



Korpusbasierte Beobachtungen zu *entlang* Eine Modellstudie

1. Vorbemerkungen

In diesem Skript wird an einem konkreten Beispiel gezeigt, wie man bestimmte linguistisch bearbeitbare Probleme mittels einer korpusbasierten Studie in Angriff nehmen kann. Als Beispieleinheit fungiert dabei *entlang*. Folgend werden in einem ersten Schritt zunächst Problemkreise um *entlang* herum identifiziert mit dem Ziel, konkret bearbeitbare Fragestellungen in Bezug auf *entlang* zu gewinnen, woraufhin dann in einem zweiten Schritt einzelne Probleme herausgegriffen werden. Diese sind schließlich der Ausgangspunkt für die in diesem Skript vorgestellten Analysen zu *entlang*. Dieses Skript hat also zwei Funktionen: eine phänomenbezogene (Analyse zu einer nicht ganz unproblematischen sprachlichen Einheit mit konkreten Ergebnissen) und eine heuristisch-didaktische (Demonstration, wie man eine linguistisch relevante Fragestellung gewinnt, und vor allem, wie man sie mithilfe von korpusbasierten Analysen bearbeiten kann). Da es viele verschiedene Möglichkeiten gibt, korpuslinguistisch zu arbeiten, sei an dieser Stelle ausdrücklich darauf hingewiesen, dass es sich bei der vorliegenden Modellstudie nicht um den Versuch handeln kann, DIE eine oder die beste usw. Methode vorzustellen. Diese gibt es nämlich nicht. Zwar ist es nicht das Anliegen dieses Skriptes, die verschiedenen Möglichkeiten korpuslinguistischen Arbeitens in der Breite aufzuzeigen. Dafür können überblicksartig Lehrbücher herangezogen werden (etwa Lemnitzer/Zinsmeister ³2014, Scherer ²2014, Perkuhn/Keibel/Kupietz 2012, Draxler 2008, Stede 2007). Aber es muss klar sein, dass es sich bei der folgenden Modellstudie lediglich um *eine* Studie handelt, die durchaus auch anders hätte durchgeführt werden können. Welche Methode jeweils gewählt wird, hängt dabei von vielen Faktoren ab, nicht zuletzt auch von dem Erkenntnisinteresse, also der Fragestellung.¹

2. Die Fragestellung

Ganz grundsätzlich geht es der Linguistik um die Erfassung, Beschreibung und Erklärung sprachlicher Sachverhalte. Zum einen haben linguistische Analysen dabei auch eine externe Motivation, bei der also von außen (etwa von nicht-linguistischen Sprachbenutzern, Wissenschaftlern anderer Bereiche usw.) Fragen an die Linguistik herangetragen werden, die die Linguistik dann zu beantworten versucht. Häufig ergeben sich Anlässe für die Bearbeitung bestimmter Probleme aber auch aus linguistisch-internen Gründen. Etwa, weil ein bestimmter Bereich noch nicht oder bisher unzureichend beschrieben wurde, weil neue Phänomene auftauchen, weil eine Theorie noch ausgebaut werden muss, damit mit ihr gut gearbeitet werden kann, weil man andere Beschreibungs- und Erklärungsansätze, die man für inadäquat hält, entkräften möchte, weil man eine neue Theorie aufstellen möchte usw. Und manchmal kann für die Entwicklung einer Fragestellung eine grundsätzliche externe Motivation postuliert werden. Das wäre etwa der Fall, wenn man sich als Linguist mit Zweifelsfällen² beschäftigt, ohne dabei aber auch konkret nach der Lösung eines bestimmten Falls gefragt worden zu sein. Hier berührt man ein sprachpraktisches Interesse, das zu einem großen Teil normativer Natur ist: Man will wissen, wie etwas richtig verwendet wird. Eine Anlaufstelle für derartige Frage-

¹ Zum wissenschaftlichen Arbeiten im Bereich der Linguistik allgemein vgl. Rothstein (2011).

² Vgl. hierzu die [Ausgabe 16/4](#) der Zeitschrift [Linguistik Online](#) und darin insbesondere [Klein \(2003\)](#).

stellungen war (von 1915 oder 1918 an kostenlos³) und ist (seit 1998 kostenpflichtig⁴) die Duden-Sprachberatung. Der Zweifelsfall-Duden (7. Aufl. 2011: „Richtiges und gutes Deutsch“, 8. Aufl. 2016: „Das Wörterbuch der deutschen Zweifelsfälle“ (hrsg. v. Mathilde Hennig)) ist hier, wenn man so will, mittlerweile zu einer Art Nebenprodukt der Duden-Sprachberatung geworden:

„Häufig sich wiederholende Sprachanfragen wurden und werden von der Dudenredaktion in einer eigenen Kartei erfaßt. [...] das so zusammengetragene Material [ist] die Grundlage für die Bearbeitung des 9. Bandes der Dudenreihe, der seit der dritten Auflage unter dem Titel ‚Richtiges und gutes Deutsch‘ angeboten wird. Es handelt sich hierbei um ein alphabetisch strukturiertes Nachschlagewerk über sprachliche Zweifelsfälle, das man als eine Art Seismograph für die Schwierigkeiten beim Sprachgebrauch der deutschen Sprache ansehen könnte.“ (Scholze-Stubenrecht 1991: 180)

Man kann so vielleicht sagen: Wenn etwas im Zweifelsfall-Duden steht, so ist dies ein Indiz dafür, dass hierüber in der Sprachgemeinschaft Zweifel bestehen und möglicherweise auch ganz konkret zu dieser Einheit Fragen bei der Sprachberatung eingegangen sind.

Eine grundlegende Motivation linguistischer Arbeiten besteht nun darin, zu überprüfen, ob Sprachbeschreibungen oder auch Normierungen gegenüber einer sich im Sprachgebrauch zeigenden Sprachwirklichkeit Bestand haben. Ob also zum Beispiel die Einträge im Zweifelsfall-Duden auch einer zu erhebenden Sprachgebrauchsrealität entsprechen, ob sie adäquat sind.⁵

Zu der uns hier interessierenden Einheit ist in der CD-Rom-Version des Bandes „Richtiges und gutes Deutsch“ (Dudenredaktion ⁷2011: s. v. *entlang*, ohne Änderung auch in Hennig ⁸2016: s. v. *entlang*) folgender Eintrag zu finden:

entlang:

1. Stellung und Rektion: Das Wort kann entweder vor (Präposition) oder nach dem Substantiv stehen (Postposition): *die Wand entlang / entlang der Wand; den Fluss entlang / entlang des Flusses* usw. Im heutigen Sprachgebrauch wird *entlang*, wenn es **nach** dem Substantiv steht, gewöhnlich mit dem Akkusativ und abgesehen vom Schweizerischen nur selten mit dem Dativ verbunden: *die Wand / das Seil entlang. Die Grenze verlief von X nach Y den Wald entlang*. Aber: ... *die dem blitzenden Strom- und Meeresufer entlang aus der Hauptstadt hinausführte nach Belem* (R. Schneider). Wenn *entlang* **vor** dem Substantiv steht, dann wird es mit dem Genitiv, seltener mit dem Dativ verbunden: *entlang **des** Flusses / dem Fluss; entlang **der** Linien / den Linien*. *Als er ... durch ein dünnes Glimmerfensterchen entlang **des** Rohres Alphateilchen hindurchschoss ...* (R. Menzel). Der Akkusativ (*entlang den Wald / das Brückengeländer*) ist bei Voranstellung veraltet.

2. Rechtschreibung: Bei *entlang* in Verbindung mit einem Verb ist sowohl Getrennt- als auch Zusammenschreibung möglich. **Zusammen** schreibt man, wenn (betontes) *entlang* Verbzusatz ist: *Wir werden am Ufer entlanglaufen / das Ufer entlanglaufen*. **Getrennt** schreibt man, wenn *entlang* Teil einer adverbialen Bestimmung ist (das Verb kann oder muss ebenfalls betont werden): *Wir werden am Ufer entlang weiterfahren*. **↑ Getrennt- oder Zusammenschreibung ↑ (1.3).**

Abbildung 1: CD-Rom Eintrag zu *entlang* im Zweifelsfall-Duden (Dudenredaktion ⁷2011, Hennig ⁸2016)

³ Scholze-Stubenrecht (1991: 178) bezieht sich bei der Datierung auf die 9. Auflage der Duden-Rechtschreibung, auf der Duden-Informationssseite (URL s. nächste Fußnote) wird auf die Auflage des Rechtschreibduden von 1915 Bezug genommen.

⁴ Vgl. hierzu: http://www.duden.de/ueber_duden/sprachberatung (aufgerufen am 06.11.2014).

⁵ Damit reiht sich diese Fragestellung grundsätzlich in die Kodexforschung ein. Vgl. hierzu auch Klein (2014).



Dieser Eintrag soll uns als Einstieg dienen. Es sind hier mindestens die folgenden Probleme (mehr oder weniger explizit) erkennbar:

1. Was kann *entlang* als Wort alles sein? Im ersten Teil („Stellung und Rektion“) ist die Rede von einer Prä- oder Postposition, also einem Kasus fordernden Fügteil. Im zweiten Teil „Rechtschreibung“ wird die Schreibung diskutiert unter der Vorgabe „in Verbindung mit einem Verb“. Es wird ersichtlich, dass damit einerseits ‚Teil eines Verbs‘⁶ gemeint ist („Verbzusatz“), andererseits aber auch „Teil einer adverbialen Bestimmung“ und damit möglicherweise wohl auch ‚Adverb‘ gemeint ist. Damit hätte man einen eigenständigen Problembereich identifiziert, der quer liegt zu der Einteilung im Zweifelsfall-Duden. Nämlich die Frage nach dem lexikalischen Status und damit verbunden der Wortart von *entlang*. Ist es kasusforderndes Fügteil, Adverb, Teil einer Adverbialbestimmung oder Bestandteil eines komplexen Verbs?
2. Für den Fall, dass es ein Fügteil ist, behandelt der Eintrag die Probleme, welchen Kasus es verlangt (Rektion) in Verbindung mit der Frage, ob es vor oder nach der gefügten Einheit steht (Stellung). Die Frage also ist: Welcher Kasus (oder welche Kasus) tritt (oder treten) auf, wenn das Fügteil *entlang* vor der so gefügten Einheit steht, und welche/r, wenn es nach der gefügten Einheit steht?
3. Dass die anderen Wortarten-Varianten unter der Frage der Getrennt-und-Zusammenschreibung diskutiert werden, ist vielleicht ein wenig unglücklich, denn Wortart und Schreibung liegen im Falle trennbarer Verben quer zueinander, könnte aber dem Auftauchen etwaiger Fragen nach der Schreibung im Rahmen der Beratung entspringen. Wenn man die Frage lexikalisch aufrollt, so versteckt sich hier das Problem, ob es Anhaltspunkte für die Entscheidung hinsichtlich des lexikalischen Status von *entlang* gibt, wenn es kein Fügteil ist, also Anhaltspunkte dafür, ob es dann ein eigenständiges Adverb oder ein nicht-eigenständiger, aber trennbarer Verbbestandteil ist.

Diese Problembereiche sind es, die wir folgend mit einer korpusbasierten Studie (und zwar im Engeren in den Abschnitten 3.3 und 3.4) angehen wollen. Dafür ist es notwendig Daten zu erheben, die erhobenen Daten auszuwerten und die Ergebnisse zu interpretieren. Welche Daten wie erhoben werden, hängt direkt von der konkreten einzelnen Fragestellung ab. Für die Hauptfrage, welche Kasus bei einer eventuellen Präposition *entlang* (also in Voranstellung) und welche Kasus bei einer eventuellen Postposition *entlang* (also in Nachstellung) vorkommen, kann z. B. in einem Korpus nach dem Vorkommen des Wortes *entlang* überhaupt (also nach Belegen mit *entlang*) gesucht werden und können diese Vorkommen (Belege) dann danach ausgewertet werden, in welchem Kasus die jeweilig folgende oder vorausgehende Nominalgruppe (NGr) steht. Dabei ist es notwendig, die auszuwertenden Belege möglichst ohne unnötige Verzerrung zu erheben. In der Studie unten werden dafür die ersten 100 Belege des chronologisch geordneten Vorkommens von *entlang* in dem DWDS-Kernkorpus ausgesucht (s. dazu im Detail weiter unten). Durch verschiedene weitere Erhebungen können gezielt weitere Teile der Fragestellung bearbeitet werden. So kann z. B. separat nach zusammengesetzten und getrennt vorkommenden Formen von *entlang* mit bestimmten Verben gesucht werden, nach dem Vorkommen von *entlang* mit bestimmten anderen Präpositionen usw. Die Nutzung eines Korpusportals bietet hier gute Möglichkeiten, die verschiedenen Aspekte der obigen Fragestellung korpusbasiert zu bearbeiten. Im Einzelnen wird dies in den folgenden Abschnitten präsentiert.

⁶ Einfache Anführungsstriche werden außerhalb von Zitaten hier in diesem Skript benutzt, um damit Bedeutungen oder Gemeintes etc. zu markieren, und doppelte, wenn es sich um zitierte Stellen handelt. Objektsprachliches wird kursiviert; manchmal wird Kursivierung auch verwendet, um etwas als erwähnten (zum Beispiel: zu definierenden) Terminus zu markieren.



Eine Bemerkung noch zu dem Terminus *korpusbasiert* und den damit zusammenhängenden Möglichkeiten der Ausrichtung von Korpus-Studien im Allgemeinen und der folgenden im Besonderen. Das Gegenstück zu *korpusbasiert* (engl.: *corpus-based*) ist *korpusgesteuert* oder *korpusgetrieben* (engl.: *corpus-driven*). Rein korpusgesteuerte Analysen sind *induktiv*. Das heißt: Sie liefern Fragen und Erkenntnisse aus den Daten heraus. Ihre Erkenntnisrichtung ist somit *bottom up*. Das heißt: Hier liefert die Analyse eines Korpus Daten, in denen sich bestimmte Phänomene zeigen und die dann die Grundlage für darauf bezogene Fragestellungen bilden – kurz gesagt: von den Daten als Grundlage (bottom) nach oben (up) zu Fragen und Ergebnissen. So zeigen sich bei solchen Analysen beispielsweise, welche Wörter mit welchen anderen Wörtern in einem Korpus statistisch signifikant⁷ häufig vorkommen (sog. Kookkurrenzen oder Kollokationen) – etwa im Vergleich mit einem anderen Korpus, in dem möglicherweise ganz andere Kookkurrenzen zutage treten. Solche quantitativen Analysen sind wichtig, können aber, wollen sie zu relevanten qualitativen Aussagen kommen, nur Teil von umfassenderen Analysen sein. Zumindest müssen die Ergebnisse induktiv gewonnener quantitativer Daten ja immer erst noch interpretiert werden, bevor man mit ihnen etwas anfangen kann. *Deduktiv*, also ‚ableitend‘, sind dagegen Analysen, bei denen zu Anfang bereits eine mehr oder weniger klare Fragestellung existiert, die man dann – und das ist korpusbasiertes Arbeiten – mit der konkret darauf ausgerichteten Untersuchung von Korpora beantworten möchte. Hier ist der Prozess als *top down* zu charakterisieren. Das heißt: von der Frage oben (top) zu den Daten hinunter (down), die man in Bezug auf die Frage erhebt und mit denen man die Frage schließlich beantworten möchte. Induktive und deduktive Verfahren werden in größeren Studien schließlich sinnvollerweise miteinander kombiniert, sodass man auch von einem spiralenartigen Erkenntnisprozess sprechen kann.

Die folgend vorgestellte Studie ist sowohl

- qualitativ (d. h. sie möchte zu relevanten beschreibenden Aussagen kommen) als auch
- quantitativ (d. h. sie stellt Häufigkeiten bestimmter Phänomene fest), sie ist weiterhin
- deduktiv (d. h. es geht ihr um die Beantwortung konkreter, vorher feststehender Fragestellungen) und sie ist
- korpusbasiert (d. h. die Fragen werden unter Zuhilfenahme der Auswertung von Daten beantwortet, die man durch darauf abgestimmte gezielte Suchen in Korpora erhält).

3. Methode (Datenerhebung, Datenauswertung), Ergebnisse und Interpretationen

3.1. Korpora und Korpusportale

Linguistisch sind Korpora zunächst einmal nichts anderes als verfügbare Mengen sprachlicher Einheiten, die irgendwann einmal von irgendwem verwendet wurden. So wäre bspw. jedes Buch mit Text ebenso ein Korpus wie etwaige Aufnahmen aller Äußerungen, die Sie und ich gestern und heute getätigt haben. Rein theoretisch ließe sich auch für das Deutsche insgesamt eine korpuslinguistische Definition aufstellen, nämlich: Gesamtmenge aller sprachlicher Äußerungen des Deutschen. Ein Korpus, das alle schriftlichen und mündlichen Äußerungen des Deutschen enthalten würde, kann es aber trivialerweise nicht geben – die Gesamtmenge wäre also virtuell. Zum einen wird längst nicht jede Äußerung konserviert und ist längst nicht jede irgendwie konservierte Äußerung auch verfügbar. Zum anderen kommen in jeder Sekunde neue Äußerungen hinzu und man kann nicht an allen Orten dieser Welt zugleich alle deutsch-

⁷ Zu quantitativen und insbesondere statistischen sowie zu allgemein empirischen Verfahren in der Linguistik vgl. Meindl (2011), Albert/Marx (2010), Hauser/Humpert (2009), Gries (2008), Albert/Koster (2002), Schlobinski (1996).



sprachigen Äußerungen konservieren. Weil das so ist, ist jedes noch so große Korpus letztlich ein Ausschnitt der sprachlichen Realität. Dies eröffnet sofort die Frage nach der Repräsentativität eines Korpus. Wenn ein Korpus jeweils nur einen Teilbereich der sprachlichen Realität abbildet, ist dieses Korpus strenggenommen auch erst einmal nur für diesen Teilbereich aussagekräftig. Wenn etwa ein Korpus der Äußerungen von Schülern und Lehrern während einer Schulstunde vorliegt, so gelten Aussagen, die man über die Analyse dieses Korpus gewinnt, auch erst einmal nur für die Äußerungen von Schülern und Lehrern dieser Klasse zu diesem Zeitpunkt an diesem Ort. Dieses Korpus kann man aber als ein Beispiel für konservierte Unterrichtskommunikation ansehen und die Ergebnisse der Analyse dieses Korpus mit denen der Analyse eines anderen Korpus vergleichen (etwa mit anderen Schülern und anderen Lehrern als Interaktanten).

Die oben gestellten Fragen zu *entlang* beziehen sich auf die deutsche Standardsprache. Die Wahl oder Zusammenstellung des Korpus sollte also so ausfallen, dass das Korpus geeignet ist, zumindest beispielhaft für die Untersuchung des Standarddeutschen herangezogen werden zu können. Auch wenn die Frage, was genau dieses Standarddeutsch ist (bzw. welche Äußerungen man als standarddeutsche ansieht, welche SprecherInnen StandardsprecherInnen sind, in welchen Situationen was für ein Standard erwartbar ist, ob es lediglich einen Standard gibt oder zumindest auch nationale Varietäten wie etwa das deutsche, das schweizerische und das österreichische Deutsch als Standard angesehen werden, und ob darüber hinaus nicht auch zwischen gesprochenem und geschriebenem Standard unterschieden werden muss usw.), so kann die Frage korpuslinguistisch eingengt werden auf: Welches für Analysen nutzbare Korpus ist als standardsprachliches und halbwegs repräsentatives akzeptiert? Glücklicherweise gibt es auf diese Frage derzeit gleich mehrere Antworten (für eine Übersicht vgl. Bartz/Radtke 2014, Scherer ²2014: 94f.). Drei im Internet verfügbare große Korpora (bzw. eigentlich Korpusportale, denn jede dieser Anlaufstellen verfügt über mehrere verschiedene Korpora, die man über diese Portale ansteuern kann) sind:

- das DeReKo (= Deutsches Referenzkorpus) des IDS (= Institut für deutsche Sprache) in Mannheim, das über das Programm COSMAS (= Corpus Search and Management System) befragbar ist, wobei dieses Programm mittlerweile sowohl herunterladbar und vom eigenen Computer aus gesteuert (COSMAS_{win}) als auch komplett im Netz genutzt werden kann (COSMAS_{web})
(COSMAS Einstiegs-URL: <http://www.ids-mannheim.de/cosmas2/>),
- die Korpora des Wortschatz-Portals der Universität Leipzig, das komplett über das Internet genutzt werden kann
(Einstiegs-URL: <http://wortschatz.uni-leipzig.de/>), und
- die Korpora des DWDS (= Digitales Wörterbuch der deutschen Sprache) der BBAW (= Berlin-Brandenburgischen Akademie der Wissenschaften), ebenfalls im Netz nutzbar
(Einstiegs-URL: <https://www.dwds.de/>).

COSMAS als das älteste Programm ist in der Anwendung am komfortabelsten und eröffnet auch mehr Nutzungsmöglichkeiten als die anderen Portale. Das Inventar an Zeichen für die Gestaltung verschiedener Suchanfragen (also die Suchsyntax) ist dort zwar nicht unbedingt komplizierter, jedoch etwas umfangreicher und weniger intuitiv als bei den anderen beiden, die ihrerseits dafür aber auch erheblich weniger Such- und Auswertungsmöglichkeiten anbieten. Und man muss sich für die Nutzung von COSMAS registrieren lassen, was aber einfach und kostenlos ist. Derzeit entwickelt das IDS im Rahmen der Projekte in OWID (= Online-Wortschatz-Informationssystem Deutsch) übrigens ebenfalls eine vereinfachte Korpusuche



(vgl. <http://www.owid.de/>), die (nach Registrierung) auch bereits online nutzbar ist. Die anderen beiden Portale sind etwas leichter (und auch ohne Registrierung) zugänglich. Man muss aber auch dort genau wissen, was man tut, wenn man sie nutzt. Und das heißt: Man muss sich die Möglichkeiten der Korpusuche in allen Fällen erarbeiten, will man die Portale sinnvoll nutzen.

Alle drei Portale stützen sich auf geschriebenes Deutsch (Zeitungen, Zeitschriften, Romane, Erzählungen usw., aber mittlerweile auch Wikipedia-Einträge und vieles mehr), sodass mit ihrer Hilfe nicht so ohne Weiteres auch gesprochene Sprache untersucht werden kann. Das derzeit größte Korpus der gesprochenen Sprache ist

- die DGD (= Datenbank für Gesprochenes Deutsch mit FOLK (= Forschungs- und Lehrkorpus) als dem ‚kleinen gesprochenen sprachlichen Bruder des großen schriftsprachlichen DeReKo) des IDS in Mannheim, die komplett im Netz genutzt werden kann (incl. Speicherung von Suchergebnissen usw.)

(Einstiegs-URL: http://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.welcome).

Um eine Einheit im Standarddeutsch zu untersuchen, ist die Korpus-Situation also einigermaßen gut: Für die meisten Erkenntnisinteressen sind Korpora verfügbar. Man braucht in der Regel also nicht erst eigens ein standardsprachliches Korpus erstellen.

Im Rahmen dieser Modellstudie wurde das DWDS als Portal gewählt. Es ermöglicht – wie wir noch sehen werden – die wichtigsten Suchen für unsere Fragen, die Ergebnisse sind schnell und gut sortiert verfügbar und die entsprechenden Fundstellen lassen sich auch im größeren Umfang herunterladen, womit sie für spätere Analysen und Nach-Analysen auf dem eigenen Computer erhalten bleiben. Für die Suche steht im DWDS eine gut verstehbare (oder *verständliche*?) und nicht zu lange Hilfeseite zur Verfügung, auf die man über den Link oben rechts „Hilfe zur Suche“ gelangt (vgl. <http://www.dwds.de/hilfe/suche/>).

3.2. Im Korpus suchen und ein Belegkorpus als Treffermenge erhalten⁸

Grundsätzlich gibt es zwei Möglichkeiten, Korpora zu nutzen, die beide in Ergänzung zueinander angewendet werden können. Entweder man befragt das Gesamtkorpus über bestimmte Suchanfragen daraufhin, wie häufig welche durch die Suchanfragen angepeilten sprachlichen Einheiten vorkommen, und belässt es für die einzelnen Suchen jeweils bei der quantitativen Auswertung. Oder man schaut sich die durch die Suchanfrage erzielten Treffer genauer an, um damit spezielle qualitative Fragestellungen zu beantworten. Die erste Methode geht schneller, ist aber dafür aber auch fehleranfälliger. Maschinen machen grundsätzlich nur das, was man ihnen sagt, und man muss stets überprüfen, ob auch im Sinne des Erkenntnisinteresses gearbeitet wurde. Wie gut eine Analyse ist, hängt gleich zu Beginn direkt davon ab, wie

⁸ Die folgend präsentierte DWDS-gestützte Studie wurde im Frühsommer 2015 durchgeführt. Die Treffermengen usw. dürften jetzt größer sein, da auch die DWDS-Korpora anwachsen. Zudem hat sich inzwischen der Internetauftritt geändert. Die nunmehr alte Version ist aber nach wie vor (Stand: Frühsommer 2017) unter <http://eins.dwds.de/> ansteuerbar. Die neue unter <https://www.dwds.de/>. Geändert hat sich vor allem der Aufbau, nicht sehr die Funktionalität. D.h. u.a.: Alle Suchanfragen, die folgend in diesem Skript vorgestellt werden, funktionieren so auch in der neuen Version. Eine Veränderung ist aber so wichtig, dass sie unbedingt hier erwähnt werden muss. Wenn man die DWDS-Startseite öffnet, in die dortige Suchmaske ein Wort eingibt und dann auf Enter drückt, öffnet sich – soweit vorhanden – zunächst der DWDS-Wörterbucheintrag. Um eine Korpusuche durchführen zu können (wie in der vorliegenden Studie) muss man vorher das Menü „Textkorpora“ anklicken bzw. aufrufen, bevor man eine Suchkette eingibt. Erst dann gelangt man in eine Ansicht, bei der die Treffer als Ansicht (einstellbar: KWIC, voll oder maximal) erscheinen. **Wenn man die Suchen der vorliegenden Studie nachvollziehen möchte, ist es also unbedingt notwendig, vor einer Sucheingabe „Textkorpora“ durch Klick auszuwählen!**

gut die Suchanfrage ist. Mit der Suchanfrage sagt man einem Programm, wie es etwas suchen soll, und damit auch, was es finden kann (und was nicht!). Man muss also erst einmal an der richtigen Suchanfrage arbeiten. Sie sollte möglichst wenig Fehlbelege als Treffer erzielen (also möglichst wenig Verwendungen, in denen nicht das Gesuchte, sondern etwas anderes gefunden wurde) und möglichst wenig an möglichen und gewünschten Belegen ausschließen (also nicht nur Verwendungen, die lediglich einen stark beschränkten Ausschnitt aus der Verwendungsbreite der fraglichen Einheit zeigen). Ein Beispiel: Wenn man sich für das Verb *entlanggehen* unter der Fragestellung interessiert, ob es sich dabei um ein trennbares Verb (= *entlang* und *gehen* können einzeln und voneinander entfernt stehen) handelt oder nicht (= in allen Verwendungen immer direkt aufeinanderfolgend und im Schriftlichen auch zusammengeschieden), so kann man dies durch die folgende einfache und naheliegende Suche nicht herausbekommen:

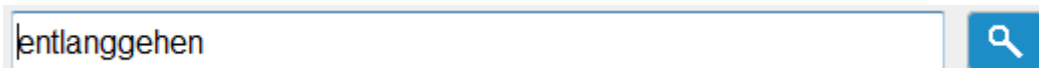


Abbildung 2: Suchanfrage *entlanggehen* im DWDS⁹

Denn die Suchanfrage¹⁰

entlanggehen

bedeutet: ‚Finde alle Stellen, in denen eine Form von *entlanggehen* vorkommt‘, wobei auch tatsächlich nur nach *entlanggeh-* + *-en, -e, -st* etc. und *entlanggegangen, entlangging, entlangzugehen* usw. gesucht wird. Das ist an sich schon erstaunlich, denn woher weiß das Programm, dass es auch nach z. B. *entlangging* suchen soll, obwohl nur

entlanggehen

eingetragen wurde? Die Antwort lautet: Die Suchsoftware ist so programmiert, dass sie *entlanggehen* als ein Verb erkennt und über grundsätzliche Informationen darüber verfügt, wie Verbformen (dieses Verbs) gebildet werden. Diese Information ist etwa auch in der sehr wichtigen Suchhilfe (ansteuerbar über [Hilfe zur Suche](#) oben rechts in dem DWDS-Fensterrahmen) zu finden. Es heißt dort (unter dem Stichwort *Expansion*) zu dieser Suchmöglichkeit:

„[...] können Sie eine beliebige Wortform eingeben (Grundform oder flektierte Form). Die Suchmaschine ermittelt zunächst das Lexem, in dessen Flexionsparadigma Ihre Eingabe enthalten ist. Gefunden werden dann alle Wortformen, die zu dem ermittelten Lexem gehören.“ (aus der DWDS-Hilfeseite; <http://www.dwds.de/hilfe/suche/>)

Wenn dies nicht so wäre, so würde die Suchmaschine ausschließlich Vorkommen des Infinitiv Präsens und der 1. und 3. Person Singular Präsens Aktiv finden, denn die Suchvorgabe (die Suchkette) sind bestimmte Buchstaben, die es ansonsten im Korpus exakt aufzufinden gälte. Selbst *entlanggehe* würde dabei nicht gefunden werden, denn diese Form ist nicht iden-

⁹ Alle Scans stammen aus der nunmehr (2017) alten DWDS-Version, die aber noch unter <http://eins.dwds.de/> erreichbar und auch voll funktionsfähig ist. Der neue Internetauftritt <https://www.dwds.de/> bringt visuelle Veränderungen mit sich, behält aber die Funktionen der älteren DWDS-Version bei, sodass sich inhaltlich nichts ändert. Ich habe daher davon abgesehen, die Scans in diesem Lehrskript zu erneuern. Wer mit dem neuen DWDS arbeitet, sollte sich dort auf die Suche begeben, wo die einzelnen Funktionen zu finden sind und wie die Zugänge dort aussehen. Das dürfte wenig Probleme bereiten.

¹⁰ Folgend werden alle Suchanfragen im Fließtext textuell in einer eigenen, eingerückten Zeile und (in Fußnoten nur) durch Courier New in der Schrifttype oder durch eine Scan-Abbildung hervorgehoben. Auf diese Weise sollen Missverständnisse in Bezug auf die Formulierung der Suchanfrage vermieden werden. Das heißt: Die Zeichen, die derart hervorgehoben als Suchanfrage erscheinen, sind allesamt Zeichen, die komplett ins Suchfeld eingegeben werden.

tisch mit *entlanggehen*. Man könnte sagen: Die Programmierer haben schon damit gerechnet, dass man bei einer einfachen Sucheingabe wie


entlanggehen

eigentlich das VERB *entlanggehen* als Lexem meint und nicht die reine Buchstabenfolge *entlanggehen*.

Über die KWIC-Ansichten (KWIC = Key Word In Context) kann man sich nun einen ersten Eindruck davon verschaffen, wie viele und welche Treffer mit dieser Suchkette gefunden wurden. Für das Kernkorpus des DWDS,

„ein zeitlich und nach Textsorten (Belletristik, Gebrauchsliteratur, Wissenschaft, Journalistische Prosa) ausgewogenes Korpus des gesamten 20. Jahrhunderts. Textgrundlage sind Werke der Literatur, wissenschaftliche Texte, Gebrauchstexte und Zeitungstexte, die in einem ausgewogenen Verhältnis zueinander stehen.“ (<http://www.dwds.de/ressourcen/korpora/>),

werden (am 11.11.2014) 72 Treffer erzielt – erstaunlich wenig. Das KWIC-Fenster zeigt davon einen Ausschnitt:



KWIC	Datum ↓	Datum ↑	Zufällig	Links	Rechts
8	1971				silarmembran , die die Cochlea entlanggehen können , vom Steig
9	1971				Und während wir den Korridor entlanggehen , sage ich noch etw
10	1969				n höchst geheimnisvollen Weg entlanggegangen , den das Licht
11	1969				macht es Spaß , auf dem Wall entlangzugehen , unter den alten
12	1966				dern an einer Fertigungsstraße entlanggehend , die gesellschaftl
13	1964				se aber - wenn man das Kabel entlanggeht - manchmal ihre Paa
14	1959				ußen Haufen nagelneuer Steine entlanggegangen bist und sie be
15	1958				l mit weiten Schritten den Weg entlanggehen und hinter dem Die
16	1957				n Seiten grüßend , einen Gang entlanggehen , eine Tür öffnen , z
17	1957				ann er jetzt verzweifelt den Flur entlangging , dem Wiedersehen i
18	1954				glücklich , an ihrer alten Mauer entlangzugehen , aber die Macht
19	1953				Wiesental an einem Fließchen entlanggeht , wird man mit Siche
20	1950				Während sie den Gang entlanggehen , sagt Doktor Linz
21	1946				n ich aber sonntags am Hafen entlanggehe oder , wie jetzt , auf
22	1946				ann der Rückweg an der Alster entlanggehen , so den Kranken:

Version: 1.1 Optionen

Abbildung 3: KWIC-Ansicht der Treffer im Kernkorpus des DWDS zur Sucheingabe *entlanggehen*


Wenn man sich alle Treffer herunterlädt (über das Diskettensymbol  oben rechts ganz links) und sich alle Treffer ansieht, was bei diesen wenigen Treffern ja noch ohne Weiteres möglich ist, bemerkt man, wofür die KWIC-Ansicht schon ein Indiz ist. Man muss dafür nicht unbedingt alle Treffer auch herunterladen. Durch einen Klick auf die jeweilige Zeile sieht man jeden einzelnen Treffer jeweils auch als Ganzes im KWIC-Fenster. Zum Beispiel den Treffer 17:



Abbildung 4: Treffer 17 bei der Suche von *entlanggehen* im DWDS

Wollte man allerdings alle Treffer so durchgehen, wäre dies arbeitsorganisatorisch möglicherweise etwas zu aufwändig (abgesehen davon, dass man die Treffer nicht gesichert hätte).

Das Resultat dieser Suche wäre: *entlanggehen* wird immer zusammenschrieben. Dieses Resultat bedeutet aber nicht auch zugleich, dass es keine trennbaren Formen (etwa: *Ich gehe die Straße entlang*) gibt, sondern es wird erzielt, weil man nur nach zusammenschriebenen Formen gesucht hat. Der Grund dafür ist: Das Suchprogramm erkennt zwar, dass man ein Verb und keine Einzelform sucht, aber es unterscheidet nicht, ob es sich dabei um ein trennbares Verb handelt oder nicht. Man bekommt also heraus, was man hineingesteckt hat. Um diese Frage beantworten zu können, muss man folglich eine andere Suchkette eingeben, z. B.:



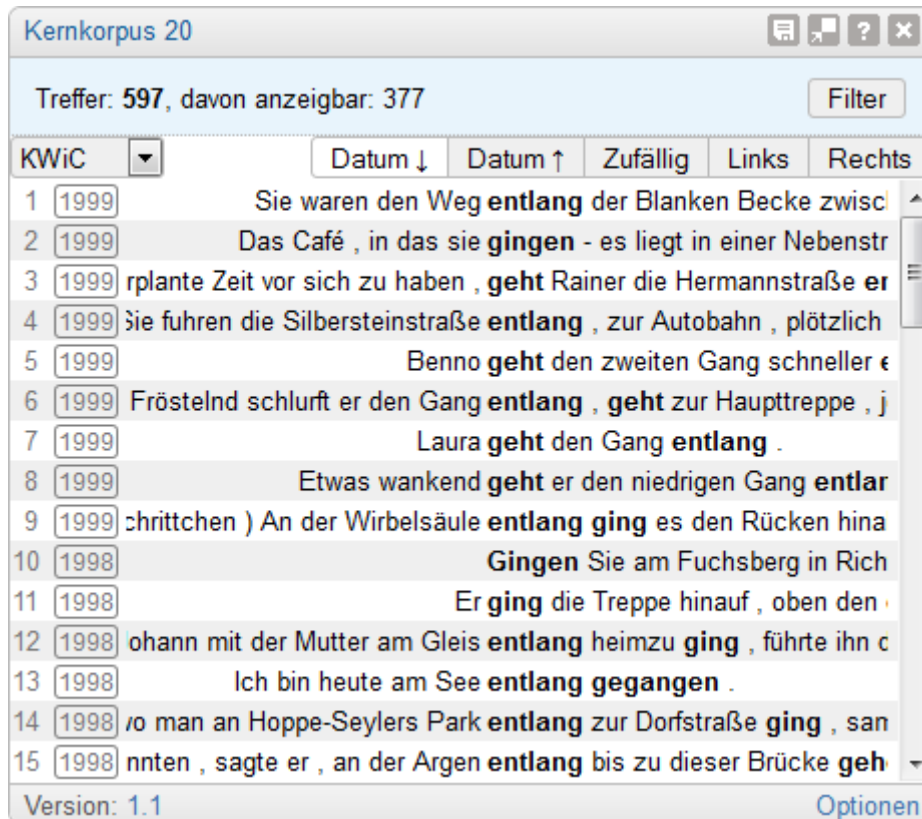
Abbildung 5: Suchanfrage *entlang* + *gehen* im DWDS

oder (gleichbedeutend):



Abbildung 6 Suchanfrage *entlang* + *gehen* im DWDS mit dem UND-Operator „&&“

Hier sucht man zwei Lexeme, die zusammen in einem Satz vorkommen. Auf diese Weise können trennbare Formen von *entlanggehen* gefunden werden, was die KWIC-Ansicht auf den ersten Blick auch bestätigt:



Kernkorpus 20

Treffer: 597, davon anzeigbar: 377

Filter

KWiC	Datum ↓	Datum ↑	Zufällig	Links	Rechts
1	1999				
2	1999				
3	1999				
4	1999				
5	1999				
6	1999				
7	1999				
8	1999				
9	1999				
10	1998				
11	1998				
12	1998				
13	1998				
14	1998				
15	1998				

Version: 1.1

Optionen

Abbildung 7: KWIC-Ansicht der Treffer im DWDS-Kernkorpus zur Sucheingabe "entlang" "gehen"

Jetzt hätte man als Ergebnismenge allerdings lediglich die getrennten Formen, und zwar wiederum deswegen, weil man diesmal nur nach den getrennten Vorkommen von *entlang* und *gehen* gesucht hat. Um beides mit nur einer Suche finden zu können, bedarf es einer etwas komplexeren Suchanfrage, bei der auch sog. logische Operatoren (UND, ODER, NICHT) verwendet werden. In unserem Fall von *entlanggehen* wäre dies eine Suche, mit der man sowohl *entlanggehen* ungetrennt als auch *entlang* und *gehen* getrennt finden kann. Dafür gibt es den ODER-Operator `||`¹¹ zwischen den Gliedern der Suchkette:

"entlang" "gehen" || entlanggehen

Abbildung 8: Suchkette, mit der man nach Vorkommen von *entlang* + *gehen* getrennt oder *entlanggehen* nicht getrennt in einem Satz suchen kann

Statt der beiden Anführungszeichen hätte man auch im ersten Glied einen entsprechenden Operator verwenden können, nämlich den UND-Operator `&&`.¹² Das Ergebnis (669 Treffer) ist das gleiche.

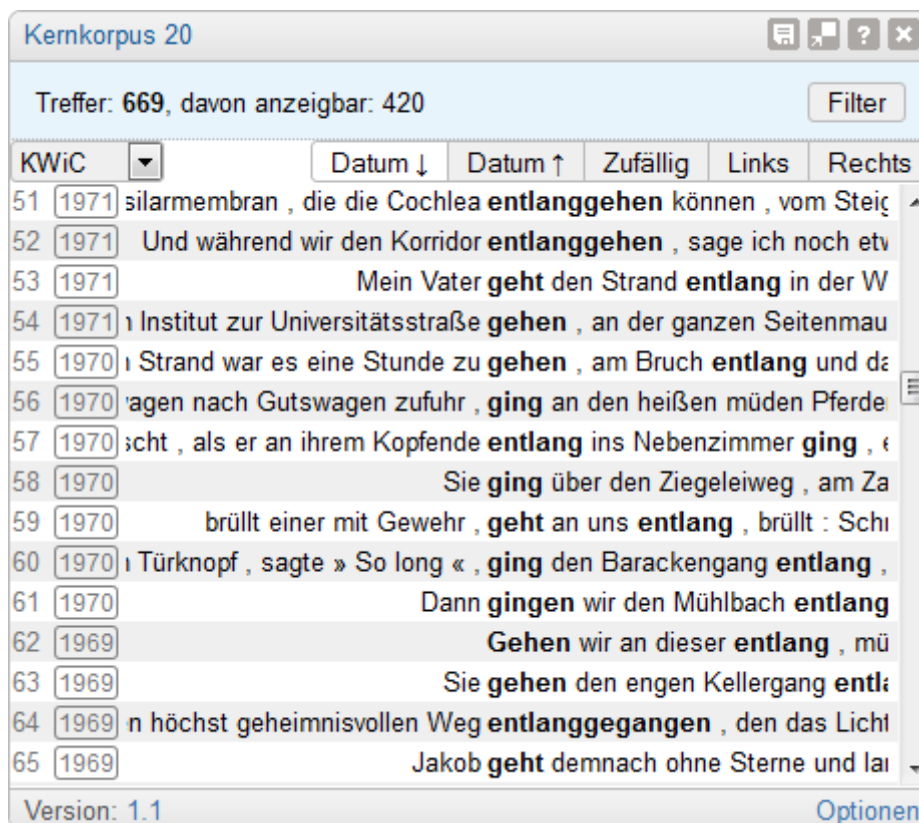
¹¹ Diesen ODER-Operator kann man z. B. direkt aus der Suchhilfe nehmen, indem man etwa auf das dortige Beispiel „Kanzler || Schröder“ klickt. Ansonsten sucht man sich das Zeichen (den Senkrechtstrich) auf der Tastatur oder als Symbol in einem Textverarbeitungsprogramm. Auf meiner Tastatur ist er bspw. in der ersten Buchstabenreihe zwischen [strg] und [y] und kann mittels gleichzeitigem Drücken von [alt gr] eingefügt werden, was man zweimal machen muss, um || zu erhalten. Der ASCII-Code für den senkrechten Strich ist: 0124. Dafür drückt man [alt] und währenddessen hintereinander die vier Ziffern 0124 (allerdings nicht über die Leiste oben auf der Tastatur, sondern über das separate Nummernfeld rechts, sofern vorhanden).

¹² Eine Übersicht über die verschiedenen anwendbaren Operatoren mit Erklärung und Beispielen findet man auf der Hilfeseite des DWDS (alt: <http://www.dwds.de/hilfe/suche/>, neu: <https://www.dwds.de/d/suche>). Eine maximal exakte Suchanfrage müsste im gegebenen Beispiel allerdings noch Einklammerungen vornehmen, weil hier ja zwei Operatoren verwendet werden und der zweite sich nicht auf das zweite und dritte Glied bezieht, sondern auf die ersten beiden zusammen als Einheit und das dritte als zweitem Glied. Die Kette müsste

entlang && gehen || entlanggehen 

Abbildung 9: Suchkette, mit der man nach Vorkommen von *entlang* + *gehen* getrennt oder *entlanggehen* nicht getrennt in einem Satz suchen kann

Über die KWIC-Ansicht erhält man wieder einen ersten Eindruck, ob man die richtigen Treffer erzielt hat. Man findet nun sowohl getrenntes als auch nicht-getrenntes Vorkommen von *entlang* und *gehen*:



Kernkorpus 20

Treffer: 669, davon anzeigbar: 420 Filter

KWiC	Datum ↓	Datum ↑	Zufällig	Links	Rechts
51	1971				
52	1971				
53	1971				
54	1971				
55	1970				
56	1970				
57	1970				
58	1970				
59	1970				
60	1970				
61	1970				
62	1969				
63	1969				
64	1969				
65	1969				

Version: 1.1 Optionen

Abbildung 10: KWIC-Ansicht der Treffer im Kernkorpus des DWDS zur Suche nach getrenntem oder nicht getrenntem Vorkommen von *entlanggehen* in einem Satz

Zur Überprüfung ließe sich die Trefferanzahl der beiden ersten Suchergebnisse, nämlich 72 für ungetrennt und 597 für getrennt, addieren. Man erhält als Summe tatsächlich genau die Anzahl des dritten Ergebnisses, nämlich 669. Mit diesem letzten Korpus kann man nun qualitativ weiterarbeiten. Es fungiert dabei als zu analysierendes Belegkorpus zu *entlanggehen* und enthält die erzielten Treffer als Belege, die man komplett oder von denen man eine Teilmenge einzeln nach bestimmten Kriterien auswerten kann. Hier beginnt jetzt die eigentliche qualitative Arbeit. Das heruntergeladene Korpus, mit dem man weiterarbeitet, ist eine txt-Datei, die mit jedem Textverarbeitungsprogramm, auch mit WORD oder dem Writer von Open Office, geöffnet und bearbeitet werden kann. Vor dem Herunterladen wird man gefragt, wie viele und welche Treffer man als Belege herunterladen möchte. Für die meisten qualitativen Analysen empfiehlt es sich, nicht nur den einen Satz, in dem die Treffer sind, auszuwählen, sondern

inhaltlich also so lauten: ‚(A und B) oder C‘ und also so aussehen: (entlang && gehen) || entlanggehen. Die Anzahl der Treffer ist aber gleich, nämlich 669. Man erkennt daran, dass das Suchprogramm wohl automatisch die Operatoren von links ausgehend weiter einschachtelt (vgl. auch Abschnitt 3.2 in der DWDS-Suchhilfe). Die Suche entlang && (gehen || entlanggehen) dagegen erzielt eine andere Trefferzahl, nämlich 597.

auch jeweils den Kotext, also die Sätze davor und danach. Das kann man tun, indem man „Maximal“ wählt.

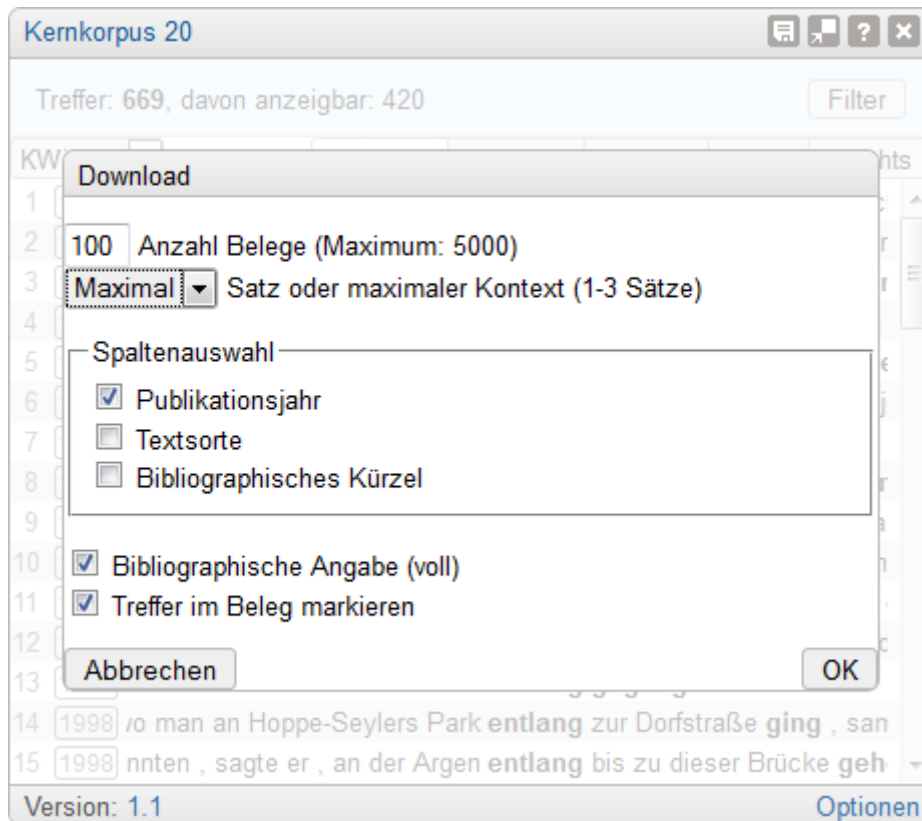


Abbildung 11: Optionsfenster beim Download (öffnet sich nach Klick auf )

3.3. Das Belegkorpus auswerten I: Kriterien und Datentabelle

Die folgend vorgestellten Belegkorpus-Auswertungen beziehen sich auf die Treffer, die man mit der einfachen Suchkette *entlang* im Kernkorpus erhält. Das sind (am 05.11.2014) insgesamt 3.376 Treffer, von denen eine Teilmenge (nämlich die ersten 100) unter den oben angesprochenen Fragestellungen ausgewertet wurden. Die Anzahl der auszuwertenden Treffer richtet sich nach verschiedenen Faktoren. Es gibt nicht für alle Studien eine gleichermaßen gültige Vorgabe, ab soundsoviel Belegen sei die Auswertungsmenge ausreichend. Zum Teil ist das Ansichts- oder Verhandlungssache (etwa zwischen Ihnen und dem Betreuer oder Auftraggeber der Studie). Grundsätzlich aber gelten die folgenden BeNIMMregeln:

- Nimm so viel wie möglich, denn je mehr Belege, desto besser.
Klarerweise ist die Analyse von mehr Fällen immer besser, weil sie sich dem allerdings unerreichbaren Total der Gesamtmenge des Vorkommens der zu untersuchenden Einheit annähert.
- Nimm nur so viel, wie du kannst, denn du hast nicht ewig Zeit.
Ebenso klar ist: Man kann nur das schaffen, was man schaffen kann. Eine sinnvolle Trefferanzahl muss berücksichtigen, dass man nicht unbegrenzt Zeit hat, die Auswertung des Korpus vorzunehmen.
- Nimm im Zweifelsfall eher weniger und analysiere dafür präziser, denn Qualität sticht.
Man braucht Zeit, um ordentliche Analysen durchzuführen. Ordentlich sind Analysen aber nur dann, wenn sie reflektiert und präzise genug durchgeführt werden. Dafür wird



es unter Umständen notwendig, weniger Daten für die Analyse zu nehmen als geplant. Jedenfalls aber sollte man die Analyse in kleineren Probedurchläufen daraufhin testen, ob die Kriterien auch geeignet sind. Das kann und wird bei der Analyse schließlich in einen spiralartigen Prozess münden: Auswertung, Kriterienkorrektur, Nachauswertung usw.

- Nimm nicht nur die passende Anzahl, sondern auch die passende Art von Belegen.

Beispielsweise wäre es unsinnig, Belege aus Romanen auszuwerten, wenn man den Gegenstand der Untersuchung als Zeitungsdeutsch bestimmt hat (was immer das dann genau sein mag).

Für die Datenauswertung eines Belegkorpus kann man bspw. eine Excel-Tabelle anlegen. Man geht dann Beleg für Beleg nach den Kriterien durch und füllt die einzelnen Zellen mit kurzen und für jeweils gleiche Antworten auch mit exakt den gleichen Zeichen aus. Meine Auswertungstabelle sieht im Ausschnitt z. B. so aus:

Nr	Jahr	Bereich	Autor	Verb	Stellung	Präp	Kasus	Wortart	Verbeil	syn Fkt. Gr
1	1999	Belletristik	Degenhardt	gehen	vor	nein	Gen/Dat	PräpGENDAT	nein	Adverbial
2	1999	Belletristik	Degenhardt	ohne	vor	nein	Akk	PräpAKK	nein	Adverbial
3	1999	Belletristik	Degenhardt	wandern	nach	nein	Akk	PostAKK	möglich	Adverbial
4	1999	Belletristik	Degenhardt	laufen	nach	nein	Akk	PostAKK	möglich	Adverbial
5	1999	Belletristik	Dückers	stolzieren	nach	nein	Akk	PostAKK	möglich	Adverbial
6	1999	Belletristik	Dückers	laufen	nach	an	Dat	Adv	möglich	Adverbial
7	1999	Belletristik	Dückers	fahren	nach	nein	Akk	PostAKK	möglich	Adverbial
8	1999	Belletristik	Dückers	schieben	nach	nein	Akk	PostAKK	möglich	Adverbial
9a	1999	Belletristik	Dückers	laufen	nach	nein	Akk	PostAKK	möglich	Adverbial
9b	1999	Belletristik	Dückers	rennen	nach	nein	Akk	PostAKK	möglich	Adverbial
10	1999	Belletristik	Dückers	fahren	nach	evtl. über	Akk	PostAKK	nein	Adverbial
11	1999	Belletristik	Dückers	schweifen	vor	an	Dat	Adv	möglich	Adverbial
12	1999	Belletristik	Dückers	gehen	nach	nein	Akk	PostAKK	möglich	Adverbial
13	1999	Belletristik	Dückers	sich schieben	nach	an	Dat	Adv	möglich	Adverbial
14	1999	Belletristik	Dückers	laufen	nach	an	Dat	Adv	möglich	Adverbial
15	1999	Belletristik	Dückers	laufen	nach	nein	Akk	Adv	möglich	Adverbial
16	1999	Belletristik	Dückers	laufen	nach	an	Dat	Adv	möglich	Adverbial
17	1999	Belletristik	Dückers	wandern	nach	evtl. über	Akk	Adv	nein	Adverbial
18	1999	Belletristik	Dückers	fahren	nach	nein	Akk	PostAKK	möglich	Adverbial
19	1999	Belletristik	Dückers	laufen	nach	an	Dat	Adv	möglich	Adverbial
20	1999	Belletristik	Dückers	fahren	nach	nein	Akk	PostAKK	möglich	Adverbial

Abbildung 12: Datentabelle – Auswertung der ersten 20 *entlang*-Belege nach einzelnen Kriterien

Zu den einzelnen Spalten (von links nach rechts nummeriert):

1. Nr: fortlaufende Nummer des Belegs
2. Jahr: Jahr, aus dem der Beleg stammt
3. Bereich: Bereich, aus dem der Beleg stammt
4. Autor: Autor (oder Zeitung etc.) des Textes, aus dem der Beleg stammt

Dies sind äußere Datenbereiche, die eher die Zusammensetzung des Korpus betreffen, aber wichtig werden können, wenn man Ergebnisse in Beziehung zueinander setzt (s. u.).

5. Verb: Verb des Satzes, in dem *entlang* vorkommt

Mit diesem Kriterium soll auswertbar werden, ob bestimmte einzelne Verben oder Verbgruppen besonders häufig mit *entlang* vorkommen. Hier werden dann auch Kandidaten für komplexe Verben mit *entlang*- ermittelt.

6. Stellung: Stellung von *entlang* in Relation zu einer NGr (= Nominalgruppe), auf die sich *entlang* bezieht



Hier hat man es bereits mit einer interpretierenden Auswertung zu tun, denn worauf sich *entlang* bezieht, kann nicht ohne Weiteres einfach durch formale (und damit von Maschinen anwendbare) Kriterien entschieden werden. Sie als Analysierende stellen einen Bezug zu einer NGr (oder einem Pronomen) her.

7. Präp: Präposition, die sich auf dieselbe NGr bezieht wie *entlang*

Dass diese Frage überhaupt als Kriterium auftaucht, ist einer ersten Sicht auf die Belege geschuldet. Denn das Vorkommen von Präpositionen bei NGr mit *entlang* spielte bei den oben formulierten Fragen keine Rolle. Hier haben wir es also mit einem eher induktiven Erkenntnisfortschritt zu tun, der – wie wir weiter unten noch sehen werden – Relevantes aufdeckt, nach dem zunächst nicht gefragt wurde.

8. Kasus: Kasus der NGr, auf die sich *entlang* bezieht

Natürlich ist es so, dass bei Vorliegen einer Präposition sich der Kasus nach dieser Präposition richtet. Interessant wird es also in den Fällen, da keine Präposition vorhanden ist. Und dann kann man beispielsweise bei der Auswertung der Datentabelle die Frage stellen, ob Stellung und Kasus zusammenhängen. Das entspricht ja gerade der oben skizzierten Fragestellung.

9. Wortart: Wortart von *entlang*

Hier haben wir es wieder mit einem interpretativen Auswertungsschritt bei der Füllung der Datentabelle zu tun. Die Wortart-Entscheidung basierte dabei auf den folgenden Prinzipien: Wenn bereits eine Präposition vorhanden ist, wird *entlang* nicht auch noch als Präposition angesehen. Höchstens – und das wird in der Tat wichtig werden – könnte man von einer komplexen Zirkumposition sprechen, die aus einer Präposition vor der NGr und *entlang* nach der NGr besteht. In allen anderen Fällen gilt: Wenn *entlang* vor der NGr steht, wird es als Präposition angesehen, wenn *entlang* danach steht als Postposition. Mitangegeben ist jeweils der Kasus der NGr. Eigentlich eine redundante Information, weil dies ja schon in Spalte 8 erhoben wurde, für eine weniger umständliche Formulierung bestimmter Auswertungsformeln in Excel aber von Vorteil.

10. Verbsart: Möglichkeit, ob *entlang* ein Verbbestandteil sein könnte

Auch dies ist natürlich wieder klar interpretativ auswertend. Da sich die Auswertung hier auf den Kern der obigen Fragestellung bezieht, wird diese Kategorie im ganzen weiter unten erläutert.

11. syn Fkt. Gr: syntaktische Funktion der NGr, auf die sich *entlang* bezieht

Hier gibt es nur drei Möglichkeiten: Adverbial (entweder als freie Angabe neben dem oder als geforderte Ergänzung zum Verb), Attribut (als nähere Bestimmung eines Nomens) oder unklar (ob Adverbial oder Attribut).

Mithilfe der nach diesen Kriterien ausgefüllten Datentabelle können nun die Einzelauswertungen vorgenommen werden.

3.4. Das Belegkorpus auswerten II: Auswertung der Datentabelle

Für die zählende Auswertung der Excel-Tabelle können verschiedene Excel-Funktionen genutzt werden. Bei größeren Datenmengen wäre auch daran zu denken, statistisch auszuwerten. Das Tabellenprogramm stellt dafür jedenfalls die wichtigsten Mittel bereit. Für die Auswertung des 100 Belege umfassenden Belegkorpus zu *entlang* reichen aber wenige, nicht-statistische Funktionen aus. Zumeist handelt es sich um Zählen, In-Beziehung-Setzen und Ordnen. Einen ersten Eindruck, welche Kriterien zusammenhangstiftend sind, kann man sich

über die Filterfunktion schaffen. Die Zeichen bei den Überschriftzellen der Tabelle zeigen an, dass die Filterfunktion eingestellt ist. Durch Klicken auf dieses Zeichen sieht man in einem Scroll-Fenster, welche verschiedenen Eingaben es in den Zellen dieser Spalte gibt, und man kann die Tabelle etwa auf bestimmte dieser Ausfüllungen reduzieren, indem man jeweils ein Häkchen in dem Scroll-Fenster setzt. Für das Zählen stehen die Funktionen Zählen-wenn (bei einem Kriterium) und Zählen-wenn-s (bei mehr als einem Kriterium) zur Verfügung. Die Benutzung von Excel zur zählenden Auswertung wird hier jetzt nicht weiter besprochen. Letztlich könnte man diese Zählerei ja auch per Hand vornehmen, was aber natürlich immer etwas stärker fehleranfällig ist.

Ausgewertet wurden Belege aus den Jahren 1999 (60), 1998 (32) und 1997 (8). Bereits bei der Auswertung der weiteren äußeren Korpusdaten zeigen sich korpusrelevante Charakteristiken.

Quelle/Art	Belletristik	Fachbuch	Zeitung	gesamt
ARCHIV	-	-	2	2
Beck	-	4	-	4
Degenhardt	4	-	-	4
Dückers	27	-	-	27
Engler	-	4	-	4
Fath	-	1	-	1
Hannover	-	2	-	2
Hars	-	1	-	1
Kurz	-	2	-	2
Lebert	3	-	-	3
Luhmann	-	3	-	3
Moers	7	-	-	7
Schwanitz	-	3	-	3
Walser	20	-	-	20
ZEIT	-	-	17	17
gesamt	61	20	19	100

Abbildung 13: Auswertung der äußeren Korpusdaten

Obwohl das Kernkorpus laut DWDS ausgewogen sein soll, überwiegen die in belletristischer Literatur erzielten Treffer (61% – nebenbei bemerkt: bei einem 100'er Belegkorpus können alle Zahlen, die sich auf das gesamte Belegkorpus beziehen, als Prozent verstanden werden). Das ist vielleicht kein Zufall. Das Kernkorpus selbst ist tatsächlich halbwegs ausgewogen. Eine Überprüfung mittels der Suche nach der exakten Wortform *die* (Suchkette: @die) macht dies deutlich. Die Wortform *die* ist nach Zählungen die am häufigsten vorkommende Wortform im Deutschen.¹³ Es ist zudem unwahrscheinlich, dass *die* in irgendeinem Bereich (Belletristik, Zeitung etc.) häufiger vorkommt als in einem anderen. Also müsste *die* bei einem ausgewogenen Korpus in etwa gleich häufig in den verschiedenen Bereichen vorkommen. Das ist auch der Fall. Wenn man die Ansicht

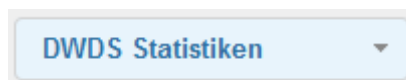


Abbildung 14: Gewählte Ansicht (wählbar in dem Fenster rechts neben dem Suchfeld)

¹³ Vgl. bereits Meier (1964 Bd. II: 112). Auch nach Zählungen innerhalb des Wortschatzportals Leipzig ist der Artikel *der* (gemeint ist der bestimmte Artikel insgesamt, nicht nur die Wortform *die*) das häufigste Wort (vgl. http://wortschatz.uni-leipzig.de/cgi-portal/de/wort_www?site=208&Wort_id=5407). In Quasthoff/Fiedler/Hallsteinsdöttir (2011: 25) sind als die drei Wortformen mit der höchsten Frequenz *der*, *die* sowie *und* angeführt.

einstellt, so kann man sich die Gesamtanzahl der Treffer pro Bereich und Jahr (den dort sog. *Wortverlauf*) anzeigen lassen:

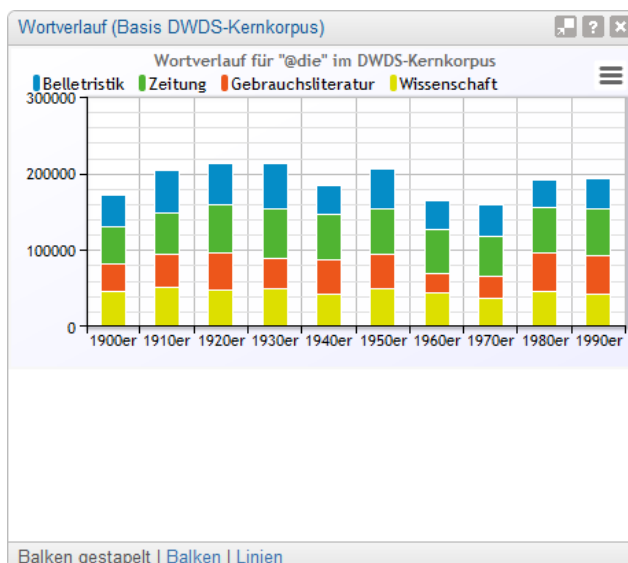


Abbildung 15: Wortverlauf von @die im DWDS-Kernkorpus (Statistik-Ansicht)¹⁴

Das Korpus dürfte also in der Tat etwa ausgewogen sein. Bei *entlang* ist dies etwas anders:

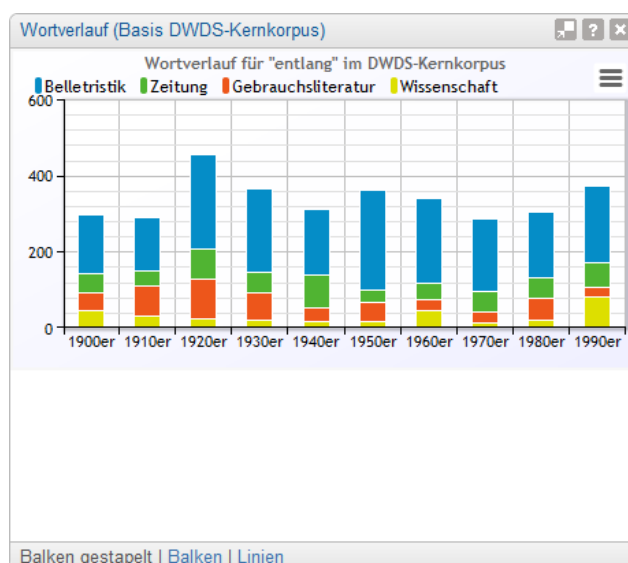


Abbildung 16: Wortverlauf von entlang im DWDS-Kernkorpus (Statistik-Ansicht)

Hier überwiegen eindeutig die belletristischen Treffer (jeweils der obere bzw. blaue Bereich der Balken). Die sich bei der Auswertung der Datentabelle zeigende Tendenz lässt sich auch hier feststellen. Ob das schon hinreicht, *entlang* als eine der Tendenz nach literatursprachliche Einheit anzusehen, ist nicht klar. Aber als eine Interpretation der Daten könnte eine solche

¹⁴ Die Unterscheidung zwischen Belletristik und Gebrauchsliteratur scheint mir dabei ein wenig missverständlich zu sein. Welche Literatur wird wozu gebraucht? Kann nicht auch Belletristik gebraucht werden und Gebrauchsliteratur belletristisch sein? Entweder zielt diese Unterscheidung auf schöngestige vs. Trivial-Literatur ab, oder aber als Gebrauchsliteratur werden nicht-literarische Gebrauchstexte (Bedienungsanleitungen, Kochrezepte usw.) verstanden. In meiner Auswertung habe ich literarische Texte jedenfalls zusammen als Belletristik geführt (man hätte es stattdessen natürlich auch einfach Literatur nennen können). Dagegen abzusetzen sind dann wissenschaftliche (etwa philosophische, sozial- oder geschichtswissenschaftliche) Fachbücher, die auch beim DWDS unter der Bezeichnung *Wissenschaft* geführt werden. Nicht-literarische Gebrauchstexte sind im Belegkorpus nicht vertreten.



These durchaus aufgestellt werden. Sie müsste dann durch zusätzliche (vielleicht ja auch experimentelle) Studien verifiziert (= ‚in ihrem Wahrheitsgehalt bestätigt‘) oder falsifiziert (= ‚entkräftet, widerlegt‘) werden.

Darüber hinaus sind mit der Auswertungstabelle nach Bereich und Autoren zugleich auch Vergleichsmöglichkeiten eröffnet. Es könnten so z. B. die hauptsächlichen Einzelvorkommen, nämlich die zwei belletristischen Vorkommen untereinander (Autoren: Dückers und Walser) und diese mit den Vorkommen in der ZEIT, verglichen werden. Vielleicht benutzt Walser *entlang* ja anders als Dückers oder/und die Autoren anders als die ZEIT-Redakteure.

Nun zur inhaltlich orientierten Auswertung mit Blick auf die obigen Fragestellungen. Die Frage nach der Stellung von *entlang* ist im Zusammenhang mit dem Vorkommen von Präpositionen zu sehen:

Stellung	zusammen mit Präpositionen						gesamt
	<i>an</i>	<i>am</i>	<i>auf</i>	evtl. <i>über</i>	<i>hinter</i>	nein	
vor	1	-	-	-	-	22	23
nach	25	8	1	2	1	39	76
Distanz nach	1	-	-	-	-	-	1
gesamt	27	8	1	2	1	61	100

Abbildung 17: Vorkommen von *entlang* hinsichtlich Stellung und Mitvorkommen von Präpositionen

Wenn *entlang* vor der NGr (Vorstellung) steht, scheint es nicht mit Präpositionen kombinierbar zu sein, möglicherweise aber mit *an* (und dann möglicherweise auch *am*). Das müsste in Nacherhebungen gezielt überprüft werden. Eine erste Suche mit Zeichenabstand nach

`near(entlang, an, 0)`

(= ‚suche alle Vorkommen von den direkt aufeinander folgenden Einheiten *entlang* und *an* (allerdings egal, in welcher Reihenfolge)‘) bzw. phrasal (hier in genau der Reihenfolge)

`"entlang an"`

ergibt in der Tat im Kernkorpus 15 Treffer (bzw. 14 bei der phrasalen Suche):

Kernkorpus 20

Treffer: 15, davon anzeigbar: 12 Filter

KWiC	Datum ↓	Datum ↑	Zufällig	Links	Rechts
1	1999				
2	1986				
3	1952				
4	1950				
5	1942				
6	1927				
7	1926				
8	1925				
9	1920				
10	1913				
11	1911				
12	1905				

Version: 1.1 Optionen

Abbildung 18: KWIC-Ansicht der Treffer bei der Suche von *near(entlang, an, 0)*

Zwar erzielt auch die analog formulierte Suche mit *auf* 22 Treffer. Aber das sind allesamt Treffer, bei denen *entlang* nach der Bezugs-NGr (im Beispiel unten: *den Gang entlang*) steht und *auf* zu Beginn einer neuen (im Beispiel unten: *auf ihr Zimmer zehn*) und beides Adverbial (im Beispiel unten: zu *zustöckelte*) ist:

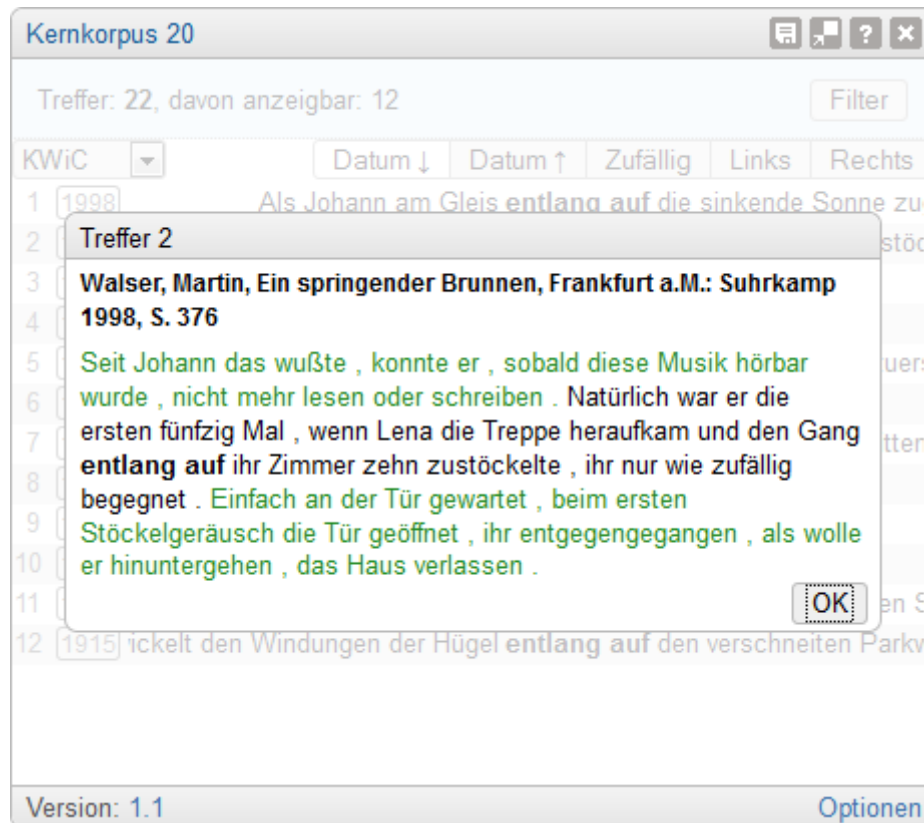


Abbildung 19: Beispiel(fehl)beleg für die Ergebnisse der Suche nach *near(entlang, auf, 0)*

Man sieht: Eine rein automatische Suche nach dem Zusammenvorkommen (der Kookkurrenz) von *entlang* und *an* sowie *entlang* und *auf* würde die syntaktische Unterschiedlichkeit dieser Vorkommen nivellieren. Es bedarf sprachkompetenter Analysierender, um die Unterschiede zu erkennen.

Was die Voranstellung von *entlang* anbetrifft, so geben die erhobenen und ausgewerteten Daten Hinweise auf ein relevantes Vorkommen von *entlang*, das als eine Konstruktion vorläufig so notiert werden kann:

- Konstruktion I: [*entlang* + NGr]

Wenn *entlang* nach der NGr steht (Nachstellung), tritt in 39 (von 76 Fällen insgesamt) keine Präposition auf, in 37 Fällen aber schon. Hier haben wir also ein 50:50-Vorkommen. Es überwiegen aber wiederum *an* und *am*, sodass man hier zwei weitere Verwendungen (oder: Grundtypen, Muster, Konstruktionen) festhalten kann:

- Konstruktion II: [*an/am* + NGr + *entlang*]
- Konstruktion III: [NGr + *entlang*]

Ein erster Zugriff hat also drei verschiedene, gut belegte Vorkommensmöglichkeiten zum Ergebnis. Im nächsten Schritt kann noch eine genauere Beschreibung dieser Typen versucht werden. Dazu schauen wir uns die erhobenen Kasus an. Interessant sind diese allerdings nur in den Fällen, da keine Präposition vorliegt. Die *an/am*-Fälle haben, das sei hier festgestellt, allesamt eine Dativ-NGr im Gefolge. Wie also steht es mit den Kasus, wenn keine Präposition vorliegt? Hier die Tabelle:



	GEN	GEN/DAT	DAT	AKK	ohne	gesamt
vor	1	17	3	1	-	22
nach	-	-	-	37	2	39
Distanz nach	-	-	-	-	-	0
gesamt	1	17	3	38	2	61

Abbildung 20: Vorkommen von *entlang* hinsichtlich der Kasus der NGr ohne Präposition

Als ein klares Ergebnis kann festgehalten werden: Akkusativ bei Nachstellung. Bei der Voranstellung ist die Sachlage nicht so klar. Das hängt zum einen damit zusammen, dass bei femininen Substantiven im Singular nicht zwischen Genitiv und Dativ unterschieden werden kann. Zum Beispiel in den folgenden beiden Fällen, also in Beleg (57), in dem die NGr mit *entlang* übrigens sehr wahrscheinlich ein Attribut zu *Ziele* ist (*Ziele entlang der ...Küste*), und Beleg (85) mit der Adverbial-Verwendung:

57.) 1999 Archiv der Gegenwart, 69, 1999

Die NATO, die seit dem 28. Mai Ziele in Jugoslawien auch von Ungarn aus bombardierte, reagierte zurückhaltend auf das vermeintliche Einlenken Belgrads und griff Ziele **entlang** der montenegrinischen Adria-Küste an.¹⁵

85.) 1998 Die Zeit 03.09.1998, 03.09.1998

Majok, dessen Frau an Malaria starb, floh mit seinem zwei Monate alten Sohn Magang vor der Armee und den Reiterhorden aus dem Nordsudan, die **entlang** der Bahnstrecke nach Wau regelrechte Menschenjagden veranstalten – verschleppte Frauen und Kinder werden in Khartum als Sklaven verkauft.

In beiden Fällen könnte es sich um Genitiv oder Dativ handeln, denn die Formen fallen ja zusammen. Die klaren Fälle (also maskulin oder neutrum oder feminin Plural) sind in dem Belegkorpus allerdings viel seltener als die unklaren. Man müsste hier nachjustieren. Über die Suchen *near(entlang, @des, 0)* und *near(entlang, @dem, 0)* kann man einen ersten Eindruck gewinnen (allerdings eben nur für NGr mit bestimmtem Artikel und maskulinem oder neutralem Nomen im Singular):

- *near(entlang, @des, 0)*
also Genitiv: 75
- *near(entlang, @dem, 0)*
also Dativ: 48

Auch die Nachjustierung ergibt also kein klares Bild. Vielleicht aber – und das ist das Ergebnis dieser Suchen – ist es ja genau so, dass die NGr bei vorangestelltem *entlang* und ohne weitere Präposition einfach im Genitiv oder im Dativ steht. Bei Nachfolgestudien müsste man hier schauen, ob es einen Unterschied macht und/oder ob es die gleichen Fälle sind, in denen Dativ oder Genitiv verwendet wird usw. Hierzu können wir zunächst einmal noch nichts sagen.

Wir sind nun in der Lage die oben herausgearbeiteten Verwendungen oder Konstruktionen hinsichtlich der Kasus zu konkretisieren:

¹⁵ Die Beispiele aus dem Belegkorpus werden hier so in den Text eingefügt, wie sie im heruntergeladenen Belegkorpus stehen. Das betrifft auch die Zählung. Lediglich die Schriftgröße und der Rand werden dem Lehrmanuskript angepasst und Leerzeilen oder gehäufte Leerzeichen getilgt. Falls die Sätze davor und danach sehr lang und nicht unbedingt zum Verständnis notwendig sein sollten, werden diese getilgt. Die Einfügung des Belegs beschränkt sich in solch einem Fall auf den Satz, in dem *entlang* enthalten ist.



- Konstruktion I: [*entlang* + NGr_{Genitiv/Dativ}]
- Konstruktion II: [*an/am* + NGr_{Dativ} + *entlang*]
- Konstruktion III: [NGr_{Akkusativ} + *entlang*]

Mit diesen verschiedenen Verwendungen scheinen weitere Auffälligkeiten verbunden zu sein. So sind die nachgestellten Konstruktionen II und III knapp sechsmal häufiger Adverbial (75 Belege) als Attribut (13), die Konstruktion I jedoch ist beinahe ausschließlich Attribut (8 Belege) und nicht Adverbial (1 Beleg) – um diesen Zusammenhang quantitativ untermauern zu können, müssten jedoch aussagekräftigere Zahlen durch Nacherhebungen beschafft werden.

Hier ist schließlich auch ein erster Rückbezug auf die obige Fragestellung möglich. Was den Zweifelsfall-Duden angeht, so sind die Beschreibungen bis auf eine durch diese Untersuchung bestätigt worden. Die eine nicht bestätigte ist, dass bei Voranstellung seltener Dativ anzutreffen wäre. In dem Belegkorpus ist Dativ häufiger (die Zahlen da sind aber nicht aussagekräftig) und eine kursorische Nachüberprüfung ergibt, dass der Genitiv zwar etwas häufiger zu sein scheint, der Dativ aber keinesfalls selten oder viel seltener auftritt, sodass man davon ausgehen muss, dass beides jedenfalls gebräuchlich ist.

Allerdings – und dies ist gegenüber dem ersten Teil des Eintrags im Zweifelsfall-Duden eine neue Erkenntnis – tritt *entlang* in Nachstellung gern mit vorangestelltem *an/am* auf. Beispiele dafür finden sich tatsächlich im zweiten Teil des Eintrags, der aber die Frage von Getrennt- oder Zusammen-Schreibung behandelt (ohne allerdings dieses auffällige Zusammenvorkommen explizit zu behandeln). Folgend wird dieser zweite große Punkt aufgegriffen – der Zusammenhang von *entlang* mit Verben.

An Verben tauchen insgesamt auf:

Verben	Anzahl
<i>gehen</i>	13
<i>laufen</i>	12
ohne ein Verb in demselben Satz	9
<i>fahren</i>	8
<i>rennen</i>	4
<i>führen, sein</i>	(2x) 3
<i>marschieren, schieben, schlendern, schlurfen, sich schlängeln, sich ziehen, wandern</i>	(7x) 2
<i>anbringen, angreifen, beobachten, bestätigen, durchdenken, eilen, falten, finden, heraufkommen, hinabgehen, identifizieren, jagen, kommen, liegen, nehmen, radeln, reorganisieren, ritzen, schleppen, schweifen, sich drücken, sich schieben, sich verhalten, stolzieren, streichen, tasten, tingeln, veranstalten, verfügen, verlaufen, vorantreiben, zugehen, zusammenziehen, zustöckeln</i>	(34x) 1
<i>gesamt</i>	100

Abbildung 21: Anzahl des Vorkommens der verschiedenen Verben in den *entlang*-Sätzen

Auffällig ist zunächst, dass die mehrmals vorkommenden Verben bis auf *sein* semantisch allesamt Verben der Bewegung und morphologisch allesamt unpräfigiert sind.



Für die Frage, mit welchem Verb *entlang*- eine Einheit bildet, kann es als ein Indiz angesehen werden – und insofern spielt die Schreibung tatsächlich eine Rolle –, wenn Zusammenschreibungen auftreten. Dies kann man für jedes Verb einzeln prüfen, indem jeweils einfach der zusammengeschrriebene Infinitiv als Suchkette eingegeben wird. Das Ergebnis lässt sich in Form einer recht langen Tabelle darstellen. In der ersten Spalte befindet sich jeweils das geprüfte Verb, in der zweiten die Anzahl im Belegkorpus, in der dritten die Information darüber, ob das DWDS in seinem Wörterbuch einen Eintrag zu dem zusammengeschrriebenen Verb hat, in der vierten die Anzahl der zusammengeschrriebenen Treffer im DWDS-Kernkorpus und in der fünften die Anzahl der zusammengeschrriebenen Treffer im DWDS-ZEIT-Korpus (das größer ist als das Kernkorpus und auch bis zur Gegenwart heranreicht). Grau eingefärbt ist oben der Teil der Tabelle mit Treffern im Kernkorpus oder in dem ZEIT-Korpus, darunter potenzielle weitere Verben der Bewegung:

Verben	Anzahl	DWDS		
		Eintrag	Anzahl in Kern	ZEIT
<i>führen</i>	3	ja	49	332
<i>fahren</i>	8	ja	42	243
<i>gehen</i>	13	ja	72	177
<i>laufen</i>	12	nein	52	148
<i>sich ziehen</i>	2	nein	23	67
<i>wandern</i>	2	nein	6	51
<i>schlendern</i>	2	nein	6	44
<i>tasten</i>	1	ja	6	26
<i>sich schlängeln</i>	2	nein	0	19
<i>streichen</i>	1	ja	16	17
<i>kommen</i>	1	ja	11	17
<i>marschieren</i>	2	nein	2	14
<i>schieben</i>	2	ja	3	12
<i>sich schieben</i>	1	nein	2	9
<i>radeln</i>	1	nein	1	7
<i>sich drücken</i>	1	nein	0	7
<i>schlurfen</i>	2	ja	1	4
<i>eilen</i>	1	nein	0	4
<i>stolzieren</i>	1	nein	0	3
<i>jagen</i>	1	nein	1	2
<i>schweifen</i>	1	nein	1	2
<i>rennen</i>	4	nein	1	1
<i>schleppen</i>	1	nein	1	0
<i>sein</i>	3	nein	0	0
<i>anbringen</i>	1	nein	0	0
<i>angreifen</i>	1	nein	0	0
<i>beobachten</i>	1	nein	0	0
<i>bestätigen</i>	1	nein	0	0
<i>durchdenken</i>	1	nein	0	0
<i>falten</i>	1	nein	0	0



<i>finden</i>	1	nein	0	0
<i>heraufkommen</i>	1	nein	0	0
<i>hinabgehen</i>	1	nein	0	0
<i>identifizieren</i>	1	nein	0	0
<i>liegen</i>	1	nein	0	0
<i>nehmen</i>	1	nein	0	0
<i>reorganisieren</i>	1	nein	0	0
<i>ritzen</i>	1	nein	0	0
<i>sich verhalten</i>	1	nein	0	0
<i>tingeln</i>	1	nein	0	0
<i>veranstalten</i>	1	nein	0	0
<i>verfügen</i>	1	nein	0	0
<i>verlaufen</i>	1	nein	0	0
<i>vorantreiben</i>	1	nein	0	0
<i>zusammenziehen</i>	1	nein	0	0
<i>zustöckeln</i>	1	nein	0	0
ohne	9			
<i>zugehen</i>	1	Interferenz mit <i>gehen</i> : <i>entlang</i> + <i>zu</i> + <i>gehen</i>		
gesamt	100			

Abbildung 22: Zusammenhang von *entlang* und möglichen verbalen Wortbildungen

Das sog. Zipf'sche Gesetz¹⁶ ist auch hier wieder bestätigt: Vier Spitzenreiter haben sich abgesetzt (*führen, fahren, gehen, laufen*), es folgen ein größeres Mittelfeld und ein größeres Schlussfeld. Nicht für alle vier Spitzenreiter-Verben hat das DWDS dabei auch einen eigenen Eintrag: *entlanglaufen* fehlt. Hier müsste man dem DWDS einen entsprechenden Eintrag vorschlagen, zumal die Anzahl im Kernkorpus höher ist als die zu *entlangführen* und *entlangfahren*, die aber das ZEIT-Korpus anführen.

Bei den Verben, bei denen eine Zusammenschreibung mit *entlang-* durch diese Suche belegt ist, gibt es grundsätzlich die Möglichkeit, dass es sich um komplexe Verben mit *entlang-* handelt. Dies kann schließlich als vierte Verwendungsmöglichkeit festgehalten werden:

- Konstruktion IV: [*entlang*VERB_{Bewegung} + (NGr_{Akk} oder *an/am*+NGr_{Dat})]

Weil es in vielen Fällen nicht ohne Weiteres entscheidbar ist, ob ein komplexes Verb vorliegt oder nicht, kann auch zur Häufigkeit der vier Verwendungen keine exakte Angabe gemacht werden. Und wegen der Überschneidungsmöglichkeiten der Verwendungen ist die Gesamtmenge beim Vorkommen der vier Konstruktionen in dem Belegkorpus auch höher als die Menge an untersuchten Belegen überhaupt, nämlich 146 (bei 100 Belegen):

Konstruktionen/Verwendungen	Anzahl
IV: [<i>entlang</i> VERB _{Bewegung} + (NGr _{Akk} oder <i>an/am</i> +NGr _{Dat})]	55
III: [NGr _{Akkusativ} + <i>entlang</i>]	37
II: [<i>an/am</i> + NGr _{Dativ} + <i>entlang</i>]	33
I: [<i>entlang</i> + NGr _{Genitiv/Dativ}]	21
gesamt	146

Abbildung 23: Vorkommenshäufigkeit der einzelnen *entlang*-Verwendungen (mit Überschneidungen)

Die Konstruktionen I, II und III ergeben zusammen 91% aller Fälle. Nicht erfasst sind damit die folgenden neun Vorkommen, die sich aber – bis auf die letzte – den einzelnen Konstruktionen noch anlehnen lassen:

¹⁶ Vgl. Perkuhn/Keibel/Kupietz (2012: 84).



- 2x *Hier entlang!*

Diese Konstruktion wurde bei der Auswertung nicht erfasst, weil *Hier* als Adverb in keinem Kasus stehen kann. Macht man sich klar, dass es sich dabei aber um ein Proadverb handelt, an dessen Stelle ein anderer Ausdruck (etwa eine NGr im Akkusativ oder eine NGr mit der Präposition *an/am*) steht, wird klar, dass dieses Vorkommen zur Konstruktion II oder III gezählt werden könnte.

- Belege mit anderen Präpositionen: 2x *über*, 1x *auf*, 1x *hinter*

Diese Belege müssten einzeln durchgesehen werden. Zum Teil liegen hier Koordinierungen vor, bei denen bspw. eine NGr mit *über* koordiniert wird mit Konstruktion III, wobei es aber möglich wäre, dass bei letzterer die bei Richtungsangaben ebenfalls den Akkusativ verlangende Präposition *über* nicht noch einmal genannt wird (siehe weiter unten den Beleg 10).

- 1x die Reihenfolge *entlang* + *an*

Dies wäre eine Variante zu Konstruktion II. Wie oben angedeutet, lassen sich für diese Konstruktion durch separate Suchen jedenfalls Belege finden.

- 1x Distanzstellung von *an* + NGr und *entlang* (Beleg 43: *An seinen Wänden zogen sich schleimige Kabel entlang.*)

Diese Verwendung ließe sich der Konstruktion II oder IV zuschlagen. Es müsste geklärt werden, ob [*an* + NGr + *entlang*] eine Konstruktion ist, die ansonsten auch in Distanzstellung auftritt.

- 1x Voranstellung mit folgender NGr im Akkusativ (Beleg: *entlang den verspiegelten Wänden die langen Eßtische* (verbloses Vorkommen innerhalb einer Aufzählung bei der Beschreibung eines Raumes))

Davon unabhängig wurden die Fälle möglicher Verbwortbildung mit *entlang-* gezählt. Die Frage, wann dies möglich ist und wann nicht, ist allerdings schwer entscheidbar und richtet sich nach verschiedenen Faktoren. Bei der Analyse der Belege zeigen sich hierbei Phänomene, die im Einzelfall mehr oder weniger stark gegen die Wortbildungshypothese sprechen. Ein paar davon sollen folgend andiskutiert werden:

- Attribuierung

1.) 1999 Degenhardt, Franz Josef, Für ewig und drei Tage, Berlin: Aufbau-Verl. 1999, S. 69

Sie waren den Weg **entlang** der Blanken Becke zwischen Knüppelweiden zum Kranichteich gegangen.

Hier ist die *entlang*-NGr Attribut zu *Weg*. Eine getrennte Umstellung wäre wohl als unkorrekt anzusehen, z. B.: **Den Weg waren sie entlang der Blanken Becke zwischen Knüppelweiden zum Kranichteich gegangen* oder **Sie waren entlang der Blanken Becke zwischen Knüppelweiden zum Kranichteich den Weg gegangen* und weitere mehr. Und zusammen dürften die Einheiten auch nicht auftreten können: **Sie waren den Weg der Blanken Becke zwischen Knüppelweiden zum Kranichteich entlanggegangen*. Damit ist klar: Überschneidungsfälle liegen immer nur in den Fällen vor, da die *entlang*-NGr nicht Attribut sondern Adverbial ist (oder zumindest sein kann).

- Reihenfolge

65.) 1998 Walser, Martin, Ein springender Brunnen, Frankfurt a.M.: Suhrkamp 1998, S. 161

Ich bin heute am See **entlang** gegangen.



67.) 1998 Walser, Martin, Ein springender Brunnen, Frankfurt a.M.: Suhrkamp 1998, S. 222

Vorne, wo der Seeweg endete, wo man an Hoppe-Seylers Park **entlang** zur Dorfstraße ging, sammelten sich alle.

87.) 1998 Die Zeit 30.07.1998, 30.07.1998

An der Scheunenwand **entlang** führt die Allee der Kugelahorne, dazwischen Eibenbüsche.

Wenn die NGr mit *entlang* vor dem Verb steht, ist es möglicherweise unwahrscheinlich, dass *entlang* zu dem Verb als Bestandteil gehört. Syntaktisch ist dies jedoch nicht ganz klar. Die grundlegende Frage wäre: Inwiefern unterscheiden sich die Bedeutungen von *gehen* + *entlang* in einem Syntagma und *entlanggehen* als einzelner Verb? Wenn z. B. im Beleg 65 kein Fall von *entlanggehen* vorliegt, so wird hier *gegangen* und dieses Gehen geschieht *am See entlang*. Möglicherweise wird dadurch die Verb­­tätigkeit als ungerichtete hervorgehoben und die Gerichtetheit kommt erst durch die *entlang*-Konstruktion ins Spiel. Es handelt sich um eine Frage, die durch weitere Studien und Analysen angegangen werden müsste.

- Reihung/Koordination

10.) 1999 Dückers, Tanja, Spielzone, Berlin: Aufbau-Verl. 1999, S. 29

Jetzt ist seine Zunge zwischen ihren Beinen, es sieht aus, als würde er sie beißen, aber es scheint ihr nicht weh zu tun. Nun legt die Frau sich auf den Typ, etwas Schaumwasser spritzt auf, sie fährt mit ihrer Zunge über seine linke, seine rechte Lende, seinen Schwanz **entlang**, herauf, herunter.

24.) 1999 Dückers, Tanja, Spielzone, Berlin: Aufbau-Verl. 1999, S. 165

Erst fährt er langsam, dann immer schneller, je weiter er sich von der Charité entfernt; die Luisenstraße **entlang**, den Pariser Platz, dann durchs Brandenburger Tor, weiter in die Straße des 17. Juni, bis in den Tiergarten.

88.) 1998 Die Zeit 10.06.1998, 10.06.1998

Ein schmaler Weg führt nach Westen, an der Küste Fehmarns **entlang**, immer das Meer zur Linken.

Würde man im Beleg 10 annehmen, dass hier nicht nur *fahren* (*über*), sondern darüber hinaus noch die komplexen Verben *entlangfahren*, *herauffahren* und *herunterfahren* vorliegen, so hätte man es mit vier verschiedenen Verben zu tun. Plausibler scheint mir aber zu sein, dass es sich in allen vier Fällen um dasselbe ‚mit der Zunge‘ ausgeführte ‚Fahren‘ als Bedeutung von *fahren* handelt, nur eben an verschiedenen Stellen und in verschiedene Richtungen. Zudem scheinen mir die letzten beiden anzunehmenden Verben *herauffahren* und *herunterfahren* lexikalisch inadäquat zu sein (was aber auch an der Metaphorik von *fährt mit ihrer Zunge* liegen kann). Bei Beleg 24 ist die Situation ähnlich. Hier hätten wir einerseits *entlangfahren* und andererseits *fahren durch*, *in* und *bis in* für dieselbe Tätigkeit. Ein wirklich hartes, zwingendes syntaktisches Entscheidungskriterium scheint es aber nicht (oder nicht so einfach) zu geben. Es bleibt bei Reihungen dieser Art die Möglichkeit bestehen, elliptische Sätze mit verschiedenen Verben anzunehmen.

Damit wird ein zweiter Rückbezug zur obigen Fragestellung möglich. Unabhängig von der Frage, ob *entlang* selbst als Präposition (also als kasusforderndes Fügteil vor einer NGr), Postposition (kasusforderndes Fügteil nach einer NGr) oder als Adverb (in Nachstellung zur NGr, die durch eine Präposition gefügt wird) erscheint, kann es sich nur dann um einen Verb­­bestandteil handeln, wenn mit der *entlang*-NGr kein Attribut vorliegt. Die Frage, ob zusammengeschieden wird oder nicht, kann nur auftauchen, wenn *entlang* direkt vor dem jeweiligen Verb steht, wie es etwa in Beleg 65 der Fall ist. In allen Fällen ist aber die Zweifelsfrage die, ob es sich um ein komplexes Verb handelt oder nicht. Wenn das stimmt, so ist bei Beleg



65 wohl die Entscheidung für Verb *gehen* + Adverbial *am See entlang* gefallen, was – wie oben angedeutet – vielleicht semantisch verschieden ist von *entlanggehen* + Adverbial *am See*. Diese Frage (thematisch reformulierbar als Frage der Inkorporation von *entlang* in bestimmte Verben) ist linguistisch keineswegs trivial und führt tief in theoretische Gefilde, die wir hier aber nicht mehr betreten können. Meine These wäre, dass es syntaktisch sinnvoll ist, bei Konstruktion II (mit der Präposition *an* bzw. *am*) von einem Adverbial mit dem Adverb *entlang* auszugehen (und nicht von einem komplexen Verb) und bei Konstruktion III (mit *entlang* als Akkusativ fordernde Postposition) wenn möglich von einem komplexen Verb. Kurzgefasst: Die Konstruktionen III und IV wachsen zu einer zusammen, wo dies möglich ist. Und wie wir oben gesehen haben, besteht jedenfalls immer dann die Möglichkeit dazu, wenn es sich bei dem Verb um ein unpräfigiertes Bewegungsverb handelt (also nicht zum Beispiel in den folgenden Fällen: **sich entlangverlaufen*, **entlangvorbeigehen*, **entlangwegrennen* usw., sehr wohl aber *entlanglaufen*, *entlanggehen*, *entlangrennen* usw.). Das Hauptergebnis der Studie liegt jedoch zunächst einmal in der Identifizierung der vier verschiedenen Verwendungen.

Übrigens zeigen sich bei den zu Anfang ermittelten Hauptlieferanten für Belege in dem Belegkorpus (Autoren: Dückers, Walser, ZEIT) gewisse Bevorzugungstendenzen:

Konstruktion/Autor	Dückers	Walser	Zeit	gesamt
IV: [<i>entlang</i> VERB _{Bewegung} + (NGr _{Akk} oder <i>an/am</i> +NGr _{Dat})]	24	7	7	38
III: [NGr _{Akkusativ} + <i>entlang</i>]	18	5	4	27
II: [<i>an/am</i> + NGr _{Dativ} + <i>entlang</i>]	7	6	3	16
I: [<i>entlang</i> + NGr _{Genitiv/Dativ}]	1	0	8	9
gesamt	50	18	22	90

Abbildung 24: Autorabhängige Tendenzen beim *entlang*-Vorkommen

Die interessanten Ergebnisse in positiver Hinsicht sind hier eingegraut. Es sind folgende Tendenzen erkennbar:

- Dückers: Konstruktion III (bei starker Inkorporation),
- Walser: zu gleichen Teilen Konstruktion III und II (bei schwächerer Inkorporation),
- beide: (fast) keine Voranstellung von *entlang*,
- ZEIT: Konstruktion I, also Voranstellung (bei schwächerer Inkorporation).

Möglicherweise deutet dies daraufhin, dass es sich bei Konstruktion III um eine insbesondere literarisch attraktive sprachliche Einheit handelt, die Konstruktion I dagegen eher journalistische Beliebtheit besitzt. Auch hier müssten weitere Daten erhoben werden, um einen etwaigen Zusammenhang dieser Art quantitativ zu erhärten.

4. Schluss und Anschlussmöglichkeiten

In der hier vorgestellten *entlang*-Studie wurden vier verschiedene Vorkommen von *entlang* in bestimmten sprachlichen Kontexten (oder auch: sprachlichen Umgebungen) ermittelt. Mit diesen vier Hauptverwendungen sind Unterschiede auf verschiedenen Ebenen verbunden, von denen die wichtigsten sind: Inkorporationsmöglichkeiten mit Bewegungsverben, Kasusverteilung bei Voran- und Nachstellung und Zusammenvorkommen mit der Präposition *an/am*. Die im Zweifelsfall-Duden aufgeworfene zweite Frage nach der Getrennt- oder Zusammenschreibung entpuppte sich dabei als Frage nach einer möglichen Inkorporation ins Verb. Eine solche Möglichkeit ist nur dann gegeben, wenn die NGr mit *entlang* kein Attribut ist. In den anderen

Fällen müssten sich an diese Studie vertiefende Betrachtungen anschließen. Dabei wären dann auch Vergleiche zum Vorkommen anderer Einheiten anzustellen, und zwar einerseits zu Vorkommen wie [*an* NGr *heran*], [*auf* NGr *zu*], [*in* NGr *hinein*], [*über* NGr *hinaus*], [*über* NGr *hinweg*], [*um* NGr *herum*], [*von* NGr *her*], [*zur* NGr *hinab*] usw., andererseits aber auch zu bestimmten Verben, z. B. mit Pronominaladverbien als erster Konstituente wie *heraufkommen*, *herbeilaufen*, *heruntergehen*, *hinaufgehen*, *voranfahren*, *vorbeifahren* usw. und weiteren ähnlichen Einheiten. Die in dieser Studie erzielten Einzelergebnisse zu *entlang* ließen sich so in einen sprachsystematisch weiteren Kontext stellen, *aus* dem *heraus* sicher noch viele Erkenntnisse *geborgen* werden können. Und natürlich müsste noch die einschlägige Literatur¹⁷ zu der hier besprochenen Einheit *entlang* recherchiert und besprochen werden, denn vielleicht ist das ja auch alles schon herausgefunden worden. Darauf wurde hier jedoch zugunsten einer etwas detaillierteren Präsentation der mithilfe des DWDS-Portals durchgeführten *entlang*-Studie als einer Modellstudie für eine der vielen Möglichkeiten korpusbasierten Arbeitens verzichtet.

5. Literatur

- Albert, Ruth und Cor J. Koster (2002): Empirie in Linguistik und Sprachlehrforschung. Ein methodologisches Arbeitsbuch. Tübingen: Narr.
- Albert, Ruth und Nicole Marx (2010): Empirisches Arbeiten in Linguistik und Sprachlehrforschung. Anleitung zu quantitativen Studien von der Planungsphase bis zum Forschungsbericht. Tübingen: Narr/Francke/Attempto.
- Bartz, Thomas und Nadja Radtke (2014): Digitale Korpora im Deutschunterricht: Didaktisches Potenzial. – In: ZGL 42/1. S. 130–143.
- COSMAS_{web} = Corpus and Search Management System im Netz (<https://cosmas2.ids-mannheim.de/cosmas2-web/>).
- COSMAS_{win} = Corpus and Search Management System zum Installieren auf externen Computern (<http://www.ids-mannheim.de/cosmas2/win-app/install/>).
- DeReKo = Deutsches Referenzkorpus des IDS Mannheim (Informationen unter: <http://www1.ids-mannheim.de/kl/projekte/korpora/>).
- DGD = Datenbank für Gesprochenes Deutsch (http://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.welcome).
- Di Meola, Claudio (1998): Semantisch relevante und semantisch irrelevante Kasusalternation am Beispiel von *entlang*. – In: ZS 17/2. S. 204–235.
- Di Meola, Claudio (2000): Die Grammatikalisierung deutscher Präpositionen. Tübingen: Stauffenburg. (= Studien zur deutschen Grammatik; 62).
- Draxler, Christoph (2008): Korpusbasierte Sprachverarbeitung. Eine Einführung. Tübingen: Narr/Francke/Attempto.
- Dudenredaktion (Hrsg.) (2011): Richtiges und gutes Deutsch. Das Wörterbuch der sprachlichen Zweifelsfälle. Unter Mitwirkung von Peter Eisenberg und Jan Georg Schneider. Dudenverlag: Mannheim/Zürich. [8. Aufl. 2016: s. Hennig (2016)]
- DWDS = Digitales Wörterbuch der deutschen Sprache (alt: <http://eins.dwds.de/>, neu: <https://www.dwds.de/>).
- GRAMMIS = Grammatisches Informationssystem des IDS Mannheim (<http://hypermedia.ids-mannheim.de/index.html>).

¹⁷ Insbesondere kommen hier bspw. Forschungsergebnisse von Claudio Di Meola ebenso ins Spiel (vgl. Di Meola 1998 und 2000: 62–68) wie die Auswertung entsprechender Abschnitte in Grammatiken oder bei Grammis (vgl. <http://hypermedia.ids-mannheim.de/index.html>, wo man *entlang* suchen kann und es dann an verschiedenen Stellen des GRAMMATischen InformationsSystems findet.)



- Gries, Stefan Th. (2008): Statistik für Sprachwissenschaftler. Göttingen: Vandenhoeck & Ruprecht.
- Hauser, Bernhard und Winfried Humpert (2009): *signifikant?* Einführung in statistische Methoden für Lehrkräfte. Zug: Klett/Kallmeyer.
- Hennig, Mathilde (Hrsg.) (⁸2016): *Das Wörterbuch der deutschen Zweifelsfälle. Richtiges und gutes Deutsch.* Berlin: Dudenverlag.
- Klein, Wolf Peter (2003): Sprachliche Zweifelsfälle als linguistischer Gegenstand. Zur Einführung in ein vergessenes Thema der Sprachwissenschaft. – In: *Linguistik Online* 16/4. (http://www.linguistik-online.org/16_03/klein.html).
- Klein, Wolf Peter (2014): Gibt es einen Kodex für die Grammatik des Neuhochdeutschen und, wenn ja, wie viele? Oder: Ein Plädoyer für Sprachkodexforschung. – In: Plewnia, Albrecht und Andreas Witt (Hrsg.): *Sprachverfall? Dynamik – Wandel – Variation* (= Jahrbuch des Instituts für deutsche Sprache 2013). Berlin/Boston: de Gruyter. S. 219-242.
- Lemnitzer, Lothar und Heike Zinsmeister (³2014): *Korpuslinguistik. Eine Einführung.* Tübingen: Narr/Francke/Attempto.
- Meier, Helmut (1964): *Deutsche Sprachstatistik.* 2 Bände. Hildesheim: Olms.
- Meindl, Claudia (2011): *Methodik für Linguisten. Eine Einführung in Statistik und Versuchsplanung.* Tübingen: Narr/Francke/Attempto.
- OWID = Online-Wortschatz-Informationssystem Deutsch (<http://www.owid.de/>).
- Perkuhn, Rainer, Holger Keibel und Marc Kupietz (2012): *Korpuslinguistik.* Paderborn: Fink/UTB.
- Quasthoff, Uwe, Sabine Fiedler und Erla Hallsteinsdóttir (Hrsg.) (2011): *Frequency Dictionary German – Häufigkeitswörterbuch Deutsch.* Leipzig: Universitätsverlag. (= *Frequency Dictionaries* 1).
- Rothstein, Björn (2011): *Wissenschaftliches Arbeiten für Linguisten.* Tübingen: Narr/Francke/Attempto.
- Scherer, Carmen (²2014): *Korpuslinguistik.* Heidelberg: Winter. (= *KEGLI* 2).
- Schlobinski, Peter (1996): *Empirische Sprachwissenschaft.* Opladen: WdV.
- Scholze-Stubenrecht, Werner (1991): Die Sprachberatungsstelle der Dudenredaktion. – In: *Deutsche Sprache* 19. S. 178-182.
- Stede, Manfred (2007): *Korpusgestützte Textanalyse. Grundzüge der Ebenen-orientierten Textlinguistik.* Tübingen: Narr/Francke/Attempto.
- Wortschatz-Portal (<http://wortschatz.uni-leipzig.de/>).



6. Anhänge

In der folgenden Tabelle sind alle in dem Skript vorkommenden Suchanfragen versammelt und darüber hinaus auch die entsprechenden COSMAS-Suchanfragen, mit denen man jeweils nach dem gleichen Vorkommen im DeReKo suchen kann. Dies ist vor allem für diejenigen gedacht, die über die in diesem Skript verhandelten Fragen einen Einstieg in die Nutzung von COSMAS vermittelt bekommen und selbst ein paar Suchanfragen dort ausprobieren möchten.

Bedeutung der Suchanfrage	Suchanfrageformulierung beim DWDS	Suchanfrageformulierung bei COSMAS
„Suche alle Vorkommen des Verbs <i>entlanggehen</i> , inkl. der Flexionsformen“	entlanggehen	&entlanggehen
„Suche alle Vorkommen der Formen von <i>entlang</i> und der Formen von <i>gehen</i> in einem Satz“	"entlang" "gehen" entlang && gehen	entlang /s0 &gehen [Anmerkung: Wenn man tatsächlich nach eventuellen Formen von <i>entlang</i> und nicht nur nach der einen Form <i>entlang</i> suchen möchte, müsste man „&“ als Operator davorsetzen.]
„Suche alle Vorkommen der Formen von <i>entlang</i> und der Formen von <i>gehen</i> in einem Satz oder der Formen von <i>entlanggehen</i> “	"entlang" "gehen" entlanggehen entlang && gehen entlanggehen (entlang && gehen) entlanggehen	(entlang /s0 &gehen) oder &entlanggehen
„Suche alle Vorkommen von den direkt aufeinander folgenden Formen von <i>entlang</i> und <i>an</i> (allerdings egal, in welcher Reihenfolge)“	near(entlang, an, 0)	entlang /w1 an
„Suche alle Vorkommen von den direkt aufeinander folgenden Formen von <i>entlang</i> und der exakten Wortform <i>des</i> (allerdings egal, in welcher Reihenfolge)“	near(entlang, @des, 0)	entlang /w1 des
„Suche alle Vorkommen von den direkt aufeinander folgenden Formen von <i>entlang</i> und der exakten Wortform <i>dem</i> (allerdings egal, in welcher Reihenfolge)“	near(entlang, @dem, 0) [Anmerkung zu den letzten drei: Mit phrasalen Suchen wie "entlang an", "entlang @des" und "entlang @dem" kann man zumindest bei direkt aufeinanderfolgenden Einheiten mit Abstand 0 auch die Suchreihenfolge festlegen.]	entlang /w1 dem [Anmerkung zu den letzten drei: Durch ein „+“ (= links vor rechts) oder „-“ (= rechts vor links) vor „w“ könnte man bei COSMAS die Reihenfolge festlegen, und zwar unabhängig vom Wortabstand.]



Im Folgenden werden über die im Text genutzten Suchketten hinaus kurz weitere Suchmöglichkeit vorgestellt, die für zusätzliche Arbeitsschritte in der Modellstudie zu *entlang* relevant sein könnten. Nicht genutzt wurden bisher nämlich Suchen, die die Tatsache ausnutzen, dass sowohl das Kernkorpus als auch das Deutsche Textarchiv (DTA) und das ZEIT-Korpus getaggt sind. Das heißt: angereichert mit zumindest (aber nicht nur) der Information über die Wortartzugehörigkeit der Tokens. So kann man gezielt suchen nach *entlang* als Prä-, Post- oder Zirkumposition und *entlang* als Adverb. Der Nachteil besteht natürlich darin, dass die Entscheidungen beim Tagging bereits getroffen sind, man also ohne Sichtung der jeweiligen Belege nicht mehr selbst herausbekommen kann, ob man nicht vielleicht besser Adverb statt Präposition oder Verbteil annimmt. Das ist auch der Grund, weshalb in der eigentlichen Studie davon kein Gebrauch gemacht wurde.

Die wortgebundene Suche funktioniert über die Verwendung von

```
with $p=
```

innerhalb der Suchkette, wobei links noch das eigentliche Suchwort und rechts (nach dem Gleichheitszeichen¹⁸) die Information eingefügt wird, um welche Wortart es sich bei dem gesuchten Wort handeln soll. Zum Beispiel sucht man mittels

```
entlang with $p=APPR
```

nach *entlang* als Präposition oder den linken Bestandteil einer zweiteiligen Zirkumposition. Mit der Suchkette

```
entlang with $p=APPO
```

erhält man Treffer, in denen *entlang* eine nachgestellte Präposition (= Postposition) ist. Falls *entlang* als rechter Teil einer Zirkumposition *an ... entlang* getaggt wurde, müsste dies schließlich mittels

```
entlang with $p=APZR
```

herauszubekommen sein. Das ist tatsächlich der Fall. Hier könnte man als nächstes überprüfen, welche Rolle *an* als linker Teil einer anzunehmenden Zirkumposition *an ... entlang* spielt:

```
(an with $p=Appr) && (entlang with $p=APZR)
```

Hier erhält man allerdings nur Treffer im DTA.

Als nächstes kann *entlang* als Adverb gesucht werden. Das kann mittels der folgenden Suchkette geschehen:

```
entlang with $p=ADV
```

Hier erzielt man aber nur wenige Treffer. Ein erster Blick auf die KWIC-Ausgabe ergibt, dass es sich hierbei entweder um Fälle von *entlang* + Präposition (z. B. *von*) oder um den Gebrauchs von *entlang* ohne eine weitere Präposition und ohne eine NGr, die im Fall von *entlang* regiert sein könnte, handelt. Letzteres sind vor allem Verwendungen von *entlang* in der Nähe von Verben, mit denen *entlang* aber nicht zusammengeschrieben auftaucht:

¹⁸ Auf der Seite <http://www.deutschestextarchiv.de/doku/pos> sind alle wichtigen Suchausdrücke hierzu aufgelistet.

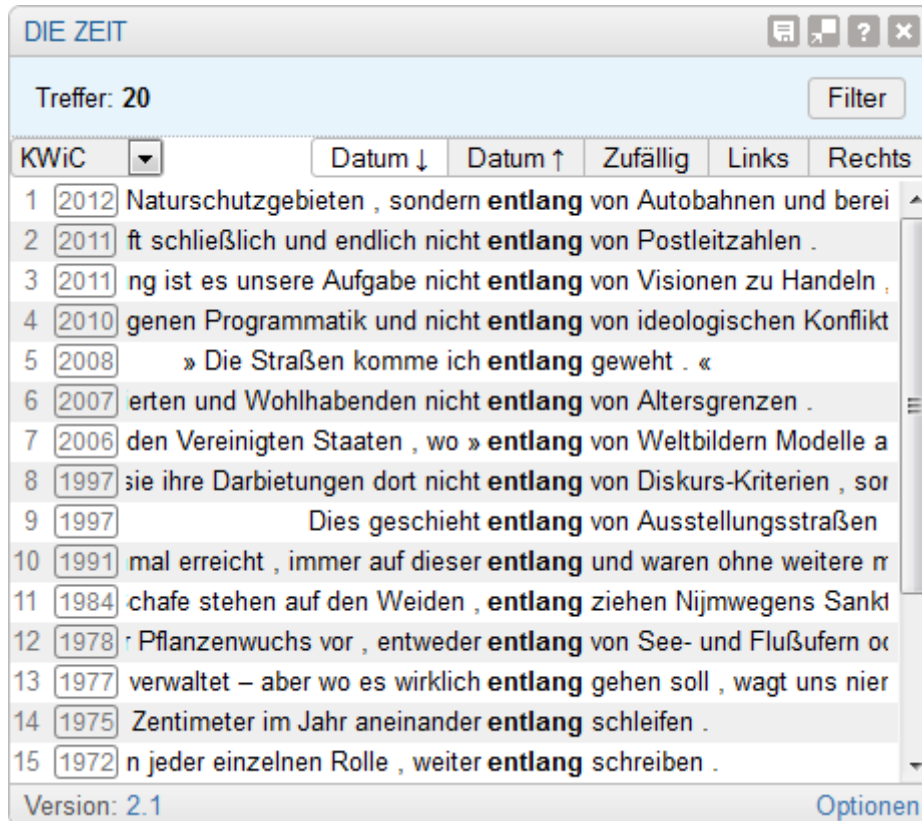


Abbildung 25: KWIC-Fenster der Treffer zu der Suche von *entlang* als Adverb

Über die Wortartensuche hinaus könnte es noch relevant sein, *entlang* auf die Position in Bezug zu Interpunktionszeichen zu suchen. So könnte man entlang etwa am Satzanfang suchen, weil es hier auf jeden Fall vor einem entsprechenden Verb in dem Satz steht. Eine solche Suche kann über

```
entlang with $.=0
```

vorgenommen werden (vgl. hierzu die DWDS-Hilfeseite im Abschnitt 4.2). Vielleicht können auch die in der Studie angesprochenen Reihungen direkt gefunden werden, wenn man *entlang* bspw. als nach einem Komma stehend sucht:

```
\, entlang
```

Diese letzten beiden Suchen könnten hilfreich sein, um genau die Verwendungen zu finden, in denen es unwahrscheinlich sein müsste, dass *entlang* ein Versteil ist, weil *entlang* vor dem Verb oder die Gruppe mit *entlang* gereiht mit anderen Gruppen vorkommt.

Wenn man direkt nach dem Vorkommen von zusammengesetzten Verben mit *entlang* als erstem Teil suchen möchte, so kann man eine Trunkierung (= Markierung einer Stelle mit einem Sternchen, das für irgendwelche Zeichen steht) vornehmen und dies mit der Information kombinieren, dass es sich insgesamt um ein Verb handeln soll, zum Beispiel um ein finites Vollverb, ein Vollverb im Infinitiv oder ein Partizip Perfekt eines Vollverbs (für alle weiteren Verb-Kodierungen s. die in der Fußnote 16 erwähnte Hilfeseite des DTA):

```
entlang* with $p=VVFİN
```

```
entlang* with $p=VVINF
```

```
entlang* with $p=VVPP
```



Derartige Suchen, wie sie in diesem zweiten Anhang des Lehrskripts vorgestellt wurden, kann man gezielt einsetzen, um spezielle Verwendungen der jeweiligen Sucheinheit zu finden. Wenn die Suchen gut überlegt sind und auch gut funktionieren (also wenige Treffer erzielen, die man eigentlich nicht erzielen wollte (sog. falsche Positive) und wenige Treffer nicht erzielt, die man eigentlich erzielen wollte (sog. falsche Negative)), so können die Trefferanzahlen direkt ausgewertet werden. Dafür ist jedenfalls immer eine wenigstens stichprobenartige Überprüfung der Treffer in den KWIC-Ansichten nötig. Oft wird man aber darüber hinaus auch qualitativ auswertende Arbeitsschritte noch an die Suchen anschließen müssen, um zu relevanten und guten Ergebnissen zu kommen. Dies wurde beispielhaft in der *entlang*-Studie oben gezeigt.

Es bleibt jetzt zunächst einmal nur noch, Ihnen viel Spaß und viel Erfolg bei eigenen korpusbasierten linguistischen Forschungsarbeiten zu wünschen. Also: Viel Spaß und viel Erfolg!