

Methodisch geht es korpuswärts! Zur Produktivität des Suffixes *-wärts*

1 Einleitung

Eine Tätigkeit wie das Bergsteigen ist, wie Bubenhofer/Rothenhäusler treffend bemerken, geprägt von bestimmten Abfolgen und Handlungsmustern z.B. der physikalischen Notwendigkeit unten zu starten, sich nach oben zu bewegen und dann wieder nach unten zu kommen: „Die Anatomie einer Bergtour scheint prototypisch also ein Unten-Hoch-Gipfel-Runter-Unten zu sein, das einer topographisch erzwungenen Sequenz folgt“ (2018: 39). Man kann also auch erwarten, dass sich in Texten, die sich mit dieser Tätigkeit befassen, sprachliche Abfolgen und Muster wiederfinden, die eine solche Bergtouren-Anatomie widerspiegeln. Dieser Gedanke bildet den Ausgangspunkt für unsere Überlegungen in diesem Beitrag. Wir gehen zunächst davon aus, dass in alpinbezogenen Texten Raum- oder Direktionaladverbien eine größere Rolle spielen als in anderen Textsorten. Dies bestätigt sich in der unterschiedlichen relativen Häufigkeit, wie in 4 näher beschrieben wird.

Um den Untersuchungsgegenstand weiter einzuschränken, führten wir zwei Keyword-Analysen im thematisch eng fokussierten Korpus Alpenwort - Korpus der Zeitschrift des österreichischen Alpenvereins durch, das von uns kontinuierlich aufgebaut und versioniert publiziert wird. Alpenwort enthält 19,9 Millionen Wörter und umfasst derzeit alle Bände der als Jahrbuch erscheinenden Zeitschrift von 1870–2010. Die Daten wurden umfangreich korrigiert und vollständig mit dem STTS-Tagset annotiert (für eine genaue Beschreibung der Korpus-Genese siehe Posch/Rampl 2020) und sind über die Plattform CQPweb für Analysen frei zugänglich (Rampl/Posch 2019). Das Korpus ist so aufgebaut, dass sich über den gesamten Zeitraum der Publikation der Zeitschrift diachrone Analysen anstellen lassen. Enthalten sind unterschiedlichste Textsorten, von wissenschaftlichen Artikeln, die eher in den

frühen Publikationen eine größere Rolle spielen, über zahlreiche Besteigungs- und Expeditionsberichte bis zu literarischen Texten und kleineren Textsorten. Allen gemein ist das Interesse an den Themen „Berg“ und „Bergsteigen“.

Dabei zeigte sich die interessante Entdeckung, dass mit dem Suffix *-wärts* gebildete Direktionaladverbien nicht gleichmäßig im Korpus verteilt sind. Aus dieser Entdeckung entstand die leitende Untersuchungsfrage des vorliegenden Beitrags, nämlich wie sich dieses Suffix *-wärts* diachron entwickelt und ob sich diese Entwicklung in verschiedenen Korpora und Textsorten gleich niederschlägt.

2 Keyness der Direktionaladverbien in Alpenwort

Um die eingangs gestellte Frage, welche Raum- und Direktionaladverbien in alpinbezogenen Texten wichtig sind, beantworten zu können, führten wir Keyword-Analysen durch. Dadurch konnten wir zu einer für die Untersuchung einerseits handhabbaren und andererseits auch relevanten Teilmenge der Raum- und Direktionaladverbien gelangen. Gabrielatos folgend betrachten wir *keyness* und die Keyword-Methode als „Weg in Texte hinein (2018: 227)“, also als eine Technik, um für linguistische Fragestellungen interessante Items überhaupt erst zu finden. In einem ersten Schritt wurden hierfür auf zwei unterschiedliche Arten *candidate key items* (CKI) berechnet, um für die weiterführenden Forschungsfragen eine Auswahl treffen zu können.

Keyitems 1: Für die erste Berechnung wurden Subkorpora der Jahrzehnte 1890 (Untersuchungskorpus, 153 Texte 1.903.643 Wörter) und 1990 (Referenzkorpus, 336 Texte, 1.687.216 Wörter) verwendet. Beide Subkorpora haben eine ähnliche Größe und sind sich auch inhaltlich ähnlich. Folgende Einstellungen in CQPweb wurden ausgewählt:

- Statistik: LogLikelihood,
- Signifikanz Cut-off: 0.01% (LL threshold = 15.14),
- Minimalfrequenz in beiden Subkorpora: 3.

Bei der manuellen Sichtung der CKI-Liste weckten schließlich einige zusammengesetzte Direktionaladverbien mit dem bisher selten untersuchten Suffix *-wärts* (6 von 46 Direktionaladverbien) unser Interesse.

Rang	Wort	Statistik
73	aufwärts	287.29
111	abwärts	20.03
399	nordwärts	72.00
478	ostwärts	61.46
715	rückwärts	40.31
753	vorwärts	38.30
988	südwärts	28.04
1022	seitwärts	26.93
1090	westwärts	25.17

Tabelle 1: CKIs aus der Keyword-Liste

LogLikelihood (LL) gilt wohl als das traditionelle Keywordmaß und findet in zahlreichen Studien Verwendung, obwohl es für die Errechnung von Keywords inzwischen nicht mehr uneingeschränkt empfohlen wird (für eine Kritik siehe u.a. Gabrielatos 2018: 234; Pojanapunya/Watson 2016: 162; Brezina/Meyerhoff 2014: 23). Wir können aus der obigen Berechnung vorerst lediglich schließen, dass es wohl einen Unterschied in der Häufigkeit der in Tabelle 1 gelisteten Direktionaladverbien mit *-wärts* zwischen den Dekaden 1890 und 1990 gibt, der statistisch signifikant erscheint ($p < 0.0001$). Das heißt im Untersuchungskorpus kommen diese Items also signifikant häufiger vor als im Referenzkorpus. Ein bekanntes Problem mit LL ist, dass auch kleine Unterschiede statistische Signifikanz erlangen können und so generell eher zu viele Key-items berechnet werden. Dies spielt hier jedoch eine untergeordnete Rolle, weil wir die CKIs nutzen, um unser Untersuchungsinteresse eingrenzen zu können.

Um nunmehr zu überprüfen, ob sich ebenfalls ein messbarer Unterschied von Direktionaladverbien mit *-wärts* über den Verlauf der gesamten Zeitspanne zeigt, wurde folgende Strategie verfolgt (vgl. auch Posch in Vorbereitung):

Keyitems 2: Die einzelnen Dekaden wurden jeweils im Vergleich zum gesamten restlichen Korpus mittels eines Effektstärkenmaßes verglichen. Wir wählten dafür ebenfalls die Standardeinstellungen von CQPweb:

- Statistik: LogRatio mit LogLikelihood Filter (angepasste LL Schwelle = 37.52, mit Šidák Korrektur für Mehrfachvergleiche),
- Signifikanz Cut-off: 0,01%,
- Minimalfrequenz in beiden Subkorpora: 3.

Die resultierenden CKIs kommen somit signifikant bis höchst signifikant häufiger in den jeweiligen Untersuchungskorpora (Dekaden) im Vergleich zum restlichen Korpus vor. Auch negative Keyitems wurden diesmal berechnet (i.e. signifikant seltenere Items).

In dieser Berechnung ist das Referenzkorpus jeweils erheblich größer als das Untersuchungskorpus (Brezina 2018: 81), aber aufgrund der thematischen Enge des Korpus sind Referenz-/Untersuchungskorpus inhaltlich sehr ähnlich. Mit dem Ergebnis dieser Berechnung lässt sich nunmehr ein chronologischer Vergleich der CKIs pro Dekade darstellen. Bei der Sichtung zeigt sich, dass mehrere Zusammensetzungen mit *-wärts* in mehreren Dekaden eine signifikante Keyness aufweisen. Die LR-Werte bewegen sich zwar auf einem eher niedrigen Niveau (von +2,83 bis -2,03), wohl weil die Bildungen insgesamt nicht sehr häufig sind, aber die gefundenen Zusammensetzungen mit *-wärts* sind bis zu ca. 8 Mal häufiger bzw. bis zu ca. 4 Mal seltener (bei *negative keyness*) im Untersuchungskorpus anzutreffen als im Referenzkorpus (siehe Abbildung 1).

Es zeigt sich die interessante Entwicklung, dass Direktionaladverbien mit *-wärts* in den anfänglichen Jahrzehnten als positive Keywords auffallen und im Laufe des 20. Jahrhunderts im Korpus Alpenwort deutlich an Bedeutung verlieren. Wir erweitern also im nächsten Schritt den Fokus generell auf das Suffix *-wärts* und möchten dessen Status und Entwicklung nun näher beleuchten.

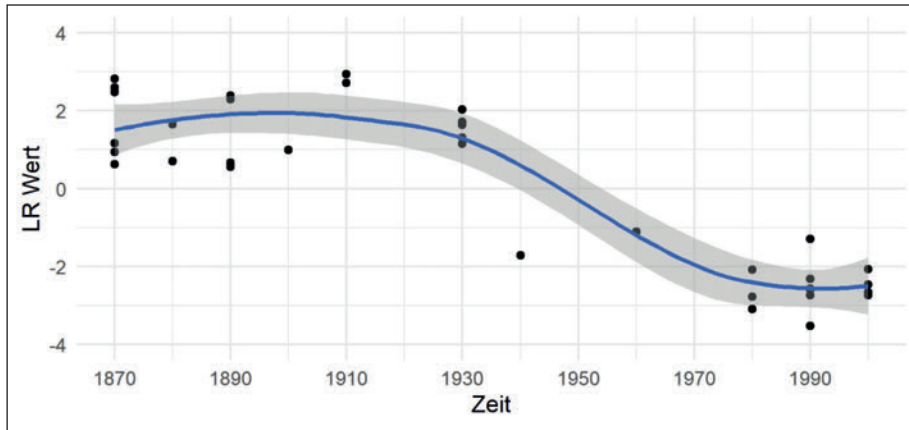


Abbildung 1: Entwicklung der Keyness von *-wärts* diachron (stat_smooth=loess)

3 Das Adverbialsuffix *-wärts*

Aus synchroner Sicht wird das Ableitungsmorphem *-wärts* als Mittel zur Bildung von Lokal- bzw. Direktionaladverbien der Richtung betrachtet, wobei eine übertragene Verwendungsweise ebenfalls möglich ist. Ein zentraler Punkt dieser Untersuchung wird sein, ob man *-wärts* wirklich noch als produktiv betrachten kann. Bevor wir näher auf diese Frage eingehen, wollen wir einen kurzen Überblick über den derzeitigen Forschungsstand zu Entwicklung und Status (Suffix vs. Suffixoid, Produktivität) geben.

Historisch betrachtet lässt sich *-wärts* auf ahd. *-wert* bzw. mhd. *-wert* (Kluge 2012: s.v. *-wärts*) zurückführen. Es ist verwandt mit dem heutigen Hilfszeitwort *werden*, wobei sich der semantische Kern schon im gemeinsamen indoeuropäischen Stamm **wert-* ‘wenden’ also ‘sich in (die betreffende) Richtung wendend’ (ibid.) finden lässt. Bereits im Grimmschen Wörterbuch wird konstatiert, dass *wärts* nur noch „in fester Verbindung mit adverbien und nomini-bus“ (Grimm, s.v. *wärts*, adv.) vorkomme, früher aber selbständig gewesen sei. Eine detaillierte Betrachtung des Suffixes bietet Graën in seiner Dissertation zu Raumadverbien (Graën 2004: 147–151). Er zeigt auf, wie das ursprünglich eigenständige Direktionaladverb *werd*, *ward* mit der Bedeutung ‘die Richtung

habend' *-wärts* im Zeitraum von 1100 bis 1350 zum produktiven Suffix zur Bildung von Richtungsadverbien wird. Er unterscheidet dabei drei Stufen:

1) Reihenbildung in Komposition mit anderen Adverbien des Kernbestandes (*abwerts, herwert, zuowert* etc.), 2) Entwicklung zum Suffix nach bestimmten Präpositionalen Wendungen (z.B. (19) *vrouwe Lachtasis daz ist mir bekant,/ di trecket mit irre hochvart / den vadem hin zu tale wart* BRUN 10449) und 3) in der Position nach vorangehenden *gegen-* oder *zuo-*Adverbialen (z.B. (21) *zv der burg wert* HERB 1375).

In 2 und 3 gibt es eine zweifache direktionale Kennzeichnung, wobei nach Wegfall der Präposition *wert* als einziges richtungsbestimmendes Merkmal bleibt und somit der Weg zum Suffix geebnet ist (wobei noch die Hinzunahme des adverb-typischen *-s* zu konstatieren ist). Graën erwähnt insgesamt 395 Belege, die Verwendung von *-wert* als Suffix setzt nach ihm in der zweiten Hälfte des 12. Jahrhunderts ein.

In der aktuellen Literatur zur deutschen Wortbildung wird *-wärts* als Suffix (u.a. bei Fleischer/Barz 2012: 369; Elsen 2014: 238; Lohde 2006: 293) bzw. Suffixoid (Kluge 2012: s.v. *-wärts*; Ganslmayer 2012: 378) bezeichnet. Henzen reiht *-wärts* in das Kapitel „Zur Ableitung übergetreten sind Adverbien mit einem Nomen als zweitem Glied, wo dieses wie ein Suffix analogisch weiterwirkt“ (Henzen 1965: 232) ein. Wir wollen an dieser Stelle nicht weiter auf die Affix/Affixoid Diskussion eingehen und verweisen auf dahingehende Untersuchungen (vgl. Nübling et al. 2017: 94–97; Elsen 2014: 145–148; Donalies 2002: 25f). Dass einige Autor*innen *-wärts* als Suffixoid bezeichnen, spiegelt aber wider, dass eine vollständige Grammatikalisierung, obwohl es heute kein eigenständiges *wärts, werts, warts* o.ä. mehr gibt, nicht stattgefunden hat. So trägt *-wärts*, ähnlich wie *-weise, -weg, -dings* oder *-maßen* (vgl. Simmler 1998), einerseits immer noch eine klar definierbare Bedeutung, andererseits kann die Reihenbildung durchaus kritisch gesehen werden (siehe folgende Kapitel). Explizit als produktiv wird das Suffix von Barz (2016: 773), Elsen (2014: 238), Dargiewicz (2012: 64), Lohde (2006: 294) und Graën (2004: 147) beschrieben, oft wird auf die Produktivität nicht näher eingegangen. Aus dem Kontext ist aber bei den meisten Autor*innen zu verstehen, dass das Suffix grundsätzlich als produktiv angesehen wird (z.B. Nübling 2016: 581; Weinrich

2005: 568; Engel 2004: 418; Schulz/Griesbach 1972: 212). Heinle behauptet für die Entwicklung vom Frühneuhochdeutschen bis ins 20. Jahrhundert eine relative Konstanz der Beleglage von *-wärts* mit stärkster Produktivität in Verbindung mit Lokaladverbien z.B. *bergab-wwärts* (Heinle 1985: 1915). Motsch bezeichnet dieses Wortbildungsmuster als „schwach aktiv“ (2004: 235f.).

Zur derzeitigen Verwendungsweise des Suffixes gibt es kaum eingehendere Untersuchungen, insbesondere nicht solche, die sich auf reale Sprachdaten berufen. Es stellt sich nun die Frage, wie produktiv das Suffix *-wärts* eigentlich (noch) ist? Wir versuchen diese Frage zu beantworten, indem wir die Produktivitätsmaße TTR und HTR in den Korpora Alpenwort, Text&Berg, DWDS, COSMASIIweb, Schweizer Textkorpus vergleichen. Gleichzeitig arbeiten wir ganz konkret praktische Probleme und Unterschiede in der Handhabung dieser sehr verschiedenen Korpora heraus, um zu zeigen, wo Schwierigkeiten bei der Vergleichbarkeit liegen können.

3.1 Produktivitätsanalyse mit TTR und HTR

Die Frage, wie morphologische Produktivität mit quantitativen Methoden zuverlässig gemessen bzw. analysiert werden kann, ist durchaus umstritten (vgl. Aftabi et al., 2021: 1–3; Schneider-Wiejowski, 2012: 43–60). Eine Möglichkeit, Produktivität zu erfassen, ist die Analyse der Type-Frequenz. Baayen (2009: 901) bezeichnet das Verhältnis von Types zu Gesamttokens als *realized productivity* und sagt, dass sich die Produktivität auf „past achievements“ beziehe. Stefanowitsch indessen definiert die Type-Token-Ratio (TTR) als Verhältnis von gefundenen Types zu gefundenen Tokens und beschreibt dieses folgendermaßen:

Likewise, observing the type frequency (i.e. the TTR) of an affix under different conditions provides information about the relationship between these conditions and the affix itself, albeit one that is mediated by the lexicon: it tells us how important the suffix in question is for the subparts of the lexicon that are relevant under those conditions (Stefanowitsch 2020: 316).

Eine weitere, von Baayen (1992: 115) definierte Größe hat sich in der korpuslinguistisch orientierten Produktivitätsanalyse als gängige Vergleichs- und Untersuchungsgröße etabliert. Es handelt sich um die *expanding productivity* (vgl. auch Baayen 2009: 902) bzw. Hapax-Token-Ratio (HTR) (Stefanowitsch 2020: 318). Dieser Wert bezeichnet das Verhältnis von Hapax Legomena zu Gesamttokenanzahl und gibt damit Aufschluss über die aktuelle Produktivität eines Suffixes.

Mit diesen Größen lassen sich sowohl die Produktivität in einem Korpus zu einer bestimmten Zeit feststellen, als auch, bei Vorliegen einer zeitlich gestaffelten Textsammlung, diachrone Produktivitätsentwicklungen nachvollziehen. Dies wurde z.B. von Schneider-Wiejowski für verschiedene Suffixe im Schweizerdeutschen gemacht (Schneider-Wiejowski 2013).

4 Der *-wärts*-Komplex in verschiedenen Korpora

Für Produktivitätsanalysen besteht Konsens, dass die Angabe der relativen Tokenfrequenz allein keine getreue Darstellung der Produktivität darstellt (vgl. Stefanowitsch 2020: 315). Dennoch ist die Betrachtung dieser Größe interessant, da sie eine Aussage über die generelle Häufigkeit der Verwendung eines Suffixes – sei es nun produktiv oder nicht – zulässt. Vergleicht man die relativen Häufigkeiten der verschiedenen Korpora, fällt auf, dass die alpinbezogenen Texte eine deutlich höhere maximale Tokenfrequenz (ca. 800 fpmw) gegenüber den anderen Korpora aufweisen (ca. 200 – 250). Es bestätigt sich hier also die eingangs getätigte Annahme, dass in Texten mit Alpinbezug direktionale Adverbien und damit auch solche mit Suffix *-wärts* häufiger verwendet werden. Außerdem zeigt sich in allen Korpora, in welchen eine zeitliche Staffelung vorliegt, dass nach einem Höchststand Anfang/Mitte des 20. Jahrhunderts die absolute und relative Tokenfrequenz bis zum Ende des 20. Jahrhunderts drastisch abnimmt. Die Größe dieser Abnahme gibt dann schon einen ersten Hinweis auf die Entwicklung der Produktivität.

4.1 Alpenwort und Text&Berg

Um dem aufgespürten Phänomen *-wärts* nachzugehen bietet sich ein näherer Blick auf das schon eingeführte Korpus Alpenwort (siehe S. 119) an sowie ein Vergleich mit dem sehr ähnlichen „Schwesterkorpus“ Text&Berg (Text+Berg-Korpus R151v01). Die Ähnlichkeit besteht sowohl inhaltlich als auch in der Zeitspanne der beiden Korpora: Text&Berg beinhaltet alle Jahrbücher des Schweizer Alpenclubs bzw. die Monatszeitschrift *Alpen* 1864–2015. (Göhring/Volk 2011). Verwendet wurden daraus alle deutschsprachigen Texte, was einer Größe von ca. 23 Mio. laufender Wortformen entspricht (Bubenhof/Rothenhäusler 2018: 41). Wie auch Alpenwort ist Text&Berg vollständig korrigiert und POS-getagged vorhanden, beide Korpora sind über CQPweb abfragbar und lassen sich deshalb ausgezeichnet vergleichen.

Wie aus Abbildung 2 ersichtlich ist, steigen TTR und HTR sowohl im Alpenwort als auch im Text&Berg Korpus im Lauf der Zeit deutlich an, was auf eine Zunahme der Produktivität hinzuweisen scheint. Zusätzlich gibt es im Alpenwort Korpus zwei ausgeprägte Spitzen, die erklärungs-würdig sind. Wie Stefanowitsch (2021: 320) zeigt, ist weder der Zusammenhang zwischen Korpusgröße und Types bzw. Hapax, noch der zwischen Treffermenge und Types bzw. Hapax linear. In beiden Fällen gilt die Regel, dass die Anzahl der Types bzw. Hapax bei geringer Tokenmenge bzw. Treffermenge zuerst stark ansteigt, bevor es zu einer gewissen Plateaubildung kommt. Das heißt, dass für den erwähnten Anstieg von TTR und HTR primär die Abnahme der Treffermenge ausschlaggebend ist, und nicht eine wirkliche Produktivitätssteigerung, bei der mit der Zunahme der TTR beziehungsweise HTR auch eine Zunahme der Treffermenge zu erwarten wäre. Die Ausreißer im Alpenwort Korpus haben zwei unterschiedliche Ursachen. Die Spitze im Jahrzehnt 1940 kann sehr gut mit der unterschiedlichen Teilkorpusgröße erklärt werden. Während in allen anderen Jahrzehnten die Tokenzahl zwischen knapp einer Million und gut zwei Millionen Tokens schwankt, ist aufgrund der durch den Zweiten Weltkrieg verursachten Publikationslücke das Teilkorpus nur ca. 300.000 Tokens groß. Dies führt aufgrund des oben beschriebenen nicht linearen Zusammenhangs zur Spitze. Wie auch Stefanowitsch (2020: 321) analysiert: „Therefore, comparing TTRs derived from samples of different sizes will always make the

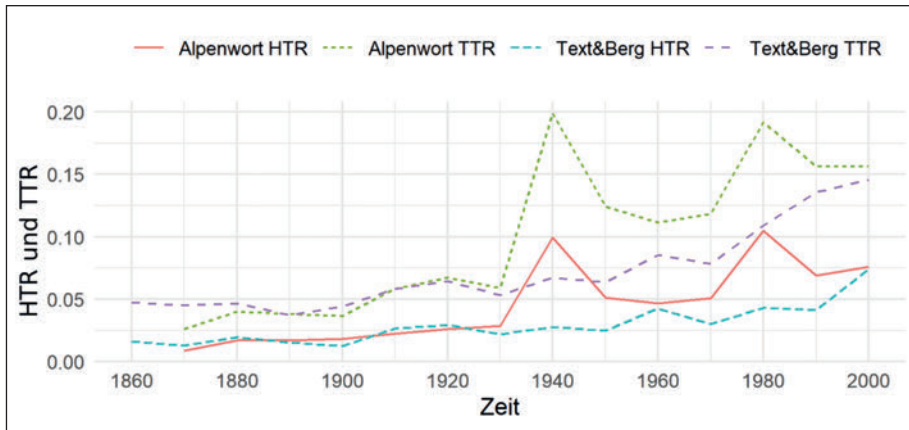
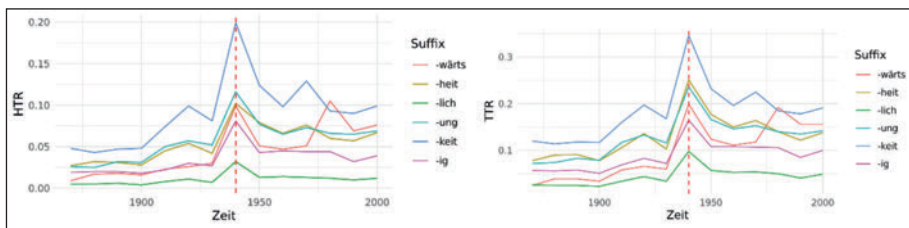


Abbildung 2: Entwicklung von TTR und HTR im Untersuchungszeitraum (Intervall 10 Jahre)

smaller sample look more productive.” Dass es sich um einen Zusammenhang mit der Korpusgröße handelt, erkennt man, wenn man die Produktivität anderer Suffixe analysiert. Hier ergibt sich ebenfalls für das Jahrzehnt 1940 ein auffälliger Produktivitätsgipfel (vgl. Abbildungen 3 und 4).

Das Teilkorpus 1980 scheint in der Tat ein Ausreißer zu sein, wobei vor allem das Jahr 1988 auffällt, das 38% der Hapax liefert (allerdings von verschiedenen Autor*innen). Insofern haben wir hier die Situation, dass bei einem Phänomen geringer Frequenz bereits kleine Unterschiede relativ große Auswirkungen haben.



Abbildungen 3 und 4: Entwicklung von TTR und HTR für verschiedene Suffixe

4.2 DWDS

Das Digitale Wörterbuch der Deutschen Sprache (DWDS, siehe Barbaresi/Geyken 2020) bietet eine ausgezeichnete Grundlage für korpuslinguistische Fragestellungen: einerseits lassen sich über die online-Werkzeuge, die auf <https://www.dwds.de/> zur Verfügung gestellt werden, bereits aussagekräftige Abfragen und Visualisierungen tätigen, andererseits bietet die Seite benutzungsfreundliche Exportschnittstellen. Über diese Exportmöglichkeiten lassen sich Abfrageergebnisse lokal weiterverarbeiten, man ist also nicht durch die Vorgaben von DWDS-online beschränkt. Abgerundet werden diese Möglichkeiten durch eine ausgezeichnete Dokumentation mit zahlreichen Abfragebeispielen. Die Referenzkorpora, die wir für unsere Analyse verwendet haben, decken einen Zeitbereich von ca. 1600–2010 ab.

Um die Daten vergleichbar zu verarbeiten und darzustellen, wurden die im Folgenden dargestellten Verarbeitungsschritte durchgeführt. Die Abfrage über das Online-Interface ist einfach und wird über eine umfassende Suchsyntax gut unterstützt. Es gibt allerdings einige Eigenheiten, die insbesondere bei Phänomenen mit geringer Okkurrenz eine Rolle spielen können und deshalb hier thematisiert werden. Erstens: Wirkt sich die fehlende verbindliche Orthographie in den ersten beiden Jahrhunderten auf das Ergebnis aus und wenn ja wie? Kann urheberrechtlich geschütztes Material ebenfalls eine Auswirkung haben? Wie auf S. 130 dargelegt, werden in der KWIC-Ansicht urheberrechtlich geschützte Inhalte nämlich nicht dargestellt, was vor allem für die jüngeren Texte deutliche Auswirkungen zeigt.

1) Abfrage und Analysevorbereitung: Bei allen drei zur Verfügung stehenden Referenzkorpora (dtak, kern, korpus21) wurden Lemmata, die auf *-wärts* enden, abgefragt (in der Abfragesyntax `$l=*wärts`). Die Abfrage über das Lemma ist in diesem Fall wichtig, weil ältere Sprachstufen eine oft stark abweichende Orthographie aufweisen (*-wertz*, *-waerts*, *-werths* etc.). Damit erhält man insgesamt 30.448 exportierbare Funde (im Gegensatz zu lediglich 26.791 Funden mit der Abfrage über die Wortform). Der Export im .csv-Format besteht aus Genre, Titel, Kontext und vor allem Jahr, so dass sich die Anzahl der Funde als Zeitfunktion abbilden lässt. Die Funde wurden dabei ebenfalls in Jahrzehnte zusammengefasst (jeweils Jahre 0–9). Die unvollstän-

digen Jahrzehnte 1590–1599 und 2010–2019 wurden entfernt, um statistische Ausreißer wegen extrem kleiner Korpusgrößen zu verhindern. Abbildung 5 zeigt die absoluten Frequenzen des Lemmas *-wärts* und es wird sichtbar, dass bis ca. 1780 ein substantieller, bis ins 19. Jahrhundert ein geringer Unterschied in der Treffermenge zwischen Lemma-Suche (blau liniert) und Wortform-Suche (rot punktiert) besteht. Die Abfrage per Lemma ist also für die DWDS-Korpora auf jeden Fall zu bevorzugen.

Nicht mit im Export sind Textstellen, die aus urheberrechtlichen Gründen nicht freigegeben wurden. So wird man zum Beispiel bei der Abfrage im kern-Korpus darauf hingewiesen, dass nur 281 von 571 Treffern, also nicht einmal die Hälfte, dargestellt werden können. Auf der Suche nach Abhilfe für diese unbefriedigende Situation wurde folgende Lösung gefunden: Die urheberrechtlich beschränkte Darstellung gilt nur für Ansichten, in denen Kontext abgerufen wird. Abfragen, die auf Wortebene basieren, werden korrekt dargestellt. Somit lässt sich mit der Abfrage `COUNT ($l=*wärts) #BY[$w, $l, date/1]`¹ eine Liste erstellen, in der sowohl die Wortform als auch das dazugehörige Lemma für jeweils ein Jahr als Summe vorhanden ist (siehe Tabelle 2).

Summe	Wortform	Lemma	Jahr
1	Abwärts	abwärts	1666
4	Abwärts	abwärts	1688
1	Anderwärts	anderwärts	1682
1	Auffwärts	aufwärts	1639
2	Auffwärts	aufwärts	1659
4	Auffwärts	aufwärts	1688
1	Auffwärts	aufwärts	1657
1	Aufwärts	aufwärts	1681
2	Aufwärts	aufwärts	1682

Tabelle 2: Wortform und Lemma pro Jahr

Mit dieser Abfrage erweitert sich das Ergebnis auf 32.443 Funde. Dies ist insofern bedeutend, als diese nicht gleichmäßig verteilt, sondern primär im 20.

¹ vgl. <https://www.dwds.de/d/korpussuche#count>

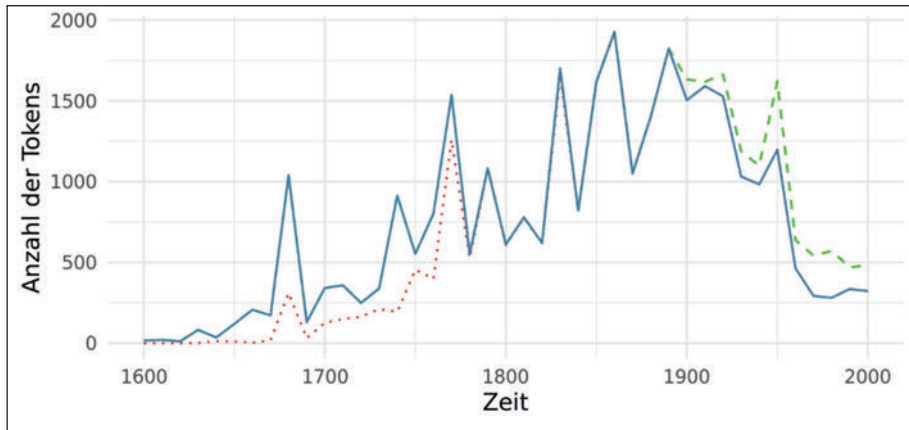


Abbildung 5: Absolute Anzahl der Tokens mit Suffix *-wärts*

Jahrhundert angesiedelt sind. In Abbildung 5 sind die zusätzlichen Funde als grün strichlierte Linie eingezeichnet. Für die weitere Verarbeitung wurden die Daten zu Jahrzehnten (wieder Jahre 0–9) akkumuliert.

2) Für weitere Berechnungen lässt sich die Gesamtkorpusgröße in der DWDS-Oberfläche einfach abfragen: Mit `COUNT(*) #BY[date/1]` erhält man eine Liste mit der absoluten Tokenanzahl pro Jahr. Diese Liste wurde wieder in Jahrzehnte zusammengefasst. Abbildung 6 zeigt die Korpusgröße als Zeitfunktion und zeigt ebenfalls, dass die Größen der jeweiligen Subkorpora sich deutlich unterscheiden (sie variieren von knapp über 100.000 bis über 15 Millionen Tokens), wobei die Korpusgröße im 20. Jahrhundert jeweils über 10 Millionen Tokens liegt und gegen 1600 hin abnimmt.

3) Zur eigentlichen Produktivitätsberechnung muss jetzt noch die Anzahl der Types sowie der Hapax Legomena berechnet werden. Aufgrund der bereits angesprochenen orthographischen Probleme wurden für diese Berechnung ebenfalls die Lemmata verwendet. Dies hat außerdem den Vorteil, dass durch die letzte Orthographiereform verursachte Probleme (*flussabwärts* vs. *flußabwärts*) ausgeglichen werden, da bekannte Lemmata in der neuen Rechtschreibung zusammengefasst werden (beides *flussabwärts*). Abbildung 7 zeigt die Anzahl der Hapax berechnet anhand der Wortformen (blau strichliert) und anhand der Lemmata (rot liniert).

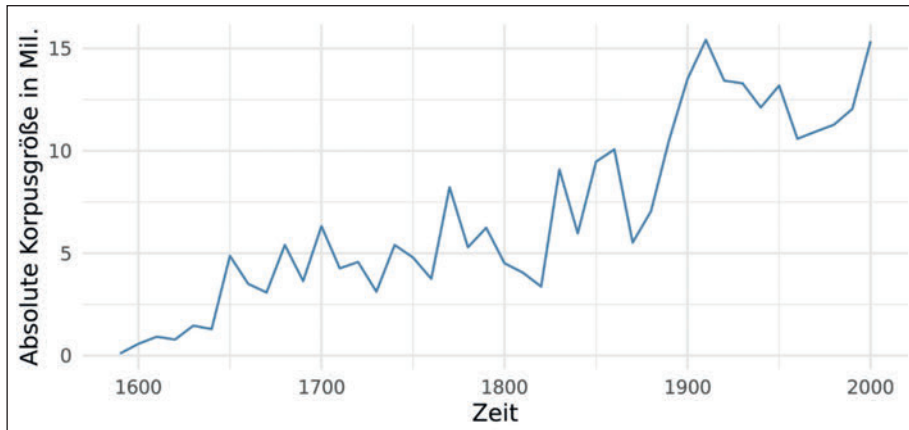


Abbildung 6: Absolute Korpusgröße im Untersuchungszeitraum (zusammengefasst in Dekaden)

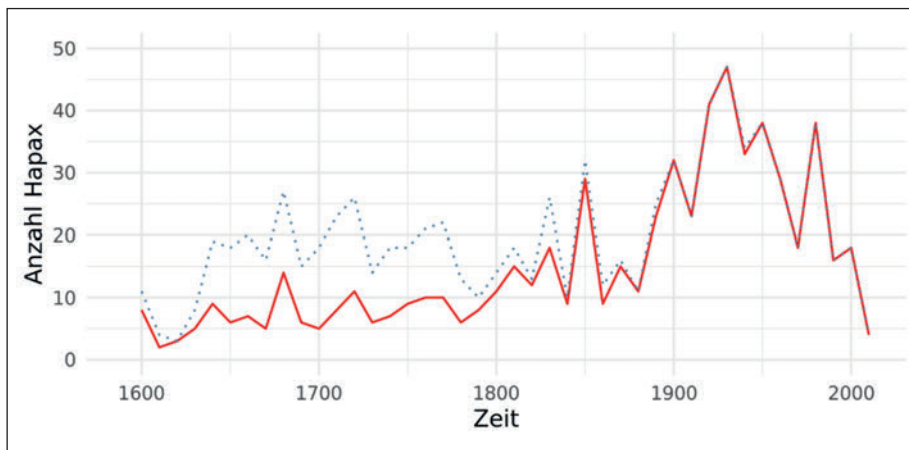


Abbildung 7: Hapax Legomena im Untersuchungszeitraum anhand Wortform bzw. Lemma

Auch für die Berechnung der Types hat die Verwendung von Lemmata große Auswirkungen im Zeitraum vor 1900. Abbildung 8 zeigt die Anzahl der Types berechnet anhand der Wortformen (grün strichliert) und anhand der Lemmata (blau liniert).

Aufbauend auf diesen Daten können nun die relative Tokenfrequenz, TTR und HTR berechnet werden. Betrachtet man die relative Tokenfrequenz

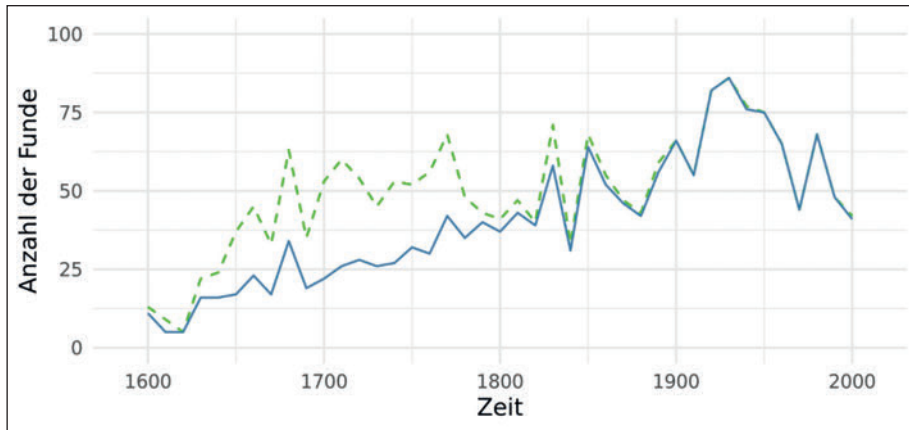


Abbildung 8: Types im Untersuchungszeitraum anhand Wortform bzw. Lemma

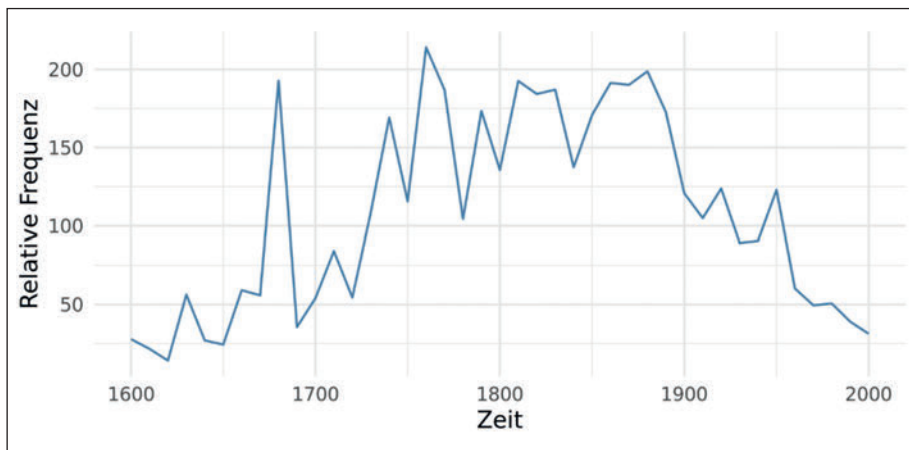


Abbildung 9: Relative Frequenz im Untersuchungszeitraum

(Abbildung 9), so lässt sich ein Ansteigen der Verwendung von Adverbien auf *-wärts* vom 17. bis zum 18. Jahrhundert erkennen. Die Spitzen zeigen an, wie stark sich einzelne Autor*innen auswirken können. So stammen z.B. von den 1041 Funden im Jahrzehnt 1680 902 Funde von nur drei Autoren (416 Johann Christoph Pinter von der Au, 348 Wolf Helmhard von Hohberg, 138 Daniel Casper von Lohenstein). Die Präferenz einzelner Personen für

diese Adverbien gepaart mit der relativ geringen Korpusgröße führt bei einem Randphänomen, wie es die Ableitung mit *-wärts* darstellt, bereits zu größeren Verzerrungen. Auch die unterschiedliche Zusammenstellung des Korpus (vgl. <https://www.deutschestextarchiv.de/doku/textauswahl>) könnte einen Einfluss auf die Ausreißer haben, wenngleich sich dies nicht so direkt feststellen lässt wie bei den Autor*innen. Ab dem Ende des 19. Jahrhunderts beobachten wir ein relativ konstantes Nachlassen und da wir hier ein deutlich größeres Korpus besitzen, können wir auch mit größerer Sicherheit vom allgemeinen Nachlassen der Verwendung dieser Adverbien sprechen.

Betrachten wir TTR und HTR (Abbildungen 10 und 11) als Maße für die Produktivität, so ergibt sich bei beiden ein ähnliches Bild: In der ersten Phase, bis in die Mitte des 17. Jahrhunderts ist das Korpus so klein, dass keine sinnvollen statistischen Aussagen getroffen werden können. Erst bei einer Korpusgröße von ca. 3 Millionen Tokens pendeln sich beide Werte auf einem stabilen, sehr niedrigen Niveau ein.

Am interessantesten ist hierbei die Beobachtung, dass auch in der Zeitspanne, in der die Verwendung von Adverbien auf *-wärts* absolut steigt, keine Zunahme von TTR oder HTR zu verzeichnen ist. Mit anderen Worten, die Produktivität steigt nicht an.

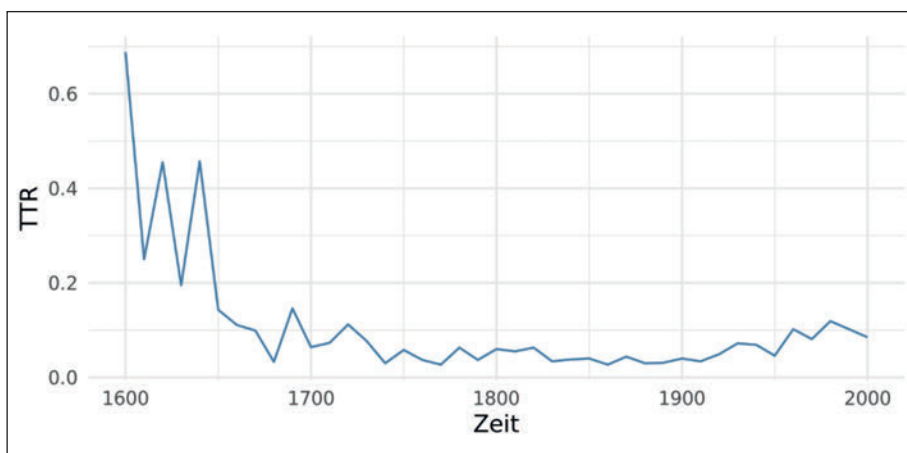


Abbildung 10: TTR im Untersuchungszeitraum

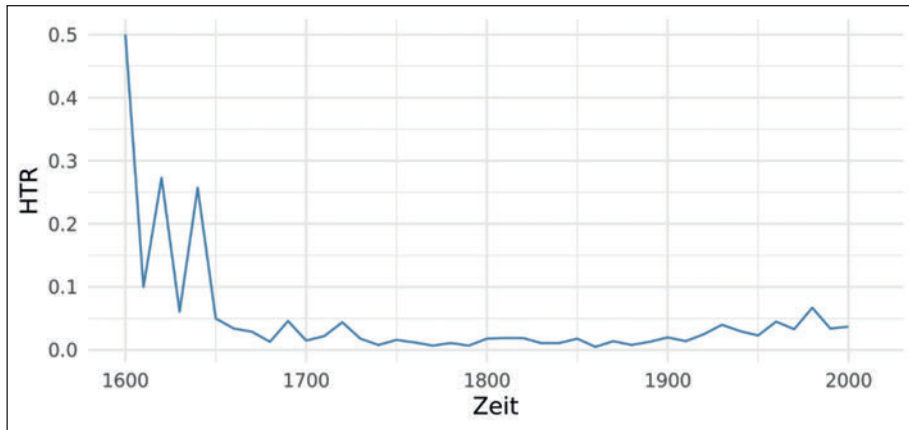


Abbildung 11: HTR im Untersuchungszeitraum

Zusammenfassend lässt sich sagen, dass das DWDS für eine diachrone Untersuchung eine ausgezeichnete Grundlage bildet. Beachtet man die Einschränkungen, die sich aus Orthographie und Urheberrecht ergeben, so kann man dennoch weitgehend automatisch statistische Aussagen über den dargebotenen Zeitraum erstellen. Für seltene Phänomene ist die Größe bzw. Kleinheit des Korpus insbesondere in der ersten Hälfte des 16. Jahrhunderts zu beachten. Weiters kann die unterschiedliche Zusammensetzung des Korpus problematisch für manche Fragestellungen sein.

4.3 DeReKo

Im Gegensatz zu den beiden thematisch fokussierten hochspezialisierten Korpora Alpenwort und Text&Berg bietet die COSMASIIweb-Schnittstelle des Instituts für Deutsche Sprache in Mannheim Zugang zu 573 unterschiedlichen Korpora. Aktuell besteht dort das verfügbare Korpus DeReKo (Deutsches Referenzkorpus, siehe Kupietz et al. 2018) aus 50,6 Mrd. (Stand 2.2.2021) laufenden Wortformen (lt. Webseite). Die enthaltenen Korpora werden laufend erweitert und gelten somit als Referenzkorpus des gegenwärtigen Deutschen. Das DeReKo ist ausdrücklich nicht als ein ausgewogenes Korpus konzipiert (Kupietz et al. 2009: 1849), was die Vergleichbarkeit der Größen TTR und

HTR mit dem DWDS oder dem Schweizer Textkorpus praktisch unmöglich macht.² Das DeReKo hat keinen Anspruch auf große zeitliche Tiefe und es kann hauptsächlich synchron gearbeitet werden. Im gesamten Korpus sind sehr unterschiedliche Teilkorpora mit ebenfalls unterschiedlichen Zeitbreiten vorhanden, so dass diese nicht methodisch sauber diachron betrachtet werden können. Die interessanteste Untersuchungsmöglichkeit im Bezug auf *-wärts-*Suffigierung wäre die lexikalische Varianz der entstehenden Wörter. Diese Untersuchung würde den hier gesetzten Schwerpunkt einer quantitativen Untersuchung sprengen. Eine diesbezügliche Studie, wieder unter Einbezug der hier verwendeten Korpora, ist in Vorbereitung.

TTR und HTR lassen sich im DeReKo zwar errechnen, die Werte sind allerdings nicht mit den anderen Ergebnissen vergleichbar, da, wie vorher bereits erwähnt, die Korpusgröße und die Trefferanzahl nicht linear sind. Zu beachten ist, dass bei den Hapax eine große Menge falscher Einträge zu finden sind, die aus OCR-Fehlern etc. stammen. Im hier vorliegenden Fall betrug die Anzahl der falschen Hapax immerhin 18.7%. Die errechnete TTR von 0.0043 und die HTR von 0.0022 kann jedoch für sich genommen wenig Auskunft über die Produktivität des Suffixes geben. Deshalb ist die Verwendung des DeReKo für Analysen dieser Art nur begrenzt tauglich.

4.4 Schweizer Textkorpus

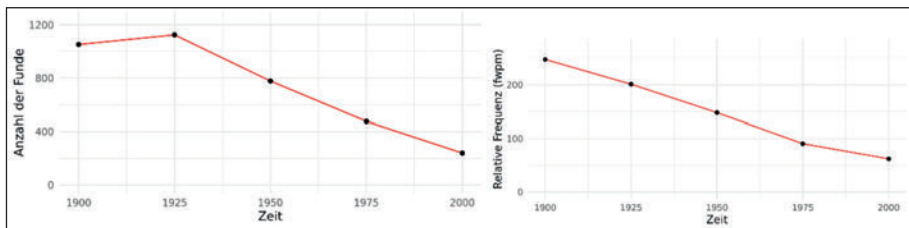
Das Schweizer Textkorpus (Bickel et al. 2009) orientiert sich in seinem Aufbau am BNC. Es bietet ein Korpus, das Gebrauchstexte, Sachtexte, Belletristik und journalistische Prosa im Bezug auf Korpusgröße und zeitlicher Distribution ausgewogen zur Verfügung stellt. Das Korpus deckt derzeit einen Zeitbereich von 1900–2018 ab und ist, bezogen auf das 21. Jahrhundert, weiterhin im Aufbau begriffen. Der Vorteil für den hier untersuchten Zeitbereich 1900–2000 ist also, dass das Korpus sowohl im Hinblick auf Textsorten als auch auf Größe sehr homogen ist. Damit können, zumindest auf das Deutsche in der Schweiz bezogen, vergleichbare Aussagen zur Produktivität ge-

2 Dies könnte sich mit der neuen Abfrageoberfläche KORAP ändern, für diese Studie wurde diese Möglichkeit jedoch nicht näher untersucht.

macht werden. Das Korpus ist mit Hilfe der DDC Query Language bequem zu benutzen, ein kleiner Nachteil ist jedoch, dass die Exportmöglichkeiten beschränkt sind (max. 500 Treffer können exportiert werden). Damit wird die Datensammlung zur Errechnung von nicht direkt im Interface abfragbaren Größen wie TTR und HTR etwas umständlicher. Die absolute Korpusgröße wird in der Dokumentation für einen Zeitraum von jeweils 25 Jahren für die einzelnen Textsorten sowie für das Gesamtkorpus angegeben. Damit lässt sich die relative Frequenz für diese Zeitbereiche prinzipiell einfach errechnen.

Die für diese Untersuchung verwendete Abfrage zum Erhalt aller Wörter auf *-wärts* ist `*wärts #SEPARATE_HITS`. Der Zusatz `#SEPARATE_HITS` ist nötig, weil ohne diesen mit der DDC Query Language nur jeweils das erste Wort eines Satzes, auf das die Suchbedingung zutrifft, gefunden bzw. angezeigt wird. Der Unterschied zwischen der Suche mit `*wärts` gegenüber `*wärts #SEPARATE_HITS` beträgt immerhin 3.382 vs. 3.622 Treffer.³

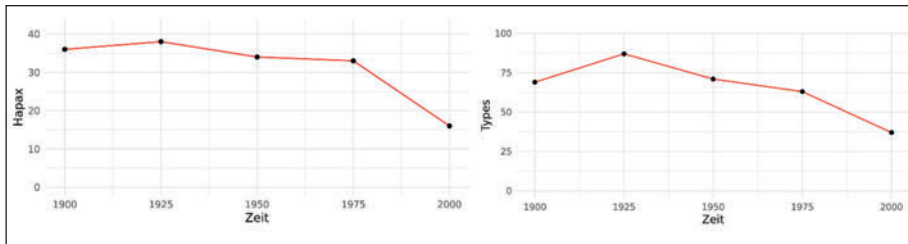
Wie aus den Abbildungen 12 und 13 zu sehen ist, fällt sowohl die absolute Zahl der Treffer als auch die relative Frequenz, ähnlich wie im Alpenwort Korpus, nach einem Hoch Anfang des 20. Jahrhunderts stetig ab. Die relative Frequenz ist mit anfänglich 200 fpmw im Bereich des DWDS.



Abbildungen 12 und 13: Anzahl der Funde und relative Frequenz im Untersuchungszeitraum

Sowohl die Anzahl der Types als auch die Anzahl der Hapax nimmt in der zweiten Hälfte des Untersuchungszeitraums ab (Abbildungen 14 und 15). Die Produktivitätsindizes TTR und HTR (Abbildung 16) nehmen jedoch im selben Zeitraum zu. Da die Teilkorpusgröße im Schweizer Textkorpus annähernd gleich bleibt, scheint dies also auf eine Produktivitätszunahme hinzu-

³ Die Untersuchung von Schneider-Wiejowski (2013) wäre dahingehend anzupassen.



Abbildungen 14 und 15: Anzahl der Hapax und Types im Untersuchungszeitraum

weisen. Dies ist, wie oben bereits erwähnt, aber insofern trügerisch, als der Zusammenhang zwischen Types bzw. Hapax und gefundenen Tokens nicht linear ist. Wie bei Stefanowitsch (2020: 320, Figure 9.1) zu sehen ist, steigt die Anzahl der Types/Hapax im Verhältnis zur Tokenmenge zuerst steil an, bevor es zu einer Plateaubildung kommt. Das heißt, dass auch bei annähernd gleicher Korpusgröße TTR und HTR nicht linear zur Treffermenge verlaufen. Je kleiner die Treffermenge, desto größer sind TTR und HTR. Der Anstieg in der zweiten Hälfte des 20. Jahrhunderts ist also stärker auf die Abnahme der Verwendung insgesamt zurückzuführen, als dass es eine wirkliche Steigerung der Produktivität gäbe. Diese Entwicklung ist also ähnlich wie in oben beschriebenen thematisch fokussierten Korpora.

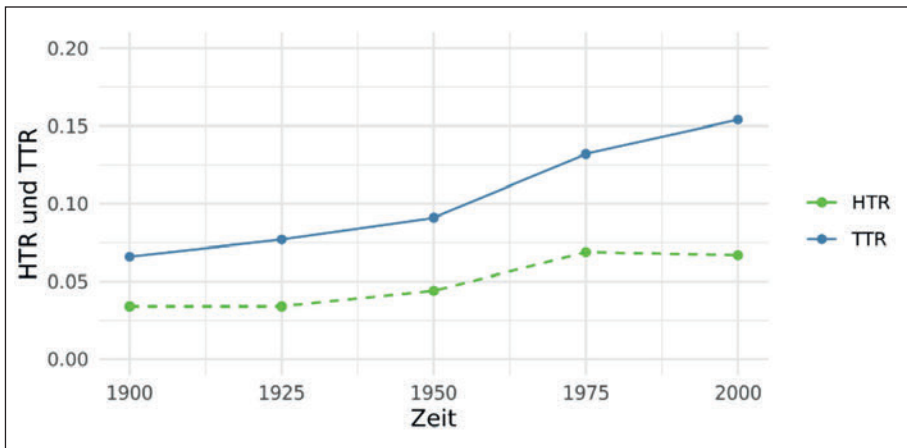


Abbildung 16: HTR und TTR im Untersuchungszeitraum

5 Zusammenfassung

Einige bei Keyword-Vergleichen im Korpus Alpenwort auffällige Bildungen mit dem Suffix *-wärts* weckten unser Interesse dieses spannende Derivationsmorphem näher zu beleuchten. Mit unserem Beitrag gehen wir einerseits mittels TTR- und HTR Berechnungen der Frage nach, ob das Suffix *-wärts* heute noch als produktiv gewertet werden kann. Andererseits wollen wir aber auch interessierten Forscher*innen darlegen, wie und ob sich für diese Fragestellung stark unterscheidende Arten von Korpora vergleichen lassen bzw. welche Fallstricke hierbei auftreten können.

Die eingangs erwähnte Hypothese, dass Direktionaladverbien in raumbezogenen Texten, wie sie in den Alpenwort und im Text&Berg Korpora zu finden sind, häufiger gebraucht werden als in anderen Texten, konnte mittels Frequenzanalyse belegt werden. In Anlehnung an die anfangs zitierten Worte von Bubenhofer/Rothenhäusler (2018) sind diese wohl Teil einer bestimmten, nennen wir es Bergtext-Anatomie. Für die Untersuchung von Direktionaladverbien bietet dies den Vorteil, in diesen Korpora eine größere Grundmenge des Untersuchungsgegenstands zu haben. Allgemeiner ausgedrückt: Thematische Korpora bieten Vorteile, weil das Auftreten eines bestimmten linguistischen Phänomens in ihnen gezielter untersucht werden kann als in ausgewogenen oder breit angelegten Referenzkorpora. Dies gilt insbesondere dann, wenn es sich um eher seltene Phänomene handelt.

Die Frage nach der Produktivität des Suffixes *-wärts* selbst muss facettiert beantwortet werden. So scheinen die Indizes TTR und HTR in allen diachron auswertbaren Korpora (Alpenwort, Text&Berg, Schweizer Textkorpus, DWDS) auf einen Anstieg der Produktivität in der zweiten Hälfte des 20. Jahrhunderts hinzuweisen. Gegen diese Interpretation spricht die Tatsache, dass sowohl die absolute als auch relative Tokenfrequenz im selben Zeitraum drastisch abnehmen und die vermeintliche Zunahme praktisch ausschließlich auf das nichtlineare Verhältnis von HTR und TTR zu gefundenen Tokens zurückzuführen ist. Eine geringe Anzahl von gefundenen Tokens führt dazu, dass geringfügige Änderungen, z.B. die auf einen Autor zurückzuführende Präferenz dieser Direktionaladverbien, zu unverhältnismäßigen Produktivitätsspitzen führen. Dies ist insbesondere in den frühen Jahrzehnten im DWDS

und anhand der Spitze im Jahrzehnt 1980 im Alpenwort Korpus gut zu beobachten. Auch die absolute Korpusgröße kann eine Rolle spielen, wie die Spitze in der Dekade 1940 im Alpenwort Korpus zeigt.

Die ernüchternde Feststellung ist, dass sich weder HTR noch TTR als Messgröße für ein seltenes Phänomen wie das hier untersuchte zu eignen scheinen. Zwei Lösungen scheinen möglich zu sein: Die erste Lösung besteht in der Vergrößerung des Korpus, so dass man sich mit HTR und TTR in einem Bereich befindet, in dem eine Abnahme der Tokenfrequenz weniger starke Auswirkungen auf die Maßzahl selbst hat (dies entspricht einer Verschiebung auf der Kurve TTR/Token bzw. HTR/Token auf einen Bereich, in dem die Kurve abflacht). Die zweite Lösung wäre, den TTR und HTR Wert abhängig von der absoluten bzw. relativen Tokenfrequenz zu korrigieren. Eine dahingehende Diskussion würde allerdings den Rahmen dieser Untersuchung sprengen. Unabhängig von diesen Problemen kann aber festgestellt werden, dass der Rückgang der Gebrauchsfrequenz des Suffixes *-wärts* im 20. Jh. über verschiedene Korpora hinweg zu beobachten ist. Dieser Rückgang scheint textsortenunabhängig zu sein. Daher wäre die Einschätzung zur Produktivität folgendermaßen zu korrigieren: De facto werden Wörter mit dem Suffix wesentlich weniger gebraucht als vor hundert Jahren und das Suffix ist derzeit nicht mehr produktiv. Eine weitere interessante Frage, die an dieser Stelle aber nicht beantwortet werden kann, ist, ob in den jüngeren Texten alternative Formulierungen anstelle der direktionalen Adverbien mit *-wärts* verwendet werden und welche diese sind.

Die vielleicht wichtigste methodische Erfahrung zur Analyse der Suffixproduktivität ist, dass die absolute Korpusgröße sowie die absolute Tokenhäufigkeit nicht zu unterschätzende Auswirkungen auf die Größen TTR und HTR haben und ein Vergleich verschiedener Korpora bzw. verschiedener Teilkorpora deshalb nur mit größter Vorsicht gemacht werden sollte. So ist beim DWDS, das eine hervorragende Quelle für diachrone Untersuchungen darstellt, aufgrund der großen Unterschiede der Größen der Teilkorpora ohne Angabe dieser schwer abzuschätzen, wie stark die jeweiligen Produktivitätsmaße von dieser Größe abhängig sind. Es ist deshalb angebracht, zur Produktivität auch die Teilkorpusgrößen anzugeben, um eine Einschätzung der Ergebnisse vornehmen zu können. Ausgewogene Korpora wie das Schweizer

Textkorpus bieten hier den eindeutigen Vorteil, dass durch die Ähnlichkeit der Teilkorpusgrößen eine dahingehende Beeinflussung ausgeschlossen werden kann. Es besteht, trotz der Verwendung eingeführter statistischer Methoden, eine gewisse Schwierigkeit verschiedene Korpora zu vergleichen.

Ein weiteres, nicht zu unterschätzendes Problem ist, dass die Korpusabfragewerkzeuge über unterschiedliche Suchsyntax verfügen. Hier obliegt es derzeit noch den einzelnen Wissenschaftler*innen, sich gewissenhaft in die einzelnen Abfragesprachen einzuarbeiten. Eine Lösung des Problems ist die Bündelung verschiedener Korpora unter einer gemeinsamen Abfrageoberfläche (z.B. ist Text&Berg auch im DWDS verfügbar). Damit ließe sich zumindest teilweise ein weiteres Problem, nämlich die Verwendung unterschiedlicher Begrifflichkeiten, lösen. Beispielsweise wird in der Dokumentation die absolute Größe des Schweizer Textkorpus in Textwörtern angegeben, welche Zeichensetzung ausblendet. Die meisten anderen Korpora haben jedoch eine Tokendefinition, die Interpunktion mit einbezieht. Dies dürfte Benutzer*innen nicht so ohne weiteres klar sein und kann dazu führen, dass Größen wie die relative Tokenfrequenz unterschiedlich berechnet werden und damit nicht direkt vergleichbar sind. Natürlich wäre es unrealistisch zu erwarten, dass alle Korpora identische Angaben und (Abfrage-)Möglichkeiten liefern. Nichtsdestotrotz wären Hinweise auf Vergleichbarkeit in den Korpusdokumentationen nützlich. Eine Übersichtsstudie über genau diese Anwendungsunterschiede stellt ein Desiderat dar.

Ziel dieses Beitrags ist unsere Leser*innen über Möglichkeiten datengeleitet ein Forschungsinteresse zu entwickeln zu unterrichten. Weiters wollten wir mit unseren quantitativen Analysen darlegen, dass das zentrale Forschungsinteresse, das Suffix *-wärts*, in deutschen Texten sinkend produktiv ist. Es ist uns besonders wichtig aufzuzeigen, dass gerade quantitative Analysen extrem vorsichtig betrieben werden müssen und wir empfehlen insbesondere ein Augenmerk auf Vergleichsgrundlagen zu legen, besonders wenn verschiedene (Arten) von Korpora und verschiedene Plattformen verwendet werden. Nicht zuletzt hoffen wir, dass die Diskussion der Schwierigkeiten sowie Ergebnisse für andere Korpusanalysen sinnvoll sind sowie zur Methodenentwicklung beitragen können.

Webseiten Korpora (Stand 04/2021)

Alpenwort CQPweb Edition. <http://sprawi-cqpweb.uibk.ac.at/CQPweb/>
DWDS – Digitales Wörterbuch der deutschen Sprache. <https://www.dwds.de/>
DeReKo – Deutsches Referenzkorpus. <https://cosmas2.ids-mannheim.de/cosmas2-web/>
Schweizer Text Korpus. <https://www.chtk.ch/index.php/de/>
Text&Berg digital. <http://textberg.ch/site/de/willkommen/>

Bibliographie

- Aftabi, S. Z./Ahangar, A. A./Mishmast Nehi, H. (2021): Derivational Suffix Productivity in Persian: A Fuzzy Analysis. *Journal of Quantitative Linguistics*, 1–25. <https://doi.org/10.1080/09296174.2021.1887575>.
- Baayen, H. (1992): A quantitative approach to morphological productivity. In: G. Booij/J.van Marle (Hg.), *Yearbook of morphology 1991*. Dordrecht: Kluwer, 109–149.
- Baayen, R. H. (2009): Corpus linguistics in morphology: Morphological productivity. In: A. Lüdeling (Hg.), *HSK: Vol. 29,2. Corpus linguistics 2: An international handbook*. Berlin: De Gruyter, 899–919.
- Barbatesi, A./Geyken, A. (2020): Die Webkorpora im DWDS - Strategien des Korpusaufbaus und Nutzungsmöglichkeiten. In: K. Marx/H. Lobin/A. Schmidt (Hg.), *Deutsch in Sozialen Medien: Interaktiv - multimodal – vielfältig*. Berlin, Boston: de Gruyter, 345–348.
- Barz, I. (2016): Die Wortbildung. In: A. Wöllstein (Hg.), *Duden in 12 Bänden: Band 4: Die Grammatik (9th Ausgabe)*, Mannheim: Bibliographisches Institut, 644–774.
- Bickel, H./Gasser, M./Häcki Buhofer, A./Hofer, L./Schön, C. (2009): Schweizer Text Korpus – Theoretische Grundlagen, Korpusdesign und Abfragemöglichkeiten. *Linguistik online* 39(3), 5–31, URL: <http://dx.doi.org/10.13092/lo.39.474>.
- Brezina, V. (2018): *Statistics in corpus linguistics: A practical guide*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316410899>.
- Brezina, V./Meyerhoff, M. (2014): Significant or random? *International Journal of Corpus Linguistics*, 19(1), 1–28. <https://doi.org/10.1075/ijcl.19.1.01bre>.

- Bubenhofner, N./Rothenhäusler, K. (2018): „Die Aussicht ist grandios!“ – Korpuslinguistische Analyse narrativer Muster in Bergtourenberichten. In: N. Eller-Wildfeuer/ P. Rössler/A. Wildfeuer (Hg.), *Alpindeutsch. Einfluss und Verwendung des Deutschen im alpinen Raum*. Regensburg: edition vulpes, 39–60.
- Bubenhofner, N./Volk, M./Leuenberger, F./Wüest, D. (Hg.) (2015): *Text+Berg-Korpus (Release 151v01): Digitale Edition des Jahrbuch des SAC 1864–1923, Echo des Alpes 1872–1924, Die Alpen, Les Alpes, Le Alpi 1925–2014, The Alpine Journal 1969–2008*. Universität Zürich: Institut für Computerlinguistik.
- Dargiewicz, A. (2012): Die Sprache lebt und verändert sich. Zu neuesten Tendenzen in der deutschen Wortbildung. *Acta Neophilologica*, 14(1), 61–76.
- Donalies, E. (2005): *Die Wortbildung des Deutschen: Ein Überblick (2., überarbeitete Auflage)*. Tübingen: Narr.
- Elsen, H. (2014): *Grundzüge der Morphologie des Deutschen (2. aktualisierte Ausgabe)*. Berlin: de Gruyter.
- Engel, U. (2004): *Deutsche Grammatik (Neubearbeitung)*: München: Iudicium.
- Fleischer, W./Barz, I. (2012): *Wortbildung der deutschen Gegenwartssprache (4., völlig neu bearbeitete Auflage)*. Tübingen: Niemeyer.
- Gabrielatos, C. (2018): Keyness Analysis: Nature, metrics and techniques. In: C. Taylor/A. Marchi (Hg.), *Corpus Approaches to Discourse: A Critical Review* Milton: Taylor and Francis, 226–258.
- Ganslmayer, C. (2012): *Adjektivderivation in der Urkundensprache des 13. Jahrhunderts: Eine historisch-synchrone Untersuchung anhand der ältesten deutschsprachigen Originalurkunden*. Berlin: De Gruyter.
- Göhring, A./Volk, M. (2011): The Text+Berg corpus: an alpine french-german parallel resource. In: V. P. Mathieu Lafourcade (Hg.), *Actes des conférences TALN 2011 et Recital 2011*. Montpellier. <https://doi.org/10.5167/uzh-48404>.
- Graën, S. (2004): *Die Raumadverbien des Mittelhochdeutschen (1050–1350): Wörterbuch und Untersuchungen*. Göttingen: Georg-August-Universität Göttingen. <http://hdl.handle.net/11858/00-1735-0000-0006-AEE4-4>.
- Grimm, J. (o. J.): *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm*. Wörterbuchnetz des Trier Center for Digital Humanities (Hg.), Version 01/21, <https://www.woerterbuchnetz.de/DWB>.
- Heinle, E.-M. (1985): Wortbildung des Neuhochdeutschen bis zur Mitte des 20. Jahrhunderts. In: W. Besch/O. Reichmann/S. Sonderegger (Hg.), *HSK: Vol. 2.2*.

Sprachgeschichte: Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung, Berlin: De Gruyter, 1911–1917.

Henzen, W. (1965): *Deutsche Wortbildung*. Halle (Saale): Niemeyer.

Kluge, F./Seebold, E. (2012): *Etymologisches Wörterbuch der deutschen Sprache: EBookPlus (25., aktualisierte und erweiterte Auflage)*. Berlin: de Gruyter. <https://doi.org/10.1515/9783110223651>.

Kupietz, M./Belica/C., Keibel/H./Witt, A. (2010): The German Reference Corpus DEREKO: A Primordial Sample for Linguistic Research. In: N. Calzolari /K. Choukri/B. Maegaard/J.Mariani/J.Odijk/S. Stelios Piperidis/M. Rosner/D. Tapias. (Hg.), *Seventh international conference on Language Resources and Evaluation (LREC)*, 1848–1854, URL: http://www.lrec-conf.org/proceedings/lrec2010/pdf/414_Paper.pdf.

Kupietz, M./Lüngen, H./Kamocki, P./Witt, A. (2018): The German Reference Corpus DeReKo: New Developments – New Opportunities. In: N. Calzolari/K. Choukri/C. Cieri/T. Declerck/S. Goggi/K. Hasida/H. Isahara/B. Maegaard/J. Mariani/H. Mazo/A. Moreno/J. Odijk/S. Piperidis/T. Tokunaga (Hg.): *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki: European Language Resources Association (ELRA), 4353–4360.

Lohde, M. (2006): *Wortbildung des modernen Deutschen: Ein Lehr- und Übungsbuch*. Tübingen: Narr.

Motsch, W. (2004): *Deutsche Wortbildung in Grundzügen*. Berlin: de Gruyter, 235f.

Nübling, D. (2016): Die nicht flektierbaren Wortarten. In: A. Wöllstein (Hg.), *Duden in 12 Bänden: Band 4: Die Grammatik (9. Auflage)*. Mannheim: Bibliographisches Institut, 579–643.

Nübling, D./Dammel, A./Duke, J./Szczepaniak, R. (2017): *Historische Sprachwissenschaft des Deutschen: Eine Einführung in die Prinzipien des Sprachwandels*. Tübingen: Narr.

Pojanapunya, P./Watson Todd, R. (2016): Log-likelihood and odds ratio: Keynes statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory*, 14(1), 133–167. <https://doi.org/10.1515/cllt-2015-0030>

Posch, C./Rampl, G. (2020): Lima or Cima? Structure recognition and OCR in building the corpus of the Austrian Alpine Club Journal. *International Journal of Corpus Linguistics*, 25(4), 489–503.

- Posch, C./Rampl, G. (2017): *Projekt Alpenwort – Korpus der Zeitschrift des Deutschen und Österreichischen Alpenvereins (1869–1998)*. Innsbruck: Institut für Sprachwissenschaft, Universität Innsbruck, URL: <http://alpenwort.at>.
- Posch, C. (2022): *Digital Linguistics. Integrating Digital Humanities, Corpus Linguistics and Critical Discourse Studies*. Habilitationsschrift. Manuskript in Vorbereitung.
- Rampl, G./Posch, C. (2019): *Alpenwort – Korpus der Zeitschrift des Deutschen und Österreichischen Alpenvereins (1869 – 1998) CQPweb Edition*. Innsbruck: Institut für Sprachwissenschaft, Universität Innsbruck, <http://sprawi-cqpweb.uibk.ac.at/CQPweb/>.
- Schneider-Wiejowski, K. (2012): *Produktivität in der deutschen Derivationsmorphologie*. Dissertation. Bielefeld: Universität Bielefeld. <https://d-nb.info/1021023833/34>.
- Schneider-Wiejowski, K. (2013): Sprachwandel anhand von Produktivitätsverschiebungen in der schweizerdeutschen Derivationsmorphologie. *Linguistik Online* 38(2), 79–90.
- Schulz, D./Griesbach, H. (1972): *Grammatik der deutschen Sprache*. München: Max Hueber.
- Simmler, F. (1998): *Morphologie des Deutschen: Flexions- und Wortbildungsmorphologie*. Berlin: Weidler.
- Stefanowitsch, A. (2020): *Corpus Linguistics: A Guide to the Methodology // Corpus linguistics: A guide to the methodology*. Berlin: Language Science Press.
- Wöllstein, A. (Hg.) (2016): *Der Duden: in zwölf Bänden; das Standardwerk zur deutschen Sprache ; Band 4. Die Grammatik*, 9., vollständig überarbeitete und aktualisierte Auflage). Berlin: Dudenverlag.
- Weinrich, H./Thurmair, M. (2007): *Textgrammatik der deutschen Sprache (4., rev. Aufl.)*. Hildesheim: Olms.

