

['stʌdi:z 'ɪn lɪŋ'ɡwɪstɪks ənd
lɪŋ'ɡwɪstɪk 'deɪtə 'saɪəns]

**Studies in Linguistics and
Linguistic Data Science**



Simon Masloch

**Über *über* und dessen
Entsprechungen im Englischen**

Linguistic Data Science Lab

Volume 3

Editor: Tibor Kiss

ISSN: 2700-8975

[ˈstʌdiːz ˈɪn lɪŋˈɡwɪstɪks ənd
lɪŋˈɡwɪstɪk ˈdeɪtə ˈsaɪəns]

**Studies in Linguistics and
Linguistic Data Science**



Editor:

Tibor Kiss

Linguistic Data Science Lab

Ruhr-Universität Bochum

Universitätsstr. 150

44801 Bochum



Studies in Linguistics and Linguistic Data Science

Linguistics is an established area of study; Linguistic Data Science is a relatively new field within linguistics that aims at combining results and methods from theoretical linguistics with methods from Data Science and Machine Learning.

The ultimate goal of Linguistic Data Science is to provide results within linguistics that have a strong empirical base and thus facilitate our growing understanding of all formal areas of linguistics (which we use in a broad term, comprising formal methods in semantics and pragmatics in addition to phonology, morphology, and syntax). We follow other areas within science and the humanities by stating that knowledge is justified belief, that belief is based on evidence, and that evidence, finally, can only be determined on the background of theoretical assumptions.

Studies in Linguistics and Linguistic Data Science will publish results that follow these leads. The publication language will be English, but sometimes results will be published in German, for the simple purpose to make these results available.

Tibor Kiss

Linguistic Data Science Lab

Über *über* und dessen Entsprechungen im Englischen

Simon Masloch

DEUTSCHES ABSTRACT

In dieser Arbeit wird ein Korpus mit Annotationen 632 deutsch-englischer Satzpaare vorgestellt, bei denen der deutsche Satz *über* enthält. Auf Basis dieser Annotationen werden Abbildungen zwischen den verschiedenen Bedeutungen von *über* und seinen englischen Entsprechungen erstellt, es werden also für jede Bedeutung *b* von *über* diejenigen (bestimmte Voraussetzungen erfüllenden) sprachlichen Elemente des Englischen bestimmt, mit denen *b* in den jeweiligen Sätzen ausgedrückt wird. Es zeigt sich, dass die meisten englischen Entsprechungen sich jeweils sehr wenigen Bedeutungen zuordnen lassen. Eine Ausnahme bildet *over*. Die Annotationen und Bedeutungs-Entsprechungs-Abbildungen können u. a. bei der automatischen Präpositionsbedeutungsdisambiguierung, beim Erstellen bilingualer Wörterbücher und zum Finden von Ausdrücken mit einer bestimmten Bedeutung genutzt werden. Diese möglichen Anwendungen werden beschrieben und es werden Änderungen am Bedeutungsinventar für *über* von Kiss et. al. (2020) vorgeschlagen. Diese Vorschläge betreffen die Emotionsgegenstandsbedeutung, die Unterscheidung zwischen den Lesarten *Bezugspunkt* und *Thema* und die Unterscheidung verschiedener Lesarten bei Traversen mit weniger als drei relevanten Dimensionen.

KEYWORDS: Präpositionen, Polysemie, Deutsch, Englisch, bilinguales Korpus

ENGLISH ABSTRACT

This paper presents a corpus consisting of 632 annotated German-English sentence pairs where the German sentence contains the preposition *über*. The annotations are used to create mappings between the different readings of *über* and its translation equivalents, which are the elements used to express the same meaning in English: Which translation equivalents are used when *über* has a specific meaning? One result is that—with the exception of *over*—each of the translation equivalents is only used with a small subset of the readings of *über*. It is shown that and how the mappings and the annotations can be useful for a variety of different purposes, including word-sense disambiguation, the creation of bilingual dictionaries, and the search for expressions with a specific meaning. Besides, the paper takes a closer look at some of the readings of *über* and their differences and suggests some modifications of Kiss et al.'s (2020) entry for it concerning the reading *object of emotion* and the distinction between the readings *topic* and *point of reference* as well as the distinction between several traversal readings.

KEYWORDS: prepositions, polysemy, German, English, bilingual corpus

Inhaltsverzeichnis

Vorwort	vi
1. Einleitung	1
1.1. Theoretischer Hintergrund	2
1.1.1. Rektion	2
1.1.2. Die Bedeutungen von <i>über</i> nach Kiss et. al. (2020)	3
1.1.3. Entsprechungen	4
1.2. Ziele und Aufbau	5
1.3. Verwandte Arbeiten	6
1.4. Verfügbarkeit der Daten und Skripte und Anmerkungen zur Darstellungsform	8
2. Die Erstellung des Korpus	8
2.1. Grundlage	8
2.1.1. CoStEP	9
2.1.2. Satzalignierung	9
2.1.3. Wortalignierung	9
2.1.4. Beispielauswahl	10
2.2. Das Annotationsschema	13
2.2.1. Metainformationen	14
2.2.2. Entsprechung	16
2.2.3. Ein Spezialfall: Alignierung mit Verb (AMV)	17
2.2.4. Bedeutung	18
2.2.5. Rektionsstatus	21
2.2.6. Formulierungsnähe	23
2.3. Durchführung der Annotation	27
3. Das Korpus	28
3.1. Vor Bearbeitung	28
3.2. Nach Bearbeitung	30
3.2.1. Der Datensatz I	31
3.2.2. Andere Datensätze	34
3.3. Die deutschen Daten	34
4. Auswertung	37
4.1. Der Datensatz I	37
4.1.1. Rektionsstatus	37
4.1.2. Formulierungsnähe	40
4.1.3. Die Bedeutungs-Entsprechungs-Abbildungen	43
4.2. Der Datensatz AMV	45
4.3. Der Datensatz 4N	46
5. Überlegungen zum Bedeutungsinventar	47
5.1. <i>Über</i> bei psychologischen Prädikaten	47
5.2. <i>Bezugspunkt</i> und <i>Thema</i>	53
5.3. Unterscheidung verschiedener Arten der Traverse mit weniger als drei relevanten Dimensionen	58
6. Anwendungen	60
6.1. Präpositionsbedeutungsdisambiguierung	60
6.2. Nutzen der Abbildungen für Lerner	63
6.3. Die Methode der Suche über Entsprechungen	65

6.4. Nutzungsmöglichkeiten der deutschen Daten	67
7. Zusammenfassung und Ausblick	67
7.1. Zusammenfassung	67
7.2. Ausblick	68
7.2.1. Mehr Daten	68
7.2.2. Mehr Sprachen	69
7.2.3. Ausführlichere Annotation	70
A. Tabellen und Abbildungen	71
B. Abkürzungsverzeichnis	93
Literatur	94

Abbildungsverzeichnis

1.	Die zwanzig am häufigsten automatisch mit <i>über</i> alignierten Worte	11
2.	Rektionsstatus der Beispiele in I im Deutschen und Englischen	37
3.	Rektionsstatus der Beispiele in I im Deutschen und Englischen nach Bereinigung	38
4.	Sankey-Diagramm: Bei welchen Lesarten kommen welche Entsprechungen vor?	44
5.	Bildschirmfoto eines Bedeutungseintrags aus (PONS, o. D. b)	64
6.	Sankey-Diagramm: Wozu wurden die automatisch alignierten Entsprechungen?	91
7.	Sankey-Diagramm: Bei welchen Lesarten kommen in IB welche Entsprechungen vor? .	92

Tabellenverzeichnis

1.	Die Spalten der Annotationsdateien	13
2.	Die möglichen Metainformationswerte	14
3.	Die bei der Annotation verwendeten Bedeutungs-Tags	19
4.	Formulierungsnähewerte	23
5.	Die Häufigkeiten der Werte von <i>Meta</i> in den deutschen Zeilen	28
6.	Die Häufigkeiten der Formulierungsnähewerte	28
7.	Kurzübersicht über die Häufigkeiten der Bedeutungswerte (de)	29
8.	Die Häufigkeiten der Bedeutungswerte (en)	29
9.	Kurzübersicht Frequenzen der korrigierten Entsprechungen	29
10.	Die Datensätze mit den nach Bedeutung getrennten Daten	31
11.	Die Häufigkeiten der Bedeutungen in I	32
13.	Die Häufigkeiten der Formulierungsnähewerte in I	32
12.	Die Häufigkeiten der Entsprechungen in I	33
14.	Die Häufigkeiten der Rektionsstatus in I	33
15.	Die Häufigkeiten der Rektionsstatus in AMV	34
16.	Die Häufigkeiten der Formulierungsnähewerte in AMV	34
17.	Die Häufigkeiten der Rektionsstatuswerte in DE	35
18.	Die Häufigkeiten der Bedeutungen in DE-DIS	36
19.	Die Häufigkeiten der Rektionsstatuswerte in DE-DIS	36
20.	Die Häufigkeiten der Rektionsstatus in I bei den Beispielen mit primärer Präposition als Entsprechung	38
21.	Rektionsstatus deutsch und Rektionsstatus englisch in bereinigtem I kreuztabuliert	39
22.	Kreuztabellen Rektionsstatus Bedeutungen	39
23.	Kreuztabellen Rektionsstatus Entsprechungen	40
24.	Die Häufigkeiten der Formulierungsnähewerte in den ungetrennten I-Daten	41
25.	Kreuztabelle von Formulierungsnähe und Entsprechungen in den nicht bedeutungsgetrennten Daten aus I	41
26.	Prozentuale Verteilungen der Formulierungsnähewerte auf die Rektionsstatus in den nicht bedeutungsgetrennten Daten aus I: links Englisch, rechts Deutsch	42
27.	Abbildungen nach Bedeutungen geordnet AMV	45
28.	Abbildungen 4N	47
29.	Die hundert am häufigsten automatisch mit <i>über</i> alignierten Worte	71
30.	Rektionsstatus für verschiedene Regentien	73
31.	Die Häufigkeiten der Bedeutungen in DE	73
32.	Die Frequenzen der Entsprechungen vor Trennung nach Bedeutung und Ausschluss auszusondernder Beispiele	74
33.	Wie oft wurde welche automatisch alignierte Entsprechung wozu korrigiert?	76
34.	Abbildungen nach Bedeutungen geordnet mit Rektionsstatus (I)	79
35.	Abbildungen nach Bedeutungen geordnet ohne Rektionsstatus (I)	81
36.	Abbildungen nach Entsprechungen geordnet mit Rektionsstatus (I)	83
37.	Abbildungen nach Entsprechungen geordnet ohne Rektionsstatus (I)	85
38.	Abbildungen für AMV nach Verben geordnet	86
39.	Abbildungen für IB nach Bedeutungen geordnet mit Rektionsstatus	87
40.	Abbildungen für IB nach Entsprechungen geordnet mit Rektionsstatus	88

Vorwort

Es handelt sich bei dem vorliegenden Text um eine leicht überarbeitete und gekürzte Version meiner Masterarbeit, welche ich im Juni 2020 an der Fakultät für Philologie der Ruhr-Universität Bochum eingereicht habe. Da für verschiedene Gruppen von Lesern unterschiedliche Abschnitte relevant sein mögen, sei darauf hingewiesen, dass die einzelnen Unterabschnitte in Abschnitt 5 und 6 nicht aufeinander aufbauen und nicht in linearer Abfolge gelesen werden müssen. Eine Lektüre der Abschnitte 1, 2.2 und 4.1.3 ist aber zu empfehlen.

Ich möchte an dieser Stelle allen herzlich danken, die mich während der Arbeit an diesem Werk mit Anmerkungen und Kommentaren unterstützt oder es Korrektur gelesen haben, insbesondere Sarah Broll. Vor allem gilt mein Dank aber meinem Betreuer, Herrn Prof. Dr. Tibor Kiss, sowie Frau Dr. Claudia Roch für viele hilfreiche Hinweise und Anmerkungen. Claudia Roch hat auch mit mir über die Kategorisierung problematischer Beispiele diskutiert und mir in ihrem Besitz befindliche Literatur überlassen. Jan Christian Redlich danke ich für technische Unterstützung.

1. Einleitung

Präpositionen sind oftmals hochgradig polysem und das gilt auch für die Präpositionen im Deutschen: Im *Handbuch für die Bestimmung und Annotation von Präpositionsbedeutungen im Deutschen* von Kiss et. al. (2020) hat etwa *auf* – lässt man regierten Gebrauch außer Acht – 21 (Unter-)Bedeutungen, *in* hat 20, *unter* 17 und *über* 24. Gleichzeitig sind Präpositionen hochfrequent: Volk (2006, S. 84) berichtet, dass von den 5,5 Millionen Token in seinem Computer-Zeitung-Korpus 540 000 Präpositionen sind und sich diese Vorkommen auf 99 Typen verteilen. Entsprechendes gilt auch für das Englische (vgl. zur Frequenz Saint-Dizier, 2006, S. 3; zur Anzahl der Bedeutungen CL Research, 2007; Litkowski & Hargraves, 2005). Aber es gibt sprachübergreifend betrachtet auch Unterschiede:

In languages that use prepositions, regularities over languages are relatively minor, even for closely-related languages in the same family, and even in concrete and well-mastered domains such as time or space. When looking at a bilingual dictionary it is easy to note that preposition translations are very complex, often involving semantic considerations, not to cite the large idiosyncratic variations. (Saint-Dizier, 2006, S. 20)

Vor diesem Hintergrund erscheint eine sprachübergreifende Untersuchung nicht nur fruchtbar, sondern notwendig. Wenn etwa die verschiedenen Bedeutungen von *über* im Englischen durch verschiedene sprachliche Elemente (darunter auch verschiedene Präpositionen) zum Ausdruck gebracht werden können, dann stellt das einerseits eine Herausforderung für Englischlerner mit deutscher Muttersprache dar (wie Saint-Diziers Kommentar zu Wörterbüchern im Zitat oben vermuten lässt¹), eröffnet aber andererseits Möglichkeiten z. B. für die automatische Präpositionsbedeutungsdisambiguierung.

In dieser Arbeit soll ein bilinguales Korpus vorgestellt werden, in dem deutsch-englische Satzpaare, bei denen der deutsche Satz *über* enthält, mit Annotationen versehen sind, die eine Untersuchung der Beziehungen zwischen den Bedeutungen von *über* und seinen englischen Entsprechungen erlauben. Die Sätze stammen aus Übersetzungen von Reden aus dem Europaparlament und sind dem Korpus CoStEP (Graën et. al., 2014) entnommen. Es wurde für 632 Vorkommen von *über* die Bedeutung und die englische Entsprechung annotiert. Des Weiteren wurde neben der Formulierungsnähe zwischen beiden Sätzen in Bezug auf den relevanten Teil – zuweilen gibt es erhebliche Unterschiede etwa in Bezug auf das interne Argument oder die durch *über* ausgedrückte Relation wird bei einer freien Übersetzung gar nicht ausgedrückt, vgl. Abschnitt 2.2.6 – auch für beide Sprachen der Rektionsstatus erfasst, also ob *über* bzw. dessen Entsprechung in der gegebenen Lesart regiert ist.

Über wurde hier gewählt, da seine Semantik ausgiebig erforscht ist (sein Kognat *over* und es selbst sind Gegenstand vieler Arbeiten in der kognitiven Linguistik, etwa von Brugman und Lakoff (1988/2006), Liamkina (2007), Meex (2001) und Tyler und Evans (2001)) und das für die Annotation verwendete Bedeutungsinventar von Kiss et. al. (2020) durch die Betrachtung von *über* in (Kiss, 2019) erprobt ist. Im Computer-Zeitung-Korpus von Volk (2006) ist *über* die deutsche Präposition mit der siebthöchsten Frequenz.

¹Wir werden allerdings in Abschnitt 6.2 sehen, dass bestehende zweisprachige Wörterbücher in Bezug auf die Behandlung von *über* verbesserungswürdig sind.

1.1. Theoretischer Hintergrund

Präpositionen drücken semantisch zweistellige Relationen aus, sind aber syntaktisch nur einstellig (vgl. u. a. Breindl, 2006; Saint-Dizier, 2006). Ich werde Kiss et. al. (2020) folgend das innerhalb der PP realisierte Argument als *internes* Argument bezeichnen, das außerhalb (etwa als NP, deren Adjunkt die PP ist, Verb oder Argument des Verbes (Kiss et. al., 2020, S. 6)) realisierte als *externes*. Ferner bezeichne *Regens* im Allgemeinen das Element, dessen syntaktisches Argument oder Adjunkt die PP ist.^{2,3}

Da sich die Betrachtungen hier auf die Präposition *über* beschränken werden, könnten Abgrenzung nach außen und innere Unterteilung der Klasse der Präpositionen für uns allenfalls in Bezug auf die englischen Entsprechungen relevant sein. Wie ich in Abschnitt 1.1.3 erläutern werde, ist eine vertiefende Untersuchung des Themas jedoch auch für diese nicht erforderlich. Wichtig ist allerdings die Unterscheidung zwischen primären und sekundären Präpositionen: Diese sind morphologisch komplex, jene nicht (Breindl, 2006, S. 938) und beide Gruppen unterscheiden sich in einer Reihe von weiteren Eigenschaften (vgl. zum Deutschen Breindl, 2006, S. 939; zum Englischen Kortmann & König, 1992, S. 682 ff.). Primäre Präpositionen sind z. B. *in*, *an* und *auf* bzw. *on* und *over*, sekundäre etwa *zugunsten* und *anstelle* bzw. *during* und *concerning*. Ich werde davon ausgehen, dass nur primäre Präpositionen regiert sein können.

1.1.1. Rektion

Man geht gemeinhin davon aus, dass bei einigen als Komplemente fungierenden PPs die Wahl der Präposition durch das Regens⁴ bedingt ist. Solche Präpositionen werden als *regiert* bezeichnet. Beispiele sind *über in verfügen über* und *auf in warten auf*.

Die in der Literatur genannten Kriterien zur Bestimmung des Rektionsstatus (Konstanz, Kasustransfer, vgl. Müller (2013, Kapitel 2.4)) sind nach aktuellem Forschungsstand allerdings problematisch (vgl. u. a. Müller, 2013, S. 50 ff.; Lerot, 1982, S. 265 f.) und daher wenig geeignet, bei einer Annotation einfach und schnell den Rektionsstatus zu bestimmen.⁵ Deswegen wurde bei der Annotation auf Wörterbücher und die Listen *über* regierender Elemente aus (Kiss et. al., 2020) zurückgegriffen, um den Rektionsstatus zu bestimmen, und nur, wenn dies zu keinem Ergebnis führte, selbst getestet (vgl. Abschnitt 2.2.5).

Entgegen einer weitverbreiteten Auffassung (vertreten durch u. a. Breindl, 2006, S. 936) nehme ich wie Kiss et. al. (2020) an, dass auch regierte Präpositionen eine Bedeutung tragen können.⁶ Lerot (1982, S. 267) argumentiert dafür mit u. a. folgendem Satzpaar:

²Der Begriff wird nicht nur für Elemente verwendet, die eine Präposition im Sinne von Abschnitt 1.1.1 regieren.

³Zu entscheiden, ob eine PP ein Argument oder ein Adjunkt ist, kann im konkreten Fall schwierig sein, da in der Literatur vorgeschlagene Tests nicht immer zu einem klaren Ergebnis führen (Müller, 2013, S. 42 f.). Für uns ist diese Unterscheidung vor allem deshalb wichtig, weil nur Präpositionen in Komplementen regiert sein können (vgl. dazu u. a. Breindl, 2006).

⁴Ich verallgemeinere hier und spreche allgemein von Regentien, obwohl in der Literatur oftmals nur Verben betrachtet werden, da ich wie Kiss et. al. (2020) davon ausgehen werde, dass auch Substantive und Adjektive Präpositionen regieren können.

⁵Abgesehen davon, dass diese Tests an sich schon nicht einfach anwendbar sind, gibt es auch ganz praktische Probleme: So ist etwa keineswegs offensichtlich, ob eine Präposition konstant ist (und also nicht mit einer anderen kommutiert, ohne dass sich die Bedeutung des Regens ändert) oder nicht, und es müsste immer erst nach Ersetzungsmöglichkeiten gesucht werden (vgl. Abschnitt 2.2.5).

⁶Tatsächlich konnte *über* bei der Annotation dann auch fast immer eine Bedeutung zugewiesen werden, wie Tabelle 30 auf S. 73 im Anhang zeigt, in der die Regentien, bei denen *über* desemantisiert vorkam, unter *d* gelistet sind. Es sei an dieser Stelle auch darauf hingewiesen, dass die Kategorisierung eines Vorkommens einer Präposition als desemantisiert seriöserweise nur vorgenommen werden kann, wenn man über eine vollständige Auflistung ihrer Bedeutungen verfügt. Man kann also nicht einfach feststellen, dass eine Präposition bedeutungslos ist, wie man (manchmal) feststellen kann, dass ein bestimmter Satz syntaktisch nicht akzeptabel ist. So könnte man naiverweise annehmen, die Präposition in *sich*

- (1) a. Peter hat ihr *zum Geburtstag* Blumen geschenkt
 b. Peter hat ihr *zum Geburtstag* gratuliert

Hier wird in der *zu*-PP jeweils ein Anlass ausgedrückt und es ist davon auszugehen, dass *zu* in (1b) regiert ist, in (1a) aber nicht. In beiden Fällen scheint die Präposition jedoch die gleiche Bedeutung zu haben.

Da auch für die Präpositionen in den englischen Sätzen der Rektionsstatus bestimmt werden soll, muss an dieser Stelle darauf eingegangen werden, dass in der englischen Grammatik bestimmte Kombinationen von Verb und Präposition zuweilen als sogenannte *phrasal verbs*⁷ (und nicht etwa Verb mit regierter Präposition) analysiert werden. Bei Quirk et. al. (1985) etwa wird eine Analyse, in der die Präposition bei z. B. *ask for* mit dem Verb eine Einheit bildet, zumindest erwogen. Huddleston und Pullum (2002) stellen dagegen zwar fest, dass es bei den regierten Präpositionen (die sie „specified“ nennen und über Konstanz definieren (S. 220, 272)) zwei Klassen gibt, nämlich die der beweglichen („mobile“, z. B. bei *refer to*) und die der starren („fixed“, z. B. bei *come across*) (S. 275), welche sich in ihrem syntaktischen Verhalten – etwa in Bezug auf die Möglichkeit der Voranstellung der P+NP-Kombination – unterscheiden; sie sind aber der Ansicht, dass sich auch bei letzteren eine PP bilde: Die Verb-Präposition-Kombination sei dort bloß *versteinert* („fossilised“ (S. 277)). Bei dieser Versteinierung scheint es sich allerdings um ein Phänomen zu handeln, das in verschiedenen Graden auftritt. So sprechen Huddleston und Pullum (2002) etwa von einem „fair amount of fossilisation“ (S. 283). Auch Quirk et. al. (1985) vertreten die Ansicht, dass die Grenze zwischen *prepositional verbs* und Verben, die frei mit PPs kombiniert werden, fließend sei (S. 1162). Ich werde nicht davon ausgehen, dass es im Englischen *phrasal verbs* gibt, und in den entsprechenden Fällen die Präposition – genau wie man es bei den deutschen Beispielen tun würde – als regiert (und gegebenenfalls desemantisiert) kategorisieren. Da es keine klaren Grenzen gibt, ist allerdings mit Klassifikationsschwierigkeiten zu rechnen.

1.1.2. Die Bedeutungen von *über* nach Kiss et. al. (2020)

Grundsätzlich unterscheiden Kiss et. al. (2020) zwischen Oberbedeutungen (wie z. B. *Spatial*, *Temporal* oder *Modal*), welche verschiedene Unterbedeutungen haben, und einfachen Bedeutungen ohne Unterbedeutungen. Unterbedeutungen der spatialen Lesart von *über* sind etwa *Traverse vertikal* (wie in *Hubert hüpfte über den Bach*) oder *achsenbezogen, vertikal* (wie in *das Bild über der Tür*). Für die Oberbedeutungen *spatial* und *temporal* liegen Entscheidungsbäume vor, in denen die Unterbedeutungen anhand bestimmter Merkmale klassifiziert sind und die bei der Annotation dazu genutzt werden können, die richtige Unterbedeutung zu finden. Die für *über* angenommenen Oberbedeutungen und einfachen Bedeutungen sind:

- (2) *Spatial, Temporal, Modal, Konditional, Bezugspunkt, Machtverhältnis, Maßangabe, Rangfolge, Thema, Über-/Unterschreitung, Zuordnung*

aufregen über sei bedeutungslos, eine genauere Betrachtung zeigt aber, dass *über* regelmäßig verwendet wird, um Emotionsgegenstände auszudrücken (vgl. Abschnitt 5.1).

⁷Dieser Begriff wird manchmal von dem des *prepositional verb* abgegrenzt und manchmal als Oberbegriff gebraucht, der auch *prepositional verbs* umfasst. Hier ist letzteres gemeint.

Die ersten vier Bedeutungen aus (2) haben Unterbedeutungen. Aus Platzgründen können hier allerdings weder alle Unterbedeutungen aufgeführt werden,⁸ noch können die einzelnen Lesarten erläutert werden. Es sei diesbezüglich auf den Eintrag zu *über* bei Kiss et. al. (2020, S. 172 ff.) verwiesen.

Da das Annotationsschema von Kiss et. al. (2020) auf Basis eines Zeitungskorpus entwickelt wurde (S. 9), ist zunächst offen, ob es sich auch auf ein Korpus mit Parlamentsreden anwenden lässt. Einer der Zwecke dieser Arbeit ist, dies zu prüfen (vgl. Abschnitt 2.3).

1.1.3. Entsprechungen

Es soll davon ausgegangen werden, dass mit *über* in einem deutschen Satz jeweils eine bestimmte semantische Relation ausgedrückt wird und dass im englischen Satz oftmals die gleiche oder eine sehr ähnliche Relation Ausdruck findet. Nun können Sprachen – und kann auch eine einzelne Sprache – sich ganz unterschiedlicher Mittel bedienen, um eine bestimmte Relation auszudrücken. Auch wenn es eine interessante Frage ist, auf welche Weise die durch *über* ausgedrückten Relationen im Englischen grundsätzlich realisiert werden können, werden wir in dieser Arbeit aus praktischen Gründen (s. dazu Abschnitt 2.2.2) vornehmlich diejenigen sprachlichen Mittel untersuchen, die sich als klare *Entsprechungen* von *über* auffassen lassen. Eine Entsprechung in unserem Sinne soll eine aus adjazenten Einheiten bestehende Wortgruppe (im Normalfall ein einzelnes Wort) sein, mit der die gleiche (oder eine sehr ähnliche) Bedeutung ausgedrückt wird wie mit *über*. So sind Entsprechungen von *über* in der Themalesart etwa *about*, *on*, *concerning*, *relating to* und *on the subject of*. Wir können uns in Bezug darauf, wie diese Worte und Wortgruppen in einer Grammatik des Englischen zu analysieren wären, einer Meinung enthalten.⁹ Es ist ein Ziel dieser Arbeit, die Entsprechungen von *über* in seinen verschiedenen Lesarten zu identifizieren.

Darüber hinaus soll aber auch der Versuch unternommen werden, die Abdeckung von Formen und Wendungen, die keine Entsprechungen von *über* sind, im Englischen aber eine Bedeutung haben können, die im Deutschen unter Verwendung von *über* ausgedrückt wird, dadurch zu erhöhen, dass etwa auch Verben erfasst werden, die im Englischen ein NP-Komplement verlangen, wo im Deutschen eine *über*-PP steht (dies soll als *Alignierung mit Verb* (AMV) bezeichnet werden, s. dazu Abschnitt 2.2.3).

Damit, dass die Entsprechungen die gleichen oder ähnliche Relationen ausdrücken wie *über* in einer bestimmten Lesart, soll nicht gesagt sein, dass es keine Unterschiede zwischen den einzelnen Entsprechungen gibt, und auch nicht, dass eine Entsprechung von *über* in einer bestimmten Bedeutung immer gebraucht werden kann, wenn diese Bedeutung ausgedrückt werden soll und im Deutschen die Verwendung von *über* möglich ist. Die Gründe dafür, dass das nicht immer möglich ist, können verschieden sein: So ist es etwa möglich, dass die Verwendung einer Entsprechung auf bestimmte Kontexte eingeschränkt ist. Die Untersuchungen von Rankin und Schiftner (2011) z. B. zeigen, dass *concerning* fast ausschließlich verwendet wird, wenn das Regens ein Nomen und das externe Argument ist und semantisch aus einem der drei folgenden Bereiche kommt: Bereitstellung von Information, Diskussion und Schließen (wissenschaftlich oder allgemein), Regeln und Vorschriften. Eine *debate on* ist auch etwas

⁸Eine Auflistung der bei der Annotation verwendeten Tags für die Bedeutungen findet sich aber in Tabelle 3 auf S. 19. Sie weicht geringfügig von der Auflistung bei Kiss et. al. (2020, S. 172) ab.

⁹Darüber, welche dieser Entsprechungen als (sekundäre) Präpositionen analysiert werden sollten, ist sich die Literatur nicht einig: Für verschiedene Positionen vgl. etwa Breindl (2006), Huddleston und Pullum (2002, Kapitel 7) und Kortmann und König (1992).

anderes als eine *debate over* (Herbst et. al., 2004, S. 206); trotzdem können sowohl *on* als auch *over* Entsprechungen von *über* im Kontext *Debatte über* sein.

Wie sich zeigt, werden einerseits Sätze in dem verwendeten bilingualen Korpus oftmals sehr frei übersetzt, sodass in vielen englischen Sätzen die im Deutschen durch *über* ausgedrückte semantische Relation nicht zu finden ist, und liegt andererseits nicht immer eine Entsprechung in unserem Sinne vor, da die Relation mit anderen Mitteln ausgedrückt wird. Deswegen wird auch die Formulierungsnähe zwischen den beiden Sätzen annotiert, wobei eine vierstufige Skala Verwendung findet, bei der der vierte Wert eben dafür steht, dass eine grundsätzlich andere Struktur vorliegt (vgl. Abschnitt 2.2.6). Eine Untersuchung entsprechender Beispiele erfolgt nur in einem engen Rahmen. Mit den anderen drei Formulierungsnähewerten werden lexikalische und strukturelle Unterschiede in Bezug auf Regens und internes Argument erfasst.

1.2. Ziele und Aufbau

Das Hauptziel dieser Arbeit ist die Erstellung von Abbildungen zwischen den verschiedenen Lesarten von *über* und seinen Entsprechungen im Englischen. Es wird in Abschnitt 4 erreicht, was aber nicht heißt, dass die Arbeit damit abgeschlossen wird.

Vorher wird in Abschnitt 2 die Erstellung des Korpus beschrieben, wobei in Unterabschnitt 2.1 das Vorgehen bei der Generierung der Annotationsdateien geschildert und in Unterabschnitt 2.2 das Annotationsschema dargestellt wird. Da das Bedeutungsinventar für *über* von Kiss et. al. (2020) in Hinblick auf eine ganz andere Fragestellung und auf Basis eines Korpus von Zeitungstexten erstellt wurde, während hier Parlamentsreden annotiert werden, soll im Zuge der Beschreibung der Durchführung der Annotation in Abschnitt 2.3 auch die Frage beantwortet werden, inwieweit es für Annotationen auf andersgearteten Korpora geeignet ist. Außerdem werden in diesem Unterabschnitt auch Probleme bei der Annotation angesprochen und Vorschläge gemacht, wie diese bei ähnlichen Annotationsvorhaben vermieden werden können. In Abschnitt 3 wird dann das Korpus der annotierten Daten beschrieben und anhand der Annotationen in verschiedene Datensätze aufgeteilt: So sind für die Erstellung von Abbildungen vornehmlich die Daten interessant, bei denen es überhaupt eine Entsprechung im oben definierten Sinne gibt.¹⁰

Die Auswertung der interessanten Datensätze erfolgt in Abschnitt 4. Dort werden für die Daten, bei denen es eine echte Entsprechung gibt, Bedeutungs-Entsprechungs-Abbildungen erstellt. Auch für englische Verben, die im Gegensatz zu ihrem eine *über*-PP verlangenden deutschen Pendant ein NP-Argument selektieren, und für sonstige Elemente, die zwar keine Entsprechungen in unserem Sinne sind, mit denen aber die Bedeutung ausgedrückt wird, zu deren Ausdruck im Deutschen *über* verwendet wird, werden Abbildungen erstellt. In diesem Abschnitt sollen aber auch andere Fragen beantwortet werden: So kann etwa – was nicht überrascht – gezeigt werden, dass es einen hochsignifikanten Zusammenhang zwischen dem Rektionsstatus im Deutschen und dem im Englischen gibt: Wenn *über* regiert ist, ist es seine Entsprechung, wenn es sich dabei um eine primäre Präposition handelt, wahrscheinlich auch. Obwohl es einen Zusammenhang zwischen Sprache und Rektionsstatus dahingehend zu geben scheint, dass die primären Präpositionen in den englischen Sätzen etwas weniger häufig regiert sind als

¹⁰Vornehmlich und nicht ausschließlich ist das der Fall, weil es auch Beispiele gibt, bei denen ein Element, das nicht als Entsprechung von *über* klassifiziert werden sollte, eine Bedeutung trägt, die einen Rückschluss auf die Bedeutung von *über* zulässt. Auch diese Fälle sollen untersucht werden.

über, kann nicht gezeigt werden, dass dieser statistisch signifikant ist. Unterschiede zwischen den Bedeutungen und Entsprechungen in Bezug auf den Rektionsstatus in beiden Sprachen werden betrachtet und es wird untersucht, ob ein Zusammenhang zwischen der Formulierungsnähe und Entsprechung, Bedeutung oder Rektionsstatus besteht.

Abschnitt 5 enthält dann Überlegungen zum Bedeutungsinventar, die auf während der Annotation der deutschen Daten aufgetretenen Problemen und Beobachtungen fußen. Im ersten Unterabschnitt geht es um die Emotionsgegenstandslesart von *über*, welche als solche bis jetzt zwar in der Literatur zu psychologischen Prädikaten beschrieben worden ist, in der Literatur zu Präpositionsbedeutungen aber nicht als eigenständige Lesart herausgestellt wird. Der zweite Unterabschnitt hat dann eine Abgrenzung der beiden Lesarten *Bezugspunkt* und *Thema* zum Ziel, deren Unterscheidung sich bei der Annotation des Öfteren als schwierig erwiesen hat. Resultat der entsprechenden Betrachtungen auf Basis von sowohl Korpusdaten als auch konstruierten Beispielen wird der Vorschlag sein, die Unterscheidung aufzugeben und nur eine Lesart anzunehmen. Mit der Lesart *Traverse < 3D*, die bei Beispielen wie *Ich fahre über Berlin nach Prag* vorliegt, befasst sich der dritte Unterabschnitt. Es soll dafür argumentiert werden, dass diese Lesart in drei aufgeteilt werden sollte.

In Abschnitt 6 sollen mögliche Anwendungen der Bedeutungs-Entsprechungs-Abbildungen und der hier vorliegenden Annotationen im Allgemeinen dargestellt werden. Das sind neben der Verwendung zur Präpositionsbedeutungsdisambiguierung (6.1) und bei der Erstellung von Einträgen in bilingualen Wörterbüchern (6.2) auch die Anwendung der neu eingeführten Methode der *Suche über Entsprechungen*. Sie wird in Abschnitt 6.3 vorgestellt und besteht darin, in einem bilingualen Korpus eine anderssprachige Entsprechung zu nutzen, um nach Vorkommen einer bestimmten Bedeutung zu suchen. Unterabschnitt 6.4 stellt dann Nutzungsmöglichkeiten für die Annotationen der deutschen Sätze vor.

Abschnitt 7 fasst die Ergebnisse zusammen und behandelt mögliche Erweiterungen des Korpus. Anhang A enthält ergänzende Tabellen und Abbildungen, Anhang B das Abkürzungsverzeichnis.

1.3. Verwandte Arbeiten

Die erste mir bekannte Arbeit, in der ein bilinguales Korpus genutzt wird, um die anderssprachigen Entsprechungen einer Präposition zu untersuchen, ist (Schmied, 1998). Schmied betrachtet dort die Übersetzungen von *with* in einem deutsch-englischen parallelen Korpus. Er unterscheidet dabei sechs verschiedene Übersetzungsmöglichkeiten und untersucht, wie oft von ihnen in verschiedenen Textsorten Gebrauch gemacht wird: *mit*, eine andere Präposition, ein Adjektiv, eine Nullübersetzung (darunter scheinen etwa unsere AMV-Fälle, vgl. Abschnitt 2.2.3, zu fallen) und eine Nichtübersetzung (die Bedeutung wird im deutschen Satz einfach nicht ausgedrückt). Die sechste Kategorie ist eine Restkategorie, in die die Beispiele zu fallen scheinen, die sich nicht anders kategorisieren lassen. Man bemerke, dass er einfach davon ausgeht, dass *mit* die prototypische deutsche Übersetzung von *with* ist, und unter den anderen Präpositionen nicht weiter unterscheidet. Es wird dann dargestellt, wie sich die Übersetzungsmöglichkeiten auf die verschiedenen Textsorten und syntaktischen Strukturen, in denen *with* auftritt, verteilen; eine Untersuchung, welche Übersetzungen bei welchen Bedeutungen gewählt werden, erfolgt aber nicht. Trotzdem vertritt Schmied die Ansicht, dass es sinnvoll sein könnte, den Eintrag für *with* in einem englisch-deutschen Wörterbuch nicht nach Bedeutungen der Präposition sondern nach prototypischer Verwendungsweisen zu ordnen, d. h. zuerst nach syntaktischer Umgebung und dann nach semantischer Kategorie von internem Argument und Regens. Das führt dann etwa dazu, dass er

in seinem Beispieleintrag eine Verwendungsweise führt, in der das Regens ein Verb und das Dependens ein nominales Element ist, das einen Grund angibt, und eine andere Verwendungsweise, in der das Dependens ebenfalls einen Grund angibt und das Regens ein Adjektiv ist, das ein Gefühl beschreibt. Im ersten Fall werden *vor* und *aus* als Übersetzungsmöglichkeiten vorgeschlagen, im zweiten *vor*. Eine Untersuchung, wie wir sie hier für *über* vornehmen wollen, hätte wahrscheinlich ergeben, dass *vor* eine der möglichen Entsprechungen von *with* in dessen kausaler Lesart ist (wobei zunächst unklar wäre, ob man diese Lesart, die oft bei psychologischen Prädikaten vorzukommen scheint, wirklich als kausal bezeichnen wollen wird – die Bedeutungen und Bedeutungsunterschiede von Präpositionen bei psychologischen Prädikaten sind für das Deutsche und Englische noch nicht hinreichend erforscht, vgl. Abschnitt 5.1). In jedem Fall ist es bei der Betrachtung anderer Präpositionen, bei denen es keine derart dominierende Entsprechung gibt, nicht sinnvoll, die Bedeutungen auszublenden und alle Präpositionen bis auf eine in einer Kategorie zusammenzufassen.

So werden bei der Untersuchung der englischen Entsprechungen der tschechischen Präpositionen *v/ve* und *po* von Klégr und Malá (2009) sowohl die Bedeutungen als auch die verschiedenen englischen Präpositionen berücksichtigt, aber auch nichtpräpositionale Entsprechungen werden untersucht. Es zeigt sich, dass es für beide Präpositionen deutlich mehr Übersetzungsmöglichkeiten gibt als die in bilingualen Wörterbüchern aufgeführten, dass aber jeweils eine Entsprechung deutlich frequenter ist als die anderen. Bei *v/ve* deckt diese auch mehr Lesarten ab als die anderen Entsprechungen, bei *po* jedoch nicht. Klégr und Malá annotieren auch die syntaktische Funktion der tschechischen PPen und finden heraus, dass es bei adverbialen PPen und solchen, die als postnominale Modifikatoren fungieren, häufiger präpositionale Entsprechungen gibt. In ihrer semantischen Analyse betrachten sie dann nur noch adverbiale PPen.¹¹ Sie nutzen in einem einsprachigen tschechischen Wörterbuch aufgeführte Bedeutungen für die Annotation und erstellen Kreuztabellen, in denen aufgeführt ist, wie oft welche Entsprechung bei welcher Bedeutung vorkommt, die sich also als Bedeutungs-Entsprechungs-Abbildungen interpretieren lassen.

Weitere relevante Arbeiten sind: (Biljetina, 2015), eine an (Klégr & Malá, 2009) angelehnte Untersuchung serbokroatischer Entsprechungen von *of*; (Cosme & Gilquin, 2008), ein Vergleich der Verwendung von *with* und *avec*; (Malá, 2017), eine Untersuchung nichtpräpositionaler englischer Übersetzungsmöglichkeiten von vier tschechische Präpositionen bzw. mit ihnen gebildeten PPen, in der aufgezeigt wird, dass diese aus einer naiven Sicht überraschenden Übersetzungen zuweilen durch typologische Unterschiede zwischen den beiden Sprachen bedingt sind,¹² (Prange & Schneider, 2018), eine Annotationstudie, bei der auf einem kleinen deutsch-englischen parallelen Korpus alle Präpositionen (109 im Deutschen, 168 im Englischen) mit einer Bedeutungsannotation versehen wurden.

Auch wenn in großen deutsch-englischen Wörterbüchern zumindest für die Einträge von *über* augenscheinlich (noch) kein Gebrauch davon gemacht wird (vgl. Abschnitt 6.2), ist der Nutzen der Verwendung paralleler bilingualer Korpora für die Erstellung von Wörterbüchern in der metalexikographi-

¹¹Die Einschränkung auf adverbiale PPen ist dadurch motiviert, dass die Präposition in ihnen semantisch nicht von einem anderen Element abhängt und sie öfter mit einer PP übersetzt wurden, weswegen davon auszugehen sei, dass eine größere strukturelle Ähnlichkeit besteht. Dem ist entgegenzusetzen, dass es gerade die vermeintlich nichtlexikalischen Verwendungsweisen sind, bei denen sich besonders interessante und bisher unentdeckte Muster finden lassen werden, weil man etwa bisher davon ausgegangen sein wird, dass sie bedeutungslos seien, nur weil sie registriert sind.

¹²Auf eine feinere (und syntaktische) Annotation entsprechender Fälle, wie Malá (2017) sie vornimmt, wurde hier verzichtet, es sei jedoch darauf verwiesen, dass das vorliegende Korpus genutzt werden kann, um relevante Beispiele zu finden, für welche dann auch schon Bedeutungsannotationen vorliegen, sodass es sich als Ausgangspunkt für ähnliche Analysen eignet.

schen Literatur grundsätzlich anerkannt (Marjanović et. al., 2018) und es gibt auch bereits beispielhafte Anwendungsversuche (vgl. u. a. Škrabal & Vavřín, 2017, und die dort zitierte Literatur; Marjanović et. al., 2018). Ein Beispiel der Verwendung eines parallelen Korpus zur Erstellung eines Lexikoneintrags für eine Präposition ist mir allerdings nicht bekannt. Zwar wären dies aufgrund der Polysemie der Präpositionen, die die Erstellung von Lexikoneinträgen erschwert, besonders interessant, allerdings besteht gerade bei ihnen das Problem, dass die automatisch erstellten Wortalignierungen (auf denen Tools wie Multilingwis (Graën et. al., 2017; Institut für Computerlinguistik, Universität Zürich, o. D.) oder Treq (Škrabal & Vavřín, 2017; Vavřín & Rosen, 2015), mit denen sich nach Häufigkeit geordnete Listen von Übersetzungsmöglichkeiten in bilingualen Korpora einfach und bequem erstellen lassen, aufbauen) oftmals inkorrekt sind.¹³ Es müsste also auf händisch annotierte Daten, wie sie hier für *über* vorliegen, zurückgegriffen werden.

1.4. Verfügbarkeit der Daten und Skripte und Anmerkungen zur Darstellungsform

Alle Annotationen sind neben den zur Auswertung¹⁴ und den zur Erstellung der Annotationsdateien verwendeten Skripten verfügbar unter https://github.com/Linguistic-Data-Science-Lab/Master-s_thesis_translational_equivalents_ueber.

Nummerierte Beispiele mit aus dem Korpus stammenden Daten haben im Weiteren das folgende Format:

(Beispielnr.) ID[: optionale Anmerkungen]

(de) Dies ist ein Satz ÜBER „über“.

(en) This is a sentence ABOUT “über”. [optional]

Die ID bezieht sich immer auf das ganze Satzpaar, die zuweilen auf sie folgenden Anmerkungen enthalten in der Regel die im Kontext relevanten Werte, mit denen das Satzpaar annotiert wurde, und ggf. Kommentare, die deren Annotation begründen. Ist der englische Satz nicht relevant für die Diskussion, wird er weggelassen. *Über* und alle Wörter, die mit ihm automatisch aligniert wurden (vgl. Abschnitt 2.1.3), stehen in Majuskeln. Alle Sätze werden so wiedergegeben, wie sie in den Daten vorhanden sind, was z. B. bedeutet, dass wegen der Tokenisierung auch vor Satzzeichen ein Leerzeichen steht.

2. Die Erstellung des Korpus

2.1. Grundlage

Die Annotationsdateien wurden auf Grundlage des CoStEP-Korpus (Graën et. al., 2014) erstellt, aus dem automatisch deutsch-englische Satzpaare extrahiert wurden, bei denen der deutsche Satz *über* ent-

¹³Eine genauere Untersuchung kann hier nicht erfolgen, aber bei *at* – einer etwas untypischeren Entsprechung von *über* – sind die ersten 16 der bei (Vavřín & Rosen, 2015) nach Suche nach *über* im deutsch-englischen Teil und Click auf „at“ aufgeführten Alignierungen falsch.

¹⁴Die statistische Auswertung der Daten erfolgte mit R-Version 3.6.3 (R Core Team, 2020). Dabei wurden folgende Pakete verwendet: *readxl* (Wickham & Bryan, 2019) zum Einlesen von Excel-Dateien, *xtable* (Dahl et. al., 2019) für die Erstellung von \LaTeX -Tabellen, *vcd* (Meyer et. al., 2020) für die Berechnung von Cramér's V , *extrafont* (Chang, 2014) für die Änderung der Schriftart in ausgegebenen Diagrammen und Abbildungen, *goolevis* (Gesmann & de Castillo, 2011) für die Erstellung von Sankey-Diagrammen, *caret* (Kuhn, 2020) für die Kreuzvalidierung in Abschnitt 6.1 und *naivebayes* (Majka, 2019) für den Bayes-Klassifikator dort, *MRCV* (Kozioł & Bilder, 2014) für Unabhängigkeitstest für Variablen, bei denen die Auswahl mehrerer Werte möglich ist.

hält. Für diese Satzpaare wurden automatisch Wortalignierungen durchgeführt (die dann später bei der Annotation zu Entsprechungen korrigiert wurden) und auf Grundlage der Alignierungen die zu annotierenden Beispiele ausgewählt.

2.1.1. CoStEP

Alle annotierten Satzpaare stammen aus CoStEP (Graën et. al., 2014), einer bereinigten Version des Europarl-Korpus (Koehn, 2005), welches Protokolle des Europäischen Parlaments enthält. Europarl bzw. CoStEP ist geeignet, da es aus professionell übersetzten, parallelen Texten in mehreren Sprachen¹⁵ besteht und frei verfügbar ist. Es ist im XML-Format verfügbar, wobei es für jeden Sitzungstag des Parlaments eine eigene Datei gibt. Die Beiträge liegen dort in den verschiedenen Sprachen vor und sind in Absätze untergliedert. URLs, Zitate etc. sind mit XML-Tags gekennzeichnet.

2.1.2. Satzalignierung

Mittels eines Python-Skripts wurde für jede der XML-Dateien aus CoStEP für beide Sprachen jeweils eine Datei erstellt, die den Text der jeweiligen Sprache aus den Beiträgen ohne diese weiteren Kennzeichnungen (aber mit dem entsprechenden Text) enthält. Dabei wurden nur Beiträge übernommen, die in beiden Sprachen vorliegen, und diese mittels PUNKT (Kiss & Strunk, 2006) unter Verwendung des NLTK (Bird & Klein, 2009) satz- und mittels des spaCy-Tokenisierers (Explosion AI, o. D.) wortweise tokenisiert. In den Ergebnisdateien befindet sich ein Satz pro Zeile und alle Worttoken sind leerzeichengetrennt. Absatzgrenzen sind gekennzeichnet, was für die Satzalignierung relevant ist. Diese wurde mit Hunalign (Varga et. al., 2007) unter Verwendung der Wörterbücher von LF Aligner (Farkas, 2019) durchgeführt, wobei in die Ergebnisdatei nur Eins-zu-eins-Zuordnungen übernommen wurden.¹⁶ Grund dafür ist, dass Wortalignierungsanwendungen – so auch das hier verwendete efmara (Östling & Tiedemann, 2016) – in der Regel Satzpaare als Ausgangsdaten benötigen. Zwar gehen auf diese Weise Sätze, die *über* enthalten, verloren, da aber ohnehin nur ein geringer Teil der potenziell annotierbaren Belege auch tatsächlich annotiert wurde, verursacht dies keinen Schaden. Mit den Satzalignierungen wurde für jede Sitzung und Sprache eine Datei erstellt, wobei in der deutschen Datei nur die Sätze enthalten sind, die eine Eins-zu-eins-Entsprechung im Englischen haben, und in der englischen Datei die Zeilennummern der Sätze mit denen ihrer deutsche Pendant übereinstimmen. Dann wurden die Dateien für die verschiedenen Sitzungen zusammengefügt, sodass eine Datei pro Sprache entstand. Aus diesen Dateien wurden diejenigen Sätze extrahiert, bei denen die deutsche Version *über* enthält, und in eine Datei pro Sprache geschrieben. Das bildete die Basis für die Wortalignierung.

2.1.3. Wortalignierung

Automatische Wortalignierungstools versuchen für Sätze in verschiedenen Sprachen Alignierungen zwischen den einzelnen Wörtern herzustellen, sodass die Wörter, die einander entsprechen, miteinander aligniert sind. Hier wurde efmara (Östling & Tiedemann, 2016) verwendet, da es Ergebnisse

¹⁵Auch wenn in dieser Arbeit ausschließlich das Deutsche und Englische betrachtet werden, ist es sinnvoll, ein Korpus zu verwenden, das auch Daten für andere Sprachen enthält, da eine Ausweitung der Annotation auf weitere Sprachen erstrebenswert wäre (vgl. Abschnitt 7.2.2).

¹⁶Hunalign wurde auf alle Sitzungstagsdateien einzeln angewandt, anstatt etwa eine große Datei mit allen Daten zu erstellen, da dadurch eine Fehlerquelle eliminiert wird: Auf diese Weise ist es ausgeschlossen, dass Sätze über Sitzungsgrenzen hinweg aligniert werden.

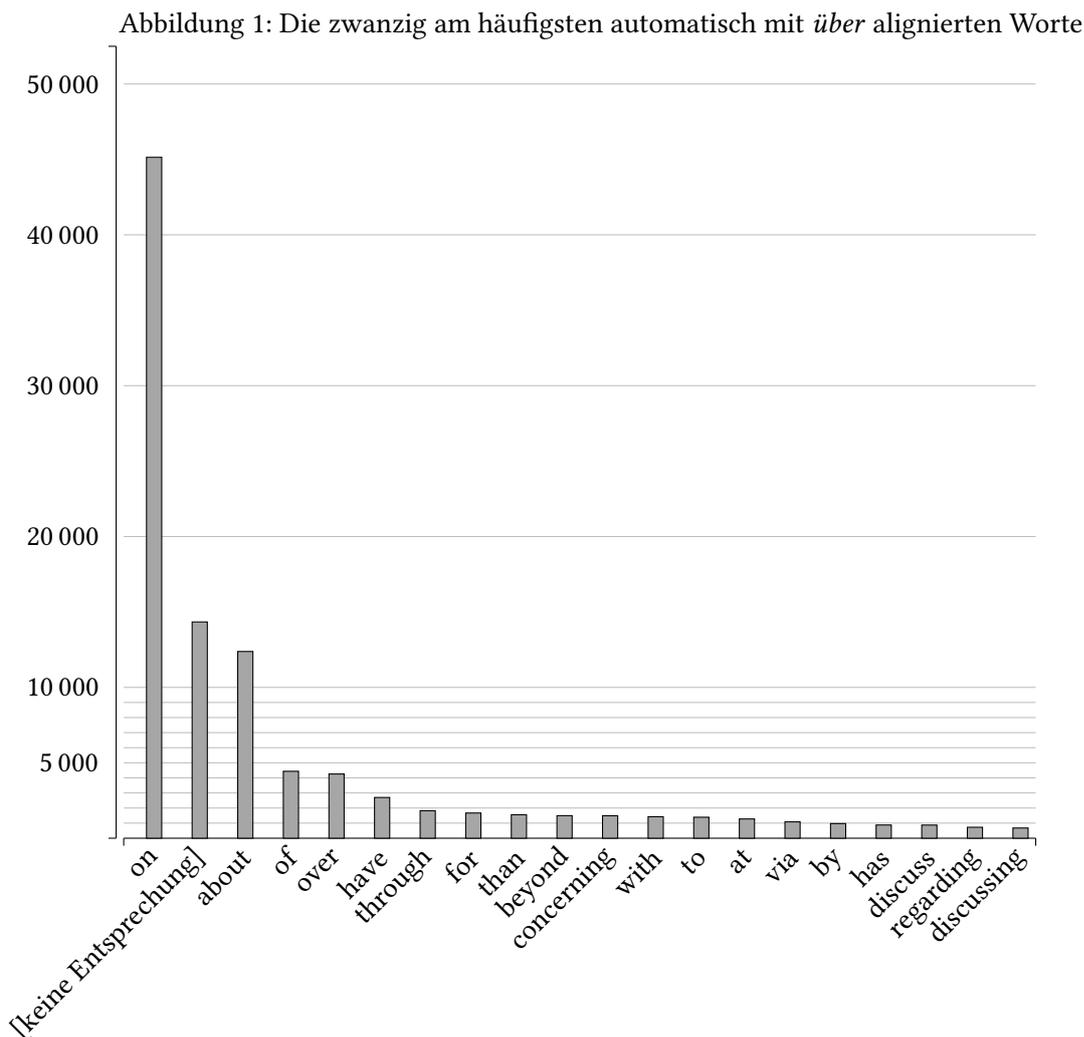
liefert, die mit denen des Standardtools GIZA++ mit den Standardeinstellungen aus (Och & Ney, 2003) vergleichbar sind, aber wesentlich schneller ist als dieses und bessere Ergebnisse und eine höhere Geschwindigkeit liefert als das ebenfalls populäre *fast_align* (Dyer et. al., 2013) (Östling & Tiedemann, 2016).¹⁷ Es wurde in zwei Punkten von den *efmaral*-Standardeinstellungen abgewichen: Die Anzahl der Sampler wurde auf vier erhöht, da eine Erhöhung der Sampleranzahl der empfohlene Weg ist, die Alignierungsqualität auf Kosten von Speicherbedarf und Verarbeitungszeit zu verbessern (Östling & Tiedemann, 2016, S. 141), das *efmaral*-Readme aber darauf hinweist, dass ein Wert, der größer als vier ist, wahrscheinlich nicht sinnvoll ist. Außerdem wurde die Anzahl der Iterationen im Vergleich zur automatisch bestimmten verdoppelt. In den Experimenten aus (Östling & Tiedemann, 2016) führt eine Erhöhung der Anzahl der Iterationen bei manchen Sprachpaaren zu einer Verbesserung der Alignierungsqualität, bei manchen zu einer leichten Verschlechterung.

2.1.4. Beispielauswahl

Insgesamt lagen 104 007 Sätze, die mindestens einmal *über* enthalten, mit englischer Übersetzung und Wortalignierung vor. In diesen wurde *über* 116 959 Mal verwendet. Diese Zahlen beziehen sich immer auf die Wortform *über*: Nicht immer handelt es sich dabei um eine Präposition, Vorkommen bei denen das nicht der Fall war, wurden aber erst bei der Annotation manuell ausgeschlossen. Das Pronominaladverb *darüber* wurde nicht betrachtet. Die automatisch alignierten Entsprechungen¹⁸ sind sehr unterschiedlich frequent: Während die häufigste Entsprechung, *on*, 45 145 Mal aligniert wurde (und damit in ca. 38,6 % der Fälle die automatisch alignierte Entsprechung ist), wurden 760 der 1497 unterschiedlichen Entsprechungstypen nur einmal mit *über* aligniert. Wie häufig die zwanzig am häufigsten alignierten Entsprechungen vorkommen, wird in Abbildung 1 veranschaulicht. Im Anhang findet sich in Tabelle 29 (S. 71) eine Auflistung der hundert am häufigsten automatisch mit *über* alignierten Worte nebst jeweiliger Alignierungshäufigkeit, Prozentanteil an der Gesamtanzahl der *über*-Vorkommen und kumulierten Prozentanteilen. Während das an zwanzigster Stelle der Liste stehende Wort *discussing* immerhin noch 676 Mal mit *über* aligniert wurde, war das beim an hundertster Stelle stehenden *given* nur 36 Mal der Fall, was einem Anteil von ca. 0,03 % entspricht. Viele der automatisch gefundenen Entsprechungen sind unerwartet: Während einige jedoch Kandidaten für AMV-Fälle (vgl. dazu die Abschnitte 1.1.3 und 2.2.3) zu sein scheinen (wie z. B. *have*, das auch tatsächlich sehr häufig als Übersetzung von *verfügen über* vorkommt. Sowohl *verfügen* als auch *über* werden mit *have* aligniert und da *have* dann mit *über*

¹⁷Man könnte meinen, dass die Wortalignierungsqualität eine untergeordnete Rolle spielt, da die Entsprechungen ohnehin manuell korrigiert werden (was allerdings wiederum der nicht ausreichenden Qualität der Wortalignierungen geschuldet ist), aber weil die zu annotierenden Beispiele nach der gefundenen englischen Entsprechung ausgewählt werden werden, ist die Güte der Alignierungen durchaus nicht unwichtig. Eine Möglichkeit, die Qualität zu erhöhen, wäre, mehrere Alignierungstools zu verwenden und dann nur die Alignierungen zu betrachten, die von allen Tools vorgeschlagen werden. Auf diese Weise behandelt Multilingwis² (Graen et. al., 2017) das FEP9-Korpus (Das Vorgehen ist meines Wissens nach noch in keiner Veröffentlichung zu Multilingwis beschrieben worden, wird aber auf der Website (Institut für Computerlinguistik, Universität Zürich, o. D.) für das FEP9-Korpus – im Gegensatz zum standardmäßig eingestellten FEP6-Korpus – genannt.). Der Nachteil bei diesem Verfahren, ist, dass der Gewinn an Precision mit einem Verlust an Recall erkauft wird: Gerade bei selteneren Entsprechungen ist die Wahrscheinlichkeit hoch, dass sie nicht von allen Tools erkannt werden. Da gerade die selteneren Entsprechung potenziell unerwartet und somit interessant sind, wurde nicht auf diese Weise verfahren. Eine andere Möglichkeit wäre gewesen, die Vereinigung der Alignierungen der Tools zu nehmen, allerdings wären dann Alignierungen mehrfach vorhanden gewesen, was die Beispielauswahl verkompliziert hätte. In Anbetracht dessen, dass die Alignierungen ohnehin korrigiert werden, hätte dieses Verfahren zu einem unangemessenen Mehraufwand geführt.

¹⁸In diesem Abschnitt wird *Entsprechung* sowohl für tatsächliche Entsprechungen im Sinne von Abschnitt 1.1.3 als auch für die automatisch mit *über* alignierten Worte verwendet. Wo eine Verwechslungsgefahr besteht, ist von *automatisch alignierten Entsprechungen* oder ähnlichem die Rede.



aligniert ist, wird es als Entsprechung gelistet), handelt es sich bei anderen (wie z. B. dem Komma oder *is*) schlicht um Alignierungsfehler.

Die Beispiele für die Annotation wurden aus den alignierten Daten nicht rein zufällig ausgewählt, denn ein solches Vorgehen hätte zur Folge gehabt, dass in ungefähr der Hälfte der Beispiele die automatisch gefundene Entsprechung *on* oder *about* ist. Beide Worte sind aber – wie vor der Annotation zu vermuten war und sich im Nachhinein bestätigt hat – ausschließlich in der Bezugspunkt- und der Themaesart (vgl. zu diesen beiden Lesarten (Kiss et. al., 2020) und Abschnitt 5.2) Entsprechungen von *über*.¹⁹ Es hätte also sehr viele gleichartige Beispiele gegeben, die für wenig Erkenntnisgewinn viele Ressourcen (in Form von für die Annotation aufgewandter Zeit) verschlungen hätten. Stattdessen wurde jeweils eine bestimmte Anzahl an Beispielen pro Entsprechung annotiert. Die Entsprechungen, für die Annotationen vorgenommen werden sollten, wurden aus den hundert am häufigsten alignierten ausgewählt.²⁰ Wie ein Blick auf Tabelle 29 (S. 71 im Anhang) zeigt, sind viele der automatisch alignier-

¹⁹Zumindest wurden für diese beiden Entsprechungen nur diese beiden Bedeutungen annotiert. Da die achsenbezogene Lesart im Korpus unterrepräsentiert ist, ist aber nicht auszuschließen, dass sich auch für sie noch Beispiele finden lassen.

²⁰Die Beschränkung auf die häufigsten hundert wurde nach Betrachtung der Liste der am häufigsten alignierten Worte und Wortgruppen getroffen: Auf der einen Seite gilt, dass die Wahrscheinlichkeit, dass es sich um eine fehlerhafte Alignierung handelt, tendenziell umso höher ist, je seltener eine vermeintliche Entsprechung automatisch aligniert wurde (und fehlerhafte Alignierungen müssen bei der Annotation manuell korrigiert werden). Auf der anderen Seite sind gerade die

ten Worte bzw. Wortgruppen offensichtlich keine Entsprechungen von *über*, sondern es liegt schlicht eine fehlerhafte Alignierung vor. Diese vermeintlichen Entsprechungen wurden nicht in die Annotationsdateien übernommen.²¹

Die Annotation wurde in drei Phasen durchgeführt, wobei es in Phase 2 und 3 im Vergleich zu Phase 1 Unterschiede in Bezug auf die Beispielauswahl gab: So wurde etwa erst während der ersten Annotationsphase festgestellt, dass Sätze mehrfach im Korpus vorhanden sind.²² Dies hat dazu geführt, dass sich auch unter den für die Annotation ausgewählten Belegen Sätze doppelt fanden (sie wurden dann aber nur einmal annotiert und nur einmal in die Datei mit allen annotierten Sätzen übernommen). In der zweiten und dritten Annotationsphase wurden dann nur Sätze in die Beispielauswahl aufgenommen, die nicht bereits vorhanden waren. Ein weiterer Unterschied zwischen der ersten Annotationsphase und den beiden anderen ist, dass nach der ersten Phase darauf verzichtet wurde, Beispiele für vermeintliche Entsprechungen zu annotieren, bei denen höchstwahrscheinlich eine Alignierung mit Verb vorliegt, also im Englischen ein Verb mit einer Objekt-NP verwendet wird, der beim deutschen Verb eine *über*-PP entspricht (vgl. Abschnitt 2.2.3). Grund dafür ist, dass die englischen Verben meist recht eindeutig einer bestimmten deutschen Konstruktion zugeordnet werden können: So ist z. B. *have* eine Entsprechung von *verfügen über*. Da eine weitere Annotation der AMV-Kandidaten also kaum Erkenntnisgewinn gebracht hätte, wurde auf sie verzichtet.²³

Decision und *discussion*, bei denen – wie die erste Annotationsphase gezeigt hat – fehlerhafte Alignierungen zu erwarten gewesen wären wurden ebenfalls nicht mehr berücksichtigt. Das gleiche gilt für *more* und *more than*, die als Entsprechungen des Adverbs *über* auftreten.

Technisch durchgeführt wurde die Beispielauswahl mittels eines Pythonskripts. Die Auswahl der zu annotierenden Belege erfolgt dort pseudozufällig (wobei das Setzen eines *random seed* die Reproduzierbarkeit garantiert) und es wird sichergestellt, dass jedes Beispiel nur einmal vorkommt, indem solange

selteneren Entsprechungen potenziell interessanter. Die hundert am häufigsten alignierten Entsprechungskandidaten zu betrachten, scheint ein guter Kompromiss zu sein, was sich im Nachhinein darin zeigt, dass einige der Entsprechungskandidaten kurz vor der Grenze (wie z. B. *around* und *throughout*) noch in vielen Fällen als Entsprechung übernommen wurden, während das bei anderen (wie etwa *access* und *thought*) nicht der Fall ist (vgl. Tabelle 33 auf S. 76 im Anhang). Da die Alignierungen manuell korrigiert werden, können auch Worte bzw. Wortgruppen, die sich nicht unter den ersten hundert finden, später als Entsprechungen vorkommen. Das ist auch tatsächlich der Fall: *as to* kommt als Entsprechung fünf Mal vor (vgl. Tabelle 32 auf S. 74 im Anhang), steht aber mit fünf automatischen Alignierungen erst auf Rang 365 in der Häufigkeitsliste. Grund dafür ist, dass *as* – wie Tabelle 33 auf S. 76 und Abbildung 6 auf S. 91 zeigen – oft zu *as to* korrigiert wurde.

²¹Es handelt sich um: *and, any, are, available, be, being, between, could, cross, does, find, hear, how, informed, is, know, must, need, one, on on, provide, put, report, 's, see, setting, take, than, that, was, way, what, where, which, will*, Doppelpunkt, Komma. Manche Übersetzungsmöglichkeiten, die auf den ersten Blick nicht sinnvoll zu sein scheinen, sind es nichtsdestotrotz, was aber erst klar wird, wenn man sich konkrete Beispiele ansieht. Vor Ausschluss eines Kandidaten wurden stets mindestens vier Beispiele betrachtet. Bei *on on* muss es sich nicht um einen Alignierungsfehler handeln, denn, wenn z. B. eine PP, die im Deutschen eine aus zwei koordinierten NPen bestehende NP enthält, im Englischen durch zwei koordinierte PPen wiedergegeben wird, wird *über* mit beiden *ons* aligniert. Trotzdem wird es ausgeschlossen, da es nur eine Variante von *on* darstellt und keine Entsprechung in unserem Sinne ist.

²²Insbesondere kurze oder formelhafte Sätze (z. B. *Nach der Tagesordnung folgt die gemeinsame Aussprache über folgende Berichte*;) können durchaus an mehreren Stellen im Korpus vorkommen. Aber auch längere Sätze können mehrfach vorhanden sein: Der Satz *Der genetische Determinismus - die Auffassung, daß alles über unsere physische und mentale Ausstattung über unsere DNA identifiziert werden kann - ist ein äußerst gefährlicher Weg* findet sich z. B. bereits in den ursprünglichen CoStEP-XML-Dateien doppelt (und zwar in der Sitzung vom 18.07.1996 in Turn 22 und 28). Die Sprecherinnen sind unterschiedlich, gehören aber beide der Fraktion der Grünen an. Solche Dopplungen sind keineswegs selten und sie treten auch häufig in Abschnitten auf, die sehr nah beieinander liegen, weswegen nicht davon auszugehen ist, dass es sich um Fehler im Korpus handelt. Denkbar wäre etwa, dass beide Sprecherinnen auf dieselbe Textvorlage zurückgegriffen haben.

²³Das betrifft: *adopting, aware, consider, considering, covering, dealing, debate, debated, debating, determine, decide, deciding, discuss, discussed, discussing, enjoy, exceed, exceeds, given, going, had, has, have, having, hold, laying, possess, transcends, using, welcome*

ein neues Zufallsbeispiel gezogen wird, bis ein noch nicht ausgewähltes getroffen wurde.²⁴ Die Ausgabe der Beispiele erfolgte pro Annotationsphase in eine TSV-Datei im in Abschnitt 2.2 beschriebenen Format, die dann in einem Tabellenkalkulationsprogramm – verwendet wurde Microsoft Excel – geöffnet und bearbeitet werden konnte. Nach Abschluss der Annotation wurden die Inhalte der drei Dateien in eine einzige überführt. Nachträgliche Korrekturen wurden dann ausschließlich in dieser durchgeführt.

2.2. Das Annotationsschema

Neben den Bedeutungen sollen auch noch weitere Eigenschaften der deutsch-englischen Satzpaare annotiert werden, die für die Anwendungen relevant sein könnten: So könnte es z. B. sein, dass der Rektionsstatus einen Einfluss darauf hat, welche englischen Entsprechungen möglich sind. Da sich die Formulierungen in den deutschen und den englischen Sätzen des Korpus mal mehr und mal weniger stark ähneln, sollte auch dieser Faktor berücksichtigt werden, denn es ist vorstellbar, dass bestimmte englische Worte oder Wortgruppen nur dann als vermeintliche Entsprechungen von *über* auftreten, wenn sich die Formulierungen recht stark unterscheiden oder sie nah beieinander sind.

Die Annotation wurde in Excel-Tabellen durchgeführt, in denen die Spalten für die verschiedenen zu annotierenden Kategorien stehen und sich in den Zeilen deutsche und englische Sätze abwechseln, sodass auf einen deutschen Satz stets dessen englisches Pendant folgt. Tabelle 1 enthält eine Übersicht über die verschiedenen Spalten der Annotationsdateien, die hier in den Zeilen aufgeführt sind. Da nicht alle Kategorien für beide Sprachen relevant sind, steht in der zweiten Spalte der Tabelle auch, für welche Sprache jeweils annotiert wurde.

Tabelle 1: Die Spalten der Annotationsdateien

Kategorie	Sprache	Erläuterung
Sprache	de, en	Sprachkürzel
ID	de, en	Satznummer in der Datei mit allen <i>über</i> -Sätzen
Sätze	de, en	Deutscher bzw. englischer Satz
Meta	de, en	Metainformationen
Bedeutung	de en	Bedeutungen nach (Kiss et. al., 2020) Deutsche Bedeutungsannotation passend?
regiert	de, en	Rektionsstatus
Formulierung	en	Wie nah ist die Formulierung in der englischen Zeile an der der deutschen?
Alignierung	de en	<i>Über</i> Die automatisch alignierte Entsprechung
Entsprechung	en	Die korrigierte Entsprechung, ggf. AMV
Kommentar	de, en	

Die ID wird für Satzpaare vergeben, ist also für den deutschen und den englischen Satz identisch und richtet sich nach der Nummer des deutschen Satzes in der Datei, in der alle *über* enthaltenden Sätze stehen (s. dazu Abschnitt 2.1.2).

²⁴Das ist natürlich nur möglich, wenn es mindestens so viele unterschiedliche Beispiele gibt, wie Belege gezogen werden sollen, was aber aufgrund der eher geringen Zahl an Belegen pro automatisch alignierter Entsprechung kein Problem darstellte. Dass sichergestellt wird, dass jeder Beleg nur einmal vorhanden ist, ist erst seit der zweiten Annotationsphase der Fall. In der ersten Phase mehrfach in die Annotationsdatei aufgenommene Sätze wurden jeweils nur einmal annotiert (s. o.).

In den folgenden Unterabschnitten werden die einzelnen Kategorien vorgestellt.

2.2.1. Metainformationen

Bei der Metainformationskategorie gibt es einige Werte, die nur für den deutschen Satz annotiert werden, und einige, die für Sätze beider Sprachen Verwendung finden. Viele der Werte orientieren sich an obligatorischen Kommentaren, die von Kiss et. al. (2020) verwendet wurden.

Tabelle 2 enthält die möglichen Werte für Metainformationsannotationen. In der zweiten Spalte ist angegeben, für welche Sprachen die Werte verwendet werden können. Ein Kreuz (×) in der dritten Spalte bedeutet, dass eine Annotation dieses Wertes den Ausschluss des Beispiels nach sich zieht, was zum sofortigen Abbruch der Annotation für das Beispiel führte. Das betrifft die Werte *adv*, *raus*, *vübz*

Tabelle 2: Die möglichen Metainformationswerte

Tag	Sprache	Raus	Erklärung
<i>adv</i>	de	×	Die PP muss mit einem nicht oder nur mit Bedeutungsänderung weglassbaren Adverb stehen (z. B. <i>über ... hinaus</i>). Auch Beispiele, in denen es sich bei dem vermeintlichen Adverb eigentlich um eine Verbpartikel handelt, die aber die gleiche Form wie ein entsprechendes Adverb hat, fallen in diese Kategorie.
<i>idi</i>	de, en		Idiom, feste Wendung oder Funktionsverbgefüge
<i>postpos</i>	de		Es handelt sich um eine Postposition
<i>raus</i>	de	×	Der Satz sollte aus einem anderen Grund nicht betrachtet werden. Entsprechende Beispiele sind mit einem erklärenden Kommentar zu versehen.
<i>sa</i>	de, en		Syntaktisch ambig. Syntaktische Ambiguität wird nur dann annotiert, wenn sich Lesarten mit unterschiedlichem Sinn ergeben, die im Kontext sinnvoll sind.
<i>tit</i>	de, en		Präposition steht in einem Titel (etwa eines Kunstwerks)
<i>tg</i>	de, en		Telegrammstil
<i>vübz</i>	de	×	Konstruktion der Form <i>von ... über ... bis</i> oder <i>von ... über ... zu</i>
<i>~P</i>	de	×	Es handelt sich nicht um eine Adposition (sondern z. B. um ein Adverb).

und *~P*²⁵. Bei allen bis auf *raus* ist der Grund für den Ausschluss, dass sie auch von Kiss et. al. (2020) ausgeschlossen wurden und ihre Bedeutungen demnach nicht im Bedeutungsinventar erfasst sind.²⁶

²⁵In den entsprechenden Fällen (wie z. B. (5)) ist *über* ein Adverb, was sich darin zeigt, dass es weglassbar ist, durch Adverbien ersetzt werden kann und nicht den Kasus der NP bestimmt (vgl. Schröder, 1986, S. 12).

²⁶Eine Erweiterung des Bedeutungsinventars, die es ermöglicht hätte, dass auch diese Fälle erfasst werden können, wäre mit erheblichen Aufwand verbunden gewesen. Eine informelle Betrachtung entsprechender Beispiele ergibt gleichwohl, dass sich auch für einige dieser Fälle sinnvolle englische Entsprechungen finden ließen: So wird *über* als Adverb oft mit *more than*, *above* oder *over* übersetzt, *über hinaus* mit *beyond*.

Beispiele für *adv*, *vübz* und *~P* sind (3), wo *hinweg* nicht weggelassen werden kann, (4) und (5):

- (3) 82755 [auszuschließen wegen *adv*]
 (de) Eine Politik ÜBER die Nachbarn hinweg ist uneuropäisch und daher in meinen Augen inakzeptabel .
- (4) 72733 [auszuschließen wegen *vübz*]
 (de) Die Zahl der Möglichkeiten , über die Förderung und die Achtung der Menschenrechte zu sprechen , hat zugenommen - von Argentinien und Bangladesh ÜBER Indien , Jordanien , Marokko bis hin zu Turkmenistan und Vietnam .
- (5) 98877 [auszuschließen wegen *~P*]
 (de) Der jährliche Gesamtumsatz der pharmazeutischen Industrie in Europa beträgt ÜBER 170 Mrd. EUR .

Neben diesen systematischen Ausschlussgründen gibt es auch eine Reihe von anderen. Um diese zu erfassen, wurde der Wert *raus* eingeführt. Da es sich um eine Restkategorie handelt, mit der ganz verschiedenartige Fälle abgedeckt werden, wird stets in einem Kommentar der genaue Grund für den Ausschluss angegeben. (6) enthält zwei Beispiele:

- (6) Beispiele für *raus*-Fälle
- a. 30512
 (de) Wir reden heute ÜBER DAPHNE , das Programm für Maßnahmen zur Bekämpfung der Gewalt gegen Kinder , Jugendliche und Frauen .
 (en) A DECISION is urgently needed on this programme , in order to ensure continuity when the present period of support comes to an end .
- b. 63560
 (de) Was mich wirklich gegen den Bericht - in seiner endgültigen Fassung - aufgebracht hat , ist das Eintreten für Standards , die weit ÜBER das Mindestmaß im Falle eines Freispruchs oder die Anwesenheit eines Psychiaters .

In (6a) liegt offensichtlich ein Fehler bei der Satzalignierung vor. Dies war insgesamt fünfmal (also in ca. 0,79 % der Beispiele) der Fall. In (6b) dagegen besteht das Problem schon im deutschen Satz, da er unvollständig ist. Zwar ist das fehlende Wort leicht zu erraten, aber da der Satz in der im Korpus befindlichen Form ungrammatisch ist, wurden für ihn keine Annotationen vorgenommen. Außerdem wäre er auch mit vorhandenem *hinausgehen* auszuschließen gewesen – dann allerdings mit dem Wert *adv*.

tit und *tg* wurden in Anlehnung an Kiss et. al. (2020, S. 428 ff.) vor der Annotation eingeführt, aber nicht ein einziges Mal verwendet. Ebenfalls nicht zur Anwendung kam *postpos*, das zur Kennzeichnung postpositionaler Verwendung von *über* – wie in dem konstruierten Beispiel in (7) – gedacht war, wofür auch eine eigene Bedeutungskategorie vorgesehen war (vgl. Abschnitt 2.2.4).

- (7) Die ganze Nacht über hat er Lieder geträllert.

2.2.2. Entsprechung

In der Entsprechungsspalte der englischen Zeile wurde zur Verringerung des Annotationsaufwands nur dann etwas eingetragen, wenn die automatisch alignierte Entsprechung nicht korrekt war. Das kann etwa daran liegen, dass die Alignierung wie in (8) grundsätzlich falsch ist oder dass wie in (9) zu wenig oder wie in (10) zu viel aligniert wurde.²⁷

(8) 22229

(de) Es ist eine schwierige Aufgabe , Berichterstatter zum Haushalt zu sein , wenn es um wirklich sensible Bereiche geht , ÜBER die das Parlament zu entscheiden hat .

(en) It is not an easy task to be rapporteur on the budget when there are really sensitive issues AROUND that the House has to decide on .

(9) 72056

(de) Der Vorsitz ist erfreut , dass das Protokoll ÜBER nicht detonierte Sprengkörper in Kraft getreten ist .

(en) The Presidency is pleased that the Protocol RELATING to unexploded ordnance has entered into force .

(10) 87031

(de) Anfangs sprach Hugo Chávez nicht ÜBER Sozialismus , sondern nur über das Recht auf eine bessere Welt .

(en) At first , Hugo Chávez did not TALK ABOUT socialism but only about the right to a better world .

Als korrekte Entsprechungen von *über* werden nicht nur prototypische (primäre) Präpositionen gelistet, sondern auch sekundäre Präpositionen und alles, was semantisch die gleiche oder eine ähnliche Relation wie *über* ausdrückt und aus adjazenten Einheiten besteht. So werden z. B. *by means of*, *relating to* und *concerning* als ganz normale Entsprechungen betrachtet. Das ist zum einen sinnvoll, da für Bedeutungen, die im Deutschen mit *über* ausgedrückt werden, im Englischen (und auch im Deutschen selbst, was aber hier nicht von Belang ist) teilweise komplexe Einheiten verwendet werden, zum anderen, da eine Abgrenzung zwischen sekundären Präpositionen und anderen Konstruktionen schwierig ist. Die Entsprechungen sollten sich auf die gleiche Weise wie prototypische Präpositionen erfassen lassen und sind dann für viele Zwecke (z. B. zur Präpositionsbedeutungsdisambiguierung) genauso hilfreich. Auch für Lerner ist es genauso nützlich zu wissen, dass *by means of* eine Übersetzung von *über* sein kann, wie zu wissen, dass das bei *over* der Fall ist. Bei Bedarf können die Entsprechungen auch im Nachhinein noch nach morphologischer Form, syntaktischem Verhalten etc. gruppiert werden.

Nimmt man an, dass bei Übersetzungen die Bedeutung des in einer Sprache Gesagten auch in der anderen Sprache ausgedrückt werden soll und dass man sich innerhalb einer Sprache, aber insbesondere auch über verschiedene Sprachen hinweg grundsätzlich unterschiedlicher Mittel bedienen kann, um Bedeutungen auszudrücken, ist vorderhand nicht ersichtlich, wieso man sich auf diese Art von Entsprechungen beschränken sollte: Aus der Lernerperspektive etwa ist es nämlich auch sicherlich ebenso

²⁷Die korrekten Entsprechungen wären *on* in (8), *relating to* in (9) und *about* in (10). Wie in Abschnitt 1.4 erwähnt, stehen die automatisch alignierten Wörter in Majuskeln.

hilfreich zu wissen, dass einer Präpositionalphrase mit *über* wie in (11) ein Teil eines Kompositums entsprechen kann.

(11) 103305

(de) Nach der Tagesordnung folgt die Aussprache über sechs Entschließungsanträge zu Guantánamo : unmittelbar bevorstehende Entscheidung ÜBER ein Todesurteil .

(en) The next item is the debate on six motions for resolution on Guantánamo : imminent death penalty decision .

Die Entscheidung, auch derartige Fälle zu behandeln, würde den Annotationsaufwand erheblich erhöhen: Nicht nur werden diese von Wortalignierungssystemen meistens nicht erkannt und müssten also sehr oft manuell korrigiert werden, sie bedürften zu ihrer angemessenen Erfassung auch der Einführung weiterer Annotationskategorien, mit denen etwa angegeben würde, um was für eine Konstruktion es sich genau handelt, und es ergäben sich in Bezug auf das Annotationsschema weitere Probleme, die gelöst werden müssten. So müsste z. B. für jede infrage kommende Konstruktion eine Literaturrecherche zu ihrer Semantik durchgeführt werden und die Zuweisung eines Formulierungsnähewertes erwiese sich als noch schwieriger als sie es ohnehin schon ist (vgl. Abschnitt 2.2.6).²⁸ Deswegen werden im Allgemeinen nur diejenigen nichtpräpositionalen Ausdrücke betrachtet, die die oben genannten Kriterien erfüllen und sich darum für unsere Zwecke wie Präpositionen behandeln lassen. Davon wird in zwei Fällen abgewichen, nämlich zum einen bei den Fällen, für die die Formulierungsnähe­kategorie *4n* vorgesehen ist (s. Abschnitt 2.2.6), und zum anderen, wenn eine Alignierung mit Verb vorliegt.

2.2.3. Ein Spezialfall: Alignierung mit Verb (AMV)

In einigen Fällen gibt es keine direkte Entsprechung von *über*, weil das englische Pendant des deutschen Wortes, das eine PP mit *über* als Komplement hat, nicht eine PP, sondern eine NP selegiert. Ein typisches Beispiel ist (12), wo *gesprochen über discussed* entspricht:

(12) 34338

(de) Ich habe mit Präsident Mugabe ÜBER AIDS in Afrika gesprochen .

(en) I have DISCUSSED AIDS in Africa with President Mugabe .

Zwar könnte im Deutschen die Präposition teilweise ebenfalls fehlen (etwa bei *diskutieren*) und im Englischen könnte auch *about* stehen,²⁹ entscheidend ist aber, dass es viele Fälle gibt, bei denen genau das erwähnte Muster vorliegt. Hier wurde in der Spalte für die Entsprechung *AMV* (für „Alignierung mit Verb“) eingetragen bzw. bei Beispielen, bei denen die Alignierung mit dem Verb noch nicht automatisch gegeben ist, *AMV [Verb]*. Später bei der Auswertung konnte dann bei den Beispielen mit „korrekt“ aligniertem Verb dieses automatisch übertragen werden. Sowohl für Lerner als auch zur Präpositionsbedeutungsdisambiguierung sind Listen der englischen Prädikate, bei denen das eine bestimmte semantische Rolle, die im Deutschen durch eine PP mit *über* ausgedrückt wird, tragende Argument als NP realisiert wird, potenziell interessant. Auch wenn das Regens der deutschen PP meistens ein Verb ist,

²⁸Bei *argumental compounds* wie in (11en) etwa ist schon innerhalb des Englischen das Verhältnis zur phrasalen Paraphrase keineswegs offensichtlich (vgl. Bauer et. al., 2013, S. 467).

²⁹Zumindest für *talk* gibt es Ansätze, die zu erklären versuchen, von welchen Faktoren die Realisierung von *about* abhängt, z. B. (Clark et. al., 2019). Da beide Formen (mit und ohne overt realisiertes *about*) Übersetzungen von *sprechen über* sein können, wird auf die Unterschiede im Englischen nicht eingegangen.

muss das nicht der Fall sein: Entscheidend ist, dass die NP im Englischen die thematische Rolle, die im Deutschen durch die PP ausgedrückt wird, durch das Verb zugewiesen bekommt.³⁰

2.2.4. Bedeutung

Bedeutungen wurden für die deutschen Vorkommen von *über* nach dem (geringfügig ergänzten) Schema von Kiss et. al. (2020) in der Bedeutungsspalte des deutschen Satzes annotiert. Wenn mehrere Bedeutungen vorlagen, wurden diese durch Kommata getrennt.³¹ Dann wurde für jede annotierte Bedeutung geprüft, ob sie auch im Englischen passend ist, wobei in der Bedeutungsspalte der Zeile des englischen Satzes nur etwas eingetragen wurde, wenn nicht alle Bedeutungen möglich waren, um den Annotationsaufwand geringer zu halten. In diesem Fall wurde eine durch Kommata getrennte Liste von *n* (nicht passend) und *p* (passend) annotiert, wobei sich ein an Stelle *i* dieser Liste stehendes *p* oder *n* auf das *i*-te Element der Bedeutungsliste bezieht. *p, n* etwa bedeutet also, dass die erste der annotierten Bedeutungen auch für das englische Beispiel passend war, die zweite aber nicht. Es ist auch möglich, dass diese Liste nur aus *n* besteht. Das ist dann der Fall, wenn es im englischen Satz ein Element gibt, das im konkreten Fall durchaus eine ähnliche Bedeutung wie *über* ausdrückt und demnach durchaus als Entsprechung zu betrachten ist, für das die gewählte Bedeutungsannotation aber unpassend ist. Ein Beispiel ist (13), wo für *über* eine Bezugspunktlesart annotiert wurde:

(13) 82895

(de) Abgesehen von Flugverspätungen betreffen die meisten Beschwerden ÜBER Fluggesellschaften , die ich erhalte , die Art und Weise , in der diese mit niedrigen Ticketpreisen locken , nur um dem Käufer praktischerweise erst ganz zum Schluss den tatsächlichen Preis mitzuteilen.

(en) Apart from flight delays , the most complaints that I receive AGAINST airlines concern the way that they lure people with what appear to be low fares and conveniently tell them the true cost only at the very end of the process.

Eine Liste der bei der Annotation verwendeten Tags findet sich in Tabelle 3. *+m* und *+z* sind keine eigenen Lesarten, sondern werden an den Lesart-Tag angehängt, wenn die Lesart metaphorisch bzw. zielbezogen ist, wobei *+z* nur bei spatialen Lesarten verwendet wird. Für Beschreibungen der einzelnen Lesarten und für Beispiele sei auf (Kiss et. al., 2020) verwiesen. An dieser Stelle soll nur auf die Abdeckungslesart (*Abd*), die zusätzlich zu denen aus (Kiss et. al., 2020) erfasst wurde, und die kausale Lesart (*4a*) eingegangen werden.

Als eigene Kategorie wurde die Abdeckungslesart eingeführt, um Beispiele wie die in (14) erfassen zu können.

³⁰Steht im Deutschen kein Verb muss allerdings – da das Regens im Englischen ja ein Verb ist – der Formulierungsnäherwert 3 (s. dazu Unterabschnitt 2.2.6) zugewiesen werden.

³¹Mehrfachannotationen von Bedeutungen wurden zum einen vorgenommen, wenn eine (syntaktische) Ambiguität vorlag, zum anderen aber auch, wenn mehrere der Bedeutungsbeschreibungen aus (Kiss et. al., 2020) passend schienen, ohne dass der Satz erkennbar ambig wäre. Letzteres war hauptsächlich bei den Bedeutungen *Bezugspunkt* und *Thema* sowie bei verschiedenen modalen Bedeutungen der Fall.

Tabelle 3: Die bei der Annotation verwendeten Bedeutungs-Tags

Tag	Bedeutung
1a	Spatial Traverse Grenzbereich
1b	Spatial Traverse proximal
1c	Spatial Traverse vertikal
1d	Spatial Traverse <3D
1e	Spatial Bedeckung
1f	Spatial Achsenbezogen
2a	Temporal Tagesteil
2b	Temporal Zeitdauer
2c	Temporal Maßeinheit
3a	Modal Instrumental
3b	Modal Informationsübermittler
3c	Modal Medial
3d	Modal Abstraktes Instrument
3e	Modal Art und Weise
4a	Konditional Kausal / Emotionsgegenstand
4b	Konditional im engeren Sinne
5	Bezugspunkt
6	Machtverhältnis
7	Maßangabe
8	Rangfolge
9	Thema
10	Überschreitung
11	Zuordnung
Abd	Abdeckung
d	Desemantisiert
+m	Metaphorisch
+z	Zielbezogen

(14) Beispiele für *Abd*

a. 87628

(de) Den Gaspipeline- und Stromnetzen , die sich ÜBER ganz Europa erstrecken , muss Vorrang eingeräumt werden .

b. 8293

(de) Ich möchte nochmals dafür plädieren , daß die Mittel gleichmäßiger ÜBER das Jahr verteilt verwendet werden ; wenn nämlich zuviel auf das Ende des Jahres verschoben wird , besteht die Gefahr , daß die Sorgfalt dadurch beeinträchtigt wird .

c. 35534

(de) Die spanische Flotte verteilt sich ÜBER eine sehr lange Liste von Ländern , unter ihnen Belize , Honduras , die Niederländischen Antillen , Panama und viele andere .

Gemein ist diesen Beispielen, dass das Referenzobjekt einen (physischen oder nichtphysischen) Bereich abdeckt, wobei diese Abdeckung im Unterschied zur Bedeckungslesart bei Kiss et. al. (2020) nicht als

vollständige Bedeckung zu verstehen sein muss, bei der „das LO [das zu verortende Objekt] das RO [Referenzobjekt] für einen möglichen Beobachter verdeckt“ (S. 176). Auch wenn eine spatiale Metaphorik meist klar erkennbar ist, wird diese Lesart nicht unter den spatialen geführt, da es, wie (14b–c) zeigen, nicht immer eine spatiale Abdeckung sein muss. In allen annotierten Belegen ist das Regens der *über*-PP mit dem Stamm *verteil* oder *erstreck* gebildet, dies ist jedoch nicht notwendigerweise der Fall: Im Artikel zu *über* im Deutschen Wörterbuch (DWB) ist eine ähnliche Bedeutungsangabe geführt: „mit dem nebenbegriff der flächenhaften erstreckung“ (Grimm & Grimm, 1936, Sp. 80). Einer der Belege ist folgendes Goethe-Zitat (Goethe, 1820/1906, S. 247 f.):

- (15) Bey Eckartsberge liegen Granitblöcke, [... (Auslassung wie im DWB)]; über ganz Thüringen sind dergleichen ausgesät

Der Tag für die kausale Lesart (4a) wird gemäß der Beschreibung der entsprechenden Lesart bei Kiss et. al. (2020) verwendet, allerdings sei darauf hingewiesen, dass hier – wie in Abschnitt 5.1 gezeigt werden soll – zwei verschiedene Lesarten unter einem Oberbegriff zusammengefasst werden. Zwar haben beide eine kausale Komponente, aber bei der wesentlich häufiger vorkommenden kann es trotzdem irreführend sein, sie ohne weiteren Kommentar als kausal zu bezeichnen, da *über* in ihr verwendet wird, um bei psychologischen Prädikaten das Emotionsgegenstandsargument auszudrücken, und sich die Emotionsgegenstandsrolle von der durch *durch* ausgedrückten Verursacherrolle unterscheidet. Deswegen wird diese Lesart im Folgenden nicht *Konditional Kausal*, sondern *Konditional Kausal / Emotionsgegenstand* genannt.

Mit regierten Präpositionen wird hier etwas anders umgegangen als bei Kiss et. al. (2020): Während dort bei regierten Präpositionen mit Restbedeutung sowohl diese Bedeutung als auch ein Tag für regierten Gebrauch und bei vollständig desemantisierten Präpositionen nur dieser annotiert wird, wurde hier eine eigene Spalte für den Rektionsstatus eingeführt (vgl. Abschnitt 2.2.5). Ist noch eine Restbedeutung vorhanden, so wird in der Bedeutungsspalte nur diese annotiert, ist das nicht der Fall wird *d* (für „desemantisiert“) verwendet.

Vor Durchführung der Annotation ist für Fälle, in denen *über* als Postposition verwendet wird, eine weitere Lesart angenommen (und ein entsprechender Tag eingeführt) worden, die sich dadurch auszeichnet, dass „über eine gewisse begrenzte Zeitdauer eine Kontinuität des Geschehens gegeben ist“ (Schröder, 1986, S. 177).³² In den betrachteten Daten gab es allerdings kein entsprechendes Beispiel.

Wie bereits oben erwähnt, wurde die Angemessenheit der Bedeutungsannotation für die englische Entsprechung für jede Bedeutung einzeln geprüft, wenn für das Deutsche mehrere Bedeutungen annotiert wurden. Die Annotation mehrerer Bedeutungen kann mehrere Ursachen haben: In (16) etwa führt eine syntaktische Ambiguität zu verschiedenen Lesarten: In der (prominenteren) Themasart ist die PP Teil der *Debatte*-NP, in der modalen Lesart Teil der VP. Doch Ambiguitäten sind nicht immer syntaktisch begründet. In (17) etwa gibt es eine Bezugspunkt- und eine Informationsübermittlerlesart, ohne dass eine PP-Anbindungsambiguität bestünde. Die Ambiguität zwischen der kausalen und der Vertikale-Traversen-Lesart wie etwa in (18) ist – auch wenn in diesem Fall beide metaphorisch sind – bei *über* systematisch (Kiss et. al., 2020, S. 188), was dann zu einer systematischen Mehrfachannotation führt. Anders gelagert ist der Fall bei (19): Hier liegen mit 3a und 3b zwei verschiedene modale Lesarten vor. Die Annotation von verschiedenen Unterbedeutungen derselben Oberbedeutung war nur bei den

³²Vgl. auch Beispiel (7) auf S. 15.

Oberbedeutungen erlaubt, für die bei Kiss et. al. (2020) kein Entscheidungsbaum vorliegt, nicht also bei der spatialen und der temporalen Oberbedeutung.

(16) 32234

(de) Doch die Politik nimmt Rache , und nachdem sie aufgrund einer Debatte ÜBER eine beschränkte Agenda durch das Fenster hinausgejagt wurde , kommt sie in Wahrheit durch die Tür wieder herein , und zwar durch die Vordertür in Gestalt des Falls Österreich , denn dort gelangt eine Partei an die Regierung , die sich auf Intoleranz , Fremdenfeindlichkeit und Formen des Rassismus zu berufen scheint .

(17) 50476

(de) Über die Rahmen , die Werkzeuge , wissen wir , dass mehr als 60 % der Bürger Informationen ÜBER das Fernsehen bevorzugen .

(18) 34223

(de) Mit dieser Einsicht und dem hehren Anspruch , die Sünden der Vergangenheit nicht zu wiederholen - schließlich ist vor knapp einem Jahr die letzte Kommission ÜBER Betrugs- und Unregelmäßigkeitsfälle zu Fall gekommen - , ist die Kommission Prodi im September angetreten .

(19) 73660

(de) Wir sollten meines Erachtens von Ihnen das Gleiche hören , wie es auf dem Bildschirm angezeigt und ÜBER den Kopfhörer angesagt wird .

2.2.5. Rektionsstatus

Der Rektionsstatus wurde für beide Sprachen annotiert (wobei bei den englischen Sätzen selbstredend nur präpositionale Entsprechungen überhaupt regiert sein können). Dabei wurde *d* eingetragen, wenn die Präposition desemantisiert ist, *r*, wenn sie regiert, aber nicht desemantisiert ist, und nichts, wenn sie nicht regiert ist. Gab es mehrere Lesarten, wurde eine durch Kommata getrennte Liste gebildet, die nötigenfalls den Wert *n* (für „nicht regiert“) enthält. In ihr bezieht sich ein an Position *i* stehender Wert auf das *i*-te Element der Bedeutungsliste. So kann etwa bei einem syntaktisch ambigen Beispiel, in dem zwei Lesarten von *über* vorhanden sind, der Rektionsstatuswert *n, r* sein, weil *über* in der zweiten Lesart regiert wäre, in der ersten aber nicht.

Präpositionen werden nur als desemantisiert eingestuft, wenn keinerlei Restbedeutung mehr erkennbar ist, d. h., wenn keine der Bedeutungen aus Tabelle 3 passt und es auch nicht der Fall zu sein scheint, dass die Bedeutung einfach nicht im Schema erfasst ist. Festzustellen, dass eine Präposition desemantisiert ist, ist keineswegs eine triviale Aufgabe: Bei *verfügen über* etwa gibt es Fälle wie (20), bei denen man eine Machtverhältnis-Lesart annotieren wollen wird, und Fälle wie (21), die man wahrscheinlich (entgegen Meex (2001)³³) als desemantisiert betrachten sollte.

³³Meex (2001) zitiert einen ähnlichen Beleg aus einem Korpus und will ihn in ihrem kognitiv-linguistischen Ansatz als Beispiel dafür verwenden, dass Besitz als Kontrolle (was unserer Machtverhältnislesart entspräche) betrachtet werden kann, was über eine KONTROLLE-IST-OBEN-Metapher die Verwendbarkeit von *über* erklären soll. Allerdings zeigen derartige Beispiele gerade, dass Besitz *nicht* immer als Kontrolle verstanden werden kann. Das zeigt sich auch darin, dass Sätze wie *Die Königin verfügt über 3000 Soldaten, aber sie verfügt nicht über sie* möglich sind. Dort ist *über* im ersten Teil desemantisiert und hat im zweiten eine Machtverhältnislesart. Wären beide Lesarten identisch, sollte der Satz widersprüchlich sein.

(20) 85769

(de) Hitlerdeutschland verfügte ÜBER 400 einsatzbereite Fallschirmjäger , Stalin über mehr als 1 Million .

(21) 70502

(de) Wir benötigen Finanzinstrumente , die ÜBER ausreichende Mittel zur Erfüllung dieser Ziele verfügen .

Eine Annotation mit *r* war bei einer nicht desemantisierten Präposition obligatorisch, wenn sie nicht austauschbar zu sein schien,³⁴ auch wenn das schwer zu prüfen ist: Da daraus, dass einem Annotator keine Alternativen einfallen, nicht folgt, dass es keine gibt, wurden im Zweifelsfall der Sketch-Engine-Wordsketch (Kilgarriff et. al., 2014)³⁵ des Regens der Präposition betrachtet. Wordsketches sind nach syntaktischen Funktionen geordnete Kollokationslisten für ein gegebenes Wort.³⁶ In ihnen sind bei Sketch Engine auch die Präpositionen, die mit dem gesuchten Wort im jeweiligen Korpus auftreten, gelistet. Diese Listen konnten genutzt werden, um nach Alternativen zu suchen (wobei im Zweifelsfall Belege geprüft wurden, um sicherzustellen, dass keine Annotationsfehler im Korpus vorlagen); dass sich dort keine Alternative findet, bedeutet jedoch ebenfalls nicht, dass es keine gäbe: Zu dem grundsätzlichen Problem, dass aus dem Nichtvorkommen einer Konstruktion in einem Korpus nicht auf deren Nichtexistenz geschlossen werden kann, kommt hinzu, dass (fast) nur monolexematische Präpositionen überhaupt als solche geführt werden. So wäre etwa nicht festzustellen, dass *mittels* eine Alternative zu *über* in instrumentaler Lesart sein kann. Dass ein Kopf dort nur mit einer bestimmten Präposition auftritt, heißt also nicht, dass diese grundsätzlich nicht austauschbar ist, auch wenn es ein Indiz dafür sein kann.

Außerdem wurde *r* annotiert, wenn das Regens in einer der Listen für regierten Gebrauch in (Kiss et. al., 2020) eingetragen ist oder es in einem Lexikon als die Präposition regierend geführt wird, was, da es erheblich einfacher und schneller festzustellen ist, vor der Nichtkommutierbarkeit überprüft wurde. Die Liste mit *über* regierenden Köpfen aus (Kiss et. al., 2020) basiert auf Listen aus (Helbig & Buscha, 2007; Hertel, 1983; Schmitz, 1981; Weinrich, 2005) und ist um einige eigene Einträge ergänzt. Da die Einträge der Liste aus unterschiedlichen Quellen stammen, bei denen wiederum oft die zur Bestimmung des Rektionsstatus verwendeten Kriterien nicht klar sind, folgt die Liste nicht unbedingt einheitlichen Kriterien und es wird auch kein Anspruch auf Vollständigkeit erhoben (Kiss et. al., 2020, S. 19). In der Praxis hat sich gezeigt, dass die Liste recht großzügig ist und *über* auch oftmals als regiert betrachtet, wo Nichtkommutierbarkeit nicht unbedingt gegeben ist. Trotz dieser Probleme wurde auf die Liste und Wörterbücher zurückgegriffen, da eine Anwendung der üblichen Rektionstests, die oftmals selbst problematisch sind (vgl. Abschnitt 1.1.1 und die dort zitierte Literatur), bei allen Beispielen erheblich aufwendiger gewesen wäre: Allein eine Überprüfung der Nichtkommutierbarkeit hätte aufgrund der oben dargestellten Schwierigkeiten erheblich mehr Zeit pro Beispiel in Anspruch genommen als die Bestimmung von Bedeutung und Formulierungsnähe. In Bezug auf das Englische kommt hinzu, dass die Durchführung von Tests muttersprachliche Intuitionen bei subtilen Unterschieden erfordert hätte. Da

³⁴Dass dieses oft verwendete Kriterium nicht unproblematisch ist, haben wir in Abschnitt 1.1.1 gesehen.

³⁵Zugriff unter (Lexical Computing, o. D.). Betrachtet wurden die Wordsketches auf Basis der Korpora enTenTen15 (für das Englische) und deTenTen13 (für das Deutsche; s. für beide (Jakubíček et. al., 2013))

³⁶Sie ähneln DWDS-Wortprofilen (Didakowski & Geyken, 2014) stark. Anfangs ist erwogen worden, diese zu verwenden, aber da sie nur Informationen für das Deutsche bereitstellen und Sketch-Engine-Wordsketches für das Deutsche eine sehr ähnliche Funktionalität bieten, wurde davon abgesehen.

der einzige Annotator kein englischer Muttersprachler war, konnte das nicht geleistet werden. Neben der Liste aus (Kiss et. al., 2020) wurde – insbesondere für das Englische, für das eine solche Liste nicht vorlag – auch auf Wörterbücher zurückgegriffen. Verwendet wurden E-VALBU (Institut für Deutsche Sprache, o. D.) für das Deutsche und (Herbst et. al., 2004) sowie in geringerem Maße Merriam Webster (Merriam Webster, o. D.), das OED (OED, o. D.) und der PONS (PONS, o. D. a) für das Englische. Dabei wurde in Bezug auf die Einstufung einer Präposition als regiert sehr großzügig vorgegangen, da das auch bei der Liste aus (Kiss et. al., 2020) der Fall ist: Auch wenn z. B. in Valenzwörterbüchern mehrere Präpositionen mit ähnlicher Bedeutung gelistet sind, mit denen das (vermeintlicherweise) regierende Element auftritt, wurde die Präposition als regiert gelistet. So werden etwa die Präpositionen in *debate on/about/over* alle als regiert geführt.

Um Einheitlichkeit³⁷ zu gewährleisten, wurden Listen mit den Köpfen, die *über* bzw. dessen englische Entsprechungen regieren bzw. nicht regieren, geführt. Diese finden sich im Anhang in Tabelle 30.

Insgesamt wurde also ein sehr enger (für *d*) und ein sehr weiter (für *r*) Rektionsbegriff verwendet.

2.2.6. Formulierungsnähe

In einem mehrsprachigen parallelen Korpus sind die einzelnen Texte alle mehr oder weniger frei übersetzt. Dies führt dazu, dass *über* nicht immer eine Entsprechung im in Abschnitt 2.2.2 definierten Sinne hat. Doch selbst wenn es eine direkte Entsprechung gibt, heißt das noch lange nicht, dass internes und externes Argument der Präposition wörtlich übersetzt sind. Hier kann es große Unterschiede geben. Auch wenn die Präpositionsbedeutung gleich zu sein scheint, sollte das Vorhandensein dieser Unterschiede erfasst werden, da etwa denkbar ist, dass es einen Zusammenhang zwischen der gewählten Entsprechung und der Formulierungsnähe gibt oder bei bestimmten Bedeutungen die Formulierungen in den beiden Sprachen einander ähnlicher sind. Dem wird in Abschnitt 4.1.2 nachgegangen werden.

Damit überhaupt eine Auswertung stattfinden kann, muss der Formulierungsnähebegriff zunächst operationalisiert werden. Dazu wurden vier Grade der Formulierungsnähe eingeführt, die in Tabelle 4 aufgeführt sind. Diese Tabelle stellt nur die zugrunde liegenden Intuitionen für die einzelnen Werte

Tabelle 4: Formulierungsnähewerte

Wert	Erläuterung
1	Keine unterschiedliche Formulierung in Bezug auf die PP und ihr Regens
2	Es gibt eine unterschiedliche Formulierung innerhalb der PP oder in Bezug auf ihr Regens, aber der Unterschied ist gering und für die Bedeutung von <i>über</i> bzw. dessen Entsprechung irrelevant.
3	Es gibt größere Unterschiede in der Formulierung innerhalb der PP oder in Bezug auf ihr Regens, aber es liegt eine Entsprechung vor und die Formulierung ist zumindest verwandt.
4	Es liegt eine grundsätzlich andere Struktur vor.
4n	Es liegt eine grundsätzlich andere Struktur vor, aber die automatisch alignierte Entsprechung ermöglicht einen Schluss auf die Bedeutung im Deutschen.

³⁷Einheitlichkeit bedeutet nicht, dass es nicht Worte geben kann, mit denen *über* je nach Lesart sowohl regiert als auch nicht regiert auftritt. Das ist etwa bei *gehen über* der Fall. Vielmehr sollte verhindert werden, dass ähnliche Beispiele unterschiedlich annotiert werden, wenn eine Abdeckung durch die Liste von Kiss et. al. (2020) und Wörterbücher nicht gegeben ist.

dar und bedarf einiger Klarstellungen und Erläuterungen: Wenn in der Erläuterungsspalte von PPen die Rede ist, so sind nicht nur Präpositionalphrasen gemeint, sondern alle Arten von Phrasen, die der deutschen *über*-PP entsprechen.³⁸ Besonders unpassend ist die Bezeichnung als PP bei den AMV-Fällen: Hier ist das Argument des Verbs gemeint, dass die thematische Rolle innehat, die im Deutschen mittels *über*-PP ausgedrückt wird, also im Regelfall eine NP.³⁹ Das betrifft ausdrücklich auch die Fälle mit Formulierungsnähe 3: Hier wird 3 annotiert, auch wenn gar keine Präposition vorhanden ist, deren Bedeutung gleich bleiben könnte. Eine Alternative wäre, bei diesen Fällen 4*n* zu annotieren. Dann allerdings wäre eine Parallelität zwischen AMV- und Nicht-AMV-Fällen nicht mehr gegeben.

Die Unterscheidung verschiedener Grade der Formulierungsnähe ist nicht theoretisch fundiert, sondern folgt allein praktischen Erwägungen: Wir interessieren uns bei diesem Annotationsvorhaben für die Bedeutungen von *über* und wollen diese sprachübergreifend untersuchen, indem wir dessen Entsprechungen im Englischen betrachten, es ist also mindestens eine Kategorie erforderlich, mit der erfasst werden kann, ob es eine Entsprechung gibt oder nicht. Darüber hinaus können auch die Unterschiede zwischen *über* und dessen Entsprechungen in Bezug auf die Syntax der Argumente und das dort verwendete lexikalische Material nicht einfach ignoriert werden. Diesbezüglich weitergehende Annotationen durchzuführen, würde den Aufwand jedoch unverhältnismäßig stark erhöhen. Die Einführung der Formulierungsnähekategorie ist ein Kompromiss, der es ermöglicht, den Annotationsaufwand geringer zu halten und trotzdem zu testen, ob es einen Zusammenhang zwischen Formulierungsnähe (d. h. für die Werte 1–3, die für die Analyse in erster Linie interessant sind, Syntax und lexikalisches Material der Argumente) und Bedeutung oder gewählter Entsprechung gibt.

Es folgen nun zunächst nähere Bestimmungen zur Unterscheidung der verschiedenen Formulierungsnähewerte. In (22–30) finden sich dann einige Beispiele.

2 umfasst alle Fälle von kleineren Unterschieden, mögen sie systematisch sein oder auch nicht. Hierunter fallen Beispiele, bei denen etwa eine definite anstelle einer indefiniten NP verwendet wurde, das Nomen im Plural statt Singular steht oder – wie in (23) – nur in einer Sprache ein Artikel vorhanden ist. Die Wortarten der Regentien müssen übereinstimmen und diese sollen mögliche Übersetzungen voneinander sein. Auch Diatheseunterschiede führten zur Annotation von 2, nicht 3. Innerhalb der PP sollten die Formulierungen einander ähneln, d. h., wenn es in der einen Sprache ein Adjunkt oder ein Argument des Kopfes des internen Arguments gibt, muss es in der anderen Sprache ein bedeutungsgleiches oder -ähnliches derselben syntaktischen Kategorie geben, wie es z. B. bei (24), nicht aber bei (26) der Fall ist, weswegen bei diesem 3 und bei jenem 2 annotiert wurde; die einzelnen Wörter sollten bedeutungsähnlich sein. Unterschiede in eventuell im internen Argument eingebetteten Präpositionalphrasen, Adjektivphrasen, Genitivattributen, Relativsätzen etc. – wie bei (24) – oder Abfolgeunterschiede zwischen ihnen sind nicht relevant, Unterschiede in Bezug auf das Vorhandensein oder die Art (z. B.

³⁸Syntaktisch betrachtet ist *RELATING to the assessment and management of environmental noise* in Beispiel 92019 keine PP, aber es ist die Entsprechung der deutschen *über*-PP. Da es sich in den meisten Fällen um PPen handelt, wird hier vereinfachend von PPen gesprochen.

³⁹Eine Ausnahme hiervon stellt z. B. (i) dar (wo als Formulierungsnähe 3 annotiert wurde):

(i) 67171

(de) Im Februar haben die Europäische Gemeinschaft , die Weltbank und das US-Außenministerium in Washington ein Treffen aller größeren internationalen Geber einberufen , um ÜBER Möglichkeiten zur Förderung wichtiger Staatsreformen zu diskutieren .

(en) In February the European Community , the World Bank and the US Department of State convened a meeting in Washington with all major international donors to DISCUSS how we can foster key governance reforms .

PP im Deutschen, AP im Englischen) solcher Phrasen – wie bei (26) – führten aber zur Annotation von 3. Auch müssen die Regentien mögliche Übersetzungen voneinander sein und derselben syntaktischen Kategorie angehören (da dies bei (25) nicht der Fall ist, wurde dort 3 annotiert). Ist das interne Argument in der einen Sprache ein Kompositum und in der anderen nicht, aber es handelt sich um eine übliche Übersetzung⁴⁰, ist das ebenfalls kein Grund, 3 zu annotieren. Ein Beispiel ist (23). 4 ist zu annotieren, wenn wie bei (28) die Strukturen gänzlich verschieden sind. Auch bei Fällen wie (29) wurde 4 annotiert. Ist die automatisch alignierte Entsprechung zwar keine Entsprechung im Sinne von Abschnitt 2.2.2, drückt aber eine Relation aus, die auch von *über* ausgedrückt wird, wurde 4n annotiert.⁴¹ In (30) etwa hat *über* eine modale Lesart und im Englischen wird *use* verwendet. Gibt es zwar größere Unterschiede in der Formulierung, aber es liegt trotzdem eine Entsprechung bei der Präposition vor, wurde 3 annotiert. So unterscheiden sich die internen Argumente bei (27) stark, aber die Präpositionen sind trotzdem eindeutig Entsprechungen voneinander. Hat die deutsche Präposition eine eindeutige Entsprechung, aber deren Bedeutung ist nicht passend (vgl. Abschnitt 2.2.4), wird in der englischen Bedeutungsspalte *n* eingetragen, aber bei der Formulierungsnähe 2.

(22) 7970: 1

(de) Auch ich möchte der Berichterstatterin zu ihrem ausgezeichneten Bericht über den zweiten Jahresbericht der Kommission ÜBER die demographische Lage in der Europäischen Union gratulieren .

(en) I , too , want to congratulate the rapporteur on her excellent report on the Commission 's second annual report ON the demographic situation in the European Union .

(23) 101062: 2, da Artikel fehlt, aber bei internem Argument übliche Übersetzung

(de) Die Katastrophe in Japan hat die Besorgnis der Menschen ÜBER die Kernenergiesicherheit geweckt .

(en) The disaster in Japan has awakened people 's concern OVER nuclear safety .

(24) 18200: 2, da eingebettete PP nicht ganz gleich, aber bedeutungsähnlich und tiefer eingebettete Unterschiede nicht mehr relevant

(de) Frau Präsidentin , erneut beschäftigt sich das Europäische Parlament mit einem Gemeinsamen Standpunkt des Rats ÜBER die Richtlinie zur Patentierung von biotechnologischen Erfindungen .

(en) Madam President , once again Parliament is tackling a common position of the Council CONCERNING the directive on patenting of inventions derived from biotechnology .

(25) 85709: 3, da Regens anderer syntaktischer Kategorie

(de) Eine gute Freundin hat mir gesagt , dass sie vor kurzem bei ihrem Besuch in den Vereinigten Staaten überrascht darüber gewesen sei , wie wenig man in den USA ÜBER die neuen Institutionen und neuen Verfahren weiß , die im Rahmen des Vertrags von Lissabon geschaffen werden .

⁴⁰Eine Übersetzung soll als üblich betrachtet werden, wenn sie in einem zweisprachigen Wörterbuch gelistet ist; ist das nicht der Fall, schließt das aber nicht aus, dass sie als üblich betrachtet werden kann.

⁴¹Motivation für die Einführung von 4n war die Idee, dass diese Fälle *über* genauso gut disambiguieren wie die als solche erfassten Entsprechungen. Die Einführung von 4n ist ein erster Schritt zur Erfassung nichtpräpositionaler und präpositionsunähnlicher Mittel, *über*-Bedeutungen auszudrücken.

- (en) A good friend said to me recently that she had been in the United States and she had been surprised by the lack of knowledge REGARDING the new institutions and the new procedures created by the Treaty of Lisbon .
- (26) 12166: 3, da Modifikatoren unterschiedlicher Kategorie (und *system* ohne Gegenstück im Deutschen)
- (de) Dies erfordert zahlreiche neue Beschlüsse , wobei einer der dringendsten sicherlich die Einführung einer Richtlinie ÜBER die Versicherung auf Gegenseitigkeit ist .
- (en) This implies many new decisions , of which one of the most urgent is certainly the establishment of a directive ON the mutual insurance system .
- (27) 63584: 3, da ganz anders, aber Bedeutung der Präpositionen gleich
- (de) Ich bin der Meinung , dass diese Probleme Anlass zu strukturellen Überlegungen ÜBER die Kontrolle des Haushalts geben müssen - mit Aufteilung der Kompetenzen auf die Kommission und die Mitgliedstaaten .
- (en) I think these problems should lead to reflection UPON the structures for exercising control in the area of the budget , with competence shared between the Commission and the Member States .
- (28) 82325: 4, da ganz anders
- (de) Dennoch ist dies eine Überlegung , ÜBER die wir nachdenken sollten .
- (en) However , it is surely an idea worthy OF consideration .
- (29) 83091: 4, da Passiv statt Adjektiv + PP und Bedeutung unpassend⁴²
- (de) Er hat einen Angriff gegen Südossetien vorgenommen , zu dem Herr Van den Brande , einer der Mitberichterstatter der Parlamentarischen Versammlung des Europarates , zuständig für die Überwachung der Probleme in der Region , erklärt hat , er sei - ich zitiere - bestürzt ÜBER die Berichte der Flüchtlinge über die massive und willkürliche Beschießung und Bombardierung von Tskhinvali und die Zerstörung von Wohngebieten .
- (en) He has launched an attack on South Ossetia , about which Mr Van den Brande , one of the co - rapporteurs of the Parliamentary Assembly of the Council of Europe , responsible for monitoring problems in the region , has declared having been - and I quote - ' shocked BY the stories of the refugees about the massive and indiscriminate shelling and bombing of Tskhinvali and the destruction of residential areas ' .
- (30) 71173: 4n
- (de) Die Kommission hat es nicht nötig , sich ÜBER das Buch zu refinanzieren .
- (en) The Commission does not need to USE the book as a means of recovering its costs .

⁴²Es handelt sich hier um ein etwas problematisches Beispiel: Der Grund, hier 4 zu annotieren, ist, dass im Englischen eine Passivkonstruktion vorliegt (was sich gerade in der Verwendung von *by* zeigt), im Deutschen aber ein Adjektiv mit PP-Komplement. *By* kann nicht als Entsprechung von *über* betrachtet werden, da die *über*-PP einen *Emotionsgegenstand* angibt, die *by*-PP aber einen *Verursacher* im Sinne Pesetskys (1995) (vgl. Abschnitt 5.1): In (29de) müssen die Berichte Gegenstand der Bestürzung sein, (29en) lässt dagegen auch eine (zugegebenermaßen wesentlich weniger prominente) Lesart zu, in der van den Brande nicht über die Berichte bestürzt ist, sondern nur seine Bestürzung durch sie ausgelöst wird – etwa, weil er gehofft hatte, die Gräueltaten blieben geheim. Man hätte sich hier wohl auch damit begnügen können, *n* für die Angemessenheit der Bedeutung zu annotieren, aber damit wäre nicht deutlich geworden, dass auch ganz unterschiedliche Formulierungen vorliegen (unterschiedliche Wortart des Regens) und für die Auswertung hätte es auch keinen Unterschied gemacht. Insgesamt gab es acht vergleichbare Fälle.

Formulierungsnähe meint ausdrücklich nicht nur Unterschiede, die dadurch zustande kommen, dass sich verschiedene Übersetzer unterschiedlich ausgedrückt haben, sondern bezieht grundsätzlich auch strukturelle Unterschiede zwischen Sprachen mit ein. Grund für dieses Vorgehen ist, dass es sehr schwer sein kann, zu entscheiden, welche Unterschiede regelhaft sind und welche nicht.⁴³

Es wurde nicht annotiert, ob der Unterschied in der Formulierung grammatischer oder lexikalischer Natur ist, denn hier ist eine Unterscheidung sehr schwer zu treffen und es gäbe viele Grenzfälle.⁴⁴

2.3. Durchführung der Annotation

Durchgeführt wurde die Annotation von mir selbst, einem deutschen Muttersprachler mit Englischkenntnissen. Bei einigen Beispielen war die Zuordnung zu einer der Kategorien schwierig. Dies betraf insbesondere (aber nicht ausschließlich) den Rektionsstatus (wie zu erwarten war, vgl. Abschnitt 1.1.1), die Unterscheidung der Bedeutungen *Bezugspunkt* und *Thema* und die Angemessenheit der Bedeutung für die englische Entsprechung, wobei auch hier Beispiele mit Bezugspunkt- oder Themasart besonders oft zu Problemen führten. Solche problematischen Beispiele wurden mit Dr. Claudia Roch diskutiert. Die Probleme bei der Unterscheidung der Bedeutungen *Bezugspunkt* und *Thema* gaben Anlass zu den in Abschnitt 5.2 dargestellten Überlegungen, deren Ergebnis ist, dass es sich nicht um zwei unterschiedliche Lesarten handelt und das Annotationsschema dahingehend überarbeitet werden sollte. Manchmal war es auch nicht einfach zu entscheiden, ob eine bestimmte Lesart im Englischen überhaupt vorliegen kann oder nicht. Bei künftigen Annotationsvorhaben mit ähnlichem Aufbau könnte es sinnvoll sein, Muttersprachler beider Sprachen die Beispiele gemeinsam bearbeiten zu lassen oder so vorzugehen, dass zuerst der Satz in der Sprache, für die die Bedeutungsannotationen vorgenommen werden, von einem Muttersprachler bearbeitet wird und dann ein Muttersprachler der anderen Sprache, der aber sehr gute Kenntnisse der ersten Sprache hat, die Angemessenheit der Bedeutungsannotation für den Satz der anderen Sprache beurteilt. Eine andere Änderung am Vorgehen, die für ähnliche Annotationsvorhaben sinnvoll sein könnte, ist, die IDs anders zu bestimmen: Die hier gewählte Lösung ermöglicht es zwar, dass auch Sätze, die noch nicht annotiert wurden, über eine ID verfügen, was gut ist, wenn man diese (wie in Beispiel (67) auf S. 67) zitieren will, aber sie erschwert es, den Kontext des Satzes zu finden, der bei der Annotation manchmal hilfreich sein kann. Es würde sich also etwa die Nummer des deutschen Satzes in der satztokenisierten Datei der deutschen CoStEP-Daten als ID anbieten.

Da das Bedeutungsinventar von Kiss et. al. (2020) für ein Annotationsvorhaben in Bezug auf eine ganz andere Fragestellung und für die Annotation eines Zeitungskorpus, welches sich erheblich vom hier verwendeten Europarl-Korpus unterscheiden dürfte, entwickelt wurde, war vor der Annotation keineswegs klar, dass sich das Annotationsschema problemlos würde anwenden lassen. Es gab in den annotierten Daten aber keine Fälle, in denen für das deutsche Beispiel nicht eine der Bedeutungen aus (Kiss et. al., 2020), *d* oder *Abd* zugewiesen werden konnte, weswegen festgestellt werden kann, dass das Bedeutungsinventar von Kiss et. al. (2020) auch für die Bedeutungsannotation von *über* in Parlamentsdebatten geeignet ist.

⁴³Eine eigene Kategorie für eindeutig regelhafte Fälle ist aufgegeben worden, da die Abgrenzung zu schwierig war und die Annotationsentscheidungen teilweise willkürlich wirkten.

⁴⁴Wenn etwa zwei Verben grundsätzlich Entsprechungen voneinander zu sein scheinen, sie aber unterschiedliche Subkategorisierungsrahmen haben, ist das dann ein lexikalischer oder ein grammatischer Unterschied? Vielleicht gibt es etwa (lexikalische) Unterschiede in der Ereignisstruktur, die (grammatische) Unterschiede in der Argumentstruktur bedingen, aber nur sehr schwer zu erkennen sind.

3. Das Korpus

In diesem Abschnitt sollen die annotierten Daten beschrieben werden. Er gliedert sich in drei Unterabschnitte. Da das Korpus so bearbeitet wurde, dass es pro Bedeutungsangabe einen Datenpunkt gibt (was für die Erstellung von Bedeutungs-Entsprechungs-Abbildungen sinnvoll ist), sollen sowohl die unbearbeiteten Daten (Unterabschnitt 3.1) als auch die bearbeiteten (Unterabschnitt 3.2) vorgestellt werden. In Unterabschnitt 3.3 werden dann ausschließlich aus den Annotationen für die deutschen Sätze bestehende Teilkorpora untersucht.

3.1. Vor Bearbeitung

Insgesamt wurden für 632 deutsch-englische Satzpaare Annotationen vorgenommen. Davon sind allerdings 82 auszuschließen, da der deutsche Satz mit einer der Metainformationskategorien annotiert wurde, die einen Ausschluss erfordern, also *über* keine Präposition ist ($\sim P$), mit einem nicht weglassbaren Adverb steht (*adv*), Teil einer Konstruktion der Form *von ... über ... bis* oder *von ... über ... zu* ist (*vübz*) oder der Satz aus anderen Gründen auszuschließen ist (*raus*). Wie häufig die verschiedenen möglichen Werte in den *Meta*-Spalten der deutschen Beispiele auftreten, zeigt Tabelle 5.

Tabelle 5: Die Häufigkeiten der Werte von *Meta* in den deutschen Zeilen

$\sim P$	<i>adv</i>	<i>idi</i>	<i>raus</i>	<i>sa</i>	<i>vübz</i>	[leer]
25	44	18	11	4	2	528

Man bemerke, dass die Anzahl der deutschen Sätze, bei denen sich bedingt durch eine syntaktische Ambiguität in der Anbindung unterschiedliche Lesarten für *über* ergeben (*sa*), gering ist.

Ebenfalls ausgeschlossen von der Betrachtung werden später alle Fälle, bei denen als Formulierungsnähe 4 eingetragen wurde, was bei 186 Belegen (also ca. 29,4%) der Fall war, wie aus Tabelle 6 zu entnehmen ist, in der die Frequenzen der verschiedenen Formulierungsnähe-Werte aufgeführt sind.⁴⁵

Tabelle 6: Die Häufigkeiten der Formulierungsnähewerte

1	2	3	4	4n	[ausgeschlossen]
96	94	152	186	22	82

Trotz der Bestrebungen, durch eine von der automatisch alignierten Entsprechung abhängige Beispielauswahl dafür zu sorgen, dass es für viele Bedeutungen ausreichend Beispiele gibt, sind die Beispiele sehr ungleichmäßig auf die verschiedenen Bedeutungen verteilt. Tabelle 7 zeigt die Frequenzen der elf häufigsten Annotationen in der deutschen Bedeutungsspalte (wobei zum Vergleich auch die Anzahl der auszuschließenden Beispiele aufgeführt ist), die restlichen Werte können Tabelle 31 auf S. 73 im Anhang entnommen werden, in welcher die Annotationen für die ausgeschlossenen Belege nicht aufgeführt sind. Eine Aufschlüsselung der bei der Annotation verwendeten Tags findet sich in Tabelle 3

⁴⁵Bei Beispielen, bei denen schon während der Annotation klar war, dass sie ausgeschlossen werden würden, weil in der deutschen *Meta*-Spalte ein entsprechender Wert eingetragen wurde, wurde nichts weiter annotiert und demnach auch kein Wert in die Zelle für die Formulierungsnähe eingetragen. Diese Fälle, bei denen die Zelle also leer ist, sind hier in der letzten Spalte geführt.

auf S. 19: So steht 5 für *Bezugspunkt* und 9 für *Thema*. Es wurde also in über der Hälfte der Belege entweder *Thema* oder *Bezugspunkt* oder beides annotiert; lässt man die auszuschließenden Belege außen vor sind es sogar über 60 %.

Tabelle 7: Kurzübersicht über die Häufigkeiten der Bedeutungswerte (de)

5	9	[ausg.]	5,9	d	4a	3d	2c	1d	Abd	3b	8	andere
147	133	82	55	42	40	31	14	12	11	8	7	50

Tabelle 8 enthält eine Übersicht über die in die englische Bedeutungsspalte eingetragenen Werte. Man erinnere sich, dass (wie in Abschnitt 2.2.4 dargestellt) *n* dafür steht, dass die deutsche Bedeutungsannotation im Englischen nicht passend ist, und *p* dafür, dass sie passend ist, *p* aber nur verwendet wird, wenn für das Deutsche mehrere Bedeutungen annotiert wurden, von denen mindestens eine nicht passend ist. Ist das nicht der Fall, bleibt die Zelle leer. Das heißt, dass sich die Anzahl der Fälle, in denen es nur eine deutsche Bedeutungsannotation gab, die auch für das Englische angemessen war, aus der Anzahl der leeren Zellen (594) abzüglich der Anzahl der auszusondernden Beispiele (82) ergibt und somit bei 512 liegt.

Tabelle 8: Die Häufigkeiten der Bedeutungswerte (en)

<i>n</i>	<i>n,n</i>	<i>n,p</i>	<i>p,n</i>	<i>p,n,n</i>	[leer]
25	2	6	4	1	594

Wie oft welche Entsprechungen in den annotierten Belegen vorkommen, ist in Tabelle 32 im Anhang (S. 74) dargestellt. Dabei geht AMV-Fällen ein „AMV“ voran und 4*n*-Fällen ein „4*n*“. Es sind auch die Werte aus der *Meta*-Spalte, die zum Ausschluss der Beispiele führen, und „4“ für die Formulierungsnähe 4 gelistet. Grund für eine derartige Vermengung unterschiedlicher Kategorien ist, dass Entsprechungen bei Beispielen, bei denen ohnehin klar war, dass sie ausgeschlossen werden würden bzw. unter die Formulierungsnähe 4 fallen, bei der Annotation teilweise nicht mehr korrigiert wurden. Es ist aber auch unabhängig davon sinnvoll, denn auf diese Weise wird ersichtlich, wie oft eine Alignierung tatsächlich übernommen wurde und sich in den tatsächlich untersuchten Daten befindet. Tabelle 9 enthält den Anfang von Tabelle 32.

Tabelle 9: Kurzübersicht Frequenzen der korrigierten Entsprechungen

4	<i>adv</i>	<i>on</i>	<i>~P</i>	<i>about</i>	<i>of</i>	<i>concerning</i>	<i>relating to</i>	<i>by means of</i>	<i>raus</i>
186	44	27	25	25	19	12	12	11	11

Neben den bereits bekannten Werten für die Formulierungsnähe 4 und die verschiedenen *Meta*-Tags sind die hohen Frequenzen von *on*, *about* und *of* auffällig: Wie kann es sein, dass es 27 Belege gibt, in denen *on* die korrigierte Entsprechung von *über* ist, wenn nur zwölf Belege mit *on* als automatisch alignierter Entsprechung bei der Beispielauswahl ausgewählt wurden? Um diese Frage zu beantworten, muss man sich ansehen, zu welchen Entsprechungen die automatisch alignierten Worte und Wortgruppen korrigiert wurden. Eine entsprechende Übersicht enthält Tabelle 33 auf S. 76 im Anhang. Dort

finden sich in der ersten Spalte die automatisch alignierten Worte bzw. Wortgruppen und in der zweiten die Entsprechungen, zu denen sie jeweils korrigiert wurden. In der dritten Spalte steht, wie häufig die automatisch alignierte Entsprechung zur korrigierten wurde. Dieselben Informationen werden im auf Basis dieser Tabelle erstellten Sankey-Diagramm⁴⁶ in Abbildung 6 auf S. 91 im Anhang auf anschaulichere Weise dargestellt: Die automatisch mit *über* alignierten Worte bzw. Wortgruppen stehen links, die korrigierten Entsprechungen bzw. die *Meta*-Werte und „4“ rechts. Je dicker die Linie ist, die von links nach rechts führt, desto mehr Fälle wurden entsprechend kategorisiert. Wie man sieht, gibt es dicke Balken, die von *report on* zu *on*, von *talk* und *talk about* zu *about* und von *speak* zu *of* führen. Das Aligierungstool efmara hat anscheinend in den genannten Fällen *über* so oft fälschlicherweise mit dem Verb oder dem Verb und der Präposition aligniert, dass diese Alignierungen unter die hundert häufigsten gekommen sind, weswegen entsprechende Beispiele für die Annotation ausgewählt wurden, wo dann die Alignierung korrigiert wurde. Man bemerke, dass es sich in allen Fällen um eine Thema- oder Bezugspunktlesart handeln dürfte: So wird auf der einen Seite der hohe Anteil dieser Lesarten in den annotierten Belegen zum Teil erklärt, auf der anderen Seite ist der Grund für die häufige fehlerhafte Alignierung dieser Worte bzw. Wortgruppen wohl in ihrer hohen Frequenz in den englischen Pendanten der *über*-Sätze im Europarl-Korpus zu suchen (und damit gerade in der hohen Frequenz von Thema- und Bezugspunktlesart bei *über* in diesem Korpus).

3.2. Nach Bearbeitung

Für die meisten Zwecke im Folgenden sind die Daten in ihrem bisherigen Zustand nicht verwendbar: So müssen die Belege mit einem zum Ausschluss führenden Metainformationswert ebenso wie die Fälle, bei denen die fürs Deutsche annotierte Bedeutung im Englischen nicht passt, und die Fälle mit dem Formulierungsnäherwert 4 auch tatsächlich ausgeschlossen werden und es stellt sich die Frage, wie mit den vorhandenen Mehrfachannotationen bei der Bedeutung umgegangen werden soll. Wie Tabelle 31 auf S. 73 im Anhang zeigt, wurde bei den nicht wegen des Metainformationswerts ausgeschlossenen Belegen insgesamt 74 Mal mehr als eine Bedeutung für das deutsche Beispiel annotiert; in 55 Fällen geht es dabei um *Thema* und *Bezugspunkt*. Diese Fälle werden in Abschnitt 5.2 gesondert betrachtet, da sich die Unterscheidung zwischen den beiden Kategorien bei der Annotation als schwierig erwiesen hat. Denkbare Möglichkeiten zum Umgang mit Mehrfachannotationen sind 1. diese als solche zu belassen und somit für statistische Zwecke als gleichwertige, von den einzelnen Bestandteilen unabhängige Kategorien zu betrachten⁴⁷ und 2. für jede Bedeutung einen eigenen Datenpunkt einzuführen. Hier wird die zweite Möglichkeit gewählt, da sich die Annotationen für die Angemessenheit der Bedeutungsannotation für den englischen Satz und die Rektionsstatusannotationen⁴⁸ jeweils auf einzelne Bedeutungen

⁴⁶Zur Erstellung des Diagramms wurde das R-Paket *googleVis* (Gesmann & de Castillo, 2011) genutzt.

⁴⁷D. h., wie oben bei der Betrachtung des Korpus vor der Bearbeitung etwa in Tabelle 7 auf S. 29 stände z. B. 5,9 gleichwertig neben 5 und 9.

⁴⁸Es kann, wenn mehrere Bedeutungen möglich sind, vorkommen, dass *über* in der einen Lesart als regiert betrachtet wird, in der anderen aber nicht:

(i) 3176

(de) Zweitens , es ist leicht , die Begeisterung ÜBER die Nationalstaaten anzuheizen , denn , wie ich in der Debatte sagte , viele Nationalstaaten entstanden durch Kriege oder durch Vorteilnahme gegenüber anderen Staaten .

(ide) ist syntaktisch ambig und *über* hat dort zwei Lesarten, nämlich eine modale, in der die Begeisterung mittels der Nationalstaaten angeheizt wird, und eine Emotionsgegenstandslesart, in der die Nationalstaaten Gegenstand der Begeisterung sind. Als regiert betrachtet wird *über* nur in der Emotionsgegenstandslesart (auch wenn nicht sicher ist, dass das sinnvoll ist, vgl. Abschnitt 5.1).

beziehen und bei der Abbildung von Entsprechungen auf Bedeutungen und andersherum die komplexeren Kategorien stören würden. Es wird also für jede Bedeutung ein eigener Datenpunkt mit den entsprechenden Annotationen für die Angemessenheit der Bedeutungsannotation für das Englische und den Rektionsstatus eingeführt, wobei die anderen Werte übernommen werden können.

Dieses Vorgehen hat zur Folge, dass Belege mit Mehrfachannotation bei der Bedeutung mehrfach in den Daten vertreten sind. Da die einzelnen Datenpunkte dann nicht unabhängig voneinander sind, sind sehr viele statistische Methoden nicht auf den Datensatz anwendbar. Interessanterweise kommt das aber nur bei den Kombinationen verschiedener modaler Lesarten und bei *Thema* und *Bezugspunkt* vor, da in allen anderen Fällen entweder die englische Entsprechung nur für eine der beiden Lesarten passt oder der Formulierungsnähewert 4 ist. In (31) beispielsweise hat das deutsche Beispiel eine temporale Lesart und eine Bezugspunktlesart, im Englischen liegt aber eindeutig nur die temporale Lesart vor.

(31) 48519

(de) Ich möchte eines kurz herausgreifen , weil ich mit dem Kollegen Harbour im Bereich des Universaldienstes einen hervorragenden Informationsaustausch ÜBER die ganze Zeit und Zusammenarbeit hatte .

(en) I would like to briefly pick out one point , as I had an excellent exchange of information THROUGHOUT the whole period of cooperation with Mr Harbour .

Nach der Aufteilung nach Bedeutungen und dem Ausschluss der ihrer Metainformationsannotation wegen auszuschließenden Beispiele gibt es weiterhin die AMV-Fälle und die Fälle mit Formulierungsnähewert 4*n*. Diese sollen aber gesondert betrachtet werden. Zu diesem Zwecke wird das Korpus nun partitioniert, sodass die in Tabelle 10 aufgeführten, paarweise disjunkten Datensätze entstehen. Der Da-

Tabelle 10: Die Datensätze mit den nach Bedeutung getrennten Daten

Name	Größe	Daten
I	258	Enthält die interessantesten Daten: Formulierungsnähewert 1–3, Bedeutung im Englischen passend, keine AMV-Fälle
4N	25	Formulierungsnähe 4 <i>n</i> , Bedeutung passend
4	220	Formulierungsnähe 4
NP	37	Formulierungsnähewert 1–3, Bedeutung im Englischen <i>unpassend</i> , keine AMV-Fälle
AMV	82	AMV-Fälle, Formulierungsnähewert 1–3, Bedeutung passend
AMVNP	4	AMV-Fälle, Formulierungsnähewert 1–3, Bedeutung <i>unpassend</i>

tensatz, der die interessantesten Fälle enthält und uns hauptsächlich beschäftigen wird, ist I. Während auch AMV und 4N Abschnitte gewidmet sind, werden 4, NP und AMVNP im Folgenden nicht betrachtet.

3.2.1. Der Datensatz I

Der Datensatz I teilt sich nach Bedeutungen wie in Tabelle 11 dargestellt auf.⁴⁹ Nach der Trennung nach Bedeutung hat sich der Anteil der Lesarten *Bezugspunkt* und *Thema* nochmals erhöht und liegt jetzt bei zusammengenommen über 64,73 % – das sind ca. 3,82 Prozentpunkte mehr als vorher. Zu bemerken

⁴⁹Man bemerke, dass es auch zwei Belege mit der Lesart *Temporal Zeitdauer* gibt, was erwähnenswert ist, da es bei (Kiss et. al., 2020, S. 179) keine gab.

Tabelle 11: Die Häufigkeiten der Bedeutungen in I

Bedeutung	Anzahl	Anteil (in %)
Thema	93	36,05
Bezugspunkt	74	28,68
Modal Abstraktes Instrument	18	6,98
Konditional Kausal / Emotionsgegenstand	16	6,20
Temporal Maßeinheit	13	5,04
Modal Informationsübermittler	9	3,49
Spatial Traverse <3D	9	3,49
Abdeckung	8	3,10
Rangfolge	5	1,94
Modal Instrumental	2	0,78
Modal Medial	2	0,78
Spatial Traverse <3D +metaph	2	0,78
Temporal Zeitdauer	2	0,78
Modal Art und Weise	1	0,39
Spatial Achsenbezogen	1	0,39
Spatial Achsenbezogen +metaph	1	0,39
Spatial Achsenbezogen +metaph +ziel	1	0,39
Spatial Traverse vertikal	1	0,39
Summe	258	100

ist ferner das völlige Fehlen von Fällen mit desemantisiertem *über*. Die 42 Fälle mit Bedeutungswert *d* und fünf mit *6,d* (wobei *6* für *Machtverhältnis* steht; vgl. Tabelle 31 auf S. 73 im Anhang), teilen sich so auf, dass zu nahezu gleichen Teilen die Formulierungsnähe *4* (23 Fälle) und ein AMV-Fall (24 Fälle) vorliegen. Dass es überhaupt 47 Fälle mit desemantisiertem *über* gibt, liegt an der hohen Frequenz von *verfügen über*, welches für 46 dieser Fälle verantwortlich ist.⁵⁰

Erwartungsgemäß dominieren *on*, *about* und andere mit *Bezugspunkt* oder *Thema* assoziierte Entsprechungen bei den Entsprechungen, deren Häufigkeiten in Tabelle 12 dargestellt sind.

Bei der Formulierungsnähe bleiben die relativen Anteile der Werte *1–3* zueinander im Vergleich zu den unbearbeiteten Daten (vgl. Tabelle 6 auf S. 28) nahezu gleich. Tabelle 13 zeigt die absoluten Häufigkeiten für die Daten in I, wo die Werte *4* und *4n* naturgemäß fehlen.

Tabelle 13: Die Häufigkeiten der Formulierungsnähewerte in I

<i>1</i>	<i>2</i>	<i>3</i>
72	71	115

⁵⁰Die einzige Ausnahme ist (i):

- (i) 20207
 (de) Wir würden verschiedene Kulturen ungerechterweise ÜBER einen Kamm scheren .
 (en) We would be bashing different cultures INTO the same unfounded shape .

Hier wurde für das Englische bei der Formulierungsnähe *4* annotiert und für das Deutsche *idi* in der *Meta*-Spalte und *d* bei Bedeutung und Rektionsstatus. Es ist zu bedenken, dass *über* innerhalb des Idioms durchaus einen Sinn haben mag und sich nur aus der Außensicht keine Bedeutung zuordnen lässt bzw. das Idiom nur als ganzes die intendierte Bedeutung trägt.

Tabelle 12: Die Häufigkeiten der Entsprechungen in 1

Entsprechung	Anzahl	Anteil (in %)
on	30	11,63
about	25	9,69
relating to	18	6,98
of	16	6,20
concerning	15	5,81
via	13	5,04
in relation to	11	4,26
upon	11	4,26
at	10	3,88
by means of	10	3,88
regarding	10	3,88
over	9	3,49
through	9	3,49
throughout	9	3,49
across	8	3,10
into	8	3,10
as to	7	2,71
in terms of	7	2,71
with	5	1,94
above	4	1,55
by	4	1,55
for	4	1,55
in	4	1,55
around	3	1,16
before	3	1,16
on the subject of	2	0,78
as regards	1	0,39
outside	1	0,39
with regard to	1	0,39
Summe	258	100

Auf die unterschiedlichen Frequenzen der Formulierungsnäherwerte und die sich daraus ergebenden Konsequenzen wird in Abschnitt 4.1.2 eingegangen.

Tabelle 14 zeigt, wie häufig *über* bzw. dessen englische Entsprechung in den deutschen und englischen Beispielen jeweils als regiert betrachtet wurde. Um die Unterschiede zwischen den beiden Sprachen in Bezug auf diese Werte geht es in Abschnitt 4.1.1.

Tabelle 14: Die Häufigkeiten der Rektionsstatus in 1

	Nicht regiert	Regiert
Deutsch	113	145
Englisch	164	94

3.2.2. Andere Datensätze

Wegen ihrer geringen Größe ist es nicht notwendig, sich die Frequenzen der Bedeutungen und Entsprechungen für die Datensätze AMV und 4N anzusehen, da die Verhältnisse aus den Bedeutungs-Entsprechungs-Abbildungen in den Abschnitten 4.2 und 4.3 leicht ersichtlich werden. In 4N ist die Formulierungsnähe per definitionem $4n$. Von den 24 Datenpunkten ist *über* in sieben registriert, die englische „Entsprechung“ ist es nie, was aber in Anbetracht der Definition des Formulierungsnähewertes $4n$ nicht verwunderlich ist, da keine präpositionale Entsprechung, die registriert werden könnte, vorliegt. Auch bei AMV ist eine Betrachtung des Rektionsstatus im Englischen der Definition nach nicht sinnvoll. Die Frequenzen der Rektionsstatuswerte für die deutschen Datenpunkte sind in Tabelle 15 aufgeführt.

Tabelle 15: Die Häufigkeiten der Rektionsstatus in AMV

<i>d</i>	<i>n</i>	<i>r</i>
24	3	56

Der hohe Anteil an desemantisierten Vorkommen ist auf *verfügen über* zurückzuführen, das oft mit Formen von *have* übersetzt wird (vgl. dazu auch Abschnitt 4.2): Beispiel (32) etwa ist ein klassischer AMV-Fall.

(32) 101575

(de) Wir müssen die digitale Kluft überwinden , sodass bis 2020 alle Unionsbürger ÜBER Zugangsmöglichkeiten zu Breitbanddiensten mit einer Geschwindigkeit von mindesten 30 Mbit/s verfügen .

(en) We must overcome the digital divide so that by 2020 , all EU citizens HAVE access to broadband services at a speed of at least 30 Mbps .

Da bei der Annotation von AMV per definitionem eine *über*-PP vorliegen muss, der im Englischen eine NP entspricht, die ein Argument eines Verbs ist, ist es sehr wahrscheinlich, dass die deutsche PP ebenfalls ein Argument ist, und aufgrund des sehr weiten Rektionsbegriffs wird hier bei vielen Argument-PPen die Präposition als registriert betrachtet. Verwunderlicher als die registrierten Vorkommen von *über* sind bei AMV also diejenigen, die es nicht sind. Auf sie wird in Abschnitt 4.2 eingegangen. Tabelle 16 zeigt die Frequenzen der verschiedenen Formulierungsnähewerte in AMV.

Tabelle 16: Die Häufigkeiten der Formulierungsnähewerte in AMV

1	2	3
24	15	44

3.3. Die deutschen Daten

Wurden in den letzten beiden Unterabschnitten die deutsch-englischen Satzpaare betrachtet, so wird es hier nun allein um die Daten für die deutschen Sätze gehen. Zwar werden diese im Folgenden nicht gesondert untersucht, aber es sollen dennoch zwei aus ihnen bestehende Subkorpora beschrieben werden,

da diese für Forschungen zu Präpositionsbedeutungen im Deutschen und zur Entwicklung automatischer Bedeutungsdisambiguierungssysteme für *über* von Nutzen sein können. Da die Formulierungsnähe nicht relevant ist, wenn nur die deutschen Daten betrachtet werden, und demnach nur die Belege ausgeschlossen werden müssen, bei denen der *Meta*-Wert dies erzwingt, sind diese Subkorpora größer als etwa *I*. Der Datensatz *DE* enthält nur die deutschen Sätze, wobei Mehrfachannotationen von Bedeutungen nicht getrennt sind. Im Datensatz *DE-DIS* dagegen entspricht jeder Datenpunkt einer Bedeutungsannotation.

DE besteht aus Annotationen für 550 deutsche Sätze, da von den 632 annotierten Sätzen 82 auszuschließen sind. In den 550 Beispielen hat *über* 18 Mal den *Meta*-Wert *idi* und viermal den *Meta*-Wert *sa*. Die Häufigkeiten der einzelnen Bedeutungsannotationen wurden bereits in Abschnitt 3.1 auf S. 29 besprochen und sind in Tabelle 31 auf S. 73 im Anhang aufgeführt. Tabelle 17 enthält eine Übersicht über die Rektionsstatuswerte in *DE*. Dass die meisten Vorkommen von *über* als regiert klassifiziert wurden, dürfte daran liegen, dass der verwendete Rektionsbegriff recht weit ist, und daran, dass der Anteil der Vorkommen mit Thema- oder Bezugspunktlesart so hoch ist, denn bei vielen entsprechenden Köpfen – wie etwa *Bericht* oder *Diskussion* (vgl. Tabelle 30 auf S. 73 im Anhang) wird *über* als regiert eingestuft.

Tabelle 17: Die Häufigkeiten der Rektionsstatuswerte in *DE*

<i>d</i>	<i>n</i>	<i>n,r</i>	<i>n,r,r</i>	<i>r</i>	<i>r,d</i>
42	176	4	1	322	5

DE-DIS enthält die gleichen Sätze wie *DE*, allerdings nicht mit einem Datenpunkt pro Satz, sondern mit einem Datenpunkt pro Bedeutungsannotation. In ihm finden sich alle deutschen Daten so, wie sie nach der in Abschnitt 3.2 beschriebenen Bearbeitung gegeben sind. D. h., *DE-DIS* ist die Vereinigung der deutschen Daten aller Datensätze aus Tabelle 10. Tabelle 18 enthält eine Übersicht über die Bedeutungsannotationen. Wie ein Vergleich mit Tabelle 11 zeigt, entsprechen die Verhältnisse in *I* nicht ganz denen in *DE-DIS*: So ist in diesem die Bezugspunktlesart frequenter, in jenem die Themalesart. Auffällig ist, dass die Machtverhältnislesart – die in den deutschen Daten zwar nur in 1,76 % der Fälle annotiert wurde, damit aber immerhin die neunthäufigste Bedeutung ist – in *I* ganz fehlt. Grund dafür ist, dass sie im Deutschen sehr oft bei *verfügen über* vorkommt, weswegen die entsprechenden Belege in *AMV* (und also nicht in *I*) enthalten sind. Dass es ein Beispiel gibt, bei dem die Lesart *Konditional Kausal / Emotionsgegenstand +metaph* vorliegt, mag auf den ersten Blick verwundern, erklärt sich jedoch daraus, dass es tatsächlich nicht um einen *Emotionsgegenstand* geht, sondern eine (obligatorische, vgl. S. 20 und eingehender S. 49) Doppelannotation mit einer metaphorischen vertikalen Traversenlesart vorliegt: Die Kommission kann in (33) nur metaphorisch über die Betrugs- und Unregelmäßigkeitsfälle zu Fall gekommen sein.

Tabelle 18: Die Häufigkeiten der Bedeutungen in DE-DIS

Bedeutung	Anzahl	Anteil (in %)
Bezugspunkt	206	32,91
Thema	190	30,35
Desemantisiert	47	7,51
Konditional Kausal / Emotionsgegenstand	41	6,55
Modal Abstraktes Instrument	38	6,07
Modal Informationsübermittler	17	2,72
Temporal Maßeinheit	15	2,40
Spatial Traverse <3D	13	2,08
Abdeckung	11	1,76
Machtverhältnis	11	1,76
Rangfolge	7	1,12
Modal Medial	6	0,96
Überschreitung	4	0,64
Spatial Traverse <3D +metaph	3	0,48
Modal Instrumental	2	0,32
Spatial Achsenbezogen	2	0,32
Spatial Achsenbezogen +metaph	2	0,32
Temporal Zeitdauer	2	0,32
Konditional Kausal / Emotionsgegenstand +metaph	1	0,16
Modal Art und Weise	1	0,16
Spatial Achsenbezogen +metaph +ziel	1	0,16
Spatial Traverse Grenzbereich	1	0,16
Spatial Traverse proximal	1	0,16
Spatial Traverse proximal +metaph	1	0,16
Spatial Traverse vertikal	1	0,16
Spatial Traverse vertikal +metaph	1	0,16
Temporal Tagesteil	1	0,16
Summe	626	100

(33) 34223

(de) Mit dieser Einsicht und dem hehren Anspruch , die Sünden der Vergangenheit nicht zu wiederholen - schließlich ist vor knapp einem Jahr die letzte Kommission ÜBER Betrugs- und Unregelmäßigkeitsfälle zu Fall gekommen - , ist die Kommission Prodi im September angetreten .

Wie oft *über* in DE-DIS regiert ist, zeigt Tabelle 19.

Tabelle 19: Die Häufigkeiten der Rektionsstatuswerte in DE-DIS

<i>d</i>	<i>n</i>	<i>r</i>
47	222	357

4. Auswertung

In diesem Abschnitt sollen die Datensätze 1, AMV und 4N, welche die interessanten Daten enthalten, ausgewertet werden. Das Hauptziel ist dabei die Erstellung von Abbildungen zwischen Bedeutungen von *über* und dessen englischen Entsprechungen, denn sie sind für fast alle der in Abschnitt 6 genannten Anwendungen relevant. Ähnliche Abbildungen werden auch für AMV und 4N erstellt. Darüber hinaus sollen aber auch Fragen in Bezug auf die Rektionsstatus und die Formulierungsnähe behandelt werden.

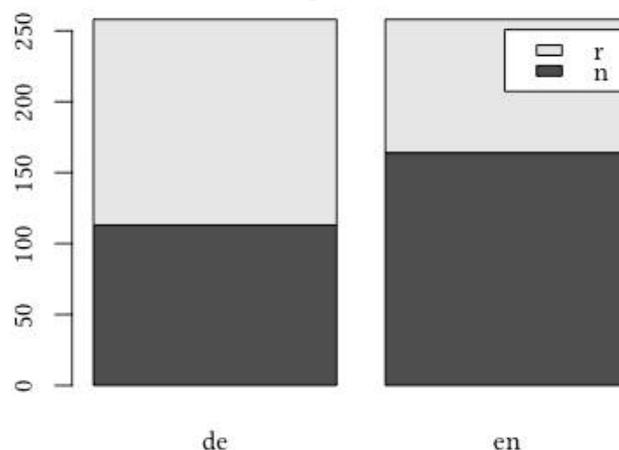
4.1. Der Datensatz 1

Der Datensatz 1 enthält diejenigen Beispiele, die für das Erstellen von Bedeutungs-Entsprechungs-Abbildungen am interessantesten sind: Die Bedeutungen sind im Englischen passend, die Formulierungsnähe ist 1, 2 oder 3 und es sind keine AMV-Fälle enthalten (vgl. Tabelle 10 auf S. 31). Bevor es in Unterabschnitt 4.1.3 um die Abbildungen geht, soll der Datensatz aber noch in Hinblick auf andere Fragestellungen untersucht werden: Unterabschnitt 4.1.2 ist den Formulierungsnähewerten gewidmet und der folgende Unterabschnitt (4.1.1) behandelt die Unterschiede im Rektionsstatus. Da nicht für beide Sprachen im gleichen Umfang Materialien zur Bestimmung des Rektionsstatus vorlagen, soll geprüft werden, ob es Unterschiede zwischen den Annotationen fürs Deutsche und denen fürs Englische gibt, denn dies würde Zweifel an der Vergleichbarkeit der Annotationen für die beiden Sprachen wecken, da vorderhand nicht ersichtlich ist, wieso Präpositionen in einer der beiden Sprachen öfter regiert sein sollten als in der anderen. Was dagegen zu erwarten ist, ist ein Zusammenhang zwischen den Rektionsstatus: Tendenziell sollte *über* regiert sein, wenn seine englische Entsprechung es ist und andersherum. Auch Unterschiede im Rektionsstatus bei verschiedenen Bedeutungen und Entsprechungen sollen betrachtet werden.

4.1.1. Rektionsstatus

Wie Tabelle 14 auf S. 33 (die hier in Abbildung 2 veranschaulicht wird) zu entnehmen ist, ist *über* bzw. dessen Entsprechung in den deutschen Belegen in 1 häufiger regiert als in den englischen.⁵¹

Abbildung 2: Rektionsstatus der Beispiele in 1 im Deutschen und Englischen



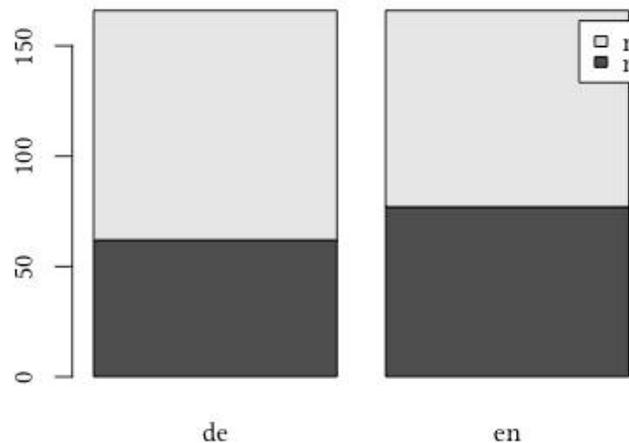
⁵¹Da die Datenpunkte in 1 nicht unabhängig voneinander sind, kann nicht getestet werden, ob dieser Unterschied signifikant ist (vgl. Abschnitt 3.2).

Dass es diese Unterschiede gibt, ist zunächst nicht verwunderlich, da es sich bei den englischen Entsprechungen nicht immer um (primäre) Präpositionen handelt und nur diese überhaupt regiert werden können. Um also die Rektionsstatusannotationen auf angemessene Weise vergleichen zu können, sollten die Fälle, bei denen im Englischen keine primäre Präposition vorliegt, ausgeschlossen werden.⁵² Um Signifikanztests anwenden zu können, wird für Belege, für die aufgrund der Trennung nach Lesarten mehrere Datenpunkte im Datensatz sind, nur ein Datenpunkt genommen. Da sich bei ambigen Beispielen die Rektionsstatus durchaus unterscheiden könnten, sollten Datenpunkte mit unterschiedlichem Rektionsstatus in einer der beiden Sprachen eigentlich auch weiterhin mehrfach vorhanden sein; es gab im Datensatz aber keine entsprechenden Daten. Auch nach dieser Bearbeitung der Daten gibt es weiterhin einen kleinen Unterschied, wie Tabelle 20 (veranschaulicht in Abbildung 3) zeigt.

Tabelle 20: Die Häufigkeiten der Rektionsstatus in *r* bei den Beispielen mit primärer Präposition als Entsprechung

	Nicht regiert	Regiert
Deutsch	62	104
Englisch	77	89

Abbildung 3: Rektionsstatus der Beispiele in *r* im Deutschen und Englischen nach Bereinigung



Dieser Zusammenhang zwischen Sprache und Rektionsstatus ist allerdings statistisch nicht signifikant: $\chi^2(1) = 2,7845$, $p = 0,09518$ ⁵³. Die Nullhypothese, dass es keinen Zusammenhang gibt, kann also nicht zurückgewiesen werden.

Eine andere, ebenfalls interessante Frage in Bezug auf den Rektionsstatus ist, ob in den Fällen, in denen *über* als regiert eingestuft wurde, selbiges auch für die englische Entsprechung der Fall war. Es ist zu erwarten, dass es einen Zusammenhang zwischen dem Rektionsstatus im Deutschen und dem im Englischen gibt: Wenn *über* regiert ist, sollte es seine englische Entsprechung tendenziell auch sein

⁵²Das betrifft Beispiele mit *as regards, as to, by means of, concerning, in relation to, in terms of, on the subject of, regarding, relating to* und *with regard to*.

⁵³Zur Berechnung der Werte wurde R verwendet (R Core Team, 2020). Angegeben werden in dieser Arbeit die Werte für Pearsons χ^2 -Test ohne Yates' Stetigkeitskorrektur, sofern nicht explizit darauf hingewiesen wird, dass das nicht der Fall ist.

und andersherum. Tabelle 21 zeigt die kreuztabulierten Werte für den Rektionsstatus im Deutschen und Englischen in den Daten, die nur die Beispiele mit primärer Präposition im Englischen enthalten und bei denen ambige Beispiele nur einmal enthalten sind. Tatsächlich ist der Zusammenhang zwischen

Tabelle 21: Rektionsstatus deutsch und Rektionsstatus englisch in bereinigtem 1 kreuztabuliert

	en n	en r
de n	52	10
de r	25	79

Rektionsstatus im Deutschen und Rektionsstatus im Englischen hochsignifikant: $\chi^2(1) = 55,915, p < 0,001$. Die Effektstärke des Zusammenhangs ist stark (Cramérs $V = 0,58$).⁵⁴

Betrachten wir nun, wie sich die verschiedenen Bedeutungen und Entsprechungen in Bezug auf den Rektionsstatus unterscheiden: Tabelle 22 enthält jeweils eine Kreuztabelle für die Bedeutung und den Rektionsstatus pro Sprache in 1, Tabelle 23 eine ähnliche Tabelle für die Entsprechungen.

Tabelle 22: Kreuztabellen Rektionsstatus Bedeutungen

	Abdeckung	Bezugspunkt	Kond. Kaus. / Em.	Modal Abst. Inst.	Modal Art u. W.	Modal Inf.üb.	Modal Inst.	Modal Medial	Rangfolge	Spatial Achse.	Spatial Achse. +m	Spatial Achse. +m +z	Spatial Trav. <3D	Spatial Trav. <3D +m	Spatial Trav. vert.	Temporal Maß.	Temporal Zeit.	Thema
de n	0	29	2	17	1	8	2	2	0	1	1	1	9	1	1	13	2	23
de r	8	45	14	1	0	1	0	0	5	0	0	0	0	1	0	0	0	70
en n	8	46	3	18	1	9	2	2	5	1	1	1	9	2	1	13	2	40
en r	0	28	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	53

In den Bedeutungen, in denen die *über*-PP normalerweise ein Adjunkt ist, ist *über* bis auf sehr wenige Ausnahmen im Deutschen (bei denen sie dann doch ein Argument ist) natürlich nicht regiert und es ist nicht überraschend, dass die Entsprechung in den englischen Sätzen dann ebenfalls nicht regiert ist, da das der *über*-PP entsprechende Element höchstwahrscheinlich ebenfalls ein Adjunkt sein wird. Das betrifft die modalen, spatialen und temporalen Lesarten. Die grundsätzlichen Unterschiede zwischen dem Deutschen und dem Englischen sind, wie oben ausgeführt, darauf zurückzuführen, dass Entsprechungen, die keine primären Präpositionen sind, nicht regiert sein können. Auffällig sind die Unterschiede zwischen den Sprachen bei den Lesarten *Abdeckung* und *Rangfolge*. Hier ist aber anzumerken, dass es im Deutschen nur wenige Prädikate sind, bei denen diese vorliegen (*stehen* und *stellen* bei der Rangfolgelesart, *verteilen* und *erstrecken* bei der Abdeckungslesart), während im Englischen verschiedene Übersetzungen gewählt wurden. In beiden Sprachen wurde die Präposition bei der Lesart *Konditional Kausal / Emotionsgegenstand* meist als regiert betrachtet. Am Ende von Abschnitt 5.1, in welchem es um die Verwendung von *über* bei psychologischen Prädikaten geht, wird eine alternative Sichtweise dargestellt werden. Dass *über* (und oft auch seine Entsprechung) bei der Thema- und der Bezugspunktlesart

⁵⁴Für die Berechnung von Cramérs V wurde das R-Paket `vcd` (Meyer et. al., 2020) benutzt. Die Interpretation des Wertes folgt Sheskin (2011, S. 535, zitiert nach Levshina (2015, S. 209)).

Tabelle 23: Kreuztabellen Rektionsstatus Entsprechungen

		about	above	across	around	as regards	as to	at	before	by	by means of	concerning	for	in	in relation to	in terms of	into	of	on	on the subject of	outside	over	regarding	relating to	through	throughout	upon	via	with	with regard to
de	n	3	2	3	0	0	4	2	0	2	10	10	4	1	0	2	6	0	7	0	1	6	3	15	9	7	2	12	1	1
	r	22	2	5	3	1	3	8	3	2	0	5	0	3	11	5	2	16	23	2	0	3	7	3	0	2	9	1	4	0
en	n	1	4	8	3	1	7	3	3	4	10	15	4	1	11	7	3	0	8	2	1	6	10	18	9	9	1	13	1	1
	r	24	0	0	0	0	0	7	0	0	0	0	0	3	0	0	5	16	22	0	0	3	0	0	0	0	10	0	4	0

so oft als regiert kategorisiert wurde, dürfte an dem weiten Rektionsbegriff liegen, der hier verwendet wurde. So sind in den Listen von Kiss et. al. (2020) etwa auch *Bericht* und *Debatte* als *über* regierend geführt.

Bei den Entsprechungen ist zu bedenken, dass diejenigen, bei denen es sich nicht um primäre Präpositionen handelt, gar nicht regiert sein können, und dass – wie wir in Abschnitt 4.1.3 sehen werden – fast alle Entsprechungen nur in sehr wenigen Lesarten Entsprechungen sind, weswegen sich die bei den Bedeutungen beobachteten Unterschiede auch hier auswirken. *For* etwa kommt in *I* nur in der temporalen Maßeinheitslesart als Entsprechung von *über* vor. Da zu erwarten ist, dass diese vor allem bei Adjunkten vorliegt, ist es nicht überraschend, dass in den Fällen mit *for* als Entsprechung weder *über* noch *for* regiert ist. In den fünf Fällen, in denen ein *über*-Token mit *across* als Entsprechung regiert ist, hat es die Abdeckungslesart, in den drei anderen Fällen eine Traversenlesart.

Zusammenfassung Die primären Präpositionen in den englischen Daten sind nicht signifikant häufiger als regiert eingestuft worden, als das in den entsprechenden deutschen Sätzen bei *über* der Fall war. Es besteht ein starker, statistisch signifikanter Zusammenhang zwischen dem Rektionsstatus von *über* und dem seiner Entsprechung, wenn es sich bei dieser um eine primäre Präposition handelt. Sowohl die Bedeutungen als auch die Entsprechungen unterscheiden sich in Hinblick auf die Rektionsstatus in beiden Sprachen.

4.1.2. Formulierungsnähe

Der Datensatz *I* enthält nur Datenpunkte mit den Formulierungsnähewerten 1, 2 und 3, da die *4n*-Daten sich in *4N* befinden und bei *4* keine wirkliche Entsprechung von *über* vorliegt. Da es bei den Daten in *I* immer eine englische Entsprechung von *über* mit passender Bedeutung gibt und uns vor allem die Beziehungen zwischen Entsprechungen und Bedeutungen interessieren, werden die Unterschiede in der Formulierungsnähe im Weiteren zumeist ausgeblendet. Vorher sollen an dieser Stelle aber noch einige Fragen geklärt werden: Wie in Abschnitt 2.2.6 beschrieben, werden mit den Formulierungsnähewerten 1–3 Unterschiede in Bezug auf Syntax und lexikalisches Material im internen Argument und Regens erfasst. Es stellt sich also die Frage, ob es Zusammenhänge zwischen Formulierungsnähe, Entsprechung,

Bedeutungen und Rektionsstatus gibt: Denkbar wäre etwa, dass zwischen Formulierungsnähe und Entsprechung ein Zusammenhang besteht, weil sich die Syntax der internen Argumente der verschiedenen Entsprechungen erheblich voneinander unterscheidet.

Da die Datenpunkte nicht unabhängig voneinander sind (vgl. Abschnitt 3.2), ist eine Anwendung von Unabhängigkeitstests (die testen, ob *Variablen* unabhängig voneinander sind) direkt auf *I* nicht möglich, weil dafür die Unabhängigkeit der Datenpunkte in der Regel eine Voraussetzung ist. Es sollen stattdessen die nicht nach Bedeutung getrennten Daten untersucht werden, was auch sachlich nicht unangemessen ist, da die Formulierungsnähe jeweils für Satzpaare und nicht für Bedeutungen annotiert wurde. Tabelle 24 zeigt die Häufigkeiten der verschiedenen Formulierungsnähewerte bei den ungetrennten Daten.

Tabelle 24: Die Häufigkeiten der Formulierungsnähewerte in den ungetrennten *I*-Daten

	1	2	3
	71	65	100
	30,08 %	27,54 %	42,37 %

Betrachten wir zuerst die Entsprechungen: Da es bei jedem englischen Satz genau eine Entsprechung und einen Formulierungsnähewert gibt, steht der Anwendung von Unabhängigkeitstest nichts im Wege. Weil bei fast allen Entsprechungen die erwarteten Werte zu klein sind, kann der χ^2 -Test zwar nicht verwendet werden, der exakte Test nach Fisher zeigt aber einen hochsignifikanten Zusammenhang ($p < 0,001$). In Tabelle 25 ist die Formulierungsnähe mit den Entsprechungen kreuztabuliert. Meistens

Tabelle 25: Kreuztabelle von Formulierungsnähe und Entsprechungen in den nicht bedeutungsgetreunten Daten aus *I*

	about	above	across	against	around	as regards	as to	at	before	by	by means of	concerning	for	in	in relation to	in terms of	into	of	on	on the subject of	outside	over	regarding	relating to	through	throughout	upon	via	with	with regard to
1	14	0	4	0	1	0	0	2	0	0	3	1	1	1	2	1	0	6	5	2	1	2	4	2	5	3	3	5	3	0
2	4	1	2	1	1	0	0	3	1	3	2	6	1	0	3	2	3	3	10	0	0	3	2	4	2	1	1	4	1	1
3	7	3	2	0	2	1	5	5	2	0	5	5	2	2	3	4	4	7	12	0	0	3	4	6	2	5	6	2	1	0

liegen pro Entsprechung nur recht wenige Belege vor, was die Interpretation erschwert, aber man sieht, dass sich etwa die Beispiele mit *about* und *on* durchaus unterschiedlich auf die Formulierungsnähewerte verteilen. Weiterhin sind in dieser Tabelle auch die Werte bei *as to* auffällig, weil dort ausschließlich der Formulierungsnähewert 3 annotiert wurde. Eine Betrachtung der entsprechenden Beispiele zeigt, dass in drei der fünf Fälle das interne Argument ein eingebetteter Satz ist, was, da das bei *über* nicht möglich ist (das Pronominaladverb *darüber* wird hier nicht betrachtet), immer zur Annotation von 3 führt. Es könnte also sein, dass die Formulierungsnäheunterschiede durch unterschiedliche syntaktische Muster bedingt sind, in denen die Entsprechungen auftreten können oder für gewöhnlich auftreten – *about* wird im Vergleich zu *as to* wahrscheinlich seltener einen eingebetteten Satz als internes

Tabelle 26: Prozentuale Verteilungen der Formulierungsnähewerte auf die Rektionsstatus in den nicht bedeutungsgetrenten Daten aus 1: links Englisch, rechts Deutsch

	<i>n</i>	<i>r</i>		<i>n</i>	<i>n,r</i>	<i>n,r,r</i>	<i>r</i>
1	27,21	34,83	1	29,79	0,00	0	31,16
2	31,29	21,35	2	30,85	66,67	0	24,64
3	41,50	43,82	3	39,36	33,33	100	44,20

Argument haben. Um dem genauer nachgehen zu können, wären tiefergehende Annotationen zur syntaktischen Umgebung der Entsprechungen erforderlich, weswegen das Problem hier ungelöst bleiben muss – ein Grund für die Einführung einer Kategorie für die Formulierungsnähe war ja gerade, dass die nicht die Entsprechung selbst betreffenden Unterschiede nicht im Fokus dieser Arbeit stehen und darum ausgeblendet werden sollten. Es wird sich jedoch in Abschnitt 6.2 zeigen, dass derartige Annotationen auch für die Erstellung sinnvoller Lexikoneinträge hilfreich sein könnten, weswegen hier ein Anknüpfungspunkt für weitere Forschungen wäre.

Bei den Bedeutungen müssen Unabhängigkeitstests verwendet werden, die mit kategorialen Variablen umgehen können, bei denen die Auswahl mehrerer Werte möglich ist. Wir benutzen zwei⁵⁵ der *Tests for Multiple Marginal Independence*, die im R-Paket MRCV (Kozioł & Bilder, 2014) implementiert sind, nämlich *bon* und *boot* (vgl. dazu (Bilder & Loughin, 2004; Bilder et. al., 2000)). Mit beiden konnte kein signifikanter Zusammenhang zwischen Bedeutung(en) und Formulierungsnähe festgestellt werden (Minimum der Bonferroni-korrigierten Ergebnisse für die verschiedenen Bedeutungen: $p = 0,1533$; Bootstrap-Ergebnisse:⁵⁶ $p.boot = 0,1788$, $p.combo.min = 0,0754$).

Ob es einen Zusammenhang zwischen Formulierungsnähe und Rektionsstatus gibt, kann wegen der fehlenden Unabhängigkeit der Datenpunkte nicht auf 1 getestet werden. Einen Test auf den nicht getrennten Daten durchzuführen, ist inhaltlich nicht sinnvoll, da der Rektionsstatus pro Bedeutung annotiert wurde. Tabelle 26 zeigt, wie sich die Beispiele mit den verschiedenen Rektionsstatuswerten prozentual auf die Formulierungsnähewerte verteilen. Es ist zu bedenken, dass alle Umstände, die eine Anwendung von Unabhängigkeitstests verhindern, auch eine Interpretation dieser Tabellen erschweren; aber es lässt sich erkennen, dass die Unterschiede zwischen den Rektionsstatus in beiden Sprachen nicht sehr groß sind und die Verteilungen grundsätzlich dem entsprechen, was man auf Basis von Tabelle 24 erwarten würde, auch wenn bei *r* öfter 1 und seltener 2 annotiert wurde als bei *n*. Dieser Trend scheint im Englischen stärker zu sein als im Deutschen.⁵⁷ Da die Unterschiede nicht sehr groß sind, ist nicht davon auszugehen, dass es einen Zusammenhang zwischen Rektionsstatus und Formulierungsnähe gibt.

⁵⁵Die dritte der dort verfügbaren Testmethoden, der Rao-Scott-Ansatz, war nicht anwendbar, weil im konkreten Fall mehr als 1800 GB Arbeitsspeicher notwendig gewesen wären, um den Test durchzuführen.

⁵⁶Beim Bootstrapping-Verfahren wurde die zu allen Null betragenden Randsummen addierte Konstante vom Standardwert 0,5 auf 0,01 herabgesetzt, da die Werte in der Tabelle grundsätzlich klein sind. Auch mit dem Standardwert ändert sich nichts an den Signifikanzniveaus, auch wenn $p.combo.min$ noch weiter sinkt ($p.boot = 0,2489$, $p.combo.min = 0,0582$).

⁵⁷Allerdings könnte es auch sein, dass das auf den Einfluss des Unterschieds zwischen den Entsprechungen zurückzuführen ist.

4.1.3. Die Bedeutungs-Entsprechungs-Abbildungen

Die Tabellen, die die Informationen darüber enthalten, wie oft bei welcher Bedeutung welche Entsprechung gebraucht wurde, befinden sich ihrer Länge wegen im Anhang. Tabelle 34 (S. 79) und Tabelle 35 (S. 81) sind dabei nach Bedeutungen geordnet, Tabelle 36 (S. 83) und Tabelle 37 (S. 85) nach englischen Entsprechungen, wobei Tabelle 34 und 36 auch Informationen zum Rektionsstatus enthalten. Da diese Tabellen recht unübersichtlich sind, werden sie in dem Sankey-Diagramm in Abbildung 4 veranschaulicht (in dem auf den Rektionsstatus keine Rücksicht genommen werden kann). Auf der linken Seite befinden sich dort die englischen Entsprechungen von *über*, auf der rechten die Bedeutungen. Beide sind mit Linien verbunden und es gilt: Je dicker die Linie, desto mehr entsprechende Datenpunkte. Die Anzahl der Datenpunkte für ein Bedeutungs-Entsprechungs-Paar ist allerdings der Art der Beispielauswahl wegen nicht aussagekräftig.

Wie zu erwarten war, lassen sich die meisten englischen Entsprechungen nur wenigen Lesarten zuordnen: So wird etwa *above* im Korpus nur bei der achsenbezogenen *spatialen* Lesart und der *Rangfolgen*-Lesart verwendet, *for* nur bei der *temporalen* Maßeinheitslesart und *via* nur in *modalen* Lesarten und bei *Traversen* mit weniger als drei relevanten Dimensionen. Die einzigen beiden englischen Entsprechungen, die viele verschiedene Bedeutungen abdecken, sind *over* und *in*. Für *in* liegen allerdings nur vier Beispiele vor und in drei davon ist sowohl es als auch *über* registriert. Für *over* dagegen gibt es im Korpus Belege mit *Bezugspunkt*-, *Emotionsgegenstand*- und *Traversenlesart* sowie mit *temporaler* und *modaler* Bedeutung und es kann kein Zweifel daran bestehen, dass sich bei einer Betrachtung weiterer Beispiele auch Belege für andere Lesarten finden ließen.⁵⁸ Eine Besonderheit stellt *of* dar, das im Korpus nur registriert eine Entsprechung von *über* ist und nur registriertem *über* entspricht.

Da bedingt durch die Gegebenheiten im gewählten Ausgangskorpus die Lesarten *Bezugspunkt* und *Thema* am häufigsten vorkommen, nehmen sie auch im Diagramm am meisten Raum ein. Auffällig ist, dass es viele englische Entsprechungen gibt, die nur mit diesen beiden Lesarten verwendet werden (*about, as to, concerning, in relation to, in terms of, into, of, on, on the subject of, regarding, relating to, with regard to*) und dass fast alle von diesen – wenn auch in unterschiedlichem Maße – auch tatsächlich für

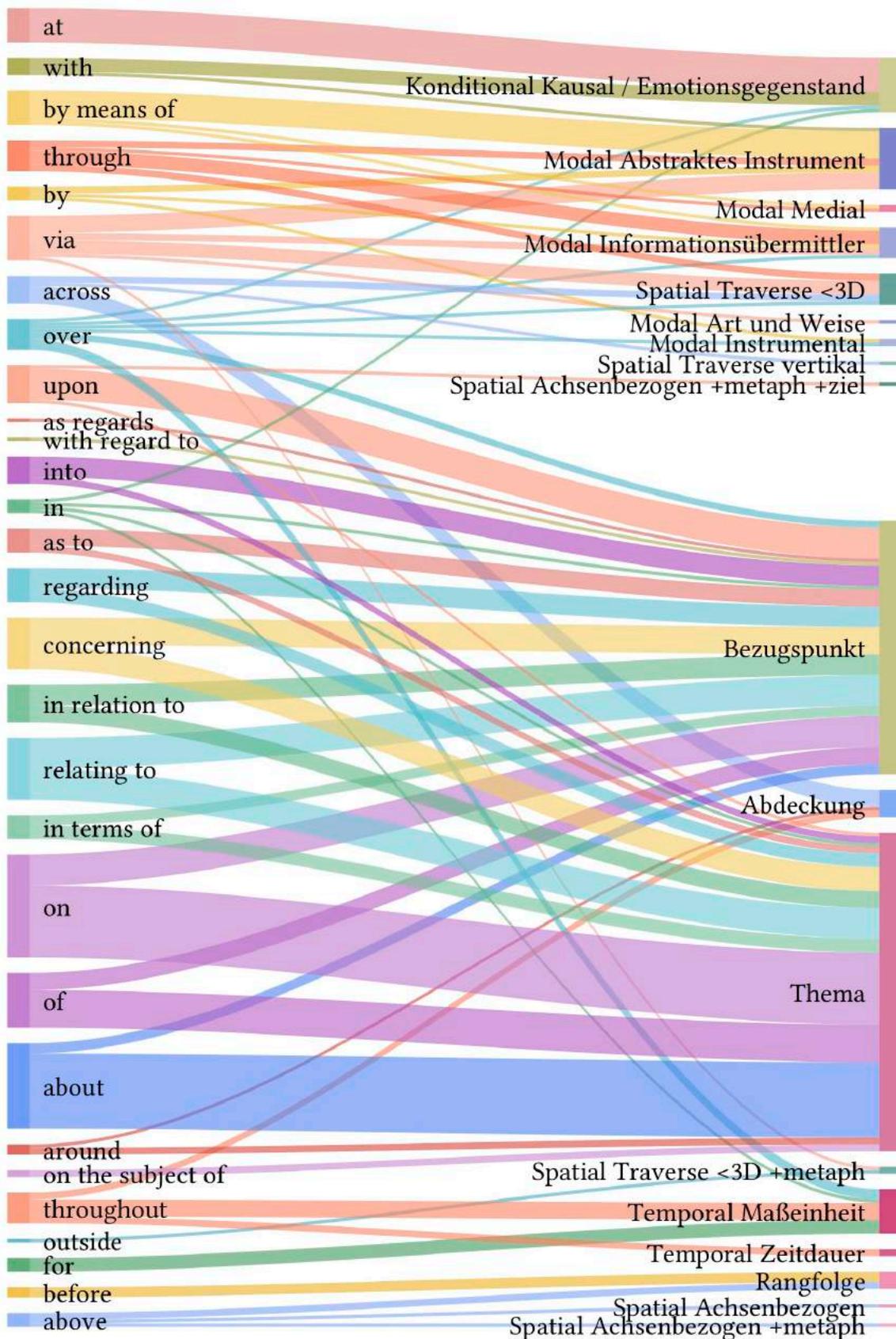
⁵⁸So liefert eine kurze Suche nach *over* in allen englischen Sätzen, bei denen im deutschen Pendant *über* vorkommt, also auch solchen, die nicht als Beispiele für die Annotation ausgewählt wurden, u. a. folgende Belege:

- (i) 130
(en) We are meant to grant discharge , even though we have no influence or effective political control over how the resources are spent .
- (ii) 164
(en) Indeed he is throwing a very unusual veil of secrecy over what is going on .
- (iii) 1125
(en) For if lawyers are not capable of ensuring that the general takes precedence over the particular , of emphasizing that which unites over that which divides , can we realistically expect this attitude from those who are not specialists in the law , which is the art of integration par excellence ?

In (i) liegt eine *Machtverhältnislesart* vor, in (ii) eine *metaphorische Bedeckungslesart* und in (iii) zweimal eine *Rangfolgelesart*.

Es kann an dieser Stelle spekuliert werden, dass der Grund dafür, dass *over* oft als gewöhnlichste Entsprechung von *über* betrachtet wird (so bezeichnet es etwa Liamkina (2007) als dessen „closest counterpart“ (S. 115)) nicht etwa in der Häufigkeit der Verwendungsfälle liegt, in denen es die zu wählende Übersetzung wäre, sondern in der Vielzahl der abgedeckten Lesarten. Auch wenn es nicht sicher gesagt werden kann, weil nicht für alle Vorkommen von *über* im Ausgangskorpus die *Alignierung* korrigiert wurde, legen doch die automatischen *Alignierungen* nahe, dass die mit Abstand häufigste Übersetzung von *über* dort *on* ist (vgl. Abbildung 1 auf S. 11), was allerdings eine spezifische Eigenart des gewählten Korpus sein mag.

Abbildung 4: Sankey-Diagramm: Bei welchen Lesarten kommen welche Entsprechungen vor?



beide benutzt werden. Ersteres folgt der oben angesprochenen Tendenz, für letzteres mag der Grund sein, dass – wie in Abschnitt 5.2 gezeigt wird – die Grenzziehung zwischen den beiden Lesarten im Annotationsschema problematisch ist. Dass bei den meisten annotierten Belegen eine Bezugspunkt- oder Themaesart vorliegt, führt also dazu, dass für diese Bedeutungen viel mehr englische Entsprechungen als für die anderen gefunden wurden. Man könnte nun vermuten, dass die Erfassung der englischen Entsprechungen für diese beiden Lesarten deswegen „vollständiger“ sei als für manch andere, allerdings würde man dabei verkennen, dass nicht klar ist, wie viele Möglichkeiten, sie auszudrücken, es im Englischen für jede der Bedeutungen überhaupt gibt.

4.2. Der Datensatz AMV

Der Datensatz AMV enthält Beispiele, bei denen im Deutschen eine *über*-PP steht, deren englisches Pendant eine NP ist, die die gleiche thematische Rolle innehat (vgl. die Abschnitte 2.2.3 und 3.2.2). Eine nach Bedeutungen geordnete Auflistung findet sich in Tabelle 27, eine nach Verben geordnete in Tabelle 38 auf S. 86 im Anhang. Wie sich zeigt, lassen sich die Verben eindeutig Lesarten zuordnen. In den Fällen, in denen *über* im Deutschen desemantisiert ist, steht dort *verfügen über*. Die englischen Übersetzungen sind dann Formen von *have*, *hold* oder *possess*.

Tabelle 27: Abbildungen nach Bedeutungen geordnet AMV

Bedeutung	Verb	regiert de	Anzahl
Abdeckung	covering	r	1
Bezugspunkt	consider	n	1
Bezugspunkt	consider	r	4
Bezugspunkt	considering	r	4
Bezugspunkt	decide	r	3
Bezugspunkt	deciding	r	4
Bezugspunkt	determine	r	3
Desemantisiert	had	d	3
Desemantisiert	has	d	6
Desemantisiert	have	d	4
Desemantisiert	having	d	3
Desemantisiert	hold	d	4
Desemantisiert	possess	d	4
Konditional Kausal / Emotionsgegenstand	enjoy	r	3
Konditional Kausal / Emotionsgegenstand	welcome	r	3
Machtverhältnis	governing	r	1
Rangfolge	transcends	r	1
Thema	debate	r	3
Thema	debated	r	4
Thema	debating	r	4
Thema	discuss	r	4
Thema	discussed	r	6

Thema	discussing	r	4
Thema	mention	r	4
Überschreitung	exceed	n	2

In Bezug auf die Abbildungstabellen zum Datensatz AMV ist anzumerken, dass es nur wenige Beispiele pro Verb gibt und insgesamt nur wenige Verben. Ersteres ist eine Folge davon, dass für automatische Alignierungen, die auf AMV-Fälle hindeuten, nur in der ersten Annotationsphase Beispiele annotiert wurden (vgl. Abschnitt 2.1.4); letzteres ergibt sich aus dem Beispielauswahlverfahren im Allgemeinen: Unter den hundert häufigsten automatischen Alignierungen für *über* war nur für diese Verben eine Annotation mit AMV sinnvoll. Es steht jedoch zu vermuten, dass sich bei einer weiteren Betrachtung der Liste mit den häufigsten Alignierungen noch weitere Kandidaten finden ließen, und für einige der in Abschnitt 6 genannten Anwendungen könnte es auch sinnvoll sein, nach ihnen zu suchen.

Da es im Englischen in den AMV-Sätzen trivialerweise keine regierte Entsprechung von *über* geben kann, ist eine Betrachtung der Rektionsstatus im Englischen uninteressant. *Über* selbst ist in den Belegen aus AMV fast immer regiert oder gar desemantisiert. In Anbetracht dessen, dass im Englischen die Verben die entsprechenden thematischen Rollen zuzuweisen vermögen, sind allerdings vor allem die Fälle, bei denen das *nicht* der Fall ist, verwunderlich. Von den drei entsprechenden Beispielen sollen hier nur zwei betrachtet werden, da sich das dritte vollkommen analog zu (35) verhält.

(34) 27511

(de) Die Kommission hatte bereits in der Agenda 2000 Überlegungen ÜBER ihre eigene Umgestaltung angesprochen .

(en) In Agenda 2000 , the Commission had already begun to CONSIDER its own reform .

(35) 5404

(de) Heute gibt es jedoch keinen Anhaltspunkt dafür , daß nach dem Jahr 2000 eine BIP-Ab-schöpfung möglich sein wird , die ÜBER der zur Zeit von den Verträgen vorgesehenen Ober-grenze von 1,27 % liegt .

(en) However , there is no reason to think today that , after the year 2000 , it will be possible to EXCEED the ceiling of 1.27 % of GDP , as currently laid down in the Treaties .

(34) ist einer der seltenen AMV-Fälle, in denen im Deutschen das Regens von *über* kein Verb ist, und man kann sich sicherlich fragen, ob nicht die Annotation des Formulierungsnähewertes 4 angebracht gewesen wäre. Allerdings ist der Unterschied nicht so groß, dass man, hätte im Englischen nicht *consider*, sondern *think about* gestanden, gesagt hätte, dass *über* nicht eine eindeutige Entsprechung (nämlich *about*) hat, weswegen 3 gewählt wurde. Bei (35) ist im deutschen Satz die spatiale Metaphorik noch klar erkennbar (auch wenn in solchen Fällen keine spatiale Lesart annotiert werden soll (Kiss et. al., 2020, S. 185)), die im englischen Pendant nicht gegeben ist.

4.3. Der Datensatz 4N

Der Datensatz 4N enthält die Beispiele, bei denen die Formulierungsnähe 4n ist, also eine grundsätzlich andere Struktur vorliegt, die automatisch alignierte „Entsprechung“ aber einen Schluss auf die Bedeutung von *über* ermöglicht. Ein Beispiel findet sich in (30) (hier wiederholt als (36)):

(36) 71173: 4n

(de) Die Kommission hat es nicht nötig, sich ÜBER das Buch zu refinanzieren.

(en) The Commission does not need to USE the book as a means of recovering its costs.

Im Gegensatz zu den AMV-Fällen kann man hier nicht sagen, dass der PP im Deutschen ein NP-Argument eines Verbs im Englischen entspricht. Trotzdem können diese Fälle für einige der im nächsten Abschnitt vorgestellten Anwendungen interessant sein.

Ein Blick auf die in Tabelle 28 aufgeführten Abbildungen zeigt, dass sich die Worte, bei denen 4n annotiert wurde, zumindest im gegebenen Korpus tatsächlich immer einer bestimmten (Ober-)Bedeutung zuordnen lassen.

Tabelle 28: Abbildungen 4N

„Entsprechung“	Bedeutung	Anzahl
dealing	Thema	4
debating	Thema	1
excess	Überschreitung	1
mention	Thema	2
subject	Thema	3
use	Modal Abstraktes Instrument	4
use	Modal Informationsübermittler	3
use	Modal Medial	2
using	Modal Abstraktes Instrument	4

5. Überlegungen zum Bedeutungsinventar

In diesem Abschnitt sollen anknüpfend an Probleme, die sich bei der Annotation ergaben (Abschnitte 5.1 und 5.2), und in Hinblick auf sich ergebende Analysemöglichkeiten (Abschnitt 5.3) Änderungen am Annotationsschema von Kiss et. al. (2020) vorgeschlagen werden. Es sei aber darauf hingewiesen, dass bei der Argumentation hier ausschließlich Daten für die Präposition *über* betrachtet werden, während bei etwaigen Änderungen am Schema aus (Kiss et. al., 2020) berücksichtigt werden müsste, wie sich diese bei anderen Präpositionen und auf den generellen Aufbau auswirken.

5.1. Über bei psychologischen Prädikaten

Bei Kiss et. al. (2020) wird *über* eine konditional-kausale Lesart zugeschrieben. Dabei drücke das interne Argument die „Ursache für das im Satz genannte Geschehen (zumeist Gemütszustände)“ (Kiss et. al., 2020, S. 182) aus. *Über* sei in den meisten Beispielen regiert, aber nicht desemantisiert. In allen Beispielen von Kiss et. al. (2020, S. 182 f.) geht es tatsächlich um Gemütszustände, allerdings wird später (S. 188) auch auf eine systematische Mehrfachannotation der spatialen und der kausalen Lesart z. B. bei *stolpern* verwiesen.

Zwar ist es richtig, dass sowohl in den Fällen, in denen es um Gemütszustände geht, als auch in den Fällen, bei denen eine spatiale Lesart vorliegt, die auch eine kausale Komponente hat, in gewisser Weise ein Kausalitätsverhältnis besteht, allerdings sind beide Lesarten voneinander zu trennen: In der Literatur zu Experiencer-Objekt-Verben im Deutschen, also Verben wie z. B. *ärgern*, die über ein

Experiencer- und ein *Stimulus*-Argument verfügen, wobei das *Experiencer*-Argument durch das Objekt ausgedrückt wird, wird *über* zuweilen als Präposition genannt, die bei entsprechenden *Experiencer*-Subjekt-Formen⁵⁹ dieser Verben die semantische Rolle *Subject Matter* einführt (Temme & Verhoeven, 2017, S. 291; Hirsch, 2018, S. 250 f.; zumindest implizit und unter einem anderen Namen bei Fanselow, 1992, S. 292 f.).⁶⁰ *Subject Matter* ist eine der beiden Rollen für den *Emotionsgegenstand* („object of emotion“), die Pesetsky (1995) der *Verursacher*-Rolle („causer“) gegenüberstellt: Pesetsky versucht, das Problem der uneinheitlichen Zuweisung von thematischen Rollen zu syntaktischen Funktionen bei Psych-Verben (bei *Experiencer*-Objekt-Verben ist der Stimulus das Subjekt und der *Experiencer* das Objekt, bei *Experiencer*-Subjekt-Verben ist es andersherum) dadurch zu lösen, dass er annimmt, dass es eine einheitliche *Stimulus*-Rolle gar nicht gibt, sondern drei verschiedene Rollen unterschieden werden müssen, nämlich *Verursacher*, *Subject Matter* und (*Emotions*-)Ziel („target of emotion“, „target“), wobei *Subject Matter* und *Ziel* auch als *Emotionsgegenstand* zusammengefasst werden, obgleich es sich durchaus um verschiedene Rollen handle. Den Unterschied zwischen *Verursacher* und *Subject Matter* erklärt Pesetsky (1995, S. 57) anhand des folgenden Beispiels:

- (37) a. John worried about the television set.
b. The television set worried John.

Während in (37a) das Fernsehgerät der Gegenstand von Johns Sorgen ist, verursacht es diese in (37b) lediglich, ohne notwendigerweise selbst deren Gegenstand zu sein, was allerdings auch nicht ausgeschlossen ist. Beide Sätze unterscheiden sich also in ihren Wahrheitsbedingungen. Pesetsky führt aus, dass (37b) im Gegensatz zu (37a) z. B. in einer Situation verwendet werden könnte, in der John ein Detektiv ist und im Wohnzimmer eines vorgeblich blinden Mannes ein Farbfernsehgerät erblickt. Als Beispiel für ein Emotionsziel gibt Pesetsky (1995, S. 56) (38):

- (38) a. Bill was very angry at the article in the *Times*. [Target]
b. The article in the *Times* angered/enraged Bill. [Causer]

Während Bill in (38a) wütend über den Artikel selbst sein muss, kann es bei (38b) auch sein, dass der Artikel Bill nur wütend macht, indem er z. B. eine Wut auf eine korrupte Regierung auslöst (Pesetsky, 1995, S. 56). *Emotionsziel* und *Subject Matter* unterscheiden sich darin, dass beim *Emotionsziel* immer eine Bewertung des Gegenstandes stattfindet (Pesetsky, 1995, S. 63). Da es sich um unterschiedliche Rollen handelt, können sie auch bei ein und demselben Prädikat gemeinsam auftreten; Pesetsky (1995, S. 63) gibt dafür u. a. folgendes Beispiel:

- (39) Sue is angry with Bill about the party.

In (39) wird mit *with Bill* das *Emotionsziel* und mit *about the party* das *Subject Matter* ausgedrückt.

In der weiteren Literatur zu psychologischen Prädikaten wird die Unterscheidung zwischen *Subject Matter* und *Emotionsziel* nicht immer so gezogen: Reinhart (2001, 2002) etwa bezieht sich zwar auf

⁵⁹Temme und Verhoeven (2017, S. 291) betrachten anscheinend das Partizip II des jeweiligen Verbs, wenn es die Realisation des Stimulus in einer PP zulässt, also Formen wie z. B. *erfreut* oder *beunruhigt*. Hirsch (2018) dagegen geht auf die möglichen Präpositionen im Zuge seiner Betrachtung von systematisch mit dem *Experiencer*-Objekt-Verb alternierende Formen, die mit dem Reflexivum *sich* gebildet werden, also etwa *sich ärgern*, ein.

⁶⁰Es gibt meines Wissens nach für das Deutsche keine systematische Untersuchung der Präpositionen, die zusammen mit psychologischen Prädikaten verwendet werden. Neben den beiden genannten Arbeiten wird das Thema auch in (Temme, 2014) berührt. Bejan (2002) geht in ihrer Arbeit zu Nominalisierungen von Psych-Verben auch auf Präpositionen ein, nimmt aber an, dass diese regiert sind (wofür sie auch auf Grundlage der Kasus der internen Argumente argumentiert), und strebt keine vollständige Auflistung der relevanten Präpositionen an.

Pesetsky (1995), benutzt aber *Subject Matter* auch als Label für den *Emotionsgegenstand* in Beispielen wie (38a).⁶¹

Auf Reinharts *Subject-Matter*-Begriff beziehen sich Temme und Verhoeven (2017). Ihnen zufolge steht bei *Experiencer-Subjekt-Formen* deutscher *Experiencer-Objekt-Verben*⁶² entweder *über* oder *von*, wobei manche Verben mit beiden Präpositionen aufträten und manche mit nur einer der beiden (S. 290 f.). Dabei könne in einer *über-PP* nur ein *Subject Matter* und kein *Verursacher* stehen.

Hirsch (2018, S. 250 f.) verknüpft *über* ebenfalls mit der *Subject-Matter*-Rolle, nennt aber mit *vor* eine weitere Präposition, die dafür verwendet wird.⁶³ Beide grenzt er von *durch* ab, das bei *Verursachern* verwendet werde. Die *Experiencer-Subjekt-Formen*, die er an dieser Stelle betrachtet, sind im Gegensatz zu denen bei Temme und Verhoeven (2017) keine *Partizipien*, sondern systematisch mit dem *Experiencer-Objekt-Verb* alternierende *Formen*, die mit *sich* gebildet werden, wie etwa *sich ärgern*.

Um der Begriffsverwirrung durch die unterschiedliche Verwendung des *Subject-Matter*-Begriffs in der Literatur entgegenzuwirken sei an dieser Stelle festgehalten, dass ich an Pesetskys Bezeichnungen grundsätzlich festhalten werde. Es ist festzustellen, dass Hirsch (2018) und Temme und Verhoeven (2017) *über* als eine Präposition betrachten, die verwendet werden kann, um einen *Emotionsgegenstand* in diesem Sinne auszudrücken.

Betrachtet man nun die Beispiele für die konditional-kausale Lesart bei *über* von Kiss et. al. (2020, S. 182 f.), welche hier als (40) aufgeführt werden, fällt auf, dass in allen Fällen eine *Emotionsgegenstandslesart* vorliegt.

- (40) a. FDP befriedigt *über Internierung* von Dealern
 b. Nichtsdestoweniger spricht blanke Verärgerung *über ungleiche Behandlung* aus seinen Erklärungen, welche sich mit einem weitreichenden arabischen und muslimischen Konsens decken. [+regiert]
 c. Im Anschluss an die allzu schnell beendete Partie liess Gerson als sichtlich enttäuschter Coach von Näfels seinem Ärger *über die Niederlage* freien Lauf und kritisierte die Terminierung des Cup-Finals unmittelbar im Anschluss an eine lange Saison. [+regiert]
 d. Ich freue mich *über die Berichtigung* im Leserbrief von Ian A. Bates, dass ab 1996 auf allen internationalen Flügen – also nicht nur über den Pazifik – ein Rauchverbot gelten soll. [+regiert]

In (40a) etwa ist die *Internierung* nicht nur Ursache, sondern Gegenstand des *Befriedigungsgefühls* der FDP. Speiste sich die *Befriedigung* z. B. nur daraus, dass die *Internierung* als Indiz für eine von der Partei gewünschte Härte in der Justiz genommen wird, wobei ihr der Umgang mit den Dealern im konkreten Fall ganz gleichgültig ist, könnte (40a) nur verwendet werden, wenn *über* etwa durch *wegen* ersetzt würde.

Da sich die *Emotionsgegenstandslesart* deutlich von anderen kausalen Lesarten unterscheidet, scheint es ratsam, sie in möglichen weiteren Ausgaben von (Kiss et. al., 2020) nicht mehr unter der gleichen Kategorie zu führen. Das muss nicht heißen, dass für *über* gar keine kausale Lesart mehr angenommen werden sollte: Kiss et. al. (2020, S. 188) schreiben bei Beispielen mit *straucheln über* oder *stolpern über*

⁶¹ Gleichwohl erwähnt sie auch die *Emotionsziel*-Rolle, die sie allerdings eher mit einem *Goal* in Beziehung setzt (Reinhart, 2001, S. 383; Reinhart, 2002, S. 273).

⁶² Gemeint zu sein scheint damit das *Partizip II* des jeweiligen Verbs, wenn es die Realisation des Stimulus in einer *PP* zulässt.

⁶³ *Vor* kommt in einer *Emotionsgegenstandslesart* in Hirschs (2018) Beispielen nur einmal vor, nämlich bei „Mir graut (es) vor dir“ (S. 57).

eine Mehrfachannotation mit der (ggf. metaphorischen) Traversen-Lesart und der kausalen Lesart vor. Unter den annotierten Sätzen gibt es nur ein entsprechendes Beispiel:

(41) 34223

(de) Mit dieser Einsicht und dem hehren Anspruch, die Sünden der Vergangenheit nicht zu wiederholen - schließlich ist vor knapp einem Jahr die letzte Kommission ÜBER Betrugs- und Unregelmäßigkeitsfälle zu Fall gekommen -, ist die Kommission Prodi im September angetreten.

Es ist naheliegend, bei derartigen Beispielen eine Mehrfachannotation mit einer kausalen Lesart vorzuschreiben, wenn es unabhängige Gründe gibt, die Existenz einer solchen anzunehmen; fallen diese Gründe aber weg, ist fraglich, ob es wirklich notwendig ist, *über* überhaupt eine kausale Lesart zuzuschreiben. Ich werde nun dafür argumentieren, dass *über* diese Lesart hat. Eine auf den ersten Blick interessante Alternative wäre, zu sagen, dass die kausale Interpretation nicht von *über* herrührt, sondern sich aus allgemeinen Schlüssen über die Natur des Stolperns und Strauchelns ergibt: In der Regel gibt es eine Ursache dafür und wird ein Gegenstand genannt, über (im Traversensinne) den man gestolpert ist und der so beschaffen ist, dass er als Ursache des Stolperns infrage kommt, kann vermutet werden, dass es sich bei diesem auch tatsächlich um die Ursache handelt. Man vergleiche (42a) mit (42b):

- (42) a. Die Bürgermeisterin ist am Montag betrunken über den Rathausplatz gestolpert.
b. Die Bürgermeisterin ist am Montag betrunken über die Türschwelle gestolpert.

In beiden Fällen ist eine Traversenlesart anzunehmen, aber bei (42a) ist die kausale Lesart (in der das Stolpern etwa durch eine Unebenheit im Rathausplatzpflaster oder ähnliches verursacht wird) wesentlich weniger prominent als bei (42b). Je eher das interne Argument der PP als Ursache infrage kommt, desto prominenter scheint die kausale Lesart zu sein. Diese Unterschiede ließen sich jedoch bei Annahme einer kausalen Lesart ebenso gut erklären, denn es handelt sich um Unterschiede in der Prominenz verschiedener Lesarten, die aber vorhanden sein müssen, wobei es unerheblich ist, ob die Lesarten bereits durch die Präposition gegeben sind oder erst inferiert werden mussten. Aus Sparsamkeitsgründen wäre dennoch die alternative Erklärung zu bevorzugen, allerdings vermag sie nicht zu erklären, warum in (43) keine kausale Interpretation möglich ist:

(43) Die Bürgermeisterin ist am Montag betrunken über der Türschwelle gestolpert.

Hier steht das interne Argument der Präposition im Dativ und die kausale Interpretation ist nicht vorhanden. Es ist nicht ersichtlich, warum die Türschwelle in (43) nicht ein ebenso guter Grund zu stolpern sein sollte wie in (42b). Geht man dagegen von einer kausalen Bedeutung bei *über* aus, kann man sagen, dass *über* in dieser mit Akkusativ steht, weswegen sie in (43) nicht verfügbar ist.

Nachdem festgestellt wurde, dass *über* eine Emotionsgegenstandslesart hat, stellt sich die Frage, ob es nur für *Emotionsziele* oder nur für *Subject Matter* oder für beides verwendet werden kann. (44) scheint eine (bedeutungserhaltende) Übersetzung von (38a) zu sein (eine emotionale Bewertung des Artikels findet auf jeden Fall statt), was zeigt, dass *über* für das *Emotionsziel* verwendet werden kann.

(44) Willi war sehr wütend über den Artikel in der *Times*.

Nicht zwangsläufig mit einer emotionalen Bewertung einher geht die Verwendung von *über* in (45), weswegen hier ein *Subject Matter* im engeren Sinne vorliegt.

(45) 63068

(de) Meine Fraktion hat sich ferner ÜBER das Verhalten einiger - von der EU leider auch noch subventionierter - NRO im Zusammenhang mit ihrer unseres Erachtens nicht sehr konstruktiven Rolle im Rahmen der WTO ein wenig erstaunt gezeigt .

Eine andere Frage ist, ob *über* im Deutschen die einzige Präposition für den *Emotionsgegenstand* ist. Sie kann hier nicht beantwortet werden, da dazu auch der Vergleich mit anderen Präpositionen, die bei psychologischen Prädikaten einen Stimulus zu kennzeichnen scheinen, nötig wäre. Zu nennen wären diesbezüglich mindestens noch *vor* (Hirsch, 2018, S. 123; Bejan, 2002), *um* (Bejan, 2002), *an* (*sich erfreuen/ergötzen an*) und *auf* (*wütend/sauer/böse auf*) sowie *für* (*sich begeistern/interessieren/faszinieren für*). Es ist zu erwarten, dass nicht bei allen Kandidaten eine Emotionsgegenstandslesart vorliegt: *Auf* etwa scheint in (46b) ein *Emotionsziel* anzugeben, allerdings wird die entsprechende PP im Gegensatz zur *über*-PP in (46a) und zu den *über*-PPen mit Emotionsgegenstandsinterpretation in den anderen Beispielen oben nicht propositional interpretiert. Dass Stimulus-Argumente propositional interpretiert werden, wird in der Literatur zuweilen behauptet.⁶⁴

- (46) a. Ümit war wütend über sich selbst.
 b. Aaron war wütend auf sich selbst.
 c. Waltraud war wütend wegen sich selbst.

Wahrscheinlich sollte die PP in (46b) eher als *Ziel* betrachtet werden, auch wenn eine weiterführende Diskussion hier nicht erfolgen kann.

Eine Klassifikation der weiteren Präpositionen, mit denen ein *Emotionsgegenstand* ausgedrückt werden kann, ist ein drängendes Forschungsdesiderat, denn entsprechende Ergebnisse wären für gleich mehrere Bereiche relevant: 1. ist eine angemessene Erfassung der Bedeutungen von Präpositionen eine grundlegende Voraussetzung für viele weitere Arbeiten zu PPen und Arbeiten, in denen PPen für syntaktische Tests genutzt werden: Wenn man etwa ein syntaktisches Phänomen untersucht, bei dem man davon ausgeht, dass die Semantik der PP eine Rolle spielt, benötigt man eine angemessene semantische Klassifikation der Präpositionen. Verbesserungen solcher Klassifikationen sind daher erstrebenswert. 2. kann eine Klassifikation entsprechender Präpositionsbedeutungen für Arbeiten zu psychologischen Prädikaten von Bedeutung sein: Temme und Verhoeven (2017)⁶⁵ und Hirsch (2018) z. B. argumentie-

⁶⁴Etwa bei Bott und Solstad (2014), was durch Hirsch (2018) aufgegriffen wird. Engelberg und Cosma (2014, S. 370 ff.) behaupten ähnliches, lassen sich aber durch ein Beispiel, bei dem sie offensichtlich nur die (prominentere) agentive Lesart erkennen, dazu verleiten, zu sagen, dass es bei NPen, deren Referent belebt ist, anders sei. Es ist allerdings nicht ganz sicher, ob sich die Behauptung für *Emotionsziele* aufrechterhalten lässt: Bei (38a) scheint es zumindest fraglich, bei (39) ist es sicherlich nicht der Fall, dass das *Emotionsziel* propositional interpretiert wird (was aber bei (44) interessanterweise durchaus geschieht). Damit könnte man auf verschiedene Arten umgehen: So wäre es einerseits möglich, die meist auf wenigen, prototypischen Beispielen aufbauende Behauptung, dass *Stimuli* propositional interpretiert werden, als falsifiziert anzusehen, andererseits könnte man sich aber auch fragen, ob es sich bei *Emotionszielen* tatsächlich um *Stimuli* handelt. Mit Reinharts weiterem Subject-Matter-Begriff könnten alle Beispiele mit propositionaler Interpretation aufgefangen werden.

⁶⁵Sie nutzen Unterschiede in der Präposition, die mit dem Partizip II eines Verbs steht, um Verben mit Subject-Matter-Argument (im weiteren, Reinhart'schen Sinne) zu identifizieren. Das ist unter anderem deshalb interessant, weil *backward binding* (die Bindung eines im Subjekt befindlichen Reflexivums durch eine Objekt-NP, für deren Zulassen Experiencer-Objekt-Verben spätestens seit (Belletti & Rizzi, 1988) bekannt sind) nach Reinhart (2001, 2002) nur bei Subject-Matter-Subjekten möglich ist. Problematisch ist bei diesem Vorgehen jedoch, dass zumindest bei einigen Verben einer der Unterschiede zwischen der EO-Form und einer reflexiven ES-Form (z. B. *sich ärgern*) darin zu bestehen scheint, dass die PP bei der ES-Form einen *Emotionsgegenstand* angibt und das Subjekt der EO-Form ein *Verursacher* ist (vgl. Fanselow, 1992, S. 292 f.). Ähnliches gilt auch für die Partizipien: Es kann einen z. B. eine Dokumentation betrüben, ohne dass man betrübt *über* sie wäre.

ren damit, dass *über* eine Subject-Matter-Präposition ist. 3. könnte ein systematischer Vergleich mit anderen Präpositionen auch Licht auf den Rektionsstatus von *über* in vielen Fällen mit Emotionsgegenstandslesart werfen. Das wiederum wäre für allgemeine Arbeiten zur Rektion relevant. So sind bei Kiss et. al. (2020, S. 187) nur vier der 31 als *über* regierend klassifizierten Adjektive keine psychologischen Prädikate. Wenn gezeigt werden könnte, dass es sich bei der Emotionsgegenstandslesart um eine ganz normale Bedeutung von *über* handelt, sich *über* dabei aber systematisch von allen anderen Präpositionen mit Emotionsgegenstandslesarten unterscheidet, müssten die entsprechenden Vorkommen von *über* entgegen Bejan (2002)⁶⁶ nicht mehr als regiert betrachtet werden.⁶⁷

Eine weitere Frage, die bei der Annotation zwar nicht unmittelbar relevant geworden ist, aber bei einer Betrachtung der Emotionsgegenstandslesart gestellt werden sollte, ist, ob diese nur bei prototypischen psychologischen Prädikaten vorkommt: Ein Beispiel für ein Verb, bei dem man *über* vielleicht ebenfalls diese Lesart wird zuschreiben wollen, ist *lachen*. Man bemerke diesbezüglich auch, dass das englische *laugh* mit den für *Emotionsgegenstände* typischen Präpositionen *at*, *about* und *over* steht.⁶⁸

Zusammenfassung *Über* kann verwendet werden, um *Emotionsgegenstände* im Sinne Pesetskys (1995) auszudrücken. Diese Lesart ist von der kausalen Lesart, wie sie bei Prädikaten wie *stolpern* anzutreffen ist, zu unterscheiden. Eine Untersuchung der bei Emotionsgegenständen im Deutschen verwendeten Präpositionen steht noch aus, ist aber in mancherlei Hinsicht wünschenswert.

⁶⁶Bejan (2002, S. 24) bemerkt, dass von derselben Wurzel abgeleitete Psych-Nomina, -Adjektive und reflexive Experiencer-Subjekt-Verben stets PPs mit der gleichen Präposition selektieren und die internen Argumente der PPs stets den gleichen Kasus tragen. Ihrer Ansicht nach sind diese Präpositionen regiert und dienen nur dem θ -Rollen-Transfer vom Kopf zur NP. Das muss aber nicht der Fall sein, wenn man davon ausgeht, dass die Präpositionen alle unterschiedliche Bedeutungen haben und die auf derselben Wurzel basierenden Formen verwandte Bedeutungen. Dann kann es sein, dass immer wieder die gleiche Präposition gewählt wird, einfach weil nur sie die passende Bedeutung hat. Man könnte etwa sagen, dass die Wurzeln jeweils eine bestimmte thematische Rolle (und nicht etwa eine PP mit einer bestimmten Präposition) verlangen oder semantisch mit einer bestimmten Art von Argument besonders kompatibel sind und immer die gleiche Präposition gewählt wird, weil nur mit ihr die entsprechende Bedeutung ausgedrückt werden kann (auf ähnliche Weise argumentiert Müller (2013, S. 50 f.) dagegen, dass regierte Präpositionen dem Kasustransfer dienen). Dazu würde auch passen, dass es einige Psych-Wurzeln gibt, bei denen die Präpositionen durchaus kommutieren und dann in paradigmatischer Opposition zueinander stehen: *sich ängstigen vor* ist nicht das gleiche wie *sich ängstigen um* (ein Beispiel, das Bejan (2002, S. 24) sogar zitiert. Trotzdem schreibt sie in beiden Fällen der PP die Thema-Rolle zu). Alternativ könnte man auch annehmen, dass *über* die Emotionsgegenstandspräposition ist (und somit in den entsprechenden Fällen nicht regiert), während die anderen Präpositionen es nicht sind und somit als regiert klassifiziert werden sollten.

⁶⁷Für das Englische ist die Situation ähnlich: Wie die Abbildungen in Abschnitt 4.1.3 zeigen, kommen im Korpus hier *at*, *in*, *over* und *with* in entsprechenden Lesarten vor. (37a) belegt, dass auch *about* möglich ist. Es ist unklar, ob es zwischen diesen Präpositionen einen Bedeutungsunterschied gibt oder ob die sie regiert werden. Alexiadou und Iordăchioaia (2014), die fürs Griechische und Rumänische zeigen, dass bestimmte Präpositionen für *Verursacher* und *Emotionsgegenstand* verwendet bzw. nicht verwendet werden, nehmen an, dass im Englischen die intransitiven Varianten alternierender Psychverben Präpositionen auf idiosynkratische Weise selektieren, wobei *from* (welches für die Kausativalternation typisch wäre) ausgeschlossen ist.

⁶⁸Das ist auch relevant für die Diskussion um Rektion bei Präpositionen: Dass sprachübergreifend Präpositionen, die keine prototypischen Übersetzungen voneinander sind, gewählt werden, um auszudrücken, worüber gelacht wird (*lachen über*, *laugh at*, *rire de*, *griech. jeláo me* ((vermeintlich) wörtlich *mit*)), scheint prima facie dafür zu sprechen, die Präpositionen als bedeutungslos zu betrachten. Wenn die *über*-PP hier aber einen *Emotionsgegenstand* angibt, ist *at* die nach den Auswertungen in Abschnitt 4.1.3 naheliegendste Entsprechung. (Interessanterweise wird das griechische *me* gerade nicht für *Emotionsgegenstände* verwendet sondern für *Verursacher* (Alexiadou & Iordăchioaia, 2014). Dies könnte – wenn man nicht annehmen will, dass entweder *me* doch bei *Emotionsgegenständen* oder *über* auch bei *Verursachern* verwendet werden kann – erklärt werden, indem man sagt, dass *jeláo me* keine genaue Übersetzung von *lachen über* ist, was dann zu prüfen wäre.)

5.2. Bezugspunkt und Thema

Wie in Abschnitt 2.3 dargestellt wurde, war die Unterscheidung zwischen den beiden Lesarten *Bezugspunkt* und *Thema* bei der Annotation nicht immer einfach zu treffen. Auch die Angemessenheit der Bedeutungsannotation für den englischen Satz war hier oftmals schwer zu beurteilen. In diesem Abschnitt soll unter Verwendung der Annotationen für die deutschen Daten der Beziehung zwischen den beiden Lesarten nachgegangen werden. Das Ergebnis der Diskussion wird sein, dass nicht davon auszugehen ist, dass *über* eine von der Themasart zu trennende Bezugspunktlesart hat.

Um die Beziehung zwischen den beiden Lesarten untersuchen zu können, ist es zunächst hilfreich, zu betrachten, wie sie in (Kiss et. al., 2020) unterschieden werden: Zur Bezugspunktlesart wird gesagt, dass es um einen „personalen oder nicht-personalen Bezugspunkt“ (S. 183) gehen könne und dass eine Ersetzbarkeit von *über* durch *bezüglich* oder *in Bezug auf* zumeist gegeben sei. Die Themasart wird dadurch von der Bezugspunktlesart abgegrenzt, dass unter sie ausschließlich Beispiele fallen sollen, bei denen eine durch das externe Argument ausgedrückte „schriftliche oder mündliche Äußerungen in Form von Büchern, Aufsätzen, Diskussionen und Gesprächen“ (S. 185) vorliegt.

Diese Abgrenzung ermöglicht die Unterscheidung mehrerer Lesarten bei Fällen wie (47), wo eine Studie als schriftliches Dokument über ein bestimmtes Thema von einer Studie als Untersuchung eines Sachverhalts zu unterscheiden ist. Diesen Bedeutungen von *Studie* scheinen auch zwei unterschiedliche Bedeutungen von *über* zu entsprechen.

(47) 47107

(de) Und beabsichtigt die Kommission , eine Studie ÜBER die Sicherheit der Kernkraftwerke in die Wege zu leiten , wie es in Deutschland der Fall ist ?

In Beispielen wie (48) und (49) kann anhand der Kriterien einfach und schnell eine Bedeutung zugewiesen werden:⁶⁹

(48) 78315: *Bezugspunkt*

(de) Bevor wir mit der Erdölförderung und Plänen für den Bau von Pipelines in der Ostsee beginnen , sollten wir ÜBER Umweltfragen nachdenken .

(49) 76698: *Thema*

(de) Als allgemeine Anmerkung zu den Hauptaspekten , die in dem Bericht enthalten sind , ÜBER den wir gerade debattieren und der von entscheidender Bedeutung für den Fischereisektor ist , kommt es unserer Meinung nach darauf an , die Unterschiede zwischen der langfristigen und mittelfristigen Verwirklichung des Prinzips der Nachhaltigkeit in den Fischereien deutlich herauszuarbeiten .

Problematisch sind dagegen Korpusbelege wie (50): Bei diesem Beispiel, das wegen der fehlerhaften Satzgrenzenalignierung ausgeschlossen wurde, wäre streng genommen *Bezugspunkt* zu annotieren gewesen, da es sich bei einer Konferenz nicht um eine Äußerung handelt.

(50) 83288

(de) Auch die im kommenden November stattfindende internationale Konferenz ÜBER die Entwicklungsfinanzierung wird in dieser Hinsicht eine ausschlaggebende Rolle spielen . Wir

⁶⁹Man bemerke jedoch, dass *über* in (48) nicht durch *bezüglich* oder *in Bezug auf* ersetzt werden kann. Entscheidend ist hier, dass es nicht um eine Äußerung geht.

dürfen bei den großen Herausforderungen , vor denen wir stehen , die Interessen der Entwicklungsländer nicht aus dem Auge verlieren !

Sinnvoller wäre allerdings eine Einordnung als *Thema*, denn, was in der *über*-PP angegeben wird, ist eben das Thema der einzelnen Beiträge und Diskussionen auf der Konferenz.

Was erhebliche Zweifel an der Sinnhaftigkeit der Unterscheidungskriterien von Kiss et. al. (2020) aufkommen lässt, ist folgendes Beispiel:

(51) 48277

(de) Frau Präsidentin , seit dem 11 . September hatten wir viele Gelegenheiten , um ÜBER die Folgen der Ereignisse jenes fatalen Tages nachzudenken und zu diskutieren , und der heutige Tag bildet dabei keine Ausnahme .

Wenn es sich bei *Thema* und *Bezugspunkt* tatsächlich um unterschiedliche Lesarten handelt und diese wie bei Kiss et. al. (2020) unterschieden werden können, sollte hier ein Zeugma vorliegen, denn *über* sollte bei *nachdenken über* wie in (48) eine Bezugspunktlesart haben, bei *diskutieren über* aber ähnlich wie in (49) eine Themasart. Tatsächlich liegt aber kein Zeugma vor.⁷⁰ Dies kann auf zweierlei Art erklärt werden:⁷¹ 1. Es gibt keinen Unterschied zwischen *Thema* und *Bezugspunkt* und die in (Kiss et. al., 2020) getroffene Unterscheidung ist rein artifiziell in dem Sinne, dass zwar konzeptuell zwei Bedeutungen unterschieden werden können, *über* in Bezug auf diese aber vage (im Sinne von unbestimmt) ist. 2. Es gibt einen Unterschied zwischen beiden Lesarten, aber in (51) liegt zweimal die gleiche Lesart vor. Daraus würde folgen, dass die Grenze zwischen den Lesarten in (Kiss et. al., 2020) falsch gezogen wird. Um zu bestimmen, ob 1. oder 2. der Fall ist, sollen einerseits Fälle betrachtet werden, bei denen eine Bezugspunktlesart vorzuliegen scheint, die sich nicht mit einer Themasart vereinigen lässt, und andererseits Fälle, die potenziell ambig sind. Wenn *über* in Bezug auf die Thema-Bezugspunkt-Unterscheidung tatsächlich ambig ist, ist zu erwarten, dass sich Beispiele konstruieren lassen, bei denen etwa im Gegensatz zu (51) tatsächlich ein Zeugma vorliegt oder die in der einen Lesart wahr und in der anderen falsch sein können, denn diese Möglichkeiten, Ambiguität von Unbestimmtheit zu unterscheiden, werden (neben anderen⁷²) in der Literatur genannt (zu Tests zur Unterscheidung von Ambiguität und Vagheit s. u. a. Cruse, 1986; Pafel & Reich, 2016; Zwicky & Sadock, 1975).

Pafel und Reich (2016) nennen den erstgenannten Test den „Zeugma-Test“ und den letztgenannten den „Wahr/Falsch-Test“ (S. 45). Wie sie funktionieren, sei hier anhand ihrer Beispiele dargestellt: Getestet werden soll, ob die beiden Ausdrücke *schwer* und *Kind* ambig oder vage (also unbestimmt) sind. *Schwer* kann sich sowohl auf das Gewicht als auch auf die Verständlichkeit beziehen, *Kind* sowohl auf männliche als auch auf weibliche Personen.

(52) Wahr/Falsch-Test (Beispiele von Pafel und Reich (2016, S. 45))

a. Das Buch ist schwer.

b. Die Müllers haben ein Kind.

⁷⁰Der Zeugma-Test wird unten ausführlicher erläutert.

⁷¹Eine dritte Erklärungsmöglichkeit wäre, zu sagen, dass *diskutieren über* ambig ist und neben einer Thema- auch eine Bezugspunktlesart hat. Das ist aber unbefriedigend, denn es stellt sich 1. die Frage, warum man überhaupt eine Themasart annehmen sollte, wenn *über* dann selbst in solch prototypischen Fällen wie mit *diskutieren* ambig ist, und 2. würde man dann erwarten, dass Sätze wie *Wir haben über die Sache diskutiert* gleichzeitig wahr (in Bezug auf die eine Lesart) und falsch (in Bezug auf die andere) sein können (vgl. die Diskussion des Wahr/Falsch-Tests unten), was aber nicht der Fall zu sein scheint.

⁷²Es gibt noch eine ganze Reihe weiterer Tests, aber die beiden genannten sind hier am ehesten anwendbar.

Der Wahr/Falsch-Test zeigt, dass bei *schwer* Ambiguität vorliegt, bei *Kind* aber Vagheit, denn (52a) kann in der einen Lesart wahr und in der anderen gleichzeitig falsch sein, (52b) dagegen ist unabhängig vom Geschlecht des Kindes wahr oder falsch.

Der Zeugma-Test beruht darauf, dass sich bei Ambiguität ein Zeugma-Effekt ergibt, wenn beide Lesarten gleichzeitig gemeint sind, bei Vagheit aber nicht. So kann etwa (53b) auch verwendet werden, wenn beide Familien Kinder unterschiedlichen Geschlechts haben.

- (53) Zeugma-Test (Beispiele von Pafel und Reich (2016, S. 45))
- a. ? Das Gewicht und das Verständnis des Buchs, beides ist schwer.
 - b. Die Müllers und die Maiers, die haben beide ein Kind.

Es mag nun scheinen, dass Korpusbelege, bei denen beide Lesarten annotiert wurden, besonders geeignet sind, lexikalisches Material zur Beispielkonstruktion für den Wahr/Falsch-Test zu liefern (wenn sie nicht gar selbst als entsprechende Beispiele fungieren können), aber in den tatsächlich vorkommenden Belegen ist – wie sich herausstellt – die Quelle der Ambiguität nicht *über*, sondern wie in (47) dessen Regens. Dass beide Lesarten annotiert wurden, ist bei folgenden Regentien der Fall:

- (54) *Abkommen, sich äußern, Bestimmung, Debatte, Dekret, Entschließung, Gesetz, Protokoll*⁷³, *Rechtsakt, Rechtsvorschrift, Richtlinie, Studie, Übereinkommen, Untersuchung, Verhandlung, Verordnung, Vertrag, Vorschlag, Zuverlässigkeitserklärung*

Meistens handelt es sich also um Nomina, die neben einem *x* auch ein Schriftstück, in dem *x* niedergelegt ist, bezeichnen können (im EU-Kontext ist dies etwa auch bei *Entschließung, Vorschlag*⁷⁴ und *Richtlinie* der Fall). Es ist bei der Annotation davon ausgegangen worden, dass der Schriftstücklesart die Themasart bei *über* entspricht und der anderen Lesart (also etwa bei *Gesetz* die Lesart *staatliche Vorschrift*) die Bezugspunktlesart. Wenn das so ist, sind diese Nomina zur Konstruktion von Beispielen für den Wahr/Falsch-Test ungeeignet, denn es ließe sich nicht mehr feststellen, ob das Verhalten des Satzes durch *über* oder das Nomen verursacht wird. Ob es tatsächlich der Fall ist, ist aber nicht unmittelbar ersichtlich: Dass bei der Schriftstücklesart des Regens die Themasart von *über* vorliegen muss, ist klar, wie auch folgendes, auf Basis von Beleg 24757 konstruiertes Beispiel zeigt:

- (55) Der spanische König unterzeichnete ein Dekret über internationale Sportaktivitäten.

Hat das Regens jedoch wie in (56) nicht die Schriftstücklesart, ist es nicht leicht zu sagen, ob neben der Bezugspunktlesart – deren Vorhandensein sich schon darin zeigt, dass Ersetzbarkeit durch *in Bezug auf* gegeben ist – nicht auch eine Themasart⁷⁵ vorhanden ist.

- (56) Der spanische König erließ ein Dekret über internationale Sportaktivitäten.

Sollte dies zutreffen, zeigt der Wahr/Falsch-Test, dass keine Ambiguität vorliegt, denn dann müsste der Satz in der einen Lesart wahr sein können, während er in der anderen falsch ist, was aber nicht der Fall ist. Damit ist freilich noch nicht gezeigt, dass es keine unabhängige Bezugspunktlesart gibt: 1. geht es um sehr subtile Unterschiede, weswegen die Beispiele schwer zu bewerten sind und also nicht nur

⁷³Gemeint ist im relevanten Beispiel (72056) nicht eine Sitzungsniederschrift, sondern (vielleicht beeinflusst durch das englische *protocol*) ein Verhaltenskodex.

⁷⁴Es finden sich Formulierungen wie „der Anfang nächsten Jahres fertiggestellte Vorschlag“ (Beispiel 51504).

⁷⁵Natürlich nicht im Sinne von Kiss et. al. (2020), sondern in einem prätheoretischen Sinne, von dem ausgehend man dann eine genauere Abgrenzung definieren könnte.

den introspektiven Urteilen einer einzelnen Person vertraut werden sollte, 2. kann es sein, dass die Beispielauswahl täuscht und es geeignetere Regentien gibt, bei denen sich dann doch eine Ambiguität zeigt,⁷⁶ und 3. könnte man auf dem Standpunkt beharren, dass *über* in (56) gar keine Themasart hat.

Um Beispiele mit Zeugma erstellen zu können, werden Regentien benötigt, bei denen *über* eindeutig die eine oder die andere Lesart zu haben scheint. Kandidaten können aus den Beispielen für die Bezugspunktlesart in (Kiss et. al., 2020, S. 183) übernommen werden, da davon auszugehen ist, dass für die Beispiele besonders klare Fälle gewählt wurden. Man findet dort *Einigung*, *informieren*, *Entscheidung*, *Kritiken* und *Einigkeit*. Regentien aus dem Korpus, die eindeutig nur eine Bezugspunktlesart zu haben scheinen, sind u. a. *abstimmen*, *beschließen* und *bestimmen*. Als Regentien, bei denen *über* eindeutig eine Themasart hat, sollen hier *sprechen*, *Gespräch* und *Gesprächsbedarf*⁷⁷ verwendet werden. (57) enthält konstruierte Beispiele für alle diese Ausdrücke und in (57h) zum Vergleich ein Beispiel, in dem *über* einmal eine modale und einmal eine Themasart haben müsste und ein Zeugma-Effekt klar erkennbar ist.

- (57) a. Es gab gestern keine Gespräche und erst recht keine Entscheidung über den Gesetzesentwurf.
- b. Die Kommissarin hatte viele Gelegenheiten, über den Gesetzesvorschlag zu informieren und zu sprechen.
- c. ? Im Parlamentsgebäude hörte man gestern viele Gespräche und einige Kritiken über die Politik der Kommissarin.
- d. Über den Gesetzesentwurf besteht noch keine Einigkeit, sondern vielmehr dringender Gesprächsbedarf.
- e. Die Abgeordneten hatten viele Gelegenheiten, über den Gesetzesvorschlag zu sprechen und abzustimmen.
- f. Die Abgeordneten hatten viele Gelegenheiten, über den Gesetzesvorschlag zu sprechen und zu beschließen.
- g. ? Der König hatte viele Gelegenheiten, über die Verwendung der Mittel zu sprechen und zu bestimmen.
- h. * Die Abgeordneten hatten viele Gelegenheiten, über den Gesetzesvorschlag die Regierung zu ärgern und zu sprechen.

Ich kann in den meisten der Beispiele aus (57) keinen Zeugma-Effekt feststellen, bewerte aber (57c) und (57g) als marginal. Im Falle von (57c) ist dies jedoch nicht auf einen Zeugma-Effekt zurückzuführen,

⁷⁶Die Regentien aus (54), bei denen es keine Schriftstücklesart gibt, sind *Debatte*, *Verhandlung* und *sich äußern*. Auch mit ihnen habe ich kein Beispiel konstruieren können, das im Wahr/Falsch-Test einen Unterschied zwischen beiden Lesarten zeigt.

⁷⁷*Gesprächsbedarf* wird hier verwendet, weil sich damit einfacher als mit *Gespräch* ein Beispiel für *Einigkeit* konstruieren lässt. Es ist allerdings auch aus einem Grund, der nichts mit der Thema-Bezugspunkt-Unterscheidung zu tun hat, interessant: In dieser Arbeit wird *über* bei Nomina wie *Gespräch* und auch *Gesprächsbedarf* als regiert betrachtet. *Gesprächsbedarf* ist allerdings ein Kompositum, bei dem der Kopf nicht *Gespräch*, sondern *Bedarf* ist, dem man kein Thema-Argument zuschreiben wollen wird. Müller (2013, S. 52) bemerkt, dass es noch andere Komposita gibt, bei denen die Präpositionswahl nicht vom Kopf abhängt. Diese Möglichkeit sei jedoch nur bei bestimmten Köpfen gegeben. Nun ist es möglich, dass es sich um eine idiosynkratische Eigenschaft einer bestimmten Klasse von Nomina handelt, und es ist sicherlich der Fall, dass das Phänomen genauer untersucht werden muss, ehe verlässliche Aussagen getroffen werden können, aber eine sehr einfache Erklärung wäre, dass die *über*-PP in derartigen Fällen gar nicht regiert ist, weil es sich eigentlich um ein Adjunkt handelt. Wenn dem so sein sollte, würde sich die Frage stellen, ob nicht auch die *über*-PPen bei *Gespräch*, *Diskussion* etc. Adjunkte (und folglich auch nicht regiert) sind.

denn ich würde auch (58), das aus dem NZZ-Korpus stammende Beispiel von Kiss et. al. (2020, S. 183), in dem eindeutig kein Zeugma vorliegt, auf die gleiche Weise bewerten.

- (58) ? Fässlers Beruf als Steuer-Treuhänder hatte zudem Kritiken *über die mögliche Kollision* mit dem Amt als Säckelmeister laut werden lassen.

Anders ist es bei (57g). Zwar klingt der Satz wesentlich besser als (57h), aber auch schlechter als die anderen Beispiele. Eine mögliche Erklärung wäre, dass *über* in *bestimmen über* gar keine Thema- oder Bezugspunkt-, sondern eine Machtverhältnislesart hat. Dafür würde auch sprechen, dass (59) wohl als ähnlich akzeptabel wie (57g) zu bewerten ist.

- (59) ? Der König hatte viele Gelegenheiten, über die Verwendung der Mittel zu sprechen und zu verfügen.

Für alle Beispiele in (57) gilt allerdings, dass ihre Bewertung recht heikel ist, weswegen es angebracht wäre, sich nicht auf die Urteile einer einzelnen Person zu verlassen und in einer experimentellen Studie zu erforschen, wie derartige Beispiele bewertet werden.⁷⁸

Da sich nicht hat nachweisen lassen, dass *über* in Bezug auf die Thema-Bezugspunkt-Unterscheidung ambig ist, sollte aus Sparsamkeitsgründen nur eine entsprechende Lesart angenommen werden, die beides umfasst. Es folgen nun einige Überlegungen dazu, was das konkret bedeutet.

Es bedeutet nicht, dass es nicht sinnvoll sein kann, eine Unterscheidung zwischen einer Thema- und einer Bezugspunktlesart zu ziehen: So wie es für bestimmte Anwendungen sinnvoll sein mag, zu entscheiden, ob sich *Kind* (bzw. eine NP, in der dieses Wort der Kopf ist) auf eine männliche oder eine weibliche Person bezieht, kann es auch bei *über* sinnvoll sein, zu entscheiden, ob es sich um einen Bezugspunkt oder ein Thema handelt und es mag sogar möglich sein, Systeme zu entwickeln, die automatisch entscheiden, was davon der Fall ist.

Eine Anwendung, für die diese Unterscheidung sinnvoll sein könnte, ist das Finden von Bedingungen dafür, wann welche englische Entsprechung angebracht ist: Es liegen hier bei weitem nicht genug Daten vor, um generalisieren zu können, aber die Daten legen nahe, dass *about* nur für Thema- und nicht für Bezugspunkt-Argumente verwendet werden kann: Wie Tabelle 37 auf S. 85 im Anhang zeigt, wird *about* nur dreimal als Entsprechung eines Bezugspunkt-*über* verwendet, aber 22 Mal als Entsprechung bei einer Themalesart. Eine Betrachtung der Beispiele mit Bezugspunktlesart in (60) zeigt, dass man *über* dort durchaus eine Themalesart zuschreiben könnte, wenn man gewillt wäre, den Begriff etwas weiter zu fassen als Kiss et. al. (2020).

- (60) a. 53395

(de) Ich habe dabei sehr viel ÜBER Somalia und Somaliland gelernt , vor allem , was das für ein komplexes Gebiet ist .

(en) As a result of that I learnt a lot ABOUT Somalia and Somaliland and I particularly learned what a complex area it is .

⁷⁸In einer Studie könnte man die Teilnehmer solche Sätze auf einer Likert-Skala bewerten lassen und diese Bewertungen mit den Bewertungen für andere Sätze mit verschiedenen Präpositionsbedeutungen vergleichen, bei denen ein Zeugma-Effekt eindeutig erwartet wird oder nicht erwartet wird. Sowohl der Zeugma-Test als auch Experimente scheinen zur Unterscheidung verschiedener Lesarten bei Präpositionen aber sehr selten eingesetzt zu werden. Die einzige mir bekannte Arbeit, die beides kombiniert, ist (Ursini & Giannella, 2016).

- b. 89626
- (de) Würden Sie uns bitte wenigstens einige Informationen ÜBER diesen integrierten Ansatz geben ?
- (en) Please would you give a clue ABOUT that integrated approach ?
- c. 26134
- (de) Nun hat sich die ganze Angelegenheit über zwei Jahre hingezogen , weil zwischen Parlament und Rat erhebliche Differenzen ÜBER die dem Parlament in diesem Prozeß zukommende Rolle entstanden sind .
- (en) The whole issue has been dragging on for two years now because of a fundamental disagreement between Parliament and the Council ABOUT the role that Parliament should be allowed to play in this process .

Zusätzlich müssten auch Fälle wie *nachdenken über* (wegen *think about*) unter den Thema-Begriff fallen. Es sei darauf hingewiesen, dass das Vorhandensein englischer Entsprechungen, die nur unter bestimmten Voraussetzungen Übersetzungen von *über* sein können, für sich genommen kein Beweis dafür ist, dass *über* in Bezug auf diese Lesarten ambig ist. Wäre *über* allerdings tatsächlich ambig, könnte es sein, dass diese Entsprechungen dabei helfen können, die Grenze zwischen den Lesarten zu ziehen.

Man mag sich fragen, wieso *über*, wenn es nur eine Bezugspunkt-Thema-Lesart hat, in dieser manchmal durch *in Bezug auf* bzw. *bezüglich* ersetzt werden kann und manchmal nicht, aber die Substituierbarkeit durch ein Synonym ist kein verlässlicher Ambiguitätstest (Cruse, 1986, S. 54 ff.).

Zusammenfassung Da keine Evidenz dafür gefunden werden konnte, dass *über* hinsichtlich der Thema-Bezugspunkt-Unterscheidung ambig ist, ist für beides aus Sparsamkeitsgründen nur eine Lesart anzunehmen.

5.3. Unterscheidung verschiedener Arten der Traverse mit weniger als drei relevanten Dimensionen

Bei Kiss et. al. (2020, S. 174 f.) wird unter *Traverse* $<3D$ nach einer allgemeinen Beschreibung zuerst die Bedeutung in (61a) erläutert. Dann wird gesagt, dass die Lesarten (61b) und (61c) ebenfalls unter diese Kategorie fallen sollen.

- (61) Unterbedeutungen von *Traverse* $<3D$
- a. Pendant zu *durch* in dessen wegbezogener Lesart, bei dem aber weniger als drei Dimensionen konzeptualisiert werden
- b. „Traversen über Objekte mit nur einer Ausdehnung in nur einer Dimension“ (Kiss et. al., 2020, S. 175)
- c. Via

Beispiele von Kiss et. al. (2020, S. 175) für Sätze, in denen *über* diese Bedeutungen hat, stehen in (62), wobei (62a) (61a) entspricht usw.

- (62) Beispiele für Unterbedeutungen von *Traverse* <3 aus (Kiss et. al., 2020, S. 175)
- a. i. Er geht über die Wiese.

- ii. Sie ziehen über deutsches Gebiet.
- iii. Er geht über die Straße.
- b. Sie gingen über die Grenze.
- c. Sie flogen über Berlin in den Süden.

Insgesamt wird *Traverse* <3 wie folgt charakterisiert:

In dieser Verwendung ist *über* weg-, aber nicht gestaltbezogen. Mittlere Teile des Weges, der durch das LO [das zu lokalisierende Objekt] definiert wird, liegen in einem Suchgebiet, welches innerhalb des ROs [Referenzobjekt] zu lokalisieren ist. Die vertikale Ausdehnung des ROs sowie die vertikale Relation zwischen LO und RO sind zu vernachlässigen, weshalb hier weniger als drei Dimensionen relevant sind. (Kiss et. al., 2020, S. 174)

Ich werde nun dafür argumentieren, dass die drei Lesarten aus (61) paarweise verschieden sind, und mich dabei des Wahr/Falsch-Tests bedienen, welcher schon im letzten Unterabschnitt verwendet wurde. Dabei werde ich davon ausgehen, dass unter (61b) auch Fälle fallen, in denen das Referenzobjekt nicht tatsächlich eindimensional ist, sondern nur als eindimensional konzeptualisiert wird. Alternativ könnte man auch die (61b)-Fälle weiterhin mit den (61a)-Fällen zusammenfassen und eine neue Lesart für Überquerungen einführen.⁷⁹

Um zu zeigen, dass *über* ambig in Bezug auf die Lesarten (61a) und (61b) ist, kann der Wahr/Falsch-Test auf (62a-iii) angewandt werden: Dieser Satz wird von Kiss et. al. (2020) als Beispiel für (61a) geführt, er hat aber auch eine Lesart, in der die Straße überquert wird. Nun ist es naheliegend, zu sagen, dass *über* einfach vage in Bezug auf die Richtung ist, in die das zu lokalisierende Objekt sich bewegt. Aber (62a-iii) kann in derselben Situation wahr sein, weil man über die Straße geht, und falsch sein, weil man sie dabei nicht überquert. Dass dem so ist, wird noch deutlicher, wenn man (63) betrachtet und bemerkt, dass der Satz nicht widersprüchlich sein muss.⁸⁰

(63) Er geht über die Straße, aber er geht nicht über die Straße.

Ein Satz ähnlicher Form soll nun genutzt werden, um (61b) und (61c) voneinander abzugrenzen. Nehmen wir an, eine Gruppe von Radfahrern befindet sich in der Nähe der deutsch-niederländischen Grenze auf deutschem Gebiet. Von ihrem Ausgangspunkt aus gibt es zwei Wege ins (ebenfalls auf deutschem Gebiet liegende) Dorf, von denen einer ein Stück weit an der Grenze entlang führt (aber keineswegs *auf* ihr verläuft). An einer bestimmten Stelle ist es möglich, diesen Weg zu verlassen und über einen abzweigenden Feldweg die Grenze zu überqueren. Die Radfahrer fahren los und nach der Ankunft im Dorf antwortet einer von ihnen auf die Frage eines Dorfbewohners, welchen Weg sie genommen hätten und ob sie in den Niederlanden gewesen seien, mit (64).

(64) Wir fuhrten über die Grenze, aber wir fuhrten nicht über die Grenze.

Das erste *über* hat die Lesart (61c), das zweite die Lesart (61b), und es liegt kein Widerspruch vor, weswegen *über* nicht vage in Bezug auf diese Lesarten sein kann.

⁷⁹Dann allerdings müsste auch das Verhältnis zur Lesart *Traverse Rand-/Grenzbereich* genauer geklärt werden. Wir wollen uns hier darauf beschränken, aufzuzeigen, dass es einen Unterschied gibt, und der Frage, *worin* er genau besteht und wie die einzelnen Bedeutungen genau charakterisiert werden können, nicht weiter nachgehen.

⁸⁰Zu dieser Art der Anwendung des Wahr/Falsch-Tests s. (Zwicky & Sadock, 1975, S. 7). Es mag im konkreten Fall hilfreich sein, das zweite *über* zu betonen.

Für die Notwendigkeit der Unterscheidung zwischen (61a) und (61c) kann auf ähnliche Weise argumentiert werden: Nehmen wir an, eine Gruppe von Jugendlichen begibt sich an die Ruhr, um dort eine Corona-Party zu feiern. Es gibt zwei Wege, über die man zur Ruhr gelangen kann: Einer von ihnen führt an der Polizeiwache vorbei, der andere an der Hundewiese. Zwar kann man die Ruhr auch erreichen, indem man quer über die Hundewiese geht, das ist allerdings nicht notwendig, da der Weg auch um die Wiese herum dorthin führt. Am nächsten Morgen findet sich Hundekot in der Wohnung der Eltern von einer der Jugendlichen. Diese tadeln nicht nur ihr verantwortungsloses Verhalten, sondern machen sie auch für die Verschmutzung der Wohnung verantwortlich. Während sie das kontaktverbotswidrige Treffen mit den Freunden aufrichtig bereut und Besserung gelobt, streitet sie vehement ab, den Hundekot in die Wohnung eingetragen zu haben. Im Verlaufe des Gesprächs äußert sie:

(65) Wir gingen über die Hundewiese zur Ruhr, aber wir gingen nicht über die Hundewiese.

In diesem Kontext kann (66) also gleichzeitig wahr und falsch sein, *über* ist also ambig in Bezug auf die beiden Lesarten.

(66) Sie gingen über die Hundewiese zur Ruhr.

Wir wollen im Folgenden die Lesart in (61a) *Traverse 2D*, die Lesart in (61b) *Traverse 1D* und die Lesart in (61c) *Traverse Via* nennen.

Es muss angemerkt werden, dass die Anwendung des Wahr/Falsch-Tests bei einem derart polysemen Lexem wie *über* problematisch ist, denn es ist nicht immer ersichtlich, welche Bedeutung dafür verantwortlich ist, dass ein Satz sowohl wahr als auch falsch gleichzeitig sein kann: Bei (63) etwa könnte man beim ersten *über* auch die Lesart *Traverse vertikal* annehmen.

Zusammenfassung Der Wahr/Falsch-Test legt nahe, dass *über* in Bezug auf die drei Unterbedeutungen von *Traverse <3D* ambig ist. Auch wenn seine Anwendung nicht unproblematisch ist, wollen wir annehmen, dass die Bedeutungen *Traverse 2D*, *Traverse 1D* und *Traverse Via* zu trennen sind.

6. Anwendungen

In diesem Abschnitt soll es um mögliche Anwendungen der Bedeutungs-Entsprechungs-Abbildungen und des Korpus im Allgemeinen gehen. Diese sind ganz unterschiedlicher Natur: So wird in Unterabschnitt 6.1 die Möglichkeit betrachtet, Entsprechungen zur Präpositionsbedeutungsdisambiguierung einzusetzen, und in Unterabschnitt 6.2 geht es darum, wie das Korpus dabei helfen könnte, die Einträge zu *über* in bilingualen Online-Wörterbüchern für Fremdsprachenlerner hilfreicher zu gestalten. Unterabschnitt 6.3 stellt dann die *Suche über Entsprechungen* vor, eine Methode, bei der anderssprachige Entsprechungen genutzt werden, um in einem bilingualen Korpus nach bestimmten Bedeutungen zu suchen. Nutzungsmöglichkeiten der Annotationen für die deutschen Sätze werden in Unterabschnitt 6.4 dargestellt.

6.1. Präpositionsbedeutungsdisambiguierung

Auf einem multilingualen Korpus wie CoStEP können Bedeutungs-Entsprechungs-Abbildungen zur Disambiguierung von Präpositionsbedeutungen eingesetzt werden: Wenn man etwa die Bedeutung von *über* bestimmen will und weiß, dass die englische Entsprechung *about* ist, kann man sich recht sicher

sein, dass eine Thema- oder eine Bezugspunktlesart vorliegt. Tendenziell gilt dabei: Je mehr Sprachen man nutzen kann, desto einfacher ist die Disambiguierung, denn es wird häufig vorkommen, dass etwa in einer einzelnen Sprache der Satz sehr frei übersetzt ist, sodass gar keine wirkliche Entsprechung vorliegt (4 ist in diesem Korpus der frequenteste Formulierungsnäherwert), oder dass die Entsprechung bei der Disambiguierung nicht hilfreich ist (wie etwa *over*). Betrachtet man mehr Sprachen liegen 1. mehr potenziell hilfreiche Daten vor und kann es 2. auch sein, dass eine Sprache eine Ambiguität auflöst, wo eine andere es nicht tut, oder dass die in zwei Sprachen gewählten Ausdrücke zwar für sich genommen ambig sind, aber nur eine Bedeutung gemeinsam haben.

Das hier vorliegende Korpus und die Bedeutungs-Entsprechungs-Abbildungen können in Disambiguierungssystemen auf verschiedene Weise eingesetzt werden. So wären etwa die für 1, AMV und 4N erstellten Abbildungen mehr oder weniger direkt nutzbar, es können aber auch unterschiedliche Teile des Korpus als Trainingsdaten für Machine-Learning-Systeme verwendet werden.⁸¹

Ein einfach gehaltenes Beispiel demonstriert den Nutzen: Wir trainieren auf den ursprünglichen, noch nicht nach Bedeutungen getrennten Daten, aber unter Ausschluss der Beispiele mit zum Ausschluss führendem *Meta*-Wert insgesamt vier sehr einfache naive Bayes-Klassifikatoren, die sich nur in den verwendeten Daten unterscheiden. Es werden zweimal Daten verwendet, bei denen die in den Abschnitten 5.2 und 5.3 vorgeschlagenen Änderungen am Bedeutungsinventar umgesetzt worden sind,⁸² und zweimal Daten, bei denen das nicht der Fall ist. Die einzige Information, die die Klassifikatoren zur Verfügung haben, ist die englische Entsprechung, wobei jeweils einmal die automatischen Alignierungen und einmal der Entsprechungseintrag (welcher bei Einträgen mit Formulierungsnäherwert 4 aber in der Regel die nicht korrigierte automatische Alignierung ist, s. S. 29) benutzt werden. Da nur wenige Daten vorliegen, führen wir eine zehnfache Kreuzvalidierung durch.⁸³ Angegeben wird im Folgenden jeweils die durchschnittliche Accuracy auf den Testdaten. Als Baseline wollen wir die durchschnittliche Accuracy betrachten, die sich ergibt, wenn bei der Kreuzvalidierung immer die in den Trainingsdaten häufigste Bedeutung gewählt wird. Das sind bei den Daten mit den vorgeschlagenen Änderungen 61,02 %, bei denen ohne 26,76 %. Der große Unterschied erklärt sich daraus, dass die drei häufigsten Bedeutungsannotationen in den unveränderten Daten *Bezugspunkt*, *Thema* und die kombinierte Annotation dieser beiden sind (vgl. Tabelle 7 auf S. 29), welche in den Daten mit den Änderungen zu einer

⁸¹ Als Trainingsdaten würde man die nicht nach Bedeutungen getrennten Daten verwenden, da man auch im Anwendungsfall möglichst alle im Kontext möglichen Bedeutungen einer Präposition kennen möchte und nicht nur eine Bedeutung pro Vorkommen.

⁸² Nennen wir den entsprechenden Datensatz 1B. Eine ausführliche Beschreibung von 1B kann hier nicht erfolgen. Es sei aber darauf hingewiesen, dass es der Aufhebung der Unterscheidung zwischen *Bezugspunkt* und *Thema* wegen nur noch vier Fälle gibt, in denen aufgrund einer Mehrfachannotation von Bedeutungen Beispiele mehrfach vertreten sind (vgl. zu dieser Problematik Abschnitt 3.2). Ansonsten unterscheidet sich 1B von 1 nur in den Annotationen für die Beispiele, die vorher den Lesarten *Spatial Traverse <3D* und *Spatial Traverse <3D +metaph* zugeordnet wurden. Diese verteilen sich nun wie folgt:

Spatial Traverse...	1D	1D +metaph	2D	Via	Via +metaph
	2	1	2	5	1

Im Anhang findet sich auf S. 87 eine Tabelle mit den nach Bedeutung geordneten Abbildungen (Tabelle 39), auf S. 88 eine Tabelle mit den nach Entsprechung geordneten Abbildungen (Tabelle 40) und auf S. 92 ein entsprechendes Sankey-Diagramm (Abbildung 7).

⁸³ Der Klassifikator wurde in R (R Core Team, 2020) mit der Funktion `naive_bayes` aus dem Paket `naivebayes` (Majka, 2019) implementiert. Dabei wurde Laplace-Smoothing mit dem Wert 0,1 angewandt (der Wert wurde so gering gewählt, weil viele Bedeutungs-Entsprechungs-Paare nur sehr selten vorkommen). Für die Aufteilung der Beispiele auf die einzelnen Gruppen bei der Kreuzvalidierung wurde eine Funktion aus dem Paket `caret` (Kuhn, 2020) genutzt. Da in dieser die Zuordnung zu den Gruppen randomisiert erfolgt, können die Accuracy-Ergebnisse geringfügig schwanken. Zur Gewährleistung der Reproduzierbarkeit wurde ein *random seed* gesetzt.

Kategorie zusammengefasst wurden. Der naive Bayes-Klassifikator erreicht auf den veränderten Daten eine Accuracy von 71,71 % mit den automatischen Alignierungen und von 74,65 % mit den Entsprechungen. Auf den unveränderten Daten sind es 50,73 % und 51,14 %. Diese Werte sollten ausschließlich in Relation zu den auf denselben Datensätzen ermittelten Baselines interpretiert werden, denn die Gegebenheiten im Testdatensatz unterscheiden sich wegen der Art der Beispielauswahl und aufgrund des Ausschlusses von Daten mit bestimmten *Meta*-Werten stark von denen im gesamten CoStEP-Korpus. Wir können aber erkennen, dass allein die Verwendung der englischen Entsprechungen zu Disambiguierungen führt, die erheblich besser als die Baseline sind, und daraus schließen, dass bereits Informationen zu Übersetzungen in nur einer weitere Sprache bei der Disambiguierung von Präpositionsbedeutungen sehr hilfreich sein können.⁸⁴ Wir sehen auch, dass die Ergebnisse auf Basis der korrigierten Entsprechungen besser sind als die auf den automatischen Alignierungen basierenden, dass der Unterschied aber nicht sehr groß ist.

Als Testdaten für Disambiguierungssysteme sind die Annotationen zwar grundsätzlich verwendbar, es muss allerdings beachtet werden, dass sich aufgrund der Art der Beispielauswahl die Verteilung der Entsprechungen und Bedeutungen von derjenigen des Europarl-Korpus höchstwahrscheinlich recht stark unterscheidet, weswegen die Ergebnisse nicht unbedingt einen Rückschluss darauf erlauben würden, wie gut das getestete System im tatsächlichen Anwendungsfall wäre.

Ein großes Hindernis für die Verwendung von Informationen über Beziehungen zwischen Ausdrücken in verschiedenen Sprachen ist, dass im realen Anwendungsfall meist keine Übersetzung vorliegen wird. Das heißt aber erstens nicht, dass es nicht sinnvoll sein kann, sie zu nutzen, wo sie vorliegen, und zweitens zeigen Gonen und Goldberg (2016), dass Informationen über sprachübergreifende Zusammenhänge bei der Präpositionsbedeutungsdisambiguierung verwendet werden können, auch wenn nur für eine Sprache Daten vorliegen: Sie benutzen in ihrem mit künstlichen neuronalen Netzen arbeitenden Ansatz einen Kontextcodierer, der aus zwei *long short-term memory networks* besteht, nämlich einem für den Kontext vor der Präposition, welches *word embeddings* der Wörter vom Satzanfang bis zum letzten Wort vor der Präposition durchgeht, und einem die *word embeddings* von hinten bis zum Wort hinter der Präposition durchgehenden. Dieser Kontextcodierer wird in verschiedene neuronale Netze (eines pro Sprache) eingebaut, die lernen, auf Basis des Kontexts die Präposition *in einer anderen Sprache* vorherzusagen, wobei die dafür erforderlichen Daten auf Basis von automatischen Wortalignierungen erstellt werden. Dann wird der dergestalt trainierte Kontextcodierer selbst in dem Netz verwendet, das die Bedeutungen vorhersagt. Der Sinn dahinter ist, dass mit den leicht zu erstellenden und darum massenhaft vorhandenen automatisch alignierten Daten Repräsentationen gelernt werden können, die für ein auf handannotierten (und darum knappen) Daten trainiertes Disambiguierungssystem nützlich sind. Mit dem hier vorgestellten Korpus liegen nun handannotierte sprachübergreifende Daten für ein Sprachpaar vor. Es wäre also z. B. möglich, auf dem Korpus (um genau zu sein: auf I, AMV und 4N) Repräsentationen für Beziehungen zwischen Entsprechungen und Abbildungen zu lernen, die dann Teil eines

⁸⁴Allein darum geht es uns hier. Es soll kein besonders leistungsfähiges System entwickelt werden und ein Vergleich mit anderen Präpositionsbedeutungsdisambiguierungssystemen ist auch gar nicht möglich, weil er nur sinnvoll wäre, wenn Resultate für die gleiche Präposition und mit demselben Bedeutungsinventar verglichen werden. Es sei aber darauf hingewiesen, dass in den Testdaten auch Sätze mit dem Formulierungsnäherwert 4 enthalten waren, bei denen gar keine brauchbare Entsprechung vorliegt, und dass Mehrfachannotationen von Bedeutungen zusammen als eine Kategorie behandelt wurden, was es – bis auf die Mehrfachannotation von *Bezugspunkt* und *Thema*, für die viele Daten vorlagen – für das System fast unmöglich macht, die entsprechenden Beispiele richtig zu kategorisieren. Auch waren Unterbedeutungen zu unterscheiden, was eine erheblich schwerere Aufgabe als die Unterscheidung einander recht unähnlicher Oberbedeutungen ist.

Disambiguierungssystemen sind, während die englische Entsprechung ein Merkmal ist, das mittels eines Kontextkodierers automatisch ermittelt wird. Es wäre zu prüfen, ob die Menge der vorhandenen Daten reicht, um einen solchen direkt zu trainieren. Falls das nicht der Fall sein sollte, könnte man ihn erst mit automatisch alignierten Daten trainieren und dann versuchen, die so gelernten Repräsentationen mithilfe der vorhandenen Daten zu automatischen Alignierungen und korrigierten Entsprechungen (und den dazugehörigen englischen Sätzen) zu verbessern. Es ist zwar vorderhand nicht ersichtlich, warum es einfacher sein sollte, eine Entsprechung einer Präposition in einer bestimmten Sprache korrekt vorherzusagen, als ihre Bedeutung zu bestimmen, und im Allgemeinen ist dies wahrscheinlich auch nicht der Fall, da sich die Bedeutungsdisambiguierung auch als Übersetzung der Präposition in eine (künstliche) Metasprache betrachten lässt, entscheidend ist aber, dass es für die Bedeutungsdisambiguierung sehr hilfreich ist, *wenn* es im konkreten Fall gelingt, die anderssprachige⁸⁵ Entsprechung zu bestimmen. Das vorliegende Korpus kann zum Training von Komponenten von Präpositionsbedeutungsdisambiguierungssystemen verwendet werden, die diese Zusammenhänge nutzen.

Zusammenfassung Sprachübergreifende Daten können bei der automatischen Präpositionsbedeutungsdisambiguierung von Nutzen sein und die hier vorliegenden Daten lassen sich zum Training entsprechender Systeme verwenden. Dass im realen Anwendungsfall nur sehr selten Übersetzungen vorliegen dürften, ist zwar ein Problem, bedeutet aber – wie Gonen und Goldberg (2016) demonstrieren – nicht, dass im Sprachvergleich gewonnene Informationen nicht nutzbringend eingesetzt werden können.

6.2. Nutzen der Abbildungen für Lerner

Beim Erlernen des Englischen sind die Präpositionen eine besonders große Herausforderung (Jarvis & Odlin, 2000, S. 551). Deswegen ist es unglücklich, dass – wie in diesem Abschnitt gezeigt werden soll – deutsch-englische Online-Wörterbücher, welche von vielen Lernern bei Unsicherheiten in Bezug auf die Wahl der Präposition als Informationsquelle genutzt werden dürften, in ihren Einträgen zu *über* einige Unvollkommenheiten aufweisen. Es soll im Folgenden gezeigt werden, dass die hier erstellten Bedeutungs-Entsprechungs-Abbildung genutzt werden können, um diese Einträge⁸⁶ zu verbessern, auch wenn dies nur innerhalb gewisser Grenzen möglich ist und einige der Probleme, die insbesondere für fortgeschrittenere Lerner relevant sind, nicht gelöst werden können. Der Wörterbucheintrag, der nun beispielhaft betrachtet werden soll, ist der Eintrag zu *über* bei PONS (o. D. a), jener bei Langenscheidt (o. D.) ist aber sehr ähnlich gestaltet.

Der PONS (o. D. a) listet für *über* als Präposition zwanzig verschiedene Bedeutungen, die fortlaufend nummeriert sind.⁸⁷ Auch wenn eine weitere Gliederung der Einträge nicht kenntlich gemacht wird, ist doch zu erkennen, dass die spatialen Bedeutungen allesamt am Anfang des Eintrags stehen und dort die Bedeutungen, bei denen *über* den Dativ regiert, denjenigen, bei denen es den Akkusativ regiert,

⁸⁵Unter Umständen könnten auch gleichsprachige Entsprechungen (also Synonyme) hilfreich sein: Ist bekannt, dass *über* durch *via* ersetzbar ist, ist davon auszugehen, dass es die Lesart *Traverse Via* oder eine modale Lesart (*Sie telefonieren via Internet*) hat.

⁸⁶Sicherlich könnten auch gedruckte Wörterbücher einen Nutzen aus den Abbildungen ziehen, ich werde mich hier aber auf Online-Wörterbücher beschränken, da sie wesentlich weniger Restriktionen etwa in Bezug auf die Länge der einzelnen Einträge unterliegen und davon auszugehen ist, dass die meisten Wörterbuchabfragen heutzutage ohnehin online (vor allem über Smartphones) stattfinden.

⁸⁷Alle Angaben in diesem Abschnitt beziehen sich auf (PONS, o. D. b).

Abbildung 5: Bildschirmfoto eines Bedeutungseintrags aus (PONS, o. D. b)



vorangehen. Ein Bildschirmfoto des Eintrags für eine dieser Bedeutungen findet sich in Abbildung 5. In den einzelnen Bedeutungseinträgen steht in der ersten Zeile zuerst die Nummer des Eintrags, dann folgt *über* und der regierte Kasus. Vor einem Doppelpunkt steht dann in Klammern eine knappe Bedeutungsangabe. In der folgenden Zeilen steht die Übersetzung (nur in einem Fall ist mehr als eine Übersetzung angegeben), danach kommen bis zu drei Beispiele, wobei in ihnen auch andere Übersetzungsmöglichkeiten vorkommen können als die in der ersten Zeile angegebene. Bei den angegebenen Übersetzungsmöglichkeiten handelt es sich ausschließlich um primäre Präpositionen.⁸⁸ Ein Click auf „10 Beispiele aus dem Internet“ führt zu einer Seite, auf der zuerst einige von der Redaktion stammende beispielhafte Übersetzungen aufgeführt sind, die sich nicht auf den speziellen Bedeutungseintrag beziehen. Bei keinem dieser Beispiele handelt es sich um einen Satz, in neun von fünfzehn Fällen wird eine Übersetzungsmöglichkeit für ein *über* regierendes Verb angegeben. Die Beispiele aus dem Internet wurden – worauf auch hingewiesen wird – nicht von der Redaktion begutachtet. Bei der Bedeckungslesart, in welcher die angegebene Übersetzung *over* ist, ist zwar sowohl *über* als auch *over* im jeweiligen Satz hervorgehoben, allerdings liegt nicht in einem einzigen der Sätze tatsächlich eine Bedeckungslesart vor (in vielen handelt es sich bei *über* gar um ein Adverb), was die Beispiele für Lerner nutzlos macht.⁸⁹

Die hier erstellten Abbildungen zwischen englischen Entsprechungen und Bedeutungen im Deutschen können genutzt werden, um die Einträge für *über* in Online-Wörterbüchern in vielerlei Hinsicht zu verbessern. Das beginnt schon bei der Gliederung der deutschen Bedeutungen: Im Annotationsschema von Kiss et. al. (2020) gibt es Ober- und Unterbedeutungen und auch bei den Unterbedeutungen stehen ähnliche Lesarten (wie etwa die verschiedenen Traversenlesarten) hintereinander. Der Eintrag für *über* wäre wesentlich übersichtlicher, wenn man nicht alle Unterbedeutungen der Reihe nach durchgehen müsste, sondern etwa die Einträge für spatiale Bedeutungen überspringen könnte, wenn man nach einer Übersetzung für *über* im instrumentalen Sinne sucht. Es könnte auf die Beschreibungen der Lesarten bei Kiss et. al. (2020) zurückgegriffen werden, welche man auf Wunsch ein- und ausblendbar sein lassen könnte (allerdings wären wohl einige Reformulierungen vonnöten, um sie allgemeinverständlich zu halten). Der Hauptvorteil der Verwendung der Abbildungen wäre jedoch, dass man wesentlich mehr Übersetzungsmöglichkeiten pro Lesart angeben könnte und für diese auch realistische Verwendungsbeispiele aus dem Korpus zur Verfügung ständen, bei denen sowohl Bedeutung als auch Entsprechung passen. Bei diesen Übersetzungsmöglichkeiten würde es sich nicht ausschließlich um primäre Präpositionen handeln, was wünschenswert ist, denn wenn sich jemand z. B. dafür interessiert, wie *über* übersetzt werden kann, wenn es um ein abstraktes Instrument geht, ist für ihn die

⁸⁸Bei Langenscheidt (o. D.) werden dagegen auch zuweilen Übersetzungsmöglichkeiten wie *on top of* angegeben, bei denen es sich nicht um primäre Präpositionen handelt, und es werden auch öfter mehrere Übersetzungsmöglichkeiten pro Lesart genannt.

⁸⁹Es steht zu vermuten, dass PONS (o. D. a) einfach zehn Beispiele, in denen das deutsche Wort die für die betrachtete Lesart angegebene Entsprechung hat, aus einem wortalignierten Korpus mit Texten aus dem Internet nimmt. Dieses Vorgehen ist für Präpositionen unangebracht, da es zu den genannten Problemen führt.

syntaktische Kategorie der Übersetzungsmöglichkeit nebensächlich – *by means of* ist genauso gut wie *via* oder *through*. Auch auf den Abbildungen für AMV basierende Listen mit Verben, die im Englischen ein NP-Komplement anstelle einer (*über*-)PP haben, könnten bei den relevanten Bedeutungseinträgen angegeben werden.

Ein Manko bei der Angabe mehrerer Übersetzungsmöglichkeiten ist, dass diese keineswegs gleichbedeutend sein müssen und Bedeutungs- oder Distributionsunterschiede zwischen den einzelnen Übersetzungsmöglichkeiten vernachlässigt zu werden drohen: So wird etwa *debate on* anders verwendet als *debate over* (Herbst et. al., 2004, S. 206), bei beiden wird man aber sagen wollen, dass mit der PP ein Thema angegeben wird. Rankin und Schiftner (2011) stellen fest, dass Anglistikstudenten mit deutscher Muttersprache an der Universität Wien bei Thema-Argumenten oftmals eine unangemessene (sekundäre) Präposition verwenden bzw. wesentlich häufiger von bestimmten sekundären Präpositionen Gebrauch machen als englische Muttersprachler. Sie zeigen außerdem, dass *concerning* nicht für alle Thema-Argumente verwendet werden kann, sondern auf bestimmte semantische Bereiche eingeschränkt ist. Es sei ein häufiger Fehler deutschsprachiger Englischler, diese Einschränkung zu missachten. Wenn viele Übersetzungsmöglichkeiten angegeben werden, sollte also auch darüber informiert werden, wann welche der verschiedenen Möglichkeiten angebracht sind. Darüber aber geben die Bedeutungs-Entsprechungs-Abbildungen keine Auskunft, sodass eine Sichtung der einschlägigen Literatur und ggf. weitere Forschungen notwendig wären.⁹⁰

Es sei an dieser Stelle ausdrücklich erwähnt, dass es bei der Nutzung von Abbildungen wie den hier erstellten nicht um eine Ersetzung, sondern nur um eine Ergänzung und Überarbeitung bestehender Wörterbucheinträge gehen kann. Das ist besonders interessant für Präpositionen wie *über*, die hochgradig polysem sind und viele Übersetzungsmöglichkeiten haben. Manche Bedeutungen von *über* sind so selten, dass sie in einem Korpus mit einer Größe, die eine Annotation mit angemessenem Aufwand zulässt, kaum oder gar nicht vorhanden sein werden, weswegen nicht damit zu rechnen ist, dass sich für alle Bedeutungen Entsprechungen und Beispiele finden lassen.

Zusammenfassung Das Korpus und die Bedeutungs-Entsprechungs-Abbildungen können dabei helfen, die Einträge zu *über* in Online-Wörterbüchern zu verbessern, indem sie Übersetzungsmöglichkeiten aufzeigen, die bis jetzt nicht aufgeführt werden, und authentische Verwendungsbeispiele für die verschiedenen Übersetzungsmöglichkeiten in den verschiedenen Bedeutungen zur Verfügung stellen. Dabei sollten sie aber durch Informationen zur Verwendungsweise der Übersetzungsmöglichkeiten ergänzt werden, die nicht unmittelbar aus ihnen selbst abgeleitet werden können.

6.3. Die Methode der Suche über Entsprechungen

Korpora wie das hier dargestellte können verwendet werden, um zu untersuchen, wie eine bestimmte Bedeutung in einer Sprache ausgedrückt werden kann. Ich möchte die entsprechende Methode *Suche über Entsprechungen* nennen und sie im Folgenden anhand eines Beispiels demonstrieren.

⁹⁰Die Unterschiede in der Distribution sekundärer Präpositionen scheinen für das Englische noch nicht gut erforscht zu sein (vgl. Rankin & Schiftner, 2011, S. 415). In manchen Fällen könnte es für ein Wörterbuch allerdings ausreichen, wenn verschiedene Beispiele verfügbar sind, in denen die Nutzer Verwendungsregularitäten selbst erkennen können, sodass keine Explizitmachung notwendig ist.

Nehmen wir an, wir interessieren uns dafür, wie ein *Thema/Bezugspunkt* (im Folgenden nur *Thema* genannt) im Englischen ausgedrückt werden kann und wie sich die Anwendungsbedingungen der einzelnen sprachlichen Mittel unterscheiden. Dann stehen wir vor dem Problem, dass eine Suche nach Bedeutungen in unannotierten Korpora nicht möglich ist. Nach Ausdrücken wie *about* zu suchen, von denen wir annehmen, dass sie dazu dienen können, Themen auszudrücken, ist keine Option, denn auf diese Weise kann man nur finden, was man ohnehin schon kennt. Was wir benutzen könnten, wäre ein Korpus mit sehr umfangreichen semantischen Annotationen, in dem eine Suche nach Themen möglich ist, allerdings ist die Erstellung solcher Korpora sehr aufwendig, weswegen man wohl auf bereits bestehende Ressourcen wie etwa FrameNet (Ruppenhofer et. al., 2016) zurückgreifen würde. Zwar ist die Gesamtgröße des FrameNet beeindruckend, aber die Größe des Datensatzes für einen einzelnen Frame (wie z. B. *Topic*) ist dann doch relativ gering (für *Topic* sind derzeit sechzehn lexikalische Einheiten gelistet (International Computer Science Institute, Berkeley, o. D.)), sodass man es wohl allenfalls als Ausgangspunkt für eine weitere Untersuchung benutzen wollen wird. Wir könnten auch englische Texte durchgehen und nur nach Thema-Bedeutungen suchen, aber auch das ist sehr aufwendig, da in den meisten der betrachteten Sätze überhaupt keine Themasart vorliegen wird. Eine Alternative dazu wäre, sich des Umwegs über eine zweite Sprache zu bedienen: Da wir wissen, dass im Deutschen ein Thema mit *über* ausgedrückt werden kann, können wir, wenn wir über ein bilinguales Korpus verfügen, im Deutschen nach Vorkommen von *über* suchen und uns bei den Fällen mit Themasart dann ansehen, ob und ggf. wie diese im entsprechenden englischen Satz ausgedrückt wird.⁹¹ Sicherlich werden auch bei diesem Vorgehen sehr viele uninteressante Sätze betrachtet werden müssen, aber es sind erheblich weniger als bei der anderen Methode, sodass der Suchaufwand bedeutend geringer ist. Diesem Vorteil stehen bei der Suche über Entsprechungen aber auch Nachteile gegenüber: Frequenzen können nur mit äußerster Vorsicht interpretiert werden, denn in einem bilingualen Korpus werden die Texte in den beiden Sprachen entweder direkte Übersetzungen voneinander sein oder auf ein Dokument in einer dritten Sprache zurückgehen und die Übersetzer könnten bei der Übersetzung durch die Syntax der Ursprungssprache beeinflusst sein, sodass etwa die Übersetzungen von *über* vielleicht häufiger als im allgemeinen Sprachgebrauch üblich Präpositionen sind.⁹² Auch besteht die Gefahr, dass die Form, nach der man in der anderen Sprache sucht, nicht genau die Bedeutung hat, für die man sich interessiert und so nicht alle interessanten Fälle gefunden werden können.

Für diesen Beispielfall ist es mit den Abbildungen aus Abschnitt 4 schon möglich, erste Aussagen zu treffen und die in Tabelle 34 auf S. 79 im Anhang aufgeführten Entsprechungen von *über* als Möglichkeiten, im Englischen ein Thema auszudrücken, aufzuführen. Aus dieser Tabelle wird auch ersichtlich, dass *of* als Kopf einer PP, die ein Thema ausdrückt, im Datensatz hier immer regiert ist. Ob das grundsätzlich gilt, wäre auf einer erweiterten Datenbasis zu prüfen.⁹³ Außerdem können die Verben aus Tabelle 27 auf S. 46 angegeben werden, die eine NP als Thema-Argument haben.

⁹¹Im konkreten Fall könnten wir natürlich die hier vorliegenden Annotationen nutzen und uns alle englischen Sätze ausgeben lassen, bei denen im Deutschen eine Themasart annotiert wurde.

⁹²Dem könnte begegnet werden, indem man zuerst mittels Suche über Entsprechungen die interessanten sprachlichen Einheiten findet und dann in einem einsprachigen Korpus nach ihnen sucht.

⁹³Die Methode der Suche über Entsprechungen kann nicht nur eingesetzt werden, wenn man sich für die Möglichkeiten, in einer Sprache eine bestimmte Bedeutung auszudrücken, im Allgemeinen interessiert, sondern auch, wenn man ein bestimmtes lexikalisches Element untersuchen will: Um unter den vielen Tokens von *of* in einem Korpus diejenigen mit Themasart zu finden, können Alignierungen mit *über* helfen (vgl. auch die Diskussion zu Disambiguierungssystemen in Abschnitt 6.4).

Zusammenfassung Entsprechungen in anderen Sprachen können dazu genutzt werden, in einem Korpus nach Bedeutungen zu suchen.

6.4. Nutzungsmöglichkeiten der deutschen Daten

Mit DE und DE-DIS liegen zwei Subkorpora mit Bedeutungsannotationen für *über* in deutschen Sätzen vor, die genutzt werden können, um Aussagen zu überprüfen, die in der theoretischen Literatur zu PPen im Deutschen getroffen werden. So behauptet etwa Breindl (2006, S. 938 f.), PPen könnten nur in temporalen und lokalen Lesarten modifiziert werden. Im Europarl-Korpus findet sich aber (67)⁹⁴:

(67) 76082

(de) So wenden einige Mitgliedstaaten wie beispielsweise die skandinavischen Länder der Volksgesundheit wegen Steuersätze an , die weit über dem Mindestsatz liegen .

Hier hat *über* eine Überschreitungslesart. Man könnte nun argumentieren, dass es sich bei dieser nur um eine metaphorisch verwendete spatiale Lesart handelt, aber dann stellte sich die Frage, wie groß die Aussagekraft der ursprünglichen Behauptung eigentlich ist, da bei vielen Lesarten von *über* Reste einer spatialen Metaphorik ausgemacht werden können (und für andere Präpositionen ähnliches gilt). Entscheidend ist hier nur, dass der Satz *prima facie* ein Problem darstellt und damit demonstriert, dass das Korpus genutzt werden kann, um Behauptungen wie die von Breindl zu prüfen.

Die deutschen Daten können auch unabhängig von den in Abschnitt 6.1 dargestellten Überlegungen als Trainingsdaten für Systeme zur Bedeutungsklassifikation genutzt werden.

7. Zusammenfassung und Ausblick

7.1. Zusammenfassung

In dieser Arbeit wurde ein Korpus mit Annotationen deutsch-englischer Satzpaare vorgestellt, das dazu genutzt werden kann, zu untersuchen, wie *über* ins Englische übersetzt wird. Auf Basis der Annotationen wurden Abbildungen zwischen Bedeutungen von *über* und seinen englischen Entsprechungen erstellt und es wurden einige mögliche Anwendungen dargestellt.

In Abschnitt 2 wurde zuerst die Erstellung der Annotationsdateien auf Basis des CoStEP-Korpus (Graën et. al., 2014) dargestellt. Aus diesem wurden deutsch-englische Satzpaare, bei denen der deutsche Satz *über* enthält, extrahiert und wortaligniert. Dann wurde das Annotationsschema präsentiert: Die automatisch alignierten Wörter und Wortgruppen wurden manuell zu Entsprechungen korrigiert und es wurden die Bedeutungen von *über* annotiert. Für die englische Entsprechung wurde annotiert, ob die Bedeutung angemessen ist. Für beide Sprachen wurde der Rektionsstatus bestimmt und es wurde die Formulierungsnähe festgestellt. Des Weiteren wurden Metaannotationen vorgenommen, die unter Umständen zum Ausschluss bestimmter Beispiele führen konnten. Bei der Durchführung der Annotation hat sich die Unterscheidung der Lesarten *Bezugspunkt* und *Thema* als besonders schwierig erwiesen. In Abschnitt 5.2 wurde die Beziehung zwischen diesen Lesarten dann unter Verwendung von Tests untersucht, mit denen Ambiguität von Unbestimmtheit unterschieden werden kann. Es hat sich keine

⁹⁴Dieser Satz stammt nicht aus dem annotierten Teil der Daten, in welchem sich allerdings ein ähnlicher Satz findet, der sich jedoch nicht gut als Beispiel eignet, da der vermeintliche Modifikator nicht eindeutig der PP zuzuordnen ist bzw. eine solche Zuordnung eine ausführliche Diskussion erfordern würde.

Evidenz dafür gezeigt, dass *über* in Bezug auf die Unterscheidung von *Thema* und *Bezugspunkt* ambig ist. Grundsätzlich hat sich das Bedeutungsinventar von Kiss et. al. (2020) aber gut für die Annotation von Parlamentsreden verwenden lassen.

In Abschnitt 3 wurden die Daten auf verschiedene Datensätze aufgeteilt und auf Basis der relevanten Datensätze wurden dann in Abschnitt 4 Abbildungen erstellt: Bei Betrachtung der Daten mit echter Entsprechung zeigt sich, dass sich die meisten englischen Entsprechungen nur sehr wenigen Lesarten von *über* zuordnen lassen. Eine Ausnahme bildet *over*, das in sehr vielen Lesarten eine mögliche Entsprechung ist. Auch bei den AMV- und 4n-Fällen lassen sich bestimmten englischen Ausdrücken eindeutig (Ober-)Bedeutungen zuordnen. Eine Betrachtung der Rektionsstatus im Datensatz I zeigt, dass ein starker, statistisch signifikanter Zusammenhang zwischen dem Rektionsstatus von *über* und dem seiner Entsprechung besteht, wenn es sich bei dieser um eine primäre Präposition handelt, und dass sich die unterschiedlichen Bedeutungen und Entsprechungen in Hinblick auf die Rektionsstatus unterscheiden. Auch konnte gezeigt werden, dass es einen signifikanten Zusammenhang zwischen der gewählten Entsprechung und der Formulierungsnähe gibt.

Veranlasst durch Beobachtungen bei der Annotation und Überlegungen zu Analysemöglichkeiten wurden in Abschnitt 5 einige Vorschläge für Änderungen am Bedeutungsinventar für *über* nach Kiss et. al. (2020) gemacht. Die Aufhebung der Unterscheidung zwischen *Bezugspunkt* und *Thema* wurde bereits oben angesprochen; anders ist es bei der Traversenlesart: Die drei verschiedene Arten von Traversen mit weniger als drei relevanten Dimensionen sollten unterschieden werden. Außerdem wurde vorgeschlagen, die Emotionsgegenstandslesart von der kausalen zu trennen und eigenständig aufzuführen. Im Zuge der Diskussion dort konnte eine Untersuchung der bei psychologischen Prädikaten verwendeten Präpositionen als Forschungsdesiderat ausfindig gemacht werden.

In Abschnitt 6.1 wurde gezeigt, dass sprachübergreifende Daten bei der automatischen Präpositionsbedeutungsdisambiguierung von Nutzen sein können. Die hier vorliegenden Daten lassen sich zum Trainieren entsprechender Systeme nutzen. Eine weitere Anwendung können die hier vorgestellten Annotationen bei der Erstellung und Bearbeitung von Wörterbüchern finden: Die Bedeutungs-Entsprechungs-Abbildungen können dabei helfen, die Einträge für *über* in bilingualen Wörterbüchern zu verbessern, welche – wie wir in Abschnitt 6.2 gesehen haben – Verbesserungspotenzial haben. Sie sollten dabei aber um weitere Informationen über die Verwendungsweise der Übersetzungsmöglichkeiten ergänzt werden. Weiterhin können Entsprechungen in anderen Sprachen dazu genutzt werden, in einem Korpus nach Bedeutungen zu suchen (Abschnitt 6.3). Die Annotationen für das Deutsche können genutzt werden, um in der Literatur zu findende Behauptungen über Präpositionen zu testen.

7.2. Ausblick

Es sollen nun einige mögliche Erweiterungen angesprochen werden. Naheliegender ist dabei die Durchführung ähnlicher Annotationsprojekte für andere Präpositionen. In den folgenden drei Unterabschnitten soll aber dafür argumentiert werden, dass auch eine Ausweitung der Annotationen für *über* in verschiedene Richtungen erstrebenswert ist.

7.2.1. Mehr Daten

Zwar wurden 632 deutsch-englische Satzpaare mit Annotationen versehen, aber aus den in den Abschnitten 2 und 3 dargestellten Gründen liegen für viele Bedeutungen von *über* keine oder nur sehr we-

nige Daten vor. Das betrifft in unterschiedlichem Maße eigentlich fast alle Lesarten außer *Thema* und *Bezugspunkt* und führt dazu, dass pro Bedeutung nur recht wenige englische Entsprechungen gefunden werden konnten. Um bei weiteren Annotationen zu verhindern, dass erneut vorwiegend Beispiele mit Thema- oder Bezugspunktlesart in den Daten vorkommen, könnte man darauf verzichten, Beispiele zu annotieren, in denen *über* automatisch mit einschlägigen Entsprechungen aligniert wurde.

7.2.2. Mehr Sprachen

Eine Ausweitung der Annotationen auf andere Sprachen als Englisch wäre mit geringem Aufwand möglich, da die bereits annotierten Daten für das Deutsche genutzt werden könnten. Es müssten dann die Übersetzungen der deutschen Sätze aus dem CoStEP-Korpus extrahiert und von einer Person annotiert werden, die der jeweiligen Sprache mächtig ist.⁹⁵

Besonders förderlich wäre eine Erhöhung der Anzahl der Sprachen für die Präpositionsbedeutungsdisambiguierung, da sich die Ergebnisse für die einzelnen Sprachen kombinieren lassen,⁹⁶ eine Ausweitung auf andere Sprachen ist auch für andere Anwendungen aus Abschnitt 6 interessant. In Bezug auf die Verbesserung von Online-Wörterbüchern etwa ist der Bedarf bei anderen Sprachen unter Umständen sogar noch größer als beim Englischen: So ist z. B. in der deutsch-griechischen Onlineversion des PONS die Paraphrase zur Bedeutung von *über*, in der es mit *ja* (*για*) übersetzt wird, „betreffend“ (PONS, o. D. c), während in der griechisch-deutschen Version *über* gar nicht als Übersetzungsmöglichkeit von *ja* gelistet ist (PONS, o. D. d). *Ja* wird aber im Griechischen nicht nur bei *Themen* und *Bezugspunkten* verwendet (wie die Paraphrase im PONS und das einzige Beispiel, das dort gegeben wird, nahelegen), es ist auch die typische Präposition für *Emotionsgegenstände* (Alexiadou & Iordăchioaia, 2014, S. 64).

Man könnte meinen, dass man die Abbildungen zwischen Bedeutungen von *über* und Entsprechungen in anderen Sprachen auch nutzen könnte, um Abbildungen zwischen den Entsprechungen der einzelnen Sprachen zu erstellen, also etwa zu sagen, dass eine Entsprechung von *über* in der achsenbezogenen Lesart im Portugiesischen in eine Entsprechung von *über* in der achsenbezogenen Lesart im Dänischen übersetzt werden können sollte, was offensichtliche Vorteile für die Erstellung bilingualer Wörterbücher mit sich brächte (man denke an die Anzahl der möglichen Sprachpaare); das ist allerdings nicht möglich, wie sich schon innerhalb des Englischen zeigt, wo – wie in Abschnitt 6.2 dargestellt – die einzelnen Entsprechungen in der Themalesart nicht in jeglicher Hinsicht äquivalent sind. Trotzdem könnte es eine Möglichkeit sein, mit geringem Aufwand Entsprechungskandidaten zu finden, die dann redaktionell geprüft werden könnten.

Des Weiteren könnten Annotationen für eine größere Anzahl an Sprachen dazu genutzt werden, das Bedeutungsinventar für *über* zu verfeinern, wie das folgende, hypothetische Beispiel zeigt: Wir betrachten die Annotationen für eine bestimmte Bedeutung von *über*. Nehmen wir an, es gibt eine Sprache S,

⁹⁵Da bei der Beispielauswahl auch ältere Texte aus dem Korpus aufgenommen wurden (die aus einer Zeit stammen, als Reden im EU-Parlament noch nicht in die heutigen Amtssprachen der EU übersetzt wurden), liegen nicht für alle Amtssprachen der EU Daten vor. In folgenden Sprachen sollten aber für fast alle Sätze Übersetzungen vorhanden sein: Dänisch, Griechisch, Englisch, Spanisch, Französisch, Finnisch, Italienisch, Niederländisch, Portugiesisch und Schwedisch (vgl. Koehn, 2005, S. 81). Es wäre dann ggf. bei der Interpretation der Ergebnisse zu berücksichtigen, dass sich die Beispielauswahl nach den automatischen Alignierungen fürs Englische gerichtet hat (vgl. Abschnitt 2.1.4).

⁹⁶So könnte es etwa sein, dass eine englische Übersetzung die möglichen Bedeutungen auf modale Lesarten und Traversenlesarten einschränkt und dann z. B. eine spanische Entsprechung mit der Traversenlesart unvereinbar ist, weswegen die modale vorliegen muss. Auch ist so die Chance höher, dass Mehrfachannotationen vorgenommen werden können, wo dies angebracht ist, da es sein kann, dass sich die Übersetzer für verschiedene Sprachen für verschiedene Bedeutungen entscheiden, wenn es nicht oder nur schwer möglich ist, die Ambiguität in ihrer Sprache zu bewahren.

in der in den relevanten Sätzen die Entsprechungen s_1 , s_2 und s_3 genutzt werden, und eine Sprache T , in der man dort die Entsprechungen t_1 und t_2 verwendet, wobei allerdings t_1 immer dort benutzt wird, wo in S s_1 oder s_2 steht, und t_2 immer dort, wo s_3 steht. Das legt nahe, dass es zwei konzeptuell unterscheidbare Bedeutungen gibt, zu deren Ausdruck in S und T auch unterschiedliche sprachliche Entitäten gebraucht werden. Es wäre dann zu prüfen, ob *über* in Bezug auf diese Bedeutungen ambig oder vage ist, und wenn ersteres der Fall ist, sollte das Annotationsschema angepasst werden.

7.2.3. Ausführlichere Annotation

Um die Unterschiede der Entsprechungen in Bezug auf die Formulierungsnähe (vgl. Abschnitt 4.1.2) zu untersuchen, könnte eine Ersetzung der Formulierungsnähe-Kategorie durch genauere Annotationen zu syntaktischen und lexikalischen Unterschieden sinnvoll sein. Dazu könnten ggf. auch automatisch erstellte und manuell korrigierte syntaktische Analysen genutzt werden. Eine Analyse dieser Unterschiede wäre für die Erstellung von Lexikoneinträgen hilfreich und für Englischlerner allgemein von Nutzen. Auch würde eine dann mögliche genauere Betrachtung der bisher weitestgehend von der Betrachtung ausgeschlossen Fälle mit Formulierungsnähewert 4 die Durchführung typologischer Vergleiche im Stile Malás (2017) erlauben.

A. Tabellen und Abbildungen

Tabelle 29: Die hundert am häufigsten automatisch mit *über* alignierten Worte

	Automatisch alignierte Entsprechung	Häufigkeit	%	kum. %
1	on	45.145	38,60	38,60
2	[keine Entsprechung]	14.337	12,26	50,86
3	about	12.378	10,58	61,44
4	of	4.437	3,79	65,23
5	over	4.258	3,64	68,87
6	have	2.701	2,31	71,18
7	through	1.820	1,56	72,74
8	for	1.670	1,43	74,17
9	than	1.556	1,33	75,50
10	beyond	1.496	1,28	76,78
11	concerning	1.488	1,27	78,05
12	with	1.426	1,22	79,27
13	to	1.392	1,19	80,46
14	at	1.277	1,09	81,55
15	via	1.085	0,93	82,48
16	by	970	0,83	83,31
17	has	875	0,75	84,06
18	discuss	871	0,74	84,80
19	regarding	727	0,62	85,42
20	discussing	676	0,58	86,00
21	,	526	0,45	86,45
22	relating	519	0,44	86,89
23	is	514	0,44	87,33
24	above	501	0,43	87,76
25	in	461	0,39	88,16
26	across	439	0,38	88,53
27	informed	407	0,35	88,88
28	discussed	320	0,27	89,15
29	as	319	0,27	89,43
30	that	306	0,26	89,69
31	into	300	0,26	89,94
32	debating	275	0,24	90,18
33	report	225	0,19	90,37
34	debate	224	0,19	90,56
35	talk	221	0,19	90,75
36	consider	203	0,17	90,93
37	what	200	0,17	91,10
38	from	198	0,17	91,27
39	means	174	0,15	91,41
40	aware	172	0,15	91,56
41	establishing	171	0,15	91,71
42	are	162	0,14	91,85
43	more	160	0,14	91,98
44	decide	149	0,13	92,11

45	had	135	0,12	92,23
46	upon	134	0,11	92,34
47	and	134	0,11	92,45
48	see	123	0,11	92,56
49	will	122	0,10	92,66
50	talk about	111	0,09	92,76
51	relation	109	0,09	92,85
52	on on	99	0,08	92,94
53	was	98	0,08	93,02
54	under	93	0,08	93,10
55	debated	93	0,08	93,18
56	put	89	0,08	93,26
57	before	83	0,07	93,33
58	governing	82	0,07	93,40
59	having	80	0,07	93,47
60	exceed	80	0,07	93,53
61	determine	80	0,07	93,60
62	how	79	0,07	93,67
63	against	79	0,07	93,74
64	welcome	78	0,07	93,80
65	considering	75	0,06	93,87
66	dealing	74	0,06	93,93
67	using	69	0,06	93,99
68	:	69	0,06	94,05
69	report on	68	0,06	94,11
70	when	66	0,06	94,16
71	more than	66	0,06	94,22
72	whether	65	0,06	94,28
73	terms	64	0,05	94,33
74	cross	64	0,05	94,39
75	covering	63	0,05	94,44
76	possess	60	0,05	94,49
77	any	60	0,05	94,54
78	excess	58	0,05	94,59
79	exceeds	57	0,05	94,64
80	adopting	57	0,05	94,69
81	available	56	0,05	94,74
82	know	51	0,04	94,78
83	speak	50	0,04	94,82
84	must	50	0,04	94,87
85	decision	50	0,04	94,91
86	hear	49	0,04	94,95
87	mention	47	0,04	94,99
88	use	45	0,04	95,03
89	outside	44	0,04	95,07
90	laying	44	0,04	95,10
91	subject	43	0,04	95,14
92	where	42	0,04	95,18
93	find	41	0,04	95,21
94	access	41	0,04	95,25
95	going	40	0,03	95,28

96	throughout	38	0,03	95,31
97	thought	37	0,03	95,35
98	one	37	0,03	95,38
99	around	37	0,03	95,41
100	given	36	0,03	95,44

Tabelle 30: Rektionsstatus für verschiedene Regentien

Deutsch	<i>r</i>	abfragen, abstimmen, Abstimmung, abwickeln, Aussprache, ausweiten, Bedauern, Begeisterung, beraten, Bericht (erstatten), berichten, Bescheid wissen, beschließen, Beschwerde, Besorgnis, besorgt, Bestimmung, bestürzt, beunruhigt, Bilanz, Debatte, debattieren, Dialog, Diskussion, diskutieren, einig, Einigung, entscheiden, Entscheidung, enttäuscht, Enttäuschung, Freude, froh, gehen, Genugtuung, Gespräch, Illusion, in Kenntnis setzen, Information, informieren, Klarheit, nachdenken, reden, sagen, schockiert, sich äußern, sich beraten, sich erstrecken, sich freuen, sich Gedanken machen, sich im klaren sein, sich verbreiten, sich wundern, spekulieren, sprechen, stehen, stellen, Überblick, Umweg, Urteil, Vereinbarung, verfügen, Verhandlung, verhängen, vermitteln, Vermutung, verteilen, Verteilung, Vertrag, zornig
	<i>n</i>	abdecken, Änderungsantrag, Anekdote, anheizen, ansagen, ansteigen, Ausschuss, Beschlussfassung, Daten, Dokument, einreisen, einsetzen, Empfehlung, erhöhen, Erklärung, Erläuterung, erreichen, Flughöhenbereich, Garantie, gehen, gelangen, Kenntnis, Leitlinien, liegen, Monopol, Nachforschung, Pilotprojekt, Protokoll, Randnummer, Rechtsakt, Rechtsvorschrift, Regelung, Richtlinie, schaffen, schmuggeln, stattfinden, stehen, Studie, Text, transportieren, Überlegung, Untersuchung, verlassen, Verordnung, Verpflichtung, Verzeichnis, Vorschlag, Vorschrift, wissen, Wut, zu Fall kommen, Zufriedenheit, zugehen
	<i>d</i>	über einen Kamm scheren, verfügen
Englisch	<i>r</i>	agreement upon, anger at, clue about, debate on, debate over, debate upon, decide upon, directive on, disagreement over, disappointed with, disappointment at, discussion about, discussion on, inform of, insight into, learn about, overview of, provision in, refer to, reflect upon, regret at, report on, say about, speak about, speak of, talk about, talk of, vote on, vote upon
	<i>n</i>	<i>alles mit</i> as to, <i>alles mit</i> in relation to, <i>alles mit</i> in terms of, <i>alles mit</i> on the subject of, <i>alles mit</i> regarding, <i>alles mit</i> relating to, anecdote from, complaint against, concern around, cover under, decision concerning, dialogue regarding, discussions concerning, enter via, enthusiasm around, extend outside, forward through, go through, investigation into, leave by, list of, place above, place before, reports concerning, rules concerning, satisfaction at, set out across, spread across, spread throughout, study into, tell over, traffic above, transport across
	<i>d</i>	before long

Tabelle 31: Die Häufigkeiten der Bedeutungen in DE

Bedeutung	Anzahl	Anteil (in %)
5	147	26,73

9	133	24,18
5,9	55	10,00
d	42	7,64
4a	40	7,27
3d	31	5,64
2c	14	2,55
1d	12	2,18
Abd	11	2,00
3b	8	1,45
8	7	1,27
3c	5	0,91
6	5	0,91
6,d	5	0,91
10	4	0,73
3b,3d	4	0,73
1d +m	3	0,55
1f	2	0,36
1f +m	2	0,36
2b	2	0,36
3b,5	2	0,36
1a	1	0,18
1b	1	0,18
1b +m	1	0,18
1c	1	0,18
1c +m,4a +m	1	0,18
1d,3a	1	0,18
1f +m +z	1	0,18
2a	1	0,18
2c,5	1	0,18
3a,3b	1	0,18
3b,3c	1	0,18
3b,6	1	0,18
3d,4a	1	0,18
3d,5,9	1	0,18
3d,9	1	0,18
3e	1	0,18
Summe	550	100

Tabelle 32: Die Frequenzen der Entsprechungen vor Trennung nach Bedeutung und Ausschluss auszu-sondernder Beispiele

Kategorie	Frequenz
4	186
adv	44
on	27
~P	25
about	25
of	19
concerning	12

relating to	12
by means of	11
raus	11
regarding	11
upon	11
via	11
at	10
through	9
throughout	9
across	8
in relation to	8
over	8
4n use	7
in terms of	7
into	7
AMV discussed	6
AMV has	6
around	6
for	6
above	5
AMV consider	5
as to	5
with	5
4n dealing	4
4n using	4
AMV considering	4
AMV debated	4
AMV debating	4
AMV deciding	4
AMV discuss	4
AMV discussing	4
AMV have	4
AMV hold	4
AMV mention	4
AMV possess	4
by	4
to	4
4n subject	3
AMV debate	3
AMV decide	3
AMV determine	3
AMV enjoy	3
AMV had	3
AMV having	3
AMV welcome	3
before	3
in	3
4n mention	2
against	2
AMV exceed	2
as regards	2

from	2
on the subject of	2
under	2
vübz	2
4n debating	1
4n excess	1
after	1
AMV covering	1
AMV governing	1
AMV transcends	1
outside	1
with regard to	1

Tabelle 33: Wie oft wurde welche automatisch alignierte Entsprechung wozu korrigiert?

Alignierung	Entsprechung	Anzahl
[keine Entsprechung]	4	7
[keine Entsprechung]	adv	2
[keine Entsprechung]	AMV have	1
[keine Entsprechung]	raus	2
about	about	12
above	~P	5
above	above	5
above	adv	2
access	4	8
across	4	2
across	across	8
across	adv	1
across	raus	1
after	4	1
after	adv	2
after	after	1
against	~P	1
against	4	9
against	against	2
around	4	4
around	around	6
around	by	1
around	on	1
as	4	6
as	as to	5
As	AMV has	1
at	4	2
at	at	10
aware	4	3
aware	of	1
before	4	7
before	adv	1
before	before	3

before	raus	1
beyond	adv	12
by	4	9
by	by	3
concerning	concerning	12
consider	AMV consider	4
considering	AMV considering	4
covering	4	2
covering	AMV covering	1
covering	raus	1
dealing	4n dealing	4
debate	AMV debate	3
debate	on	1
debated	AMV debated	4
debating	4n debating	1
debating	AMV debating	3
decide	4	2
decide	AMV decide	2
deciding	4	1
deciding	AMV deciding	3
decision	4	1
decision	as regards	1
decision	raus	1
determine	AMV determine	3
discuss	AMV discuss	4
discussed	AMV discussed	4
discussing	AMV discussing	4
discussion	4	2
discussion	of	2
down	4	3
enjoy	4	1
enjoy	AMV enjoy	3
establishing	4	11
establishing	adv	1
exceed	adv	1
exceed	AMV exceed	2
exceed	raus	1
exceeds	~P	2
exceeds	adv	2
excess	~P	5
excess	4n excess	1
excess	adv	2
for	4	6
for	for	6
from	4	8
from	adv	1
from	from	2
from	raus	1
governing	4	11
governing	AMV governing	1
had	4	1

had	AMV had	3
has	4	1
has	AMV has	3
have	4	1
have	AMV have	3
having	4	1
having	AMV having	3
hold	AMV hold	4
in	4	7
in	AMV discussed	1
in	AMV has	1
in	in	3
into	4	5
into	into	7
means	4	1
means	by means of	11
mention	4	2
mention	4n mention	2
mention	AMV mention	4
more	~P	4
more than	~P	4
of	4	4
of	of	8
on	4	1
on	on	11
outside	adv	10
outside	on	1
outside	outside	1
over	~P	3
over	adv	1
over	over	8
possess	AMV possess	4
regarding	4	1
regarding	regarding	11
relating	relating to	12
relation	4	4
relation	in relation to	8
report on	on	8
speak	4	1
speak	about	1
speak	of	6
subject	4	1
subject	4n subject	3
subject	AMV discussed	1
subject	on	1
subject	on the subject of	2
talk	4	1
talk	about	5
talk	of	2
talk about	4	1
talk about	about	7

terms	4	1
terms	in terms of	7
thought	4	6
thought	to	2
through	4	3
through	through	9
throughout	adv	1
throughout	raus	2
throughout	throughout	9
to	4	8
to	to	2
to	vübz	2
transcends	adv	3
transcends	AMV transcends	1
under	~P	1
under	4	9
under	under	2
upon	raus	1
upon	upon	11
use	4	3
use	4n use	7
use	AMV has	1
using	4n using	4
via	4	1
via	via	11
welcome	4	1
welcome	AMV welcome	3
when	4	3
when	adv	2
when	AMV consider	1
when	AMV debating	1
when	on	1
whether	4	4
whether	AMV decide	1
whether	AMV deciding	1
whether	on	2
with	4	5
with	on	1
with	with	5
with	with regard to	1
within	4	3
within	as regards	1

Tabelle 34: Abbildungen nach Bedeutungen geordnet mit Rektionsstatus (i)

Bedeutung	Entsprechung	regiert de	regiert en	Anzahl
Abdeckung	across	r	n	5
Abdeckung	around	r	n	1
Abdeckung	throughout	r	n	2

Bezugspunkt	about	n	r	2
Bezugspunkt	about	r	r	1
Bezugspunkt	as regards	r	n	1
Bezugspunkt	as to	n	n	2
Bezugspunkt	as to	r	n	3
Bezugspunkt	concerning	n	n	5
Bezugspunkt	concerning	r	n	3
Bezugspunkt	in	r	r	1
Bezugspunkt	in relation to	r	n	6
Bezugspunkt	in terms of	r	n	3
Bezugspunkt	into	n	n	3
Bezugspunkt	into	n	r	2
Bezugspunkt	into	r	r	1
Bezugspunkt	of	r	r	5
Bezugspunkt	on	n	n	1
Bezugspunkt	on	n	r	2
Bezugspunkt	on	r	n	3
Bezugspunkt	on	r	r	3
Bezugspunkt	over	r	r	2
Bezugspunkt	regarding	n	n	3
Bezugspunkt	regarding	r	n	3
Bezugspunkt	relating to	n	n	7
Bezugspunkt	relating to	r	n	2
Bezugspunkt	upon	n	r	1
Bezugspunkt	upon	r	r	8
Bezugspunkt	with regard to	n	n	1
Konditional Kausal / Emotionsgegenstand	at	n	n	1
Konditional Kausal / Emotionsgegenstand	at	n	r	1
Konditional Kausal / Emotionsgegenstand	at	r	n	2
Konditional Kausal / Emotionsgegenstand	at	r	r	6
Konditional Kausal / Emotionsgegenstand	in	r	r	1
Konditional Kausal / Emotionsgegenstand	over	r	r	1
Konditional Kausal / Emotionsgegenstand	with	r	r	4
Modal Abstraktes Instrument	by	n	n	1
Modal Abstraktes Instrument	by	r	n	1
Modal Abstraktes Instrument	by means of	n	n	8
Modal Abstraktes Instrument	through	n	n	2
Modal Abstraktes Instrument	via	n	n	5
Modal Abstraktes Instrument	with	n	n	1
Modal Art und Weise	via	n	n	1
Modal Informationsübermittler	by	r	n	1
Modal Informationsübermittler	by means of	n	n	1
Modal Informationsübermittler	over	n	n	1
Modal Informationsübermittler	through	n	n	4
Modal Informationsübermittler	via	n	n	2
Modal Instrumental	by	n	n	1
Modal Instrumental	over	n	n	1
Modal Medial	by means of	n	n	1
Modal Medial	through	n	n	1
Rangfolge	above	r	n	2
Rangfolge	before	r	n	3

Spatial Achsenbezogen	above	n	n	1
Spatial Achsenbezogen +metaph	above	n	n	1
Spatial Achsenbezogen +metaph +ziel	upon	n	n	1
Spatial Traverse <3D	across	n	n	2
Spatial Traverse <3D	over	n	n	1
Spatial Traverse <3D	through	n	n	2
Spatial Traverse <3D	via	n	n	4
Spatial Traverse <3D +metaph	outside	n	n	1
Spatial Traverse <3D +metaph	via	r	n	1
Spatial Traverse vertikal	across	n	n	1
Temporal Maßeinheit	for	n	n	4
Temporal Maßeinheit	in	n	n	1
Temporal Maßeinheit	over	n	n	3
Temporal Maßeinheit	throughout	n	n	5
Temporal Zeitdauer	throughout	n	n	2
Thema	about	n	r	1
Thema	about	r	n	1
Thema	about	r	r	20
Thema	around	r	n	2
Thema	as to	n	n	2
Thema	concerning	n	n	5
Thema	concerning	r	n	2
Thema	in	r	r	1
Thema	in relation to	r	n	5
Thema	in terms of	n	n	2
Thema	in terms of	r	n	2
Thema	into	n	r	1
Thema	into	r	r	1
Thema	of	r	r	11
Thema	on	n	n	1
Thema	on	n	r	3
Thema	on	r	n	3
Thema	on	r	r	14
Thema	on the subject of	r	n	2
Thema	regarding	r	n	4
Thema	relating to	n	n	8
Thema	relating to	r	n	1
Thema	upon	r	r	1

Tabelle 35: Abbildungen nach Bedeutungen geordnet ohne Rektionsstatus (i)

Bedeutung	Entsprechung	Anzahl
Abdeckung	across	5
Abdeckung	around	1
Abdeckung	throughout	2
Bezugspunkt	about	3
Bezugspunkt	as regards	1
Bezugspunkt	as to	5
Bezugspunkt	concerning	8

Bezugspunkt	in	1
Bezugspunkt	in relation to	6
Bezugspunkt	in terms of	3
Bezugspunkt	into	6
Bezugspunkt	of	5
Bezugspunkt	on	9
Bezugspunkt	over	2
Bezugspunkt	regarding	6
Bezugspunkt	relating to	9
Bezugspunkt	upon	9
Bezugspunkt	with regard to	1
Konditional Kausal / Emotionsgegenstand	at	10
Konditional Kausal / Emotionsgegenstand	in	1
Konditional Kausal / Emotionsgegenstand	over	1
Konditional Kausal / Emotionsgegenstand	with	4
Modal Abstraktes Instrument	by	2
Modal Abstraktes Instrument	by means of	8
Modal Abstraktes Instrument	through	2
Modal Abstraktes Instrument	via	5
Modal Abstraktes Instrument	with	1
Modal Art und Weise	via	1
Modal Informationsübermittler	by	1
Modal Informationsübermittler	by means of	1
Modal Informationsübermittler	over	1
Modal Informationsübermittler	through	4
Modal Informationsübermittler	via	2
Modal Instrumental	by	1
Modal Instrumental	over	1
Modal Medial	by means of	1
Modal Medial	through	1
Rangfolge	above	2
Rangfolge	before	3
Spatial Achsenbezogen	above	1
Spatial Achsenbezogen +metaph	above	1
Spatial Achsenbezogen +metaph +ziel	upon	1
Spatial Traverse <3D	across	2
Spatial Traverse <3D	over	1
Spatial Traverse <3D	through	2
Spatial Traverse <3D	via	4
Spatial Traverse <3D +metaph	outside	1
Spatial Traverse <3D +metaph	via	1
Spatial Traverse vertikal	across	1
Temporal Maßeinheit	for	4
Temporal Maßeinheit	in	1
Temporal Maßeinheit	over	3
Temporal Maßeinheit	throughout	5
Temporal Zeitdauer	throughout	2
Thema	about	22
Thema	around	2
Thema	as to	2
Thema	concerning	7

Thema	in	1
Thema	in relation to	5
Thema	in terms of	4
Thema	into	2
Thema	of	11
Thema	on	21
Thema	on the subject of	2
Thema	regarding	4
Thema	relating to	9
Thema	upon	1

Tabelle 36: Abbildungen nach Entsprechungen geordnet mit Rektionsstatus (i)

Entsprechung	Bedeutung	regiert de	regiert en	Anzahl
about	Bezugspunkt	n	r	2
about	Bezugspunkt	r	r	1
about	Thema	n	r	1
about	Thema	r	n	1
about	Thema	r	r	20
above	Rangfolge	r	n	2
above	Spatial Achsenbezogen	n	n	1
above	Spatial Achsenbezogen +metaph	n	n	1
across	Abdeckung	r	n	5
across	Spatial Traverse <3D	n	n	2
across	Spatial Traverse vertikal	n	n	1
around	Abdeckung	r	n	1
around	Thema	r	n	2
as regards	Bezugspunkt	r	n	1
as to	Bezugspunkt	n	n	2
as to	Bezugspunkt	r	n	3
as to	Thema	n	n	2
at	Konditional Kausal / Emotionsgegenstand	n	n	1
at	Konditional Kausal / Emotionsgegenstand	n	r	1
at	Konditional Kausal / Emotionsgegenstand	r	n	2
at	Konditional Kausal / Emotionsgegenstand	r	r	6
before	Rangfolge	r	n	3
by	Modal Abstraktes Instrument	n	n	1
by	Modal Abstraktes Instrument	r	n	1
by	Modal Informationsübermittler	r	n	1
by	Modal Instrumental	n	n	1
by means of	Modal Abstraktes Instrument	n	n	8
by means of	Modal Informationsübermittler	n	n	1
by means of	Modal Medial	n	n	1
concerning	Bezugspunkt	n	n	5
concerning	Bezugspunkt	r	n	3
concerning	Thema	n	n	5
concerning	Thema	r	n	2
for	Temporal Maßeinheit	n	n	4
in	Bezugspunkt	r	r	1

in	Konditional Kausal / Emotionsgegenstand	r	r	1
in	Temporal Maßeinheit	n	n	1
in	Thema	r	r	1
in relation to	Bezugspunkt	r	n	6
in relation to	Thema	r	n	5
in terms of	Bezugspunkt	r	n	3
in terms of	Thema	n	n	2
in terms of	Thema	r	n	2
into	Bezugspunkt	n	n	3
into	Bezugspunkt	n	r	2
into	Bezugspunkt	r	r	1
into	Thema	n	r	1
into	Thema	r	r	1
of	Bezugspunkt	r	r	5
of	Thema	r	r	11
on	Bezugspunkt	n	n	1
on	Bezugspunkt	n	r	2
on	Bezugspunkt	r	n	3
on	Bezugspunkt	r	r	3
on	Thema	n	n	1
on	Thema	n	r	3
on	Thema	r	n	3
on	Thema	r	r	14
on the subject of	Thema	r	n	2
outside	Spatial Traverse <3D +metaph	n	n	1
over	Bezugspunkt	r	r	2
over	Konditional Kausal / Emotionsgegenstand	r	r	1
over	Modal Informationsübermittler	n	n	1
over	Modal Instrumental	n	n	1
over	Spatial Traverse <3D	n	n	1
over	Temporal Maßeinheit	n	n	3
regarding	Bezugspunkt	n	n	3
regarding	Bezugspunkt	r	n	3
regarding	Thema	r	n	4
relating to	Bezugspunkt	n	n	7
relating to	Bezugspunkt	r	n	2
relating to	Thema	n	n	8
relating to	Thema	r	n	1
through	Modal Abstraktes Instrument	n	n	2
through	Modal Informationsübermittler	n	n	4
through	Modal Medial	n	n	1
through	Spatial Traverse <3D	n	n	2
throughout	Abdeckung	r	n	2
throughout	Temporal Maßeinheit	n	n	5
throughout	Temporal Zeitdauer	n	n	2
upon	Bezugspunkt	n	r	1
upon	Bezugspunkt	r	r	8
upon	Spatial Achsenbezogen +metaph +ziel	n	n	1
upon	Thema	r	r	1
via	Modal Abstraktes Instrument	n	n	5
via	Modal Art und Weise	n	n	1

via	Modal Informationsübermittler	n	n	2
via	Spatial Traverse <3D	n	n	4
via	Spatial Traverse <3D +metaph	r	n	1
with	Konditional Kausal / Emotionsgegenstand	r	r	4
with	Modal Abstraktes Instrument	n	n	1
with regard to	Bezugspunkt	n	n	1

Tabelle 37: Abbildungen nach Entsprechungen geordnet ohne Rektionsstatus (I)

Entsprechung	Bedeutung	Anzahl
about	Bezugspunkt	3
about	Thema	22
above	Rangfolge	2
above	Spatial Achsenbezogen	1
above	Spatial Achsenbezogen +metaph	1
across	Abdeckung	5
across	Spatial Traverse <3D	2
across	Spatial Traverse vertikal	1
around	Abdeckung	1
around	Thema	2
as regards	Bezugspunkt	1
as to	Bezugspunkt	5
as to	Thema	2
at	Konditional Kausal / Emotionsgegenstand	10
before	Rangfolge	3
by	Modal Abstraktes Instrument	2
by	Modal Informationsübermittler	1
by	Modal Instrumental	1
by means of	Modal Abstraktes Instrument	8
by means of	Modal Informationsübermittler	1
by means of	Modal Medial	1
concerning	Bezugspunkt	8
concerning	Thema	7
for	Temporal Maßeinheit	4
in	Bezugspunkt	1
in	Konditional Kausal / Emotionsgegenstand	1
in	Temporal Maßeinheit	1
in	Thema	1
in relation to	Bezugspunkt	6
in relation to	Thema	5
in terms of	Bezugspunkt	3
in terms of	Thema	4
into	Bezugspunkt	6
into	Thema	2
of	Bezugspunkt	5
of	Thema	11
on	Bezugspunkt	9
on	Thema	21
on the subject of	Thema	2

outside	Spatial Traverse <3D +metaph	1
over	Bezugspunkt	2
over	Konditional Kausal / Emotionsgegenstand	1
over	Modal Informationsübermittler	1
over	Modal Instrumental	1
over	Spatial Traverse <3D	1
over	Temporal Maßeinheit	3
regarding	Bezugspunkt	6
regarding	Thema	4
relating to	Bezugspunkt	9
relating to	Thema	9
through	Modal Abstraktes Instrument	2
through	Modal Informationsübermittler	4
through	Modal Medial	1
through	Spatial Traverse <3D	2
throughout	Abdeckung	2
throughout	Temporal Maßeinheit	5
throughout	Temporal Zeitdauer	2
upon	Bezugspunkt	9
upon	Spatial Achsenbezogen +metaph +ziel	1
upon	Thema	1
via	Modal Abstraktes Instrument	5
via	Modal Art und Weise	1
via	Modal Informationsübermittler	2
via	Spatial Traverse <3D	4
via	Spatial Traverse <3D +metaph	1
with	Konditional Kausal / Emotionsgegenstand	4
with	Modal Abstraktes Instrument	1
with regard to	Bezugspunkt	1

Tabelle 38: Abbildungen für AMV nach Verben geordnet

Verb	Bedeutung	regiert de	Anzahl
consider	Bezugspunkt	n	1
consider	Bezugspunkt	r	4
considering	Bezugspunkt	r	4
covering	Abdeckung	r	1
debate	Thema	r	3
debated	Thema	r	4
debating	Thema	r	4
decide	Bezugspunkt	r	3
deciding	Bezugspunkt	r	4
determine	Bezugspunkt	r	3
discuss	Thema	r	4
discussed	Thema	r	6
discussing	Thema	r	4
enjoy	Konditional Kausal / Emotionsgegenstand	r	3
exceed	Überschreitung	n	2
governing	Machtverhältnis	r	1

had	Desemantisiert	d	3
has	Desemantisiert	d	6
have	Desemantisiert	d	4
having	Desemantisiert	d	3
hold	Desemantisiert	d	4
mention	Thema	r	4
possess	Desemantisiert	d	4
transcends	Rangfolge	r	1
welcome	Konditional Kausal / Emotionsgegenstand	r	3

Tabelle 39: Abbildungen für IB nach Bedeutungen geordnet mit Rektionsstatus

Bedeutung	Entsprechung	regiert de	regiert en	Anzahl
Abdeckung	across	r	n	5
Abdeckung	around	r	n	1
Abdeckung	throughout	r	n	2
Konditional Kausal / Emotionsgegenstand	at	n	n	1
Konditional Kausal / Emotionsgegenstand	at	n	r	1
Konditional Kausal / Emotionsgegenstand	at	r	n	2
Konditional Kausal / Emotionsgegenstand	at	r	r	6
Konditional Kausal / Emotionsgegenstand	in	r	r	1
Konditional Kausal / Emotionsgegenstand	over	r	r	1
Konditional Kausal / Emotionsgegenstand	with	r	r	4
Modal Abstraktes Instrument	by	n	n	1
Modal Abstraktes Instrument	by	r	n	1
Modal Abstraktes Instrument	by means of	n	n	8
Modal Abstraktes Instrument	through	n	n	2
Modal Abstraktes Instrument	via	n	n	5
Modal Abstraktes Instrument	with	n	n	1
Modal Art und Weise	via	n	n	1
Modal Informationsübermittler	by	r	n	1
Modal Informationsübermittler	by means of	n	n	1
Modal Informationsübermittler	over	n	n	1
Modal Informationsübermittler	through	n	n	4
Modal Informationsübermittler	via	n	n	2
Modal Instrumental	by	n	n	1
Modal Instrumental	over	n	n	1
Modal Medial	by means of	n	n	1
Modal Medial	through	n	n	1
Rangfolge	above	r	n	2
Rangfolge	before	r	n	3
Spatial Achsenbezogen	above	n	n	1
Spatial Achsenbezogen +metaph	above	n	n	1
Spatial Achsenbezogen +metaph +ziel	upon	n	n	1
Spatial Traverse 1D	across	n	n	2
Spatial Traverse 1D +metaph	outside	n	n	1
Spatial Traverse 2D	over	n	n	1
Spatial Traverse 2D	through	n	n	1
Spatial Traverse vertikal	across	n	n	1

Spatial Traverse Via	through	n	n	1
Spatial Traverse Via	via	n	n	4
Spatial Traverse Via +metaph	via	r	n	1
Temporal Maßeinheit	for	n	n	4
Temporal Maßeinheit	in	n	n	1
Temporal Maßeinheit	over	n	n	3
Temporal Maßeinheit	throughout	n	n	5
Temporal Zeitdauer	throughout	n	n	2
Thema	about	n	r	3
Thema	about	r	n	1
Thema	about	r	r	21
Thema	around	r	n	2
Thema	as regards	r	n	1
Thema	as to	n	n	2
Thema	as to	r	n	3
Thema	concerning	n	n	7
Thema	concerning	r	n	5
Thema	in	r	r	1
Thema	in relation to	r	n	8
Thema	in terms of	n	n	2
Thema	in terms of	r	n	5
Thema	into	n	n	3
Thema	into	n	r	2
Thema	into	r	r	2
Thema	of	r	r	16
Thema	on	n	n	2
Thema	on	n	r	3
Thema	on	r	n	5
Thema	on	r	r	17
Thema	on the subject of	r	n	2
Thema	over	r	r	2
Thema	regarding	n	n	3
Thema	regarding	r	n	7
Thema	relating to	n	n	9
Thema	relating to	r	n	3
Thema	upon	n	r	1
Thema	upon	r	r	8
Thema	with regard to	n	n	1

Tabelle 40: Abbildungen für IB nach Entsprechungen geordnet mit Rektionsstatus

Entsprechung	Bedeutung	regiert de	regiert en	Anzahl
about	Thema	n	r	3
about	Thema	r	n	1
about	Thema	r	r	21
above	Rangfolge	r	n	2
above	Spatial Achsenbezogen	n	n	1
above	Spatial Achsenbezogen +metaph	n	n	1
across	Abdeckung	r	n	5

across	Spatial Traverse 1D	n	n	2
across	Spatial Traverse vertikal	n	n	1
around	Abdeckung	r	n	1
around	Thema	r	n	2
as regards	Thema	r	n	1
as to	Thema	n	n	2
as to	Thema	r	n	3
at	Konditional Kausal / Emotionsgegenstand	n	n	1
at	Konditional Kausal / Emotionsgegenstand	n	r	1
at	Konditional Kausal / Emotionsgegenstand	r	n	2
at	Konditional Kausal / Emotionsgegenstand	r	r	6
before	Rangfolge	r	n	3
by	Modal Abstraktes Instrument	n	n	1
by	Modal Abstraktes Instrument	r	n	1
by	Modal Informationsübermittler	r	n	1
by	Modal Instrumental	n	n	1
by means of	Modal Abstraktes Instrument	n	n	8
by means of	Modal Informationsübermittler	n	n	1
by means of	Modal Medial	n	n	1
concerning	Thema	n	n	7
concerning	Thema	r	n	5
for	Temporal Maßeinheit	n	n	4
in	Konditional Kausal / Emotionsgegenstand	r	r	1
in	Temporal Maßeinheit	n	n	1
in	Thema	r	r	1
in relation to	Thema	r	n	8
in terms of	Thema	n	n	2
in terms of	Thema	r	n	5
into	Thema	n	n	3
into	Thema	n	r	2
into	Thema	r	r	2
of	Thema	r	r	16
on	Thema	n	n	2
on	Thema	n	r	3
on	Thema	r	n	5
on	Thema	r	r	17
on the subject of	Thema	r	n	2
outside	Spatial Traverse 1D +metaph	n	n	1
over	Konditional Kausal / Emotionsgegenstand	r	r	1
over	Modal Informationsübermittler	n	n	1
over	Modal Instrumental	n	n	1
over	Spatial Traverse 2D	n	n	1
over	Temporal Maßeinheit	n	n	3
over	Thema	r	r	2
regarding	Thema	n	n	3
regarding	Thema	r	n	7
relating to	Thema	n	n	9
relating to	Thema	r	n	3
through	Modal Abstraktes Instrument	n	n	2
through	Modal Informationsübermittler	n	n	4
through	Modal Medial	n	n	1

through	Spatial Traverse 2D	n	n	1
through	Spatial Traverse Via	n	n	1
throughout	Abdeckung	r	n	2
throughout	Temporal Maßeinheit	n	n	5
throughout	Temporal Zeitdauer	n	n	2
upon	Spatial Achsenbezogen +metaph +ziel	n	n	1
upon	Thema	n	r	1
upon	Thema	r	r	8
via	Modal Abstraktes Instrument	n	n	5
via	Modal Art und Weise	n	n	1
via	Modal Informationsübermittler	n	n	2
via	Spatial Traverse Via	n	n	4
via	Spatial Traverse Via +metaph	r	n	1
with	Konditional Kausal / Emotionsgegenstand	r	r	4
with	Modal Abstraktes Instrument	n	n	1
with regard to	Thema	n	n	1

Abbildung 6: Sankey-Diagramm: Wozu wurden die automatisch alignierten Entsprechungen?

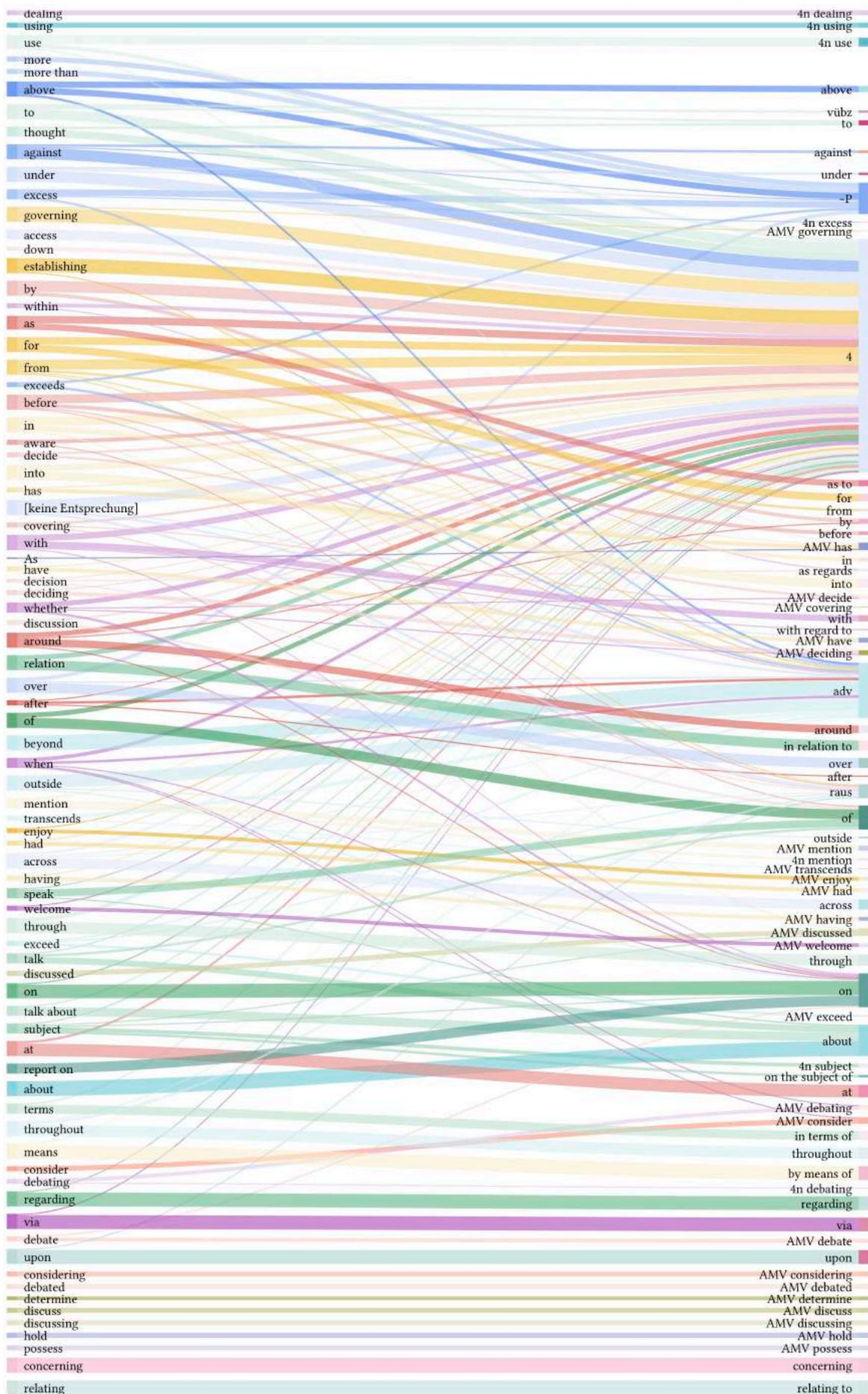
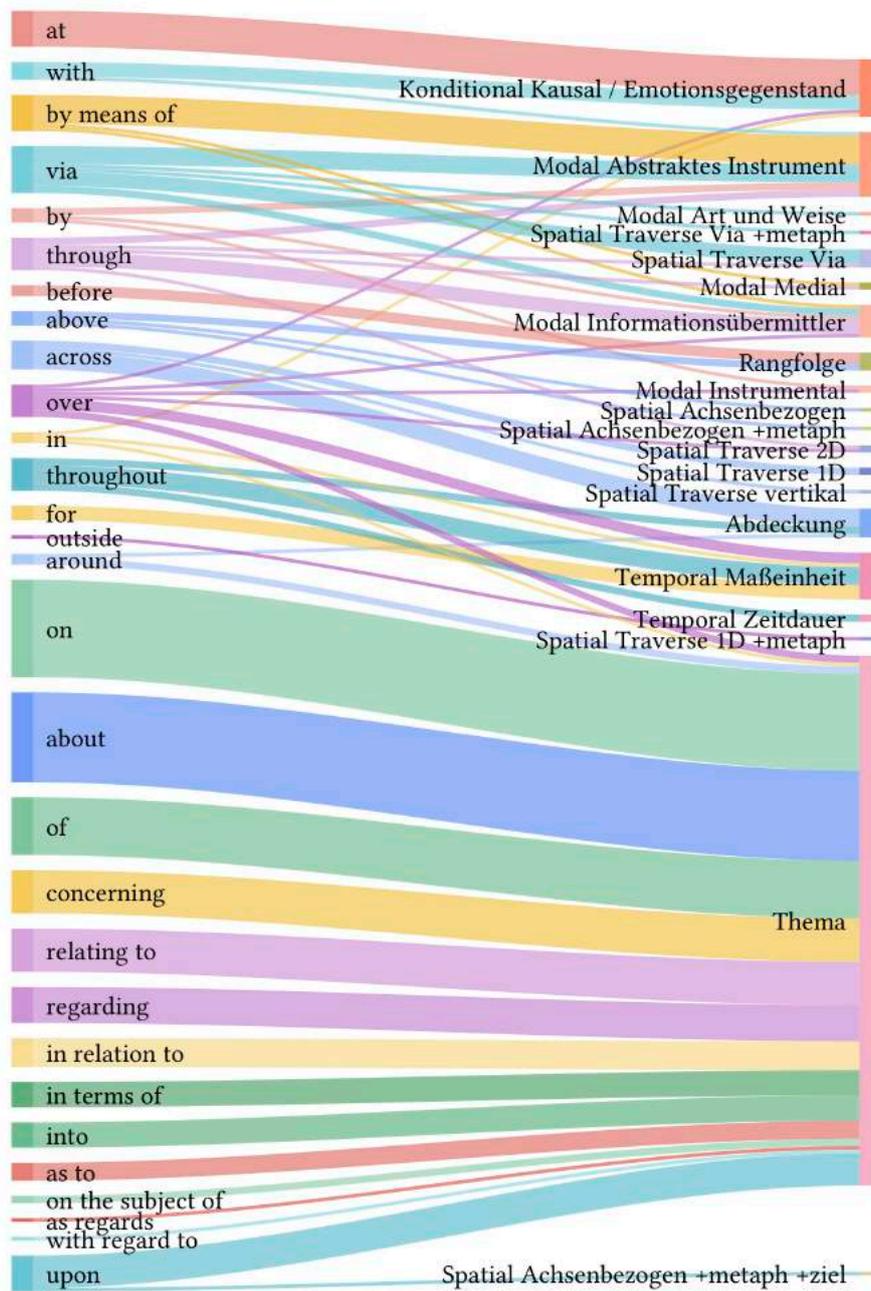


Abbildung 7: Sankey-Diagramm: Bei welchen Lesarten kommen in IB welche Entsprechungen vor?



B. Abkürzungsverzeichnis

	Bedeutung
Abd	Bedeutung: Abdeckung
AMV	Alignierung mit Verb
AMV	Datensatz: AMV-Fälle, Formulierungsnähewert 1–3, Bedeutung passend
AMVNP	Datensatz: AMV-Fälle, Formulierungsnähewert 1–3, Bedeutung unpassend
d	Bedeutung: desemantisiert
d	Rektionsstatus: desemantisiert
DE	Datensatz: Annotationen für die deutschen Belege (ohne die auszuschließenden). Bei Mehrfachannotationen nicht getrennt
DE-DIS	Datensatz: Annotationen für die deutschen Belege (ohne die auszuschließenden). Bei Mehrfachannotationen getrennt
I	Datensatz: Formulierungsnähewert 1–3, Bedeutung im Englischen passend, keine AMV-Fälle
IB	Datensatz: Wie I, aber mit Annotationen, in denen die in den Abschnitten 5.2 und 5.3 vorgeschlagenen Änderungen umgesetzt sind
n	Rektionsstatus: nicht regiert
NP	Datensatz: Formulierungsnähewert 1–3, Bedeutung im Englischen <i>unpassend</i> , keine AMV-Fälle
r	Rektionsstatus: regiert
1a	Bedeutung: Spatial Traverse Grenzbereich
1b	Bedeutung: Spatial Traverse proximal
1c	Bedeutung: Spatial Traverse vertikal
1d	Bedeutung: Spatial Traverse <3D
1e	Bedeutung: Spatial Bedeckung
1f	Bedeutung: Spatial Achsenbezogen
2a	Bedeutung: Temporal Tagesteil
2b	Bedeutung: Temporal Zeitdauer
2c	Bedeutung: Temporal Maßeinheit
3a	Bedeutung: Modal Instrumental
3b	Bedeutung: Modal Informationsübermittler
3c	Bedeutung: Modal Medial
3d	Bedeutung: Modal Abstraktes Instrument
3e	Bedeutung: Modal Art und Weise
4	Datensatz: Formulierungsnähe 4
4a	Bedeutung: Konditional Kausal / Emotionsgegenstand
4b	Bedeutung: Konditional im engeren Sinne
4N	Datensatz: Formulierungsnähe 4n, Bedeutung passend
5	Bedeutung: Bezugspunkt
6	Bedeutung: Machtverhältnis
7	Bedeutung: Maßangabe
8	Bedeutung: Rangfolge
9	Bedeutung: Thema
10	Bedeutung: Überschreitung
11	Bedeutung: Zuordnung
+m	Bedeutung: metaphorisch
+z	Bedeutung: zielbezogen

Literatur

- Alexiadou, A. & Iordăchioaia, G. (2014). The psych causative alternation. *Lingua*, 148, 53–79. <https://doi.org/10.1016/j.lingua.2014.05.010>
- Bauer, L., Lieber, R. & Plag, I. (2013). *The Oxford Reference Guide to English Morphology*. Oxford University Press.
- Bejan, C. (2002). Psych Nominalizations in English and German. *The Annals of Ovidius University of Constanța: Philology Series*, 8, 19–33.
- Belletti, A. & Rizzi, L. (1988). Psych-verbs and θ -theory. *Natural Language & Linguistic Theory*, 6(3), 291–352. <https://doi.org/10.1007/BF00133902>
- Bilder, C. R. & Loughin, T. M. (2004). Testing for Marginal Independence between Two Categorical Variables with Multiple Responses. *Biometrics*, 60(1), 241–248. <https://doi.org/10.1111/j.0006-341X.2004.00147.x>
- Bilder, C. R., Loughin, T. M. & Nettleton, D. (2000). Multiple Marginal Independence Testing for Pick Any/C Variables. *Communications in Statistics – Simulation and Computation*, 29(4), 1285–1316. <https://doi.org/10.1080/03610910008813665>
- Biljetina, J. L. (2015). Serbian equivalents of the English preposition *of*: A contrastive corpus analysis. *Journal for Languages & Literatures of the Faculty of Philosophy in Novi Sad/Zbornik za Jezike i Književnosti Filozofskog Fakulteta u Novom Sadu*, 5, 81–91. <https://doi.org/10.19090/zjik.2015.5>
- Bird, S. & Klein, E. (2009). *Natural Language Processing with Python*. O'Reilly.
- Bott, O. & Solstad, T. (2014). From Verbs to Discourse: A Novel Account of Implicit Causality. In B. Hemforth, B. Mertins & C. Fabricius-Hansen (Hrsg.), *Psycholinguistic Approaches to Meaning and Understanding across Languages* (S. 213–251). Springer International Publishing. https://doi.org/10.1007/978-3-319-05675-3_9
- Breindl, E. (2006). Präpositionalphrasen. In V. Ágel, L. M. Eichinger, H.-W. Eroms, P. Hellwig, H. J. Heringer & H. Lobin (Hrsg.), *Dependenz und Valenz: Ein internationales Handbuch der zeitgenössischen Forschung* (S. 936–950). De Gruyter.
- Brugman, C. & Lakoff, G. (2006). Cognitive topology and lexical networks. In D. Geeraerts (Hrsg.), *Cognitive Linguistics: Basic Readings* (S. 109–140). Mouton De Gruyter. (Original erschienen 1988)
- Chang, W. (2014). *extrafont: Tools for using fonts* [R package version 0.17]. <https://CRAN.R-project.org/package=extrafont>
- CL Research. (2007). *The Preposition Project*. Verfügbar 11. Mai 2020 unter https://www.clres.com/cgi-bin/onlineTPP/find_prep.cgi
- Clark, M., Khaled, N., Kohn, M. & Armoskaite, S. (2019). Let's talk emotions: A case study on affective grammar. *Glossa: A Journal of General Linguistics*, 4(1). <https://doi.org/10.5334/gjgl.678>
- Cosme, C. & Gilquin, G. (2008). Free and bound prepositions in a contrastive perspective. In S. Granger & F. Meunier (Hrsg.), *Phraseology: An interdisciplinary perspective* (S. 259–274). John Benjamins. <https://doi.org/10.1075/z.139>
- Cruse, D. A. (1986). *Lexical Semantics*. Cambridge University Press.
- Dahl, D. B., Scott, D., Roosen, C., Magnusson, A. & Swinton, J. (2019). *xtable: Export Tables to LaTeX or HTML* [R package version 1.8-4]. <https://CRAN.R-project.org/package=xtable>
- Didakowski, J. & Geyken, A. (2014). From DWDS corpora to a German word profile – methodological problems and solutions. *OPAL – Online publizierte Arbeiten zur Linguistik*, 2, 39–47.

- Dyer, C., Chahuneau, V. & Smith, N. A. (2013). A Simple, Fast, and Effective Reparameterization of IBM Model 2. *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 644–648. <https://www.aclweb.org/anthology/N13-1073>
- Engelberg, S. & Cosma, R. (2014). Subjektsätze als alternative Argumentrealisierungen im Deutschen und Rumänischen: Eine kontrastive quantitative Korpusstudie zu Psych-Verben. In R. Cosma, S. Engelberg, S. Schlotthauer, S. L. Stanescu & G. Zifonun (Hrsg.), *Komplexe Argumentstrukturen: Kontrastive Untersuchungen zum Deutschen, Rumänischen und Englischen* (S. 339–420). De Gruyter. <https://doi.org/10.1515/9783110343229>
- Explosion AI. (o. D.). *spaCy*. Verfügbar 11. Mai 2020 unter <https://spacy.io/>
- Fanselow, G. (1992). „Ergative“ Verben und die Struktur des deutschen Mittelfelds. In L. Hoffmann (Hrsg.), *Deutsche Syntax: Ansichten und Aussichten* (S. 276–303). de Gruyter. <https://doi.org/10.1515/9783110622447-012>
- Farkas, A. (2019). *LF Aligner*. Verfügbar 11. Mai 2020 unter <https://sourceforge.net/projects/aligner/>
- Gesmann, M. & de Castillo, D. (2011). googleVis: Interface between R and the Google Visualisation API. *The R Journal*, 3(2), 40–44. https://journal.r-project.org/archive/2011-2/RJournal_2011-2_Gesmann+de-Castillo.pdf
- Goethe, J. W. (1906). Brief an A. T. v. Preen vom 18. April 1820. *Goethes Briefe (Bd. 32)* (S. 246–248). Hermann Böhlau Nachfolger. (Original erschienen 1820)
- Gonen, H. & Goldberg, Y. (2016). Semi Supervised Preposition-Sense Disambiguation using Multilingual Data. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2718–2729. <https://www.aclweb.org/anthology/C16-1256>
- Graën, J., Batinic, D. & Volk, M. (2014). Cleaning the Europarl Corpus for Linguistic Applications. In J. Ruppenhofer & G. Faaß (Hrsg.), *Proceedings of the 12th edition of the KONVENS conference* (S. 222–227). Universitätsverlag Hildesheim. <http://nbn-resolving.de/urn:nbn:de:gbv:hil2-opus-2893>
- Graën, J., Sandoz, D. & Volk, M. (2017). Multilingwis² – Explore Your Parallel Corpus. In J. Tiedemann (Hrsg.), *Proceedings of the 21st Nordic Conference of Computational Linguistics* (S. 247–250). Linköping University Electronic Press. <https://www.aclweb.org/anthology/W17-0231.pdf>
- Grimm, J. & Grimm, W. (Hrsg.). (1936). über. In *Deutsches Wörterbuch*. Hirzel. <http://www.woerterbuchnetz.de/DWB?bookref=23,72,65>
- Helbig, G. & Buscha, J. (2007). *Deutsche Grammatik: Ein Handbuch für den Ausländerunterricht* (10. Aufl.). Langenscheidt.
- Herbst, T., Heath, D., Roe, I. F. & Götz, D. (2004). *A Valency Dictionary of English*. Mouton De Gruyter.
- Hertel, V. (1983). Präpositionen in fixierten Fügungen. *Beiträge zur Erforschung der deutschen Sprache*, 3, 58–76.
- Hirsch, N. (2018). *German psych verbs: insights from a decompositional perspective* (Diss.). Humboldt-Universität zu Berlin. Berlin. https://edoc.hu-berlin.de/bitstream/handle/18452/20345/dissertation_hirsch_nils.pdf
- Huddleston, R. & Pullum, G. K. (2002). *The Cambridge Grammar of the English Language*. Cambridge University Press.

- Institut für Computerlinguistik, Universität Zürich. (o. D.). *Multilingwis*. Verfügbar 11. Mai 2020 unter <https://pub.cl.uzh.ch/projects/sparcling/multilingwis2.demo/>
- Institut für Deutsche Sprache. (o. D.). *Wörterbuch zur Verbvalenz*. <https://doi.org/10.14618/evalbu>
- International Computer Science Institute, Berkeley. (o. D.). *FrameNet Data Frame Index*. Verfügbar 12. Mai 2020 unter <https://framenet.icsi.berkeley.edu/fndrupal/frameIndex>
- Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P. & Suchomel, V. (2013). The TenTen corpus family. In A. Hardie & R. Love (Hrsg.), *Corpus Linguistics 2013: Abstract Book* (S. 125–127). UCREL. <http://ucrel.lancs.ac.uk/cl2013/doc/CL2013-ABSTRACT-BOOK.pdf>
- Jarvis, S. & Odlin, T. (2000). Morphological Type, Spatial Reference, and Language Transfer. *Studies in Second Language Acquisition*, 22, 535–556. <https://doi.org/10.1017/S0272263100004034>
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1, 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Kiss, T. (2019). Determiner omission in German prepositional phrases. *Journal of Linguistics*, 55(2), 305–355. <https://doi.org/10.1017/S0022226718000312>
- Kiss, T., Müller, A., Roch, C., Stadtfeld, T., Börner, K. & Duzy, M. (2020). Ein Handbuch für die Bestimmung und Annotation von Präpositionsbedeutungen im Deutschen. *Studies in Linguistics and Linguistic Data Science*, 2, 1–440. <https://ldsl.rub.de/media/pages/research/resources/sllds/a5e84c2af0-1607105394/sllds002.pdf>
- Kiss, T. & Strunk, J. (2006). Unsupervised Multilingual Sentence Boundary Detection. *Computational Linguistics*, 32(4), 485–525. <https://doi.org/10.1162/coli.2006.32.4.485>
- Klégr, A. & Malá, M. (2009). English Equivalents of the Most Frequent Czech Prepositions. In M. Mahlberg, V. González-Díaz & C. Smith (Hrsg.), *Proceedings of the Corpus Linguistics Conference CL2009. University of Liverpool, 20.–23. Juli 2009*. <http://ucrel.lancs.ac.uk/publications/CL2009/>
- Koehn, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. *The Tenth Machine Translation Summit: Proceedings of the conference*, 79–86. <http://www.statmt.org/europarl/>
- Kortmann, B. & König, E. (1992). Categorical reanalysis: The case of deverbal prepositions. *Linguistics*, 30, 671–697.
- Koziol, N. & Bilder, C. R. (2014). *MRCV: Methods for Analyzing Multiple Response Categorical Variables (MRCVs)* [R package version 0.3-3]. <https://CRAN.R-project.org/package=MRCV>
- Kuhn, M. (2020). *caret: Classification and Regression Training* [R package version 6.0-86]. <https://CRAN.R-project.org/package=caret>
- Langenscheidt. (o. D.). *Deutsch-Englisch Wörterbuch*. PONS. Verfügbar 11. Mai 2020 unter <https://de.langenscheidt.com/deutsch-englisch/>
- Lerot, J. (1982). Die verbregierten Präpositionen in Präpositionalobjekten. In W. Abraham (Hrsg.), *Satzglieder im Deutschen: Vorschläge zur syntaktischen, semantischen und pragmatischen Fundierung* (S. 261–291). Gunter Narr.
- Levshina, N. (2015). *How to do Linguistics with R: Data exploration and statistical analysis*. John Benjamins.
- Lexical Computing. (o. D.). *Sketch Engine*. Verfügbar 11. Mai 2020 unter <https://www.sketchengine.eu/>
- Liamkina, O. (2007). Semantic Structure of the German Spatial Particle *über*. *Journal of Germanic Linguistics*, 115–160. <https://doi.org/10.1017/S1470542707000050>

- Litkowski, K. C. & Hargraves, O. (2005). The Preposition Project. *ACL-SIGSEM Workshop on “The Linguistic Dimensions of Prepositions and Their Use in Computational Linguistic Formalisms and Applications”*, 171–179. <https://www.clres.com/online-papers/sigsem2005.pdf>
- Majka, M. (2019). *naivebayes: High Performance Implementation of the Naive Bayes Algorithm in R* [R package version 0.9.7]. <https://CRAN.R-project.org/package=naivebayes>
- Malá, M. (2017). Non-prepositional English correspondences of Czech prepositional phrases: From function words to functional sentence perspective. In T. Egan & H. Dirdal (Hrsg.), *Cross-linguistic Correspondences: From lexis to genre* (S. 199–217). John Benjamins.
- Marjanović, S., Stosic, D. & Miletic, A. (2018). A Sample French-Serbian Dictionary Entry based on the ParCoLab Parallel Corpus. In J. Čibej, V. Gorjanc, I. Kosem & S. Krek (Hrsg.), *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts* (S. 423–435). Znanstvena založba Filozofske fakultete Univerze v Ljubljani. <https://doi.org/10.4312/9789610600961>
- Meex, B. (2001). The spatial and non-spatial senses of the German preposition *über*. In H. Cuyckens & B. Zawada (Hrsg.), *Polysemy in Cognitive Linguistics: Selected papers from the Fifth International Cognitive Linguistics Conference, Amsterdam, 1997* (S. 1–35). John Benjamins.
- Merriam Webster. (o. D.). *Merriam-Webster.com*. Verfügbar 11. Mai 2020 unter <https://www.merriam-webster.com/>
- Meyer, D., Zeileis, A. & Hornik, K. (2020). *vcd: Visualizing Categorical Data* [R package version 1.4-6].
- Müller, A. (2013). *Spatiale Bedeutungen deutscher Präpositionen: Bedeutungsdifferenzierung und Annotation* (Diss.). Ruhr-Universität Bochum. Bochum. <https://linguistics.rub.de/forschung/arbeitsberichte/11.pdf>
- Och, F. J. & Ney, H. (2003). A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1), 19–51. <https://doi.org/10.1162/089120103321337421>
- OED. (o. D.). *OED Online*. Oxford University Press. Verfügbar 11. Mai 2020 unter <https://www.oed.com/>
- Östling, R. & Tiedemann, J. (2016). Efficient Word Alignment with Markov Chain Monte Carlo. *The Prague Bulletin of Mathematical Linguistics*, 106, 125–146. <https://doi.org/10.1515/pralin-2016-0013>
- Pafel, J. & Reich, I. (2016). *Einführung in die Semantik: Grundlagen – Analysen – Theorien*. Metzler.
- Pesetsky, D. M. (1995). *Zero syntax: Experiencers and Cascades*. MIT Press.
- PONS. (o. D. a). *PONS Online-Wörterbuch*. Verfügbar 11. Mai 2020 unter <https://de.pons.com/>
- PONS. (o. D. b). *über*. In *PONS Deutsch-Englisch-Online-Wörterbuch*. Verfügbar 12. Mai 2020 unter <https://de.pons.com/%C3%BCbersetzung/deutsch-englisch/%C3%BCber?bidir=1#dict>
- PONS. (o. D. c). *über*. In *PONS Deutsch-Griechisch-Online-Wörterbuch*. Verfügbar 17. Mai 2020 unter <https://de.pons.com/%C3%BCbersetzung/deutsch-griechisch/%C3%BCber>
- PONS. (o. D. d). $\gamma\iota\alpha$. In *PONS Griechisch-Deutsch-Online-Wörterbuch*. Verfügbar 17. Mai 2020 unter <https://de.pons.com/%C3%BCbersetzung/griechisch-deutsch/%CE%B3%CE%B9%CE%B1>
- Prange, J. & Schneider, N. (2018). Preposition Supersenses in German–English Parallel Data. http://prange.jakob.georgetown.domains/p/2018_masc/preposition-supersenses-german-new.pdf
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Longman.
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>

- Rankin, T. & Schiffner, B. (2011). Marginal prepositions in learner English. *International Journal of Corpus Linguistics*, 16(3), 412–434. <https://doi.org/10.1075/ijcl.16.3.07ran>
- Reinhart, T. (2001). Experiencing Derivations. *Proceedings of SALT*, 11, 365–387. <https://doi.org/10.3765/salt.v11i0.2845>
- Reinhart, T. (2002). The Theta System – an overview. *Theoretical Linguistics*, 28(3), 229–290. <https://doi.org/10.1515/thli.28.3.229>
- Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., Baker, C. F. & Scheffczyk, J. (2016). *FrameNet II: Extended Theory and Practice*. <https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf>
- Saint-Dizier, P. (2006). Introduction to the Syntax and Semantics of Prepositions. In P. Saint-Dizier (Hrsg.), *Syntax and Semantics of Prepositions* (S. 1–25). Springer.
- Schmied, J. (1998). Differences and similarities of close cognates: English *with* and German *mit*. In S. Johansson & S. Oksefjell (Hrsg.), *Corpora and Cross-linguistic Research: Theory, Method, and Case Studies* (S. 255–275). Rodopi.
- Schmitz, W. (1981). *Der Gebrauch der deutschen Präpositionen* (9. Aufl.). Hueber.
- Schröder, J. (1986). *Lexikon deutscher Präpositionen*. VEB Verlag Enzyklopädie.
- Sheskin, D. J. (2011). *Handbook of Parametric and Nonparametric Statistical Procedures*. Chapman and Hall/CRC Press.
- Škrabal, M. & Vavřín, M. (2017). The Translation Equivalents Database (Treq) as a Lexicographer’s Aid. In I. Kosem, C. Tiberius, M. Jakubíček, J. Kallas, S. Krek & V. Baisa (Hrsg.), *Electronic lexicography in the 21st century: Proceedings of eLex 2017 conference* (S. 124–137). Lexical Computing CZ s.r.o. <https://elex.link/elex2017/proceedings-download/>
- Temme, A. (2014). German psych-adjectives. In A. Machicao y Priemer, A. Nolsa & A. Sioupi (Hrsg.), *Zwischen Kern und Peripherie: Untersuchungen zu Randbereichen in Sprache und Grammatik* (S. 131–156). De Gruyter. <https://doi.org/10.1524/9783050065335>
- Temme, A. & Verhoeven, E. (2017). Backward binding as a psych effect: A binding illusion? *Zeitschrift für Sprachwissenschaft*, 36(2), 279–308. <https://doi.org/10.1515/zfs-2017-0011>
- Tyler, A. & Evans, V. (2001). Reconsidering Prepositional Polysemy Networks: The Case of *over*. *Language*, 77(4), 724–765. <https://doi.org/10.1353/lan.2001.0250>
- Ursini, F. & Giannella, A. (2016). On the polysemy of Spanish spatial Ps. *Borealis: An International Journal of Hispanic Linguistics*, 5(2), 253–312. <https://doi.org/10.7557/1.5.2.3633>
- Varga, D., Halácsy, P., Kornai, A., Nagy, V., Németh, L. & Trón, V. (2007). Parallel corpora for medium density languages. In N. Nicolov, K. Bontcheva, G. Angelova & R. Mitkov (Hrsg.), *Recent Advantages in Natural Language Processing IV: Selected Papers from RANLP 2005* (S. 247–258). John Benjamins. <https://doi.org/10.1075/cilt.292>
- Vavřín, M. & Rosen, A. (2015). *Treq*. Verfügbar 14. Mai 2020 unter <https://treq.korpus.cz/>
- Volk, M. (2006). German prepositions and their kin: A survey with respect to the resolution of PP attachment ambiguities. In P. Saint-Dizier (Hrsg.), *Syntax and Semantics of Prepositions* (S. 83–100). Springer.
- Weinrich, H. (2005). *Textgrammatik der deutschen Sprache*. Georg Olms.
- Wickham, H. & Bryan, J. (2019). *readxl: Read Excel Files* [R package version 1.3.1]. <https://CRAN.R-project.org/package=readxl>

Zwicky, A. M. & Sadock, J. M. (1975). Ambiguity tests and how to fail them. In J. P. Kimball (Hrsg.), *Syntax and Semantics: Volume 4* (S. 1–36). Academic Press.

['stʌdi:z 'ɪn lɪŋ'ɡwɪstɪks ənd
lɪŋ'ɡwɪstɪk 'deɪtə 'saɪəns]

**Studies in Linguistics and
Linguistic Data Science**

Volume 3

**Simon Masloch. Über über und
dessen Entsprechungen im
Englischen**

Editor: **Tibor Kiss**

ISSN: 2700-8975