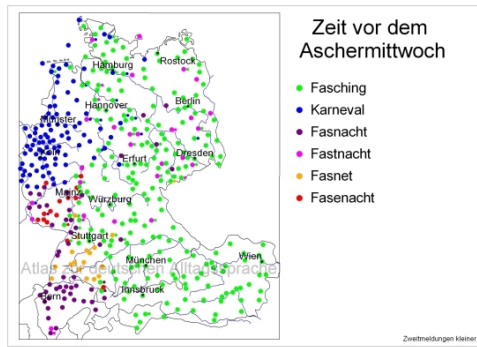


<https://www.atlas-alltagssprache.de/runde-4/f24a-g/> [26. Nov. 2015]

Diatopische Variation

Diatopische Variation in einer Sprache ist sprachliche Variation im Sprachraum der Sprache.

Bei diatopischen sprachlichen Variablen sind die Varianten nicht gleichförmig über den Sprachraum verteilt:



<https://www.atlas-alltagssprache.de/runde-2/f03/> [13. Jan. 2021]

Diatopische Varietäten

Diatopische Varietäten einer Sprache sind Ausprägungen der Sprache, deren System durch bestimmte diatopische Varianten charakterisiert ist.

Diatopische Varietäten

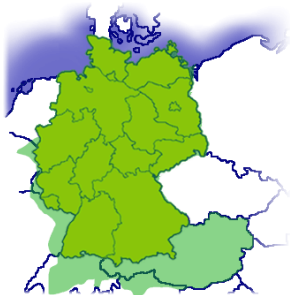
Diatopische Varietäten einer Sprache sind Ausprägungen der Sprache, deren System durch bestimmte diatopische Varianten charakterisiert ist.

In der modernen variationslinguistischen Literatur unterscheidet man zwei Arten diatopischer Varietäten des Deutschen:

Dialekte: traditionelle diatopische Ausprägungen des Deutschen.

Regiolekte: standardnahe diatopische Ausprägungen des Deutschen.

Dialekte

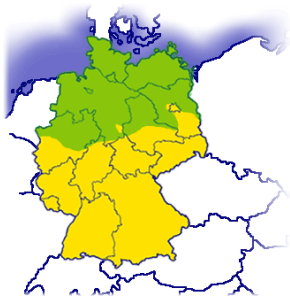


nach https://wiki-de.genealogy.net/Datei:Lokal_Dialekt_Dialekte.png
[25. Okt. 2013]

Die deutschen Dialekte zerfallen in zwei große Dialektgruppen:

- *niederdeutsche* Dialekte und
- *hochdeutsche* Dialekte.

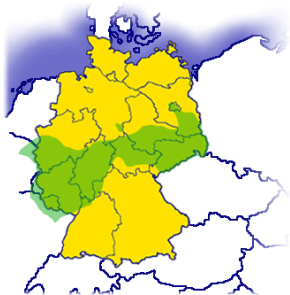
Dialekte



nach https://wiki-de.genealogy.net/Datei:Lokal_Dialekt_Niederdeutsch.png [25. Okt. 2013]

Niederdeutsche Dialekte spricht man in Norddeutschland.

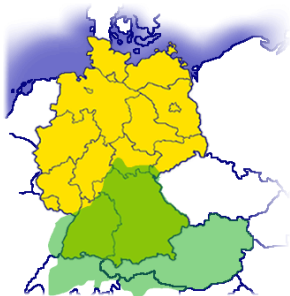
Dialekte



[https://wiki-de.genealogy.net/Datei:
Lokal_Dialekt_Mitteldeutsch.png](https://wiki-de.genealogy.net/Datei:Lokal_Dialekt_Mitteldeutsch.png) [25. Okt. 2013]

*Hochdeutsche Dialekte gliedern sich in **mitteldeutsche Dialekte** und *oberdeutsche Dialekte**

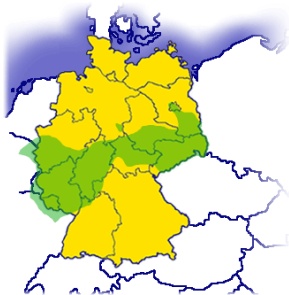
Dialekte



https://wiki-de.genealogy.net/Datei:Lokal_Dialekt_Oberdeutsch.png
[25. Okt. 2013]

Hochdeutsche Dialekte gliedern sich in mitteldeutsche Dialekte und oberdeutsche Dialekte

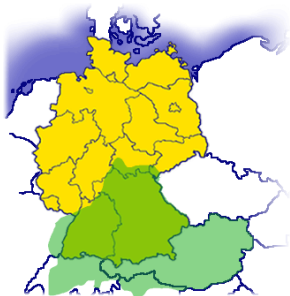
Dialekte



[https://wiki-de.genealogy.net/Datei:
Lokal_Dialekt_Mitteldeutsch.png](https://wiki-de.genealogy.net/Datei:Lokal_Dialekt_Mitteldeutsch.png) [25. Okt. 2013]

Mitteldeutsche Dialekte wiederum gliedern sich in *westmitteldeutsche Dialekte* und *ostmitteldeutsche Dialekte*.

Dialekte



https://wiki-de.genealogy.net/Datei:Lokal_Dialekt_Oberdeutsch.png
[25. Okt. 2013]

Zu den oberdeutschen Dialekten gehören *alemannische Dialekte* und *bairische Dialekte*.

Standardvarietät

Die Standardvarietät des Deutschen (*Standarddeutsch*, ‚Hochdeutsch‘, ‚Schriftdeutsch‘) entwickelte sich seit dem 16. Jahrhundert, und zwar zunächst als schriftliches Register auf der Basis von:

1. der (ostmittel)deutschen Bibelübersetzung von Martin Luther (1545);
2. der (ostmittel)deutschen Kanzleisprache.

Standardvarietät

Die Standardvarietät des Deutschen (*Standarddeutsch*, ‚Hochdeutsch‘, ‚Schriftdeutsch‘) entwickelte sich seit dem 16. Jahrhundert, und zwar zunächst als schriftliches Register auf der Basis von:

1. der (ostmittel)deutschen Bibelübersetzung von Martin Luther (1545);
2. der (ostmittel)deutschen Kanzleisprache.

Erst später entwickelten sich auch mündliche Ausprägungen des Standarddeutschen durch den Einfluss von:

1. der (nieder)deutschen Bühnenaussprache von Theodor Siebs (1898);
2. Rundfunk und Fernsehen.

Regiolekte

In Regiolekten ist das gesprochene Standarddeutsch durch den dialektalen Hintergrund der Sprecher oft *dialektal gefärbt*:

- niederdeutsch gefärbtes Standarddeutsch aus Schleswig-Holstein
- ostmitteldeutsch gefärbtes Standarddeutsch aus Sachsen-Anhalt
- westmitteldeutsch gefärbtes Standarddeutsch aus Nordrhein-Westfalen
- alemannisch gefärbtes Standarddeutsch aus Baden-Württemberg
- bairisch gefärbtes Standarddeutsch aus Bayern.

Plurizentrik

Ammon (1995) versteht das Deutsche als eine *plurizentrische* Sprache und unterteilt den deutschen Sprachraum in mehrere nationale *Zentren*:

Vollzentren: Deutschland, Österreich, deutschsprachige Schweiz

Halbzentren: Ostbelgien, Luxemburg, Liechtenstein, Südtirol

Ammon, Ulrich (1995). *Die deutsche Sprache in Deutschland, Österreich und der Schweiz: Das Problem der nationalen Varietäten*. Berlin: de Gruyter.

Plurizentrik

Ammon (1995) versteht das Deutsche als eine *plurizentrische* Sprache und unterteilt den deutschen Sprachraum in mehrere nationale *Zentren*:

Vollzentren: Deutschland, Österreich, deutschsprachige Schweiz

Halbzentren: Ostbelgien, Luxemburg, Liechtenstein, Südtirol

Klassifikationskriterium ist hierbei der Grad der Kodifizierung der diatopischen Varietäten (insbesondere durch entsprechende Wörterbücher).

Ammon, Ulrich (1995). *Die deutsche Sprache in Deutschland, Österreich und der Schweiz: Das Problem der nationalen Varietäten*. Berlin: de Gruyter.

Pluriarealität

Wolf (1994), Elspaß und Dürscheid (2017) und andere betrachten das Deutsche hingegen als *pluriareale* Sprache, deren Sprachraum in rein sprachlich bestimmte *Areale* gegliedert ist.

Wolf, Norbert R. (1994). Rezension von Pollak, Wolfgang (1992), *Was halten die Österreicher von ihrem Deutsch? Eine sprachpolitische und soziosemiotische Analyse der sprachlichen Identität der Österreicher*, Wien: ÖGS/ISSS. *Zeitschrift für Dialektologie und Linguistik* 61, 66–76.

Elspaß, Stephan und Christa Dürscheid (2017). Areale grammatische Variation in den Gebrauchsstandards des Deutschen. In *Grammatische Variation: Empirische Zugänge und theoretische Modellierung*, hg. v. Marek Konopka und Angelika Wöllstein, Institut für Deutsche Sprache: Jahrbuch 2016, Berlin: de Gruyter, 85–104.

Pluriarealität

Wolf (1994), Elspaß und Dürscheid (2017) und andere betrachten das Deutsche hingegen als *pluriareale* Sprache, deren Sprachraum in rein sprachlich bestimmte *Areale* gegliedert ist.

Häufig fallen standardsprachliche Arealgrenzen mit Dialektgrenzen zusammen.

Wolf, Norbert R. (1994). Rezension von Pollak, Wolfgang (1992), *Was halten die Österreicher von ihrem Deutsch? Eine sprachpolitische und soziosemiotische Analyse der sprachlichen Identität der Österreicher*, Wien: ÖGS/ISSS. *Zeitschrift für Dialektologie und Linguistik* 61, 66–76.

Elspaß, Stephan und Christa Dürscheid (2017). Areale grammatische Variation in den Gebrauchsstandards des Deutschen. In *Grammatische Variation: Empirische Zugänge und theoretische Modellierung*, hg. v. Marek Konopka und Angelika Wöllstein, Institut für Deutsche Sprache: Jahrbuch 2016, Berlin: de Gruyter, 85–104.

Arealgliederung im „Variantenwörterbuch“



Ammon et al. (2016: LIII)

Ammon, Ulrich et al. (2016). *Variante Wörterbuch des Deutschen: Die Standardsprache in Österreich, der Schweiz und Deutschland, Liechtenstein, Luxemburg, Ostbelgien und Südtirol sowie Rumänien, Namibia und Mennonitensiedlungen*. 2. Aufl. Berlin: Walter de Gruyter.

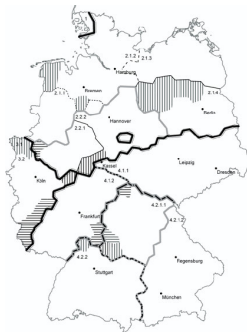
Arealgliederung in der „Variantengrammatik“



Datenerhebung (2018)

Datenerhebung (2018). In Variantengrammatik des Standarddeutschen: Ein Online-Nachschlagewerk. Verfasst von einem Autorenteam unter der Leitung von Christa Dürscheid, Stephan Elspaß und Arne Ziegler. <http://mediawiki.ids-mannheim.de/VarGra/index.php/Datenerhebung> [26. Juli 2019].

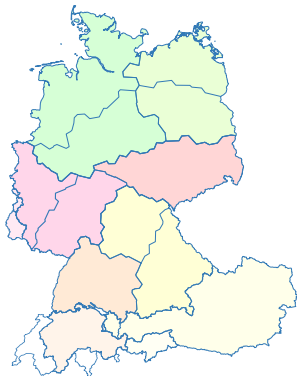
Arealgliederung von Alfred Lameli



Lameli (2013: 194)

Lameli, Alfred (2013). *Strukturen im Sprachraum: Analysen zur arealtypologischen Komplexität der Dialekte in Deutschland*. Linguistik – Impulse & Tendenzen 54. Berlin: de Gruyter.

Arealgliederung im DWDS und im ZDL-Regionalkorpus



Zur Erstellung dieser Grafik wurde Kartennaterial von www.regionsportal.de verwendet.
Die Kartengrenzen in Deutschland entsprechen sich auch der Staatsgliederung bis zum 31.12.2011.
Die Kartengrenzen der deutschsprachigen Schweiz beruhen auf Daten des Bundesamts für Statistik.

nach <https://www.dwds.de/d/regionalangaben> [21. Feb. 2023]

Nolda, Andreas, Adrien Barbaresi und Alexander Geyken (2023). Korpora für die lexikographische Beschreibung diatopischer Variation in der deutschen Standardsprache. In *Korpora in der germanistischen Sprachwissenschaft: Mündlich, schriftlich, multimedial*, hg. v. Arnulf Deppermann et al., Institut für Deutsche Sprache: Jahrbuch 2022, Berlin: de Gruyter, 29–52.

Gliederung

- 1 Diatopische Variation im Deutschen
- 2 Design des ZDL-Regionalkorpus
- 3 User-Interface des ZDL-Regionalkorpus
- 4 Daten zur diatopischen Variation im ZDL-Regionalkorpus

Desiderata für ein Korpus zur Untersuchung diatopischer Variation

Bickel, Hofer und Suter (2015: 544):

1. „[...] die Textbasis [muss] gezielt nach national und regional zuordnenbaren Texten abgesucht [...] werden können.“

Bickel, Hans, Lorenz Hofer und Sandra Suter (2015). Variantenwörterbuch des Deutschen (VWB) – NEU: Dynamik der deutschen Standardvariation aus lexikografischer Sicht. In *Regionale Variation des Deutschen: Projekte und Perspektiven*, hg. v. Roland Kehrein, Alfred Lameli und Stefan Rabanus, Berlin: de Gruyter, 541–562.

Desiderata für ein Korpus zur Untersuchung diatopischer Variation

Bickel, Hofer und Suter (2015: 544):

1. „[...] die Textbasis [muss] gezielt nach national und regional zuordnenbaren Texten abgesucht [...] werden können.“
2. „Das Korpus sollte [...] möglichst nur neuere und neuste standardsprachliche Texte enthalten.“

Bickel, Hans, Lorenz Hofer und Sandra Suter (2015). Variantenwörterbuch des Deutschen (VWB) – NEU: Dynamik der deutschen Standardvariation aus lexikografischer Sicht. In *Regionale Variation des Deutschen: Projekte und Perspektiven*, hg. v. Roland Kehrein, Alfred Lameli und Stefan Rabanus, Berlin: de Gruyter, 541–562.

Desiderata für ein Korpus zur Untersuchung diatopischer Variation

Bickel, Hofer und Suter (2015: 544):

1. „[...] die Textbasis [muss] gezielt nach national und regional zuordnenbaren Texten abgesucht [...] werden können.“
2. „Das Korpus sollte [...] möglichst nur neuere und neuste standardsprachliche Texte enthalten.“
3. „Das Korpus sollte groß genug sein, um auch bei selteneren Mehrwortverbindungen oder kleinräumigen Varianten aussagekräftige Treffermengen zu liefern.“

Bickel, Hans, Lorenz Hofer und Sandra Suter (2015). Variantenwörterbuch des Deutschen (VWB) – NEU: Dynamik der deutschen Standardvariation aus lexikografischer Sicht. In *Regionale Variation des Deutschen: Projekte und Perspektiven*, hg. v. Roland Kehrein, Alfred Lameli und Stefan Rabanus, Berlin: de Gruyter, 541–562.

Desiderata für ein Korpus zur Untersuchung diatopischer Variation

Bickel, Hofer und Suter (2015: 544):

1. „[...] die Textbasis [muss] gezielt nach national und regional zuordnenbaren Texten abgesucht [...] werden können.“
2. „Das Korpus sollte [...] möglichst nur neuere und neuste standardsprachliche Texte enthalten.“
3. „Das Korpus sollte groß genug sein, um auch bei selteneren Mehrwortverbindungen oder kleinräumigen Varianten aussagekräftige Treffermengen zu liefern.“
4. „Hilfreich wäre zudem ein zuverlässiges Wortartentagging [...].“

Bickel, Hans, Lorenz Hofer und Sandra Suter (2015). Variantenwörterbuch des Deutschen (VWB) – NEU: Dynamik der deutschen Standardvariation aus lexikografischer Sicht. In *Regionale Variation des Deutschen: Projekte und Perspektiven*, hg. v. Roland Kehrein, Alfred Lameli und Stefan Rabanus, Berlin: de Gruyter, 541–562.

Desiderata für ein Korpus zur Untersuchung diatopischer Variation

Bickel, Hofer und Suter (2015: 544):

1. „[...] die Textbasis [muss] gezielt nach national und regional zuordnenbaren Texten abgesucht [...] werden können.“
2. „Das Korpus sollte [...] möglichst nur neuere und neuste standardsprachliche Texte enthalten.“
3. „Das Korpus sollte groß genug sein, um auch bei selteneren Mehrwortverbindungen oder kleinräumigen Varianten aussagekräftige Treffermengen zu liefern.“
4. „Hilfreich wäre zudem ein zuverlässiges Wortartentagging [...].“
5. „Schließlich wäre es wünschenswert, [...] dass mindestens absolute und relative Frequenzen einer Variante bzw. ihrer Formen in den Vollzentren des Deutschen automatisiert erhoben werden können.“

Bickel, Hans, Lorenz Hofer und Sandra Suter (2015). Variantenwörterbuch des Deutschen (VWB) – NEU: Dynamik der deutschen Standardvariation aus lexikografischer Sicht. In *Regionale Variation des Deutschen: Projekte und Perspektiven*, hg. v. Roland Kehrein, Alfred Lameli und Stefan Rabanus, Berlin: de Gruyter, 541–562.

Desiderata für ein Korpus zur Untersuchung diatopischer Variation

Bickel, Hofer und Suter (2015: 544):

1. „[...] die Textbasis [muss] gezielt nach national und regional zuordnenbaren Texten abgesucht [...] werden können.“
2. „Das Korpus sollte [...] möglichst nur neuere und neuste standardsprachliche Texte enthalten.“
3. „Das Korpus sollte groß genug sein, um auch bei selteneren Mehrwortverbindungen oder kleinräumigen Varianten aussagekräftige Treffermengen zu liefern.“
4. „Hilfreich wäre zudem ein zuverlässiges Wortartentagging [...].“
5. „Schließlich wäre es wünschenswert, [...] dass mindestens absolute und relative Frequenzen einer Variante bzw. ihrer Formen in den Vollzentren des Deutschen automatisiert erhoben werden können.“

„Ein linguistisches Korpus, das alle diese Wünsche erfüllt, gibt es zur Zeit [2015] noch nicht.“

Bickel, Hans, Lorenz Hofer und Sandra Suter (2015). Variantenwörterbuch des Deutschen (VWB) – NEU: Dynamik der deutschen Standardvariation aus lexikografischer Sicht. In *Regionale Variation des Deutschen: Projekte und Perspektiven*, hg. v. Roland Kehrein, Alfred Lameli und Stefan Rabanus, Berlin: de Gruyter, 541–562.

Eigenschaften des ZDL-Regionalkorpus

Das ZDL-Regionalkorpus schließt die von Bickel, Hofer und Suter (2015) konstatierte Lücke:

1. Es enthält Artikel aus Lokal- und Regionalressorts von Tageszeitungen aus Deutschland (sowie künftig aus Österreich und der deutschsprachigen Schweiz), die mit Metadaten zu Land und sprachlichem Areal versehen sind.
2. Es deckt den Zeitraum ab 1993 ab und wird monatlich aktualisiert.
3. Es umfasst aktuell 32,4 Mio. Artikel mit insgesamt 9,4 Mrd. Tokens..
4. Es ist lemmatisiert und PoS-getaggt.
5. Das User-Interface stellt diverse Recherchetools zur Abfrage und Visualisierung absoluter und relativer Frequenzen zur Verfügung.

Eigenschaften des ZDL-Regionalkorpus

Das ZDL-Regionalkorpus schließt die von Bickel, Hofer und Suter (2015) konstatierte Lücke:

1. Es enthält Artikel aus Lokal- und Regionalressorts von Tageszeitungen aus Deutschland (sowie künftig aus Österreich und der deutschsprachigen Schweiz), die mit Metadaten zu Land und sprachlichem Areal versehen sind.
2. Es deckt den Zeitraum ab 1993 ab und wird monatlich aktualisiert.
3. Es umfasst aktuell 32,4 Mio. Artikel mit insgesamt 9,4 Mrd. Tokens..
4. Es ist lemmatisiert und PoS-getaggt.
5. Das User-Interface stellt diverse Recherchertools zur Abfrage und Visualisierung absoluter und relativer Frequenzen zur Verfügung.

Im Unterschied zu den Projektkorpora des *Variantenwörterbuchs des Deutschen* und der *Variante Grammatik* ist das ZDL-Regionalkorpus öffentlich verfügbar (nach Anmeldung).

Eigenschaften des ZDL-Regionalkorpus

Das ZDL-Regionalkorpus schließt die von Bickel, Hofer und Suter (2015) konstatierte Lücke:

1. Es enthält Artikel aus Lokal- und Regionalressorts von Tageszeitungen aus Deutschland (sowie künftig aus Österreich und der deutschsprachigen Schweiz), die mit Metadaten zu Land und sprachlichem Areal versehen sind.
2. Es deckt den Zeitraum ab 1993 ab und wird monatlich aktualisiert.
3. Es umfasst aktuell 32,4 Mio. Artikel mit insgesamt 9,4 Mrd. Tokens..
4. Es ist lemmatisiert und PoS-getaggt.
5. Das User-Interface stellt diverse Recherchertools zur Abfrage und Visualisierung absoluter und relativer Frequenzen zur Verfügung.

Im Unterschied zum Projektkorpus des *Variantenwörterbuchs des Deutschen* ist das ZDL-Regionalkorpus lemmatisiert und PoS-getaggt.

Eigenschaften des ZDL-Regionalkorpus

Das ZDL-Regionalkorpus schließt die von Bickel, Hofer und Suter (2015) konstatierte Lücke:

1. Es enthält Artikel aus Lokal- und Regionalressorts von Tageszeitungen aus Deutschland (sowie künftig aus Österreich und der deutschsprachigen Schweiz), die mit Metadaten zu Land und sprachlichem Areal versehen sind.
2. Es deckt den Zeitraum ab 1993 ab und wird monatlich aktualisiert.
3. Es umfasst aktuell 32,4 Mio. Artikel mit insgesamt 9,4 Mrd. Tokens..
4. Es ist lemmatisiert und PoS-getaggt.
5. Das User-Interface stellt diverse Recherchertools zur Abfrage und Visualisierung absoluter und relativer Frequenzen zur Verfügung.

Im Unterschied zu den Zeitungskorpora des *DeReKo* des Instituts für Deutsche Sprache enthält das ZDL-Regionalkorpus nur Artikel aus Lokal- und Regionalressorts und stellt spezifische Recherchertools zur Verfügung.

Areale und Zeitungen

Areale im ZDL-Regionalkorpus:



Zur Erstellung dieser Grafik wurde Kartenmaterial von www.regionalsprache.de verwendet.
Die Anwaltsnamen orientieren sich an der Dialektgliederung bei [Lamell \(2012: 104\)](http://Lamell (2012: 104)).

Areal	Artikel	Tokens
D-Nordwest	5,4 Mio.	1,3 Mrd.
D-Nordost	2,1 Mio.	0,6 Mrd.
D-Mittelwest	11,1 Mio.	3,4 Mrd.
D-Mittelost	5,6 Mio.	1,6 Mrd.
D-Südwest	3,5 Mio.	1,1 Mrd.
D-Südost	4,5 Mio.	1,5 Mrd.
gesamt	32,4 Mio.	9,4 Mrd.

Areale und Zeitungen

Zeitungen im ZDL-Regionalkorpus:



Zur Erstellung dieser Grafik wurde Kartenmaterial von www.regionalgeschichte.de verwendet.
Die Anrangsnummern orientieren sich an der Dialektgliederung bei Luwitt (2012: 104).

Areal	Zeitung	Zeitraum
D-Nordwest	Hamburger Abendblatt	ab 1999
	Kieler Nachrichten	ab 2017
	Neue Osnabrücker Zeitung	ab 2012
	Neue Westfälische	ab 2003
D-Nordost	Berliner Morgenpost	ab 1999
	Norddeutsche Neueste Nachrichten	ab 2012
	Der Prignitzer	ab 2012
	Schweriner Volkszeitung	ab 2004
	Der Tagesspiegel	ab 2005
D-Mittelwest	Aachener Zeitung	ab 2003
	Allgemeine Zeitung (Mainz)	ab 2002
	Frankfurter Rundschau	ab 1995
	Rhein-Zeitung	ab 1997
	Saarbrücker Zeitung	ab 1993
D-Mittelost	Döbelner Allgemeine Zeitung	ab 2011
	Dresdner Neueste Nachrichten	ab 2011
	Leipziger Volkszeitung	ab 1997
	Thüringer Allgemeine	ab 2000
D-Südwest	Badische Zeitung	ab 2003
	Reutlinger General-Anzeiger	ab 2007
	Südkurier	ab 1999
D-Südost	Fränkischer Tag	ab 2005
	Landshuter Zeitung	ab 2014
	Mittlbayerische	ab 2014
	Münchener Merkur	ab 2016
	Süddeutsche Zeitung	ab 2005

Gliederung

- 1 Diatopische Variation im Deutschen
- 2 Design des ZDL-Regionalkorpus
- 3 User-Interface des ZDL-Regionalkorpus**
- 4 Daten zur diatopischen Variation im ZDL-Regionalkorpus

Suche und Treffer

Korpusbelege ZDL-Regionalkorpus (ab 1993)

Fasching

Korpus: **Start:** **Ende:** **Areale:**

D-Nordwest D-Mittelwest D-Südwest
 D-Nordost D-Mittelost D-Südost

Anzeige: KWIC voll maximal **Sortierung:** **Treffer pro Seite:**

1-50 von 71471 Treffern

1: Fränkischer Tag, 18.02.2023
Fasching in Trossenfurt-Tretzendorf

2: Fränkischer Tag, 18.02.2023
Fasching Die Narren aus dem Landkreis haben sich nicht nur organisatorisch zusammengeschlossen, sondern auch tänzerisch.

3: Fränkischer Tag, 18.02.2023
Gerlinde Stache vom TanzRhythmus Hirschaid, die zugleich auch Sitzungspräsidentin im Häschaader **Fasching** ist, Silvia Heckes von der Tanzgarde Schammelsdorf sowie der Vorsitzende des Hallstadter Faschingsvereins, Manuel Reitberger, stießen dazu.

<https://www.dwds.de/r/?corpus=regional>

Histogrammansicht

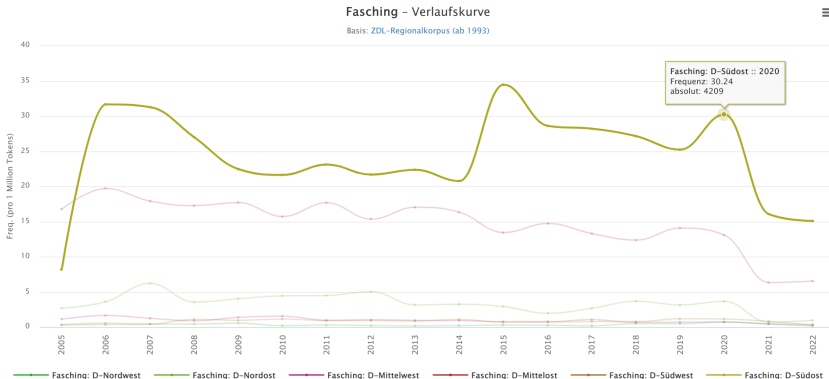
DWDS – Verlaufskurven – Basis: ZDL-Regionalkorpus (ab 1993)

Fasching

Korpus: ZDL-Regionalkorpus

Ansicht: geglättet

Zeitraum: 2005 2022 Areale:



Verteilung über Areale

Verteilung über Areale

[Karte anzeigen](#)

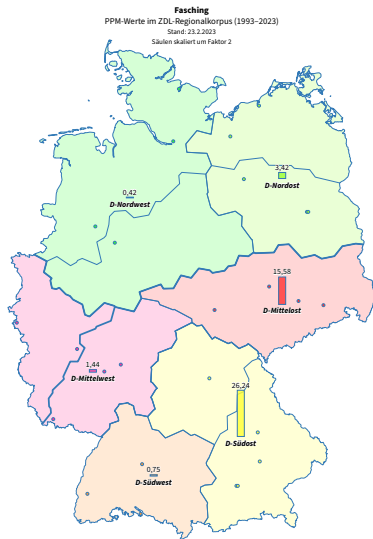
[Tabelle als CSV](#)

[Tabelle als Markdown](#)

Areal	Treffer	PPM	Anteil PPM
D-Südost	38579	26,24	54,85 %
D-Mittelost	24591	15,58	32,56 %
D-Mittelwest	4823	1,44	3,01 %
D-Nordost	2115	3,41	7,13 %
D-Südwest	806	0,75	1,57 %
D-Nordwest	557	0,42	0,88 %

PPM (*parts per million*):
Treffer-Tokens pro Million Tokens
im jeweiligen Areal
im Abfrage-Zeitraum

Verteilung über Areale



Zur Erstellung dieser Grafik wurde Kartenmaterial von www.regionalsprache.de verwendet.
Die Arealgrenzen orientieren sich an der Dialektgliederung bei Lameli (2013: 194).

PPM (*parts per million*):
Treffer-Tokens pro Million Tokens
im jeweiligen Areal
im Abfrage-Zeitraum

Verteilung über Zeitungen (Auszug)

Verteilung über Zeitungen

[Karte anzeigen](#)[Tabelle als CSV](#)[M+ Tabelle als Markdown](#)

Zeitung	Areal	Treffer	PPM
Thüringer Allgemeine	D-Mittelost	14520	18,11
Mittelbayerische	D-Südost	12997	32,88
Fränkischer Tag	D-Südost	10715	23,04
Leipziger Volkszeitung	D-Mittelost	9267	13,25
Münchner Merkur	D-Südost	8640	19,94
Landshuter Zeitung	D-Südost	5975	36,22
Saarbrücker Zeitung	D-Mittelwest	2209	2,44
Schweriner Volkszeitung	D-Nordost	1442	5,25
Rhein-Zeitung	D-Mittelwest	1217	0,98
Frankfurter Rundschau	D-Mittelwest	1032	2,95
Döbelner Allgemeine Zeitung	D-Mittelost	550	17,45
Reutlinger General-Anzeiger	D-Südwest	431	3,66
Südkurier	D-Südwest	306	0,39
Allgemeine Zeitung (Mainz)	D-Mittelwest	296	0,63
Berliner Morgenpost	D-Nordost	270	1,63
Hamburger Abendblatt	D-Nordwest	261	1,26
Dresdner Neueste Nachrichten	D-Mittelost	254	5,35
Süddeutsche Zeitung	D-Südost	252	11,71
Neue Westfälische	D-Nordwest	183	0,19
Der Prignitzer	D-Nordost	173	3,68
Norddeutsche Neueste Nachrichten	D-Nordost	151	3,64
Kieler Nachrichten	D-Nordwest	96	1,90

PPM (*parts per million*):
Treffer-Tokens pro Million Tokens
in der jeweiligen Zeitung
im Abfrage-Zeitraum

Verteilung über Zeitungen



Zur Erstellung dieser Grafik wurde Kartenmaterial von www.regionalsprache.de verwendet.
Die Arealgrenzen orientieren sich an der Dialektgliederung bei Lameli (2013: 194).

PPM (*parts per million*):
Treffer-Tokens pro Million Tokens
in der jeweiligen Zeitung
im Abfrage-Zeitraum

DWDS-API

ZDL-Regionalkorpus/Webmonitor: Verteilungen

Die in der Korpusuche zum [ZDL-Regionalkorpus](#) bzw. Korpus [Webmonitor](#) gelisteten Tabellen für die Verteilung von Treffern über Areale bzw. Zeitungen lassen sich unter dem URL <https://www.dwds.de/api/ppm> mit folgenden Parametern abfragen:

field	data type	description
<code>q</code> , <code>query</code>	string	DDC query string (cf. documentation Korpusuche – Suchmaschine und Suchabfragesprache)
<code>corpus</code>	string	valid values: <code>regional</code> , <code>regional_de_ch</code> , <code>webmonitor</code>
<code>by</code>	string	grouping criterion, valid values: <code>areal</code> , <code>corpus</code>
<code>format</code>	string	output format, valid values: <code>csv</code> , <code>json</code>

Die Ausgabe ist dann wie folgt:

field (JSON)	column (CSV)	description
<code>areal</code>	<code>Areal</code>	region, areal
<code>corpus_title</code>	<code>Zeitung</code>	newspaper title (not available for corpus <code>webmonitor</code>)
<code>corpus</code>	<code>Korpus</code>	internal corpus name
<code>hits</code>	<code>Treffer</code>	absolute number of hits
<code>ppm</code>	<code>PPM</code>	parts per million
<code>ppm_rel</code>	<code>Anteil PPM</code>	percentage of total sum of all ppm values

Beispielabfragen

Abfrage	Erläuterung
https://www.dwds.de/api/ppm?q=Haus&corpus=regional&by=areal&format=csv	Suche nach <i>Haus</i> im ZDL-Regionalkorpus, gruppiert nach Areal, Ausgabeformat CSV

<https://www.dwds.de/d/api>

DDC-Suchabfragesprache

Korpussuche – Suchmaschine und Suchabfragesprache

Die DWDS-Korpussuche bietet ein umfangliches Spektrum an Möglichkeiten für Anfragen an unsere [Textsammlungen](#) (Korpora). Entsprechend reichhaltig (und manchmal auch komplex) ist die Abfragesprache, die mittels der Software [DDC](#) realisiert wurde.

Hinweis

Alle Suchbeispiele in dieser Dokumentation sind anklickbar und führen Sie direkt zu Korpusergebnissen.
Für Fragen zur Korpussuche schreiben Sie uns gerne eine [Nachricht](#).

Schnellübersicht DDC

In dieser Schnellübersicht werden die wichtigsten Konzepte der Abfragesprache für Suchen in den DWDS-Korpora gelistet.

Sucheingabe	Bemerkung	Beispieldreffer
Haus	lemmbasierte Suche	<i>Haus, Hauses, Häuser, Häusern, ...</i>
@Haus	exakte Wortform	<i>Haus</i>
Haus*	Präfixsuche	<i>Haus, Hausmeister, ...</i>
*haus	Suffixsuche	<i>Elternhaus, zuhaus, ...</i>
haus	Infixsuche	<i>hausfraulich, Verlagshaus, Schauspiel, ...</i>
/ha[mu]s?t/	regulärer Ausdruck	<i>schaut, Hochamt, Goldhamster, ...</i> [h , gefolgt von m oder u , danach optional s , dann t]

<https://www.dwds.de/d/korpussuche>

Gliederung

- 1 Diatopische Variation im Deutschen
- 2 Design des ZDL-Regionalkorpus
- 3 User-Interface des ZDL-Regionalkorpus
- 4 Daten zur diatopischen Variation im ZDL-Regionalkorpus

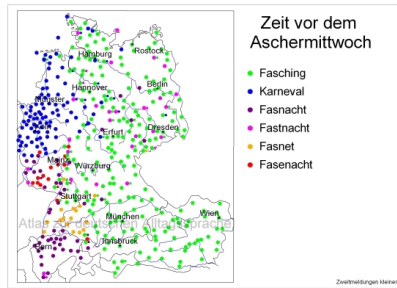
Bezeichnung für die Zeit vor Aschermittwoch: „Fasching“

ZDL-Regionalkorpus:



Zur Erstellung dieser Grafik wurde Kartematerial von www.regionalsprache.de verwendet.
Die Abwägungen orientieren sich an der Dialektgliederung bei Lamski (2013: 194).

Atlas zur deutschen Alltagssprache:



<https://www.atlas-alltagssprache.de/runde-2/f03/>
[13. Jan. 2021]

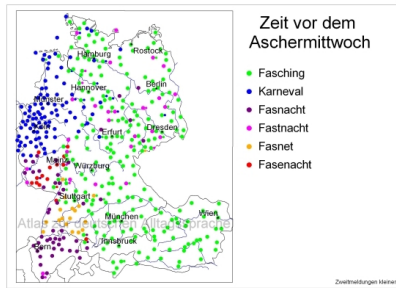
Bezeichnung für die Zeit vor Aschermittwoch: „Karneval“

ZDL-Regionalkorpus:



Zur Erstellung dieser Grafik wurde Kartenmaterial von www.regionalsprache.de verwendet.
Die Abwägungen orientieren sich an der Dialektgliederung bei Lamel (2013: 194).

Atlas zur deutschen Alltagssprache:



<https://www.atlas-alltagssprache.de/runde-2/f03/>
[13. Jan. 2021]

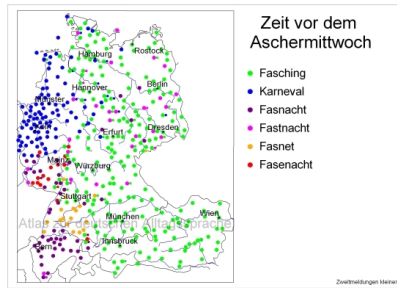
Bezeichnung für die Zeit vor Aschermittwoch: „Fas(t)nacht“-Varianten

ZDL-Regionalkorpus:



Zur Erstellung dieser Grafik wurde Kartematerial von www.regionalsprache.de verwendet. Die Abwägungen orientieren sich an der Dialektgliederung bei Lamski (2013: 194).

Atlas zur deutschen Alltagssprache:



<https://www.atlas-alltagssprache.de/runde-2/f03/>
[13. Jan. 2021]

Weitere Daten

Aufgabe

Formulieren Sie Suchabfragen im ZDL-Regionalkorpus zu den folgenden sprachlichen Variablen und vergleichen Sie Ihre Recherche-Ergebnisse mit einschlägigen Karten im „Atlas zur deutschen Alltagssprache“:

1. *Bezeichnung für ein Behältnis für Schreibutensilien*

- „Federtasche“
- „Federmappe“
- „Federmäppchen“
- ...

2. *Bezeichnung für das laufende Jahr*

- „dieses Jahr“
- „heuer“

3. *Genus von „Joghurt“*

- Maskulinum
- Neutrum

4. *Auxiliar für das Perfekt von Verben wie „stehen“ oder „sitzen“*

- „haben“
- „sein“

5. eine weitere Variable Ihrer Wahl

<https://www.dwds.de/r/?corpus=regional>

<https://www.dwds.de/d/korpussuche>

<https://www.atlas-alltagssprache.de/suchregister/>