

Stereotype in Webkorpora: Strategien zur Suche in sehr großen Datenmengen

Uwe Quasthoff (Leipzig) und Erla Hallsteinsdóttir (Odense)

Abstract

Research on stereotypes (as well as on similar constructs such as prejudice, image, identity, etc.) goes back around a century. Most research on stereotypes belongs to the field of psychological, social and cultural studies, with a strong focus on cognitive and socio-cultural phenomena. Since Uta Quasthoff's pioneering analysis in 1973, there have been several linguistic studies of stereotypes, i. e. on linguistic aspects of how stereotypes are realized in language use. Linguistic research on stereotypes uses methods from disciplines such as sociolinguistics, ethnography, discourse and conversation analysis. Discursive psychology and cognitive linguistics and single studies on stereotypes have used corpus-linguistic methods to explore stereotypes in language use. Nevertheless, we state that no large-scale empirical studies have yet investigated the linguistic realization of stereotypes by using empirical data from large text corpora. In this article, we explore the potential of corpus linguistic approaches in the research on the occurrences of stereotypes in written language. By using an analysis of co-occurrences, we aim to describe the relation between associative semantic stereotypes and co-occurrences in a corpus. Our hypothesis is that the linguistic construction of stereotypes as their realization in language use reproduces stereotypic representations of cognitive conceptualizations and meaning patterns, thus, stereotypic meanings should be detectable in statistical patterns in large corpora.

1 Einleitung

Während Arbeiten zu soziokulturellen und kognitiven Aspekten von Stereotypen in der Forschungsliteratur gut vertreten sind, gibt es nur einzelne Studien zur Realisierung von Stereotypen im Sprachgebrauch. Wir wissen zwar einiges darüber, welche Stereotype sich auf welche Art und Weise in der Sprache manifestieren **können** (cf. z. B. Pümpel-Mader 2010), jedoch kaum etwas darüber, in welchem Umfang Stereotype in welcher sprachlichen Form tatsächlich produziert werden. Stereotype werden über Sprache tradiert und die sprachliche Konstruktion von Stereotypen, d. h. ihre Manifestation im Sprachgebrauch trägt – so unsere Hypothese – maßgeblich dazu bei, dass stereotype Generalisierungen als kognitive Konzeptualisierungen und Bedeutungsmuster in einer Kultur wiedergegeben und von den Sprechern reproduziert werden können.

Sprachwissenschaftliche Untersuchungen zu Stereotypen können übergeordnet zwei Traditionen zugeordnet werden, die sich beide unterschiedlicher Methoden aus Disziplinen wie der Soziolinguistik, der Diskurs- und Konversationsanalyse, der Ethnographie, der kognitiven Linguistik und Korpuslinguistik bedienen (cf. z. B. in Harrington et al. 2008). Die erste Tradition befasst sich mit Stereotypen im Sprachgebrauch bestimmter Gruppen von Menschen. Hier wird der Fokus häufig auf gruppenspezifische Unterschiede im Sprachgebrauch dieser Gruppen gelegt (cf. z. B. Farris/Davis/Compton 2014; Hellinger/Bußmann 2003; Lakoff 1975; Motschenbacher 2010; Kendall/Tannen 1997).

Die zweite Tradition fokussiert darauf, wie Stereotype im Sprachgebrauch realisiert werden. Seit der wegweisenden Arbeit von Uta Quasthoff (1973) sind einzelne Arbeiten (cf. z. B. Grzybek 1990; Heinemann 1998; Kashima/Fiedler/Freytag 2008) erschienen, u. a. zu Themen wie Stereotype in Wörterbüchern (cf. Heier 2016; Nübling 2008; Prechter 1999), Altersstereotype in Zeitungen (cf. Wien 2005), sprachliche Realisierung von Geschlechterstereotypen in Kontaktanzeigen (cf. Gottburgsen 1995), Form und Funktion von Personenstereotypen (cf. Pümpel-Mader 2010), Frauenstereotype in literarischen Werken (cf. Stocker 2005), nationale Stereotype, das Bild einer Nation (cf. die Beiträge in Süßmuth 1996) in ausgewählten Texten oder Medien (cf. Agger/Bentikow/Hedetoft 1990; Dąbrowska 1999; Demleitner 2009; Langer 2003; Müller 2005), Stereotype in der gesprochenen Sprache (cf. Wenzel 1978) und mehr (cf. Bewanger/Kosta 2005; Desmarais/Bruce 2009).

In einzelnen Studien zu Stereotypen sind korpuslinguistische Methoden eingesetzt worden (cf. Harrington et al. 2008). Es gibt aber zur sprachlichen Realisierung von Stereotypen bisher keine groß angelegten empirischen Untersuchungen, die Stereotype mit qualitativen und quantitativen Analysen von Daten in sehr großen Korpora untersucht haben. In diesem Beitrag diskutieren wir anhand von Daten aus den Leipziger Wortschatzkorpora, welche Strategien zur Suche von nationalen Stereotypen in großen Mengen an Sprachdaten angewendet werden können. Wir zeigen Beispiele dafür, wie wir Kookkurrenzanalysen bei der Auswertung solcher Daten eingesetzt haben. Abschließend problematisieren wir am Beispiel deutsch-dänischer Nationalstereotype die Verwendung von Sprachdaten aus Korpora der geschriebenen Sprache in der Erforschung der sprachlichen Manifestation von Stereotypen.

2 Sprachliche Stereotype

Stereotype definieren wir im SMiK-Projekt¹ als generalisierende Denk- und Handlungsmuster, die unsere Wahrnehmung der Welt, Kategorisierungen und Auffassungen von uns selbst und anderen prägen. Es handelt sich hierbei um eine sehr weite Definition, die u. U. auch Phänomene einschließt, die in der Forschung mit unterschiedlichen Termini bezeichnet werden, wie z. B. Nationalimages, Vorurteile etc.²

¹ Das Projekt *Nationale Stereotype und Marketingstrategien in der interkulturellen deutsch-dänischen Kommunikation* (SMiK; www.stereotypenprojekt.eu [13.02.2016]) wurde gefördert durch INTERREG4A Syddanmark-Schleswig-K.E.R.N. mit Mitteln des Europäischen Fonds für regionale Entwicklung; www.interreg4a.de/wm390752 [13.02.2016].

² Eine ausführliche terminologische Diskussion würde den Rahmen dieses Beitrages sprengen, stellvertretend sei daher auf die Beiträge von z. B. Allen, Langer, Nielsen und Ossenbaur/Baur in diesem Heft verwiesen.

Die Phänomene, die in Nationalstereotypen prädiert werden, so wie sie im SMiK-Projekt erfasst wurden, lassen sich u. a. mit den Fragestellungen der Fragebogenerhebung zu *typisch deutsch – typisch dänisch* folgendermaßen eingrenzen (cf. ausführlich in Hallsteinsdóttir 2015; Hofmann/Hallsteinsdóttir in diesem Heft):

Mit Frage 1 *Schreiben Sie bitte die ersten drei Wörter auf, die Ihnen zu Dänemark/Deutschland einfallen* werden Assoziationen (assoziatives Wissen) und Phänomene, die dem jeweiligen anderen Land (intuitiv) zugeordnet werden, so registriert, wie sie traditionell in der Erfassung nationaler Stereotype erhoben werden (cf. z. B. Bolten 2006).

Die Fragen 2 a) *Was kennzeichnet eine typische Dänin/Deutsche?* und b) *Was kennzeichnet einen typischen Dänen/Deutschen?* erfassen Eigenschaften, die den Menschen aus dem jeweils anderen Land zugeschrieben werden (cf. zur Eigenschaftenzuordnung als Methode zur Erfassung von nationalen Stereotypen auch Ossenbaur/Baur in diesem Heft).

Mit den Fragen 3 a) *Was ist für Sie typisch dänisch/deutsch? – (eher) positiv* und b) *Was ist für Sie typisch dänisch/deutsch? – (eher) negativ* werden die subjektiv mit einer Nationalitätenbezeichnung als *typisch* verknüpften Phänomene erfasst und zugleich deren positive oder negative Zuordnungen. Hier werden sowohl Eigenschaften (das Typische) aber auch Phänomene (als Träger von typischen Eigenschaften) registriert.

Die vierte und letzte Frage bezieht sich auf das spezifische Feld der verkaufsfördernden und somit von den Probanden als positiv bewerteten Eigenschaften, die in der Vermarktung von Produkten aus dem jeweils anderen Land eingesetzt werden könnten:

Eine große dänische/deutsche Firma bittet Sie, eine Werbekampagne in Deutschland/Dänemark zu machen. Welche typischen dänischen/deutschen Eigenschaften würden Sie in einer solchen Kampagne einsetzen, um ein dänisches/deutsches Produkt in Deutschland/Dänemark zu verkaufen? Sie können die Frage gerne mit der Angabe eines konkreten Produktes beantworten.

Diese Fragen beziehen sich primär auf Stereotypisierungen zum Land, zu den Menschen und zu mit ihnen verknüpften Eigenschaften des jeweils anderen Landes. Die Art der Fragestellung impliziert jedoch auch die Angabe anderer Phänomene (Antworten zur Frage 1 und Frage 3), die potenziell als Gegenstand von Stereotypisierungen fungieren.

Zunächst sind alle in den Antworten der Fragebogenerhebung³ kollektiv vorkommenden Stereotype übergeordnet als kulturelle bzw. kulturell bedingte Stereotype einzuordnen. Als sprachliche Stereotype definieren wir diejenigen davon bzw. andere, in der Fragebogenerhebung ggf. nicht erfassten Stereotype, die sich zudem in der Sprache als assoziative semantische Stereotype manifestieren und im Sprachgebrauch realisiert werden (können).

Die meisten Definitionen von sprachlichen Stereotypen folgen der psychologischen und sozialwissenschaftlichen Stereotypenforschung. Darin werden Stereotype über ihren Inhalt als

³ Da die Ergebnisse der Fragebogenerhebung in Hallsteinsdóttir (2015) und Hofmann/Hallsteinsdóttir (in diesem Heft) umfassend beschrieben werden, verzichten wir hier auf eine ausführlichere Darstellung.

semantische oder kognitive Einheiten sehr offen definiert (cf. Dąbrowska 1999: 85; Wenzel 1978: 97)⁴.

Aus einer sprachwissenschaftlichen Perspektive betrachtet, beziehen sich die Ergebnisse aus psychologischen, psycholinguistischen und soziokulturell orientierten Untersuchungen zu Stereotypen auf die kognitive Leistung von Sprache, d. h. es handelt sich um Aspekte der soziokulturellen und sprachlichen Kompetenz der Sprecher als ein Teil ihrer kulturellen Wissensvorräte (cf. Bolten 2007). Daher haben auch Resultate aus diesen Untersuchungen durchaus eine große Relevanz für die sprachwissenschaftliche Stereotypenforschung, insbesondere wenn es darum geht, wie Stereotype wahrgenommen und verstanden werden. Sie ermöglichen jedoch keine verlässlichen Aussagen über das Vorkommen, die sprachliche Konstruktion oder die Funktion von Stereotypen im Sprachgebrauch.

Bei den sprachlichen Stereotypen wird eine Konventionalität in der Zuordnung von Ausdruck und Inhalt vorausgesetzt. Wir beziehen uns hier auf den Begriff des *assoziativen semantischen Stereotyps* von Kilian (2001, 2003, 2015; cf. auch Putnam 1975: 169, 1985 und ausführlich in Hallsteinsdóttir 2016). Assoziative semantische Stereotype sind **implizite**, im Sprachzeichen verankerte, konventionalisierte Bestandteile der Wortbedeutung und sie gehören somit zum Bedeutungspotenzial eines Wortes. Man kann sie sich in etwa so vorstellen (cf. ausführlich in Kilian 2015):

Das Sprachzeichen *tysk* (Deutsch) hat als Substantiv eine aus Lauten [ˈtʏsg]⁵ bzw. Buchstaben <tysk> bestehende Formseite und eine Inhaltseite, mit der konventionell die Kernbedeutungen (1) „eine westgermanische Sprache, die als offizielle Sprache in Deutschland, Österreich und der Schweiz gesprochen wird“ und (2) „ein Unterrichtsfach, das deutsche Sprache, Literatur u. v. m. umfasst“⁶ verbunden wird. Diese Bedeutungen werden u. a. durch die Merkmale (1) „Sprache“, „westgermanisch“, „Deutschland, Österreich, Schweiz“ und (2) „Unterrichtsfach“ spezifiziert. Zusätzlich zu diesen Bedeutungsmerkmalen gehören assoziative semantische Stereotype zum Bedeutungspotenzial von *tysk*.

Assoziative semantische Stereotype sind laut Kilian (2003: 166) generalisierende Vorstellungen, die im Laufe der Zeit von den Sprechern kollektiv mit dem Wort assoziiert bzw. bereits als Teil der Wortbedeutung lexikalisiert worden sind. Aus den Fragebogenuntersuchungen im SMiK-Projekt (ausführlich in Hallsteinsdóttir 2015; cf. auch Hallsteinsdóttir 2016; Hofmann/Hallsteinsdóttir in diesem Heft; Müller 2016; Müller/Hallsteinsdóttir 2016) wissen wir, dass in den dänischen Antworten auf die Frage *Was ist für Sie typisch deutsch – (eher) negativ?* häufig *die (deutsche) Sprache* genannt wird. Ob dies grundsätzlich für die deutsche Sprache und/oder für die Sprache als Unterrichtsfach gilt, kann anhand der Daten nicht klar unterschieden werden. Zu *Sprache* werden in den Antworten zusätzlich u. a. die Eigenschaften „hässlich“, „schwer“, „hart“, angeführt. Die kollektive Nennung dieser bewertenden Vorstellungen über Eigenschaften einer Sprache (*tysk*), die dem Phänomen (der Sprache) selbst nicht

⁴ Der Stereotypenbegriff aus der Phraseologie bezieht sich dagegen auf bestimmte Arten von lexikalisierten Mehrworteinheiten (cf. z. B. Fleischer 1997: 247; Grzybek 1990).

⁵ Dänische Aussprache cf. Den Danske Ordbog online (<http://ordnet.dk/ddo/ordbog?query=tysk&select=tysk,1>) [28.12.2015].

⁶ Übersetzung der lexikographischen Erklärungen aus: Den Danske Ordbog online: www.ordnet.dk [25.09.2015]; übers. durch die Verf.

entnommen worden sein können, ist ein Indikator dafür, dass sie stereotypisierend mit dem Wort *tysk* assoziiert werden. Das Vorkommen in den Antworten auf die Frage nach dem typisch Negativen lässt zudem darauf schließen, dass mit dem Wort *tysk* zudem konventionell eine stereotypisierende Negativbewertung assoziiert wird.

Assoziative semantische Stereotype gehören zum Bedeutungspotenzial von Sprachzeichen, sie sind somit prinzipiell implizit, sie können aber im Sprachgebrauch expliziert werden. Ob diese sprachliche Realisierung von Stereotypen implizit oder explizit erfolgt, ist jedoch nicht vorhersagbar. Die Implizitheit bzw. Explizitheit bezieht sich in der linguistischen Stereotypenforschung⁷ (cf. Dąbrowska 1999; Quasthoff 1973; Wenzel 1978) auf die Art der Versprachlichung eines Stereotyps (cf. Dąbrowska 1999: 90). Implizite Stereotype sind solche, die „nicht explizit geäußert werden, sondern implizit in den Äußerungen enthalten sind“ (Wenzel 1978: 63) und über die thematische und textuelle Struktur erschlossen werden. Ob implizite Stereotype realisiert werden, hängt davon ab, ob der Leser/Hörer das Bedeutungspotenzial der assoziativen semantischen Stereotype (er-)kennt und es im jeweils aktuellen Kontext aktiviert. Explizite Stereotype werden durch ihre eindeutige Verbalisierung in einem konkreten Kontext manifestiert und u. a. durch Prädikationen wie *Deutsche sind höflich, eine typische Dänin ist blond* sprachlich markiert (cf. Kilian in diesem Heft). Explizite Stereotype sind ein objektiv erkennbarer Bestandteil der Textbedeutung im Sprachgebrauch, während implizite Stereotype ein Teil des sprachsystematischen Bedeutungspotenzials eines sprachlichen Zeichens sind, dessen Aktivierung beim Textverstehen dem Leser/Hörer überlassen wird.

Diese Ausführungen sollen zeigen, dass die Erfassung von sprachlichen Stereotypen mit gewissen methodologischen Herausforderungen verbunden ist. Dies ist keine neue Erkenntnis, denn bereits Wenzel (1978: 17) weist auf die Illusion einer objektiven Analyse hin: Die Untersuchung von Stereotypen beruhe, so Wenzel, „auf dem Verstehen von Sinnzusammenhängen“. Den Untersuchenden selbst mit ihren Erfahrungen – Lebenswelten (cf. Bolten 2007) und Weltansichten (cf. Kilian 2001, 2003) – kommt daher eine aktive Rolle bei der Identifizierung von Stereotypen im Sprachgebrauch zu. Dies beeinflusst auch im SMiK-Projekt die wissenschaftlichen Analysen dahingehend, dass zwar keine Objektivität, sehr wohl aber intersubjektive Überprüfbarkeit der Befunde möglich ist.

Wie oben beschrieben, definieren wir im SMiK-Projekt Stereotype als durch Sprache und Kultur vorgegebene Muster. Wir schließen uns mit dieser Definition der Auffassung an, dass Musterhaftigkeit ein zentrales Phänomen von Sprache ist, die „das Wesen der Sprachproduktion und -rezeption auf allen Ebenen ausmacht“ (Steyer 2013: 9). Dass sich die Musterhaftigkeit in Form von typischen Mustern des Sprachgebrauchs als wiederkehrende sprachliche Verwendungsweisen in Korpora (cf. Steyer 2013: 9) erfassen und analysieren lassen, hat Steyer (2013) am Beispiel von festen Wortverbindungen zeigt.

Im Folgenden werden wir anhand von Korpusdaten zu Wortformen mit {dän} und {tysk} (deutsch) als Komponenten untersuchen, inwieweit sich sprachliche Stereotype in Form von

⁷ Hier gibt es einen begrifflichen Unterschied zur soziologischen und psychologischen Stereotypenforschung, in der implizite Stereotype als unbewusstes Wissen und explizite Stereotype als bewusste Stereotypisierung verstanden werden. Beide Arten sind Bestandteile der sprachlich-kulturellen Kompetenz der Sprecher.

Mustern im Sprachgebrauch in großen Korpora nachweisen lassen. Wir konzentrieren uns dabei auf die Erfassung und Beschreibung der sprachlichen Heterostereotype, die Deutsche über Dänen und Dänen über Deutsche haben.

3 Suche nach Stereotypen im Korpus

Bei der Suche nach Stereotypen in Korpora geschriebener Sprache befinden wir uns in der typischen Situation des Information-Retrieval, wie wir es aus der Suche im Web kennen:

- Wir suchen nach Belegen (hier: für Stereotype). Insgesamt bilden diese Belege die sog. Treffermenge.
- Die Treffermenge ist typischerweise klein (1.000 bis 10.000 Belege).
- Unsere Korpora sind hingegen so groß (10 bis 100 Millionen Belege), so dass eine manuelle Suche ausscheidet und Algorithmen die Suche steuern müssen.

Eine automatische Suche liefert eine Ergebnismenge, die nur im Idealfall mit der Treffermenge übereinstimmt. Erstrebt werden:

- ein hoher Recall, d. h. ein möglichst großer Teil der Treffermenge wird gefunden, und
- eine akzeptable Präzision, d. h. der Anteil der Nicht-Treffer in der Ergebnismenge soll möglichst gering sein.

Allerdings sind Recall und Präzision eine Art Gegenpole. Ein Hoher Recall sorgt meist für geringe Präzision und umgekehrt. Als praktikablen Ansatz für linguistische Probleme wählt man häufig einen höchstmöglichen Recall, solange die Präzision im unteren Prozent-Bereich bleibt. Dann kann man davon ausgehen, dass fast alle Treffer auch gefunden wurden. Und durch manuelle Inspektion lassen sich Nicht-Treffer mit vertretbarem Aufwand aus der Ergebnismenge entfernen.

Im SMiK-Projekt wurden die Korpora der Leipzig Corpora Collection (LCC s. a., cf. Goldhahn/Eckart/Quasthoff 2012) für die Suche nach Stereotypen verwendet. Das deutschsprachige Wortschatz-Korpus enthielt zum Zeitpunkt der Suche ca. 260 Millionen Sätze (cf. Quasthoff/Fiedler/Hallsteinsdóttir 2011) und das dänischsprachige Korpus rund 27 Millionen Sätze. Die darin verwendeten Texte sind Zeitungstexte, Wikipedia-Texte sowie zufällig ausgewählte Webseiten aus Deutschland und Dänemark. Aus diesen wurde zunächst der reine Text extrahiert, dieser anschließend in Sätze zerlegt. Diese Sätze wurden auf das Vorhandensein von sprachlichen Zeichen und musterbasiert auf Wohlgeformtheit überprüft⁸. Außerdem werden Satzdoubletten entfernt (cf. ausführlich zum Bearbeitungsverfahren u. a. Biemann et al. 2013). Anschließend werden Wortkookkurrenzen mit dem Log-Likelihood-Maß berechnet. Wortkookkurrenzen sind Paare von Wörtern, die signifikant häufig gemeinsam auftreten, und zwar als unmittelbare Nachbarn (sog. Nachbarschaftskookkurrenzen) oder gemeinsam im Satz (sog. Satz-kookkurrenzen). Die Signifikanz solcher Kookkurrenzen lässt sich als Assoziationsstärke zwischen den entsprechenden Wörtern interpretieren (cf. Quasthoff 2009: 152 sowie ausführlich zur Kookkurrenzberechnung Heyer/Quasthoff/Wittig 2006: 134–149). Neben reinen Stichwortlisten bieten sich Wortkookkurrenzen als Kandidaten bei der Suche nach Ste-

⁸ Um z. B. Reste von Programmiersprachen, URLs und anderes nicht-sprachliches Material zu entfernen.

reotypen an. Solche Wortpaare beschreiben oft einen semantischen Zusammenhang, der stereotyper Natur sein kann, aber natürlich nicht sein muss. Deswegen ist es wichtig, die gemeinsam auftretenden Wörter in eine solche Reihenfolge zu bringen, welche die interessanten Wortpaare möglichst bevorzugt. Dieser Forderung recht nahe kommt das oben erwähnte Log-Likelihood-Maß (cf. Dunning 1993), welches vom mathematischen Gesichtspunkt aus misst, wie sehr das wiederholte gemeinsame Auftreten zweier Wörter davon entfernt ist, nur zufällig zu sein. Interessanter ist jedoch die psycholinguistische Eigenschaft (cf. Rapp 2002), dass das Log-Likelihood-Maß sich ähnlich wie die menschliche Assoziationsstärke verhält: Die statistisch ermittelten Satzkookkurrenzen zu einem gegebenen Wort entsprechen recht gut den Ergebnissen aus Stimulus-Response-Experimenten mit dem entsprechenden Wort als Stimulus. Wenn also mit einem Wort starke Stereotype verbunden sind, sollten diese sich auch in der Menge der Kookkurrenzen zu diesem Wort wiederfinden.

Für die Untersuchung der sprachlichen Manifestation von deutsch-dänischen Heterostereotypen wurden aus dem deutschen und dem dänischen Wortschatz-Korpus zwei SMiK-Teilkorpora extrahiert. Diese bestehen aus Sätzen, die Wörter mit der Buchstabenkombination *tysk* (DÄ-SMiK 2013) und *dän* (D-SMiK 2013) beinhalten. Das unbereinigte dänische *tysk*-SMiK-Korpus enthält 176.327 Sätze (0,65 % des Gesamtkorpus) und das unbereinigte deutschsprachige *dän*-SMiK-Korpus 218.264 Sätze (0,08 % des Gesamtkorpus). Hier lässt sich bereits ein Unterschied in der Frequenz feststellen: es gibt im dänischen Korpus verhältnismäßig deutlich mehr Sätze mit *tysk*-Wörtern als *dän*-Wörter im deutschen Korpus. Dies lässt darauf schließen, dass **deutsche** Themen häufiger in dänische Medien vorkommen als **dänische** Themen in den deutschen Medien.

Die Relation zwischen der Anzahl von Wortformen und ihrer absoluten Häufigkeit ähnelt sich hingegen in den beiden Sprachen. Tabelle 1 zeigt die Verteilung der Belege:

	Dänisches *tysk*-Korpus				Deutsches *dän*-Korpus			
	Anzahl Wortformen	Prozentualer Anteil der Wortformen	Anzahl Belege	Prozentualer Anteil der Belege	Anzahl Wortformen	Prozentualer Anteil der Wortformen	Anzahl Belege	Prozentualer Anteil der Belege
1 Beleg	1.479	63,15	1.479	0,82	1.798	69,23	1.798	0,82
2–9 Belege	654	27,92	2.276	1,29	633	24,37	1.476	0,96
10–99 Belege	170	7,26	5.004	2,84	133	5,12	3.490	1,6
Über 100 Belege	39	1,67	167.568	95,03	33	1,27	210.882	96,62
Insgesamt	2.342	100	176.327	100	2.597	100	218.264	100

Tabelle 1: Verteilung der Belege in den SMiK-Korpora

Die Suche im deutschen Korpus ergab insgesamt 2.597 Wortformen, die die Buchstabenkombination *dän* enthalten. Davon kommen 1.796 nur einmal und 309 nur zweimal vor. Nur 166 Wortformen weisen zehn oder mehr Belege auf. Insgesamt 33 Wortformen davon haben

eine Frequenz von 100 oder mehr. Diese 33 Wortformen finden sich jedoch in fast 97 % der Belege wieder. Es handelt sich dabei um folgende 23 echte {dän}-Wortformen mit jeweils mehr als 100 Belegen:

Dänemark (85443), *dänischen* (39826), *dänische* (27136), *Dänen* (18732), *Däne* (9296), *Dänemarks* (6512), *dänischer* (3786), *Dänische* (2778), *Dänischen* (2245), *Dänin* (2084), *Dänisch* (1509), *dänisch* (1077), *deutsch-dänischen* (902), *dänisches* (824), *dänischem* (521), *Däninnen* (494), *deutsch-dänische* (459), *Dänischer* (407), *Dänisches* (253), *Dänenkönig* (218) *Deutsch-Dänischen* (204), *Dänemark-Spiel* (129), *Süddänemark* (106).

Zusätzlich sind folgende 10 **unechte** *dän*-Wortformen bzw. Eigennamen mit dem Morphem {dän} unter den Wortformen zu finden, die mehr als 100 Belege aufweisen:

mondänen (2036), *Däniken* (1239), *mondäne* (1195), *Dänischenhagen* (390), *mondäner* (310), *Dänischburg* (223), *mondänes* (187), *Dänikens* (134), *Dänenweg* (127), *Mondäne* (100).

Zu diesen 10 *dän*-Wortformen gehören 5.941 Belege. Diese *dän*-Belege machen jedoch nur 2,72 % des dänischen SMiK-Teilkorpus aus, so dass wir davon ausgehen, dass sie die Kookkurrenzberechnung nicht beeinflussen.

Insgesamt wurden im dänischen Korpus 2.342 unterschiedliche Wortformen mit der Buchstabenkombination *tysk* extrahiert. Davon kamen 1.479 Wortformen nur einmal und 314 nur zweimal vor. Insgesamt 209 Wortformen haben 10 Belege oder mehr. Folgende 39 Wortformen, die alle eindeutig *tysk*-Wörter sind, kommen mehr als 100 Mal vor:

tyske (,deutsche‘, 55025), *Tyskland* (,Deutschland‘, 53929), *tysk* (,deutsch‘, 23040), *tyskerne* (,die Deutschen‘, 7916), *Tysklands* (,Deutschlands‘, 4023), *tyskere* (,Deutsche‘, 2760), *Tyske* (,Deutsche‘, 2513), *Tysk* (,Deutsch‘, 2480), *Tyskerne* (,die Deutschen‘, 1675), *tyskeren* (,der Deutsche‘, 1540), *tysker* (,Deutscher‘, 1336), *Nordtyskland* (,Norddeutschland‘ 1141), *tyskernes* (,der Deutschen‘, 1139), *dansk-tyske* (,deutsch-dänische‘, 1113), *Vesttyskland* (,Westdeutschland‘, 1009), *Østtyskland* (,Ostdeutschland‘, 717), *nordtyske* (,norddeutsche‘, 612), *østtyske* (,ostdeutsche‘, 603), *tyskland* (,Deutschland‘, 521), *Tyskeren* (,der Deutsche‘, 440), *tysktalende* (,deutschsprechende‘, 412), *Sydtyskland* (,Süddeutschland‘, 377), *Nazityskland* (,Nazideutschland‘, 337), *vesttyske* (,westdeutsche‘, 323), *sydtyske* (,süddeutsche‘, 298), *tyskromerske* (,deutsch-römische‘, 258), *tysksprogede* (,deutschsprachige‘, 254), *Nazi-Tyskland* (,Nazi-Deutschland‘, 202), *dansk-tysk* (,dänisch-deutsch‘, 191), *dansk/tyske* (,dänisch/deutsche‘, 178), *fransk-tyske* (,französisch-deutsche‘, 155), *Tyskernes* (,der Deutschen‘, 149), *platttysk* (,plattdeutsch‘, 144), *Tyskere* (,Deutsche‘, 140), *tyskerens* (,des Deutschen‘, 139), *østtysk* (,ostdeutsch‘, 133), *tyskfødte* (,deutschgeborene‘, 120), *tyskundervisningen* (,Deutschunterricht‘, 116), *østtyskere* (,Ostdeutsche‘, 111).

Die Auswertung der Ausgangswörter zeigt, dass die morphembasierte Extrahierung in beiden Sprachen eindeutige Ergebnisse liefert. Die Teilkorpora enthalten nur sehr wenige Belege mit Wörtern, die die Buchstabenkombination *dän* bzw. *tysk* enthalten, die sich nicht den Morphemen {dän} und {tysk} zuordnen lassen, z. B. *partyskjorten* ‚das Festhemd‘ (1), *partysko* ‚Festschuh(e)‘ (4). Wir stellen zusammenfassend fest, dass wir mit der Suche nach den Buchstabenkombination *dän* und *tysk* sowohl einen hohen Recall an Wortformen mit

den entsprechenden Morphemen erreicht haben, als auch, dass der Anteil der Nicht-Treffer in den Daten sehr niedrig ist und demnach eine akzeptable Präzision der Suche gegeben ist.

4 Kookkurrenzanalysen

Eine der Arbeitshypothesen im SMiK-Projekt, die als ein Ausgangspunkt für die Suche nach Stereotypen in großen Korpora diene, beinhaltet, dass explizite sprachliche Stereotype im Korpus über Kookkurrenzen sichtbar werden, d. h. dass Kookkurrenzen als Stereotypenindikator fungieren können (cf. Hallsteinsdóttir 2016). Um diese Hypothese verifizieren zu können, wurden sowohl wortformbasierte Kookkurrenzen (z. B. zu Wortformen wie *Dänen*, *Dänemark*, *dänische*, *tysker*, *Tyskland*, *tyske*) als auch morphembasierte Kookkurrenzen zu den Morphemen {dän} und {tysk} berechnet.

4.1 Wortformenbasierte Kookkurrenzen

Bei den wortformbasierten Untersuchungen zu Kookkurrenzen haben wir sowohl Satz- als auch Nachbarschaftskookkurrenzen verwendet. Die Auswertung von Satzkookkurrenzen entspricht etwa der folgenden Fragestellung: Welche Wörter sind am auffälligsten in deutschen Sätzen, in denen Wortformen der Ausgangswörter, die das Morphem {dän} (z. B. *Dänemark* oder *dänisch*) beinhalten, vorkommen? Analog für Dänisch gilt die Fragestellung: Welche Wörter sind am auffälligsten in dänischen Sätzen, die ein Ausgangswort mit dem Morphem {tysk} (z. B. *Tyskland* oder *tysk*) enthalten? Hier liegen bei den Kookkurrenzberechnungen zunächst also keinerlei grammatische oder lexikalische Selektionsbeschränkungen vor, wie wir sie bei der Berechnung der unmittelbaren linken und rechten Nachbarn vorfinden. Wir können als Satzkookkurrenzen daher Wörter verschiedener Wortarten ebenso erwarten wie Wörter, die in verschiedenen semantischen Relationen zu den Ausgangswörtern stehen.

Im Falle von Nachbarschaftskookkurrenzen ist die Situation etwas eingeschränkter, denn hier beeinflussen u. a. syntaktische Kombinationsmöglichkeiten die Auswahl der möglichen Kookkurrenzpartner. In beiden Sprachen Dänisch und Deutsch werden z. B. Nominalphrasen der Form ADJ-NOM gebildet. Die Suche nach typischen rechten Nachbarn zu den Adjektiven *dänisch** (im Deutschen) bzw. *tysk** (im Dänischen) können also Substantive sein, welche die Länderbezeichnung als typische Eigenschaft tragen. Umgekehrt gehen wir davon aus, dass als linke Nachbarn der Länder- und Nationalitätsbezeichnungen *Dänemark** und *Däne*/Dänin** bzw. *Tyskland** (‘Deutschland’) und *tysker** (‘Deutsche(r)’) Adjektive stehen, die Land und Leute genauer spezifizieren. Aufgrund unterschiedlicher Flexionsformen, deren Existenz hier mit einem Stern angedeutet wird, müssten für eine genaue Berechnung der signifikanten Kookkurrenzen zu dem jeweiligen Wort die Wortformen zusammengeführt und die Kookkurrenzen nicht wortformen- sondern wortbasiert auf der Basis aller flektierten Formen eines Wortes berechnet werden. Flexion und somit unterschiedliche Flexionsformen sind im Deutschen und im Dänischen zwar unterschiedlich ausgeprägt, aber in beiden Sprachen für die lexikalischen Wortarten gegeben. Die Zusammenführung von Wortformen zu einem Lemma stellt daher für beide Sprachen eine Herausforderung dar. Es stellt sich zudem bei der manuellen Analyse der einzelnen *dän*- und *tysk*-Wortformen heraus, dass unterschiedliche Flexionsformen eines Lemma durchaus unterschiedliche Selektionspräferenzen haben. Wir illust-

rieren diese Präferenzen hier kurz anhand der Wortformen *Dänemark* und *Dänemarks* sowie *Tyskland* („Deutschland“) und *Tysklands* („Deutschlands“).

4.1.1 Satzkookkurrenzen

Ein Blick auf die Satzkookkurrenzen zu *Tyskland* und *Dänemark* zeigt, dass insbesondere Substantive, v. a. andere Ländernamen, als Satzkookkurrenzen vorkommen: Zu *Tyskland* zeigen⁹ *Frankreich*, *Schweden*, *Holland*, *Dänemark*, *England*, *Italien*, *USA*, *Norwegen* und *Großbritannien* die höchste Signifikanz, gefolgt von *Spanien*, *Österreich*, *Belgien*, *Schweiz*, *Finnland*, *Polen*, *Russland* und *Japan*. Zusätzlich zu den Ländernamen findet man die Wortformen *lande* („Länder“), *VM* („WM“), *konserter* („Konzerte“), *nabolande* („Nachbarländer“), *gennem* („durch“), *mellem* („zwischen“), *udlandet* („das Ausland“), *sydlige* („südliche“) und *eksport* („Export“).

Zu *Dänemark* gibt es mehr oder weniger dieselben Länder als Kookkurrenzen, d. h. *Schweden*, *Deutschland*, *Norwegen*, *Frankreich*, *Finnland*, *Großbritannien*, *Österreich*, *Niederlande*, *Belgien*, *Italien*, *Schweiz*, *Spanien*, *Polen*, *Irland*, *Holland*, *USA*. Zudem gibt es einzelne Konjunktionen wie *und* und *oder*, Präpositionen wie *in*, *aus*, *gegen* und *nach* sowie die Wörter *Ländern*, *Staaten*, *Bundestrainer*, *Löw*, *Titelverteidiger*, *Olympiasieger*, *Prozent*, *Urlaub* und *Königreich*.

Bei den signifikantesten Satzkookkurrenzen zu den Wortformen *Dänemarks* bzw. *Tysklands* überwiegen ebenfalls Substantive. Es handelt sich jedoch z. T. um inhaltlich andere Begriffe und Wörter.

Zur Wortform *Deutschlands* sind es neben Länderbezeichnungen im Genitiv (z. B. *Schwedens*, *Norwegens*, *Deutschlands*), auch Namen von Personen (*Rassmussen*, *Olsen*, *Margrethe*, *Frederik*) und geographische Phänomene (*Kopenhagen*, *Hauptstadt*, *Westküste*, *Nordspitze*, *Küsten*, *Norden*) sowie Bezeichnungen für gesellschaftliche oder politische Rollen wie *Ministerpräsident*, *Kronprinz*, *Königin*, *Nationaltrainer*. Weitere Satzkookkurrenzen deuten auf einen inhaltlichen Bezug zu historischen Ereignissen, wie z. B. *Besetzung*, *Geschichte*. Unter den signifikantesten Kookkurrenzen finden sich weiterhin einzelne Adjektive wie *größte*, *älteste*, *schönsten* und *zweitgrößte*.

Zwischen den Satzkookkurrenzen zu den Wortformen *Tyskland* und *Tysklands* gibt es ähnliche Unterschiede. Unter den signifikantesten Kookkurrenzen findet man zwar einzelne Länderbezeichnungen, meist im Genitiv, wie *Frankrigs* („Frankreichs“), *Tysklands*, *Storbritanniens* („Großbritanniens“), *Italiens*, *Ruslands* („Russlands“), *Grækenland* („Griechenland“) aber es überwiegen Personen und Berufsbezeichnungen wie *Merkel*, *Angela*, *kansler* („Kanzler/-in“¹⁰), *forbunds-kansler* („Bundeskanzler/-in“), *præsident* („Präsident/-in“), *Sarkozy*, *udenrigsminister* („Außenminister/-in“), *finansminister* („Finanzminister/-in“), *Schäuble*, *Schröder*, *Westerwelle*, *premierminister* („Ministerpräsident/-in“), *forsvarsminister* („Verteidigungsminister/-in“), *Wulff*, *forbundspræsident* („Bundespräsident/-in“). Weitere signifikante Kookkur-

⁹ Es handelt sich natürlich um die dänischsprachigen Namen, auch wenn hier die deutschen Ländernamen angegeben werden.

¹⁰ Im dänischen wird nicht zwischen weiblicher und männlicher Form unterschieden, daher werden, soweit möglich und sinnvoll, im Folgenden beide Formen in der deutschen Übersetzung angegeben.

renzen sind *genforening* (‚Wiedervereinigung‘), *hovedstad* (‚Hauptstadt‘), *historie* (‚Geschichte‘), *nederlag* (‚Niederlage‘), *landets* (‚des Lands‘), *grænser* (‚Grenzen‘), *vinregion* (‚Weinregion‘) und *rolle* (‚Rolle‘). Unter den signifikantesten Kookkurrenzen sind ebenfalls folgende Adjektive: *næststørste* (‚nächstgrößte‘), *ældste* (‚älteste‘), *førende* (‚führende‘), *økonomiske* (‚ökonomische‘), *bedste* (‚beste‘), *politiske* (‚politische‘), *smukkeste* (‚hübscheste‘), *fælles* (‚gemeinsames‘), *nordligste* (‚nördlichste‘) und *konservative* (‚konservative‘).

Bei den linken und den rechten Nachbarn zu *Tyskland/Tysklands* und *Dänemark/Dänemarks* zeichnet sich ein ähnliches Bild mit Abweichungen in den Kookkurrenzpartnern der einzelnen Wortformen ab.

4.1.2 Linke Nachbarschaftskookkurrenzen

Als signifikante linke Nachbarn von *Tyskland* fungieren grammatische Wörter wie Präpositionen *i*, *fra*, *til*, *mod*, *gennem*, *mellem*, *over* sowie einzelne Konjunktionen *og* (‚und‘) und Abkürzungen *f.eks.*, *bl.a.* Am häufigsten aber sind Adjektive wie *nordlige* (‚nördliche‘), *sydlige* (‚südliche‘), *østlige* (‚östliche‘), *sydvestlige* (‚südwestliche‘), *nordøstlige* (‚nordöstliche‘), *vestlige* (‚westliche‘), *nazistiske* (‚nazistische‘), *genforenede* (‚wiedervereinte‘), *forenede* (‚vereinte‘), *delte* (‚geteilte‘), *hele* (‚ganze‘), *krigshærgede* (‚durch Krieg verwüstet‘). Weiterhin kommen Substantive vor, die als Teil eines spezifizierenden Genitivs identifizierbar sind wie *Hitlers*, *mellemligstidens* (‚der Zwischenkriegszeit‘), *nazitidens* (‚der Nazizeit‘), *efterkrigstidens* (‚der Nachkriegszeit‘), *30ernes* (‚der Dreißiger‘) bzw. Substantive ohne Kasusmarkierung *Forbundsrepublikken* (‚die Bundesrepublik‘), *feriecenter* (‚Ferienzentrum‘), *hotel* (‚Hotel‘), *naboland* (‚Nachbarland‘), *hjemlandet* (‚das Heimatland‘), *ferie* (‚Urlaub‘), *miniferie* (‚Miniurlaub‘), *weekendophold* (‚Wochenendaufenthalt‘) und *møder* (‚Treffen‘). Einzelne Verben wie *forlade* (‚verlassen‘) und *erklære* (‚erklären‘) kommen auch vor.

Bei der Wortform *Tysklands* dominieren als linke Nachbarn ebenfalls grammatische Wörter. Als deutlich häufigere Kookkurrenzpartner als bei der Wortform *Tyskland* treten Verben auf, wie (hier vereinheitlicht in der Infinitivform) *sige* (‚sagen‘), *kritisere* (‚kritisieren‘), *opleve* (‚erleben‘), *opfordre* (‚auffordern‘), *træde* (‚treten‘), *afvise* (‚abweisen‘), *støtte* (‚unterstützen‘), *løse* (‚lösen‘), *være* (‚sein‘), *vende* (‚umkehren‘), *lade* (‚lassen‘), *møde* (‚treffen‘), *besejre* (‚besiegen‘), *følge* (‚folgen‘), *ligge* (‚liegen‘), *finde* (‚finden‘) und *meine* (‚meinen‘). Außerdem kommen einzelne Substantive wie *Forbundsrepublikken* (‚die Bundesrepublik‘), *meter* (‚Meter‘) und Adjektive *nazistiske* (‚nazistische‘), *genforenede* (‚wiedervereinte‘), *fascistiske* (‚faszistische‘) vor.

Linke Nachbarn von *Dänemark* sind grammatische Wörter wie *in*, *gegen*, *aus*, *nach* und *von* sowie Substantive aus dem Bereich Sport wie *Europameister*, *Titelverteidiger*, *Olympiasieger*, *Ex-Europameister*, *WM-Dritten*, *EM-Dritten*, *WM-Teilnehmer*, *Eishockey-Zwerg* und andere Substantive wie *Staate*, *Königreich*, *Nachbarland*, *Mutterland*, *Heimat*, *Heimatland*, *Gastgeber*, *Gastgeberland*, *Richtung*, *Urlaubsring*, *Beispiel* und *Außenseiter*. Als linke Nachbarn von *Dänemarks* finden wir hingegen insbesondere Substantive, die dem geographischen Bereich zuordbar sind wie *Norden*, *Westen*, *Nordspitze*, *Nordwesten*, *Süden*, *Stadt*, *Hauptstadt*, *Küste*, *Westküste*, *Südgrenze*, *Grenzen* und *Teil* sowie Wörter aus dem historisch-politischen Bereich: *Besetzung*, *Geschichte*, *Regierung*, *Vertretungen*, *Partei*, *Gerichtshof*,

Vertreter, Botschaften, Beitritt, Auswahl, Antrag, Ausnahme und *Punkt* samt einzelne grammatische Wörter wie *von, außerhalb, und, sagte* und *zugunsten*.

4.1.3 Rechte Nachbarschaftskookkurrenzen

Sehr deutlich sind die Unterschiede in den Nachbarschaftskookkurrenzen zu den Wortformen *Tyskland* und *Tysklands* bei den rechten Kookkurrenzpartnern. Während *Tysklands* bevorzugt Substantive und Adjektive als Kookkurrenzpartner selektiert, finden wir bei *Tyskland* insbesondere Verben und grammatische Wörter. Zur Wortform *Tysklands* überwiegen Bezeichnungen für offizielle, politische Berufe, Wörter, die einen Bezug zur deutschen Geschichte haben sowie aus der Wirtschaft:

Tysklands:

- *kansler* (,Kanzler/-in‘), *forbundskansler* (,Bundeskanzler/-in‘), *udenrigsminister* (,Außenminister/-in‘), *finansminister* (,Finanzminister/-in‘), *forsvarsminister* (,Verteidigungsminister/-in‘), *indenrigsminister* (,Innenminister/-in‘), *økonomiminister* (,Finanzminister/-in‘), *justitsminister* (,Justizminister/-in‘), *miljøminister* (,Umweltminister/-in‘), *Angela Merkel*, *præsident* (,Präsident/-in‘), *forbundspræsident* (,Bundespräsident/-in‘), *ambassadør* (,Botschafter/-in‘), *rigskansler* (,Reichskanzler‘), *regering* (,Regierung‘), *parlament* (,Parlament‘),
- *historie* (,Geschichte‘), *genforening* (,Wiedervereinigung‘), *deling* (,Teilung‘), *besættelse* (,Besatzung‘), *angreb* (,Angriff‘), *kapitulation* (,Kapitulation‘), *genoprustning* (,Wiederaufrüstung‘),
- *ffjender* (,Feinde‘), *kampe* (,Kämpfe‘), *nederlag* (,Niederlage‘), *sammenbrud* (,Zusammenbruch‘), *overfald* (,Überfall‘), *deltagelse* (,Teilnahme‘), *ansfører* (,Anführer/-in‘),
- *landstræner* (,Nationaltrainer/-in‘),
- *hovedstad* (,Hauptstadt‘), *grænser* (,Grenzen‘), *DAX-indeks* (,DAX-Index‘), *økonomi* (,Ökonomie‘), *svar* (,Antwort‘), *samling* (,Sammlung‘), *rolle* (,Rolle‘), *skæbne* (,Schicksaal‘), *side* (,Seite‘),
- *største* (,größte‘), *ældste* (,älteste‘), *bedste* (,beste‘), *førende* (,führende‘), *næststørste* (,nächstgrößte‘), *højeste* (,höchste‘), *nye* (,neue‘), *økonomiske* (,ökonomische‘), *nordligste* (,nördlichste‘), *kommunistiske* (,kommunistische‘), *smukkeste* (,schönste‘), *øverste* (,oberste‘), *mindste* (,kleinste‘), *kendteste* (,bekannteste‘), *store* (,große‘), *politiske* (,politische‘), *højest* (,höchste‘), *socialdemokratiske* (,sozialdemokratische‘), *betingelsesløse* (,bedingungslose‘), *tredjestørste* (,drittgrößte‘), *konservative* (,konservative‘), *fineste* (,feinste‘), *vigtigste* (,wichtigste‘), *næstbedste* (,nächstbeste‘), *historiske* (,historische‘).

Tyskland:

- *have* (,haben‘), *være* (,sein‘), *ville* (,wollen‘), *afskaffe* (,abschaffen‘), *rejse* (,reisen‘), *invadere* (,besetzen/‘einfallen‘), *stå* (,stehen‘), *vinde* (,gewinnen‘), *ligge* (,liegen‘), *skulle* (,sollen‘), *stige* (,steigen‘), *blive* (,bleiben‘), *annektere* (,annektieren‘), *forbyde* (,verbieten‘), *kunne* (,können‘), *føre* (,führen‘), *koste* (,kosten‘), *få* (,bekommen‘), *genvinde* (,wiedergewinnen‘), *besætte* (,besetzen‘), *betale* (,bezahlen‘), *udstede* (,ausstellen‘), *opleve* (,erleben‘), *underskrive* (,unterschreiben‘), *okkupere* (,okkupieren‘), *toppe* (,top-

- pen‘), *vise* (,zeigen‘), *forvente* (,erwarten‘), *offentliggøre* (,veröffentlichen‘), *gå* (,gehen‘), *bruge* (,verwenden‘);
- *og* (,und‘), *eller* (,oder‘), *samt* (,sowie‘), *for* (,für‘/„vor‘), *under* (,unter‘), *via*, *efter* (,nach‘), *igen* (,nochmals‘), *rundt* (,rund‘), *tilsammen* (,zusammen‘), *krig* (,Krieg‘), *biludlejning* (,Autovermietung‘), *tandlæge* (,Zahnarzt/-ärztin‘), *lokalitet* (,Ort‘), *mini-biografi* (,Mini-Biographie‘), *død* (,Tod‘), *anno*.

Als signifikante rechte Nachbarn von *Dänemark* erscheinen, wie bei *Tyskland*, insbesondere Verben (*haben*, *sein*, *geben*, *werden*, *fahren*, *reisen*, *importieren*, *nominieren*, *liegen*, *wollen*, *kommen*, *schleusen*, *gehören*, *zusammenbasteln*, *gelten*, *spielen*, *verbringen*, *leben*) und grammatische Wörter (*und*, *oder*, *mit*, *zurück*, *sowie*, *statt*, *zu*, *am*, *ab*). Es kommen auch Adjektive und Partizipformen von Verben vor (*gehörenden*, *geboren*, *ansässige*, *stammende*, *lebende*, *startende*, *glücklichstes*, *beheimatet*, *tierärztliches*), jedoch nur vereinzelte Substantive (*FKK-Angebot*, *Urlaubsland*).

Als signifikante rechte Nachbarn von *Dänemarks* finden wir, ebenfalls wie bei *Tysklands*, Substantive, die vorwiegend dem öffentlichen Bereich, Politik und Sport zuzuordnen sind, sowie Adjektive:

- *Ministerpräsident*, *Regierung*, *Außenminister*, *Regierungschef*, *Premier*, *Parlament*, *Nationalbank*, *Justizminister*, *Verkehrsminister*, *Königin*, *Kronprinz*, *Kronprinzessin*, *Hauptstadt*, *Bevölkerung*, *Trainer*, *Nationaltrainer*, *Coach*, *Nationalcoach*, *Keeper*, *Rekordtor-schütze*, *Fußballer*, *Fußball-Nationaltrainer*, *Torwart*;
- *größter*, *bedeutendste*, *zweitgrößter*, *höchster*, *rechtsliberaler*, *neuer*.

Es zeigen sich also am Beispiel *Dänemark/Dänemarks* und *Tyskland/Tysklands* bemerkenswerte Abweichungen in den signifikanten Kookkurrenzen zu den einzelnen Wortformen eines lexikalischen Wortes. Diese Abweichungen findet man auch in den Kookkurrenzpartnern der anderen hochfrequenten Wortformen zu *dänisch** und *Däne** bzw. *tysk** und *tysker**. Die Kookkurrenzen jeder einzelner Wortform können demnach höchstens als tendenziell für das Lexem angesehen werden. Die Kookkurrenzen des Lexems können hingegen nur durch die summative Berechnung der statistischen Kookkurrenzen aller seiner Formen festgestellt werden.

Wir zeigen im Folgenden die morphembasierte Berechnung der statistisch signifikanten Kookkurrenzen zu *dän*- und *tysk*-Wortformen. Mit dem Ergebnis können wir die lexikalische Nachbarschaft für das Gesamtvorkommen dieser Wörter im Korpus aufzeigen. Mit diesem Verfahren wollen wir testen, ob die assoziativen semantischen Stereotype erfasst und sichtbar gemacht werden können, die mit dem lexikalischen Morphem {dän} bzw. {tysk} verbunden und in den Texten im Korpora realisiert werden. Durch die morphembasierte Kookkurrenzberechnung soll zudem eine bessere Vergleichbarkeit der lexikalisch-semantischen Relationen in den deutschen und dänischen Sprachdaten erreicht werden, indem formbedingte sprachspezifische Abweichungen vermieden werden, die z. B. durch unterschiedliche syntaktische Strukturen, Verwendung von Wortarten bzw. Flexionsformen entstehen könnten.

4.2 Morphembasierte Kookkurrenzen

Bei den Kookkurrenzuntersuchungen zu den Morphemen {dän} und {tysk} haben wir sowohl Satz- als auch Nachbarschaftskookkurrenzen verwendet. Als Ausgangspunkt bleibt die oben bereits gestellte Frage nach den statistisch auffälligsten Wörtern in den deutschen und dänischen Sätzen, die die Morpheme {dän} (z. B. *Dänemark* oder *dänisch*) und {tysk} (z. B. *Tyskland* oder *tysk*) enthalten. Tabelle 1 zeigte uns, dass sich in jeder Sprache etwa 95 % der Belege auf nur wenige Wortformen verteilt, die die Länderbezeichnungen (*Tyskland*, *Dänemark*), die Nationalitätsbezeichnung (*tysk*, *dänisch*) und die Menschen (*tyskere*, *Dänen*) umfasst.

Im vorigen Abschnitt haben wir festgestellt, dass es in den Kookkurrenzpartnern der einzelnen Wortformen eines lexikalischen Wortes große Unterschiede geben kann. Uns interessieren insbesondere die lexikalisch-semantischen Relationen zu den *dän*- und *tysk*-Wörtern und deshalb haben wir, um den Einfluss von morphosyntaktischen und/oder anderen, formbedingten Selektionsbeschränkungen zu umgehen, die Kookkurrenzberechnung morphembasiert durchgeführt, d. h. die statistisch auffälligen Nachbarwörter werden für alle *dän*- und *tysk*-Wörter summiert. Wir können daher sowohl als Satzkookkurrenzen als auch als linke und rechte Nachbarn Wörter verschiedener Wortarten ebenso erwarten wie Wörter, die in verschiedenen semantischen Relationen zu den Ausgangswörtern stehen.

Bei der folgenden Darstellung der Ergebnisse verwenden wir ein Verfahren zur Einordnung der Nachbarwörter, mit dem durch einen Kontextvergleich die Unterschiede und Ähnlichkeiten in den bevorzugten Kookkurrenzpartnern in jeder Sprache sichtbar werden (cf. Dalmas et al. 2015). Wir konzentrieren uns hierbei auf die Satzkookkurrenzen und die rechten Nachbarschaftskookkurrenzen, da es sich bei den linken Nachbarschaftskookkurrenzen vorwiegend um grammatische Wörter (Funktionswörter) handelt, die keine inhaltliche Interpretation ermöglichen und somit als Stereotypenindikatoren ausscheiden.

In Dalmas et al. (2015) wird ein korpusbasiertes Verfahren zur Bestimmung von bevorzugten Kombinationen der quasi-synonymen Adjektive *vorzüglich* und *vortrefflich* angewandt. Anhand von Kookkurrenzen in den Daten des deutschsprachigen Korpus der Leipzig Corpora Collection und mithilfe der in Dalmas et al. (2015: 13–14) verwendeten Parameter „lexikalische Präferenzen“, „thematische Domänen“, „Stilebenen“, „Diskurspraktiken“ und „Medialität“ unterscheiden die Autoren den Gebrauch der beiden als identisch aufgefassten Adjektive. Wir modifizieren das Verfahren dahingehend, dass wir es für die Bestimmung von lexikalischen Präferenzen und thematischen Domänen der in der Funktion von Nationalbezeichnungen vorkommenden Wörter mit den Morphemen {tysk} und {dän} mit dem Ziel einsetzen, die bevorzugten Kookkurrenzpartner der *dän*- und *tysk*-Wörter und somit die Kontextualisierung der beiden Nationalitäten in den Leipziger Korpora zu ermitteln.

In beiden Sprachen ist es möglich, dass *dänisch* bzw. *tysk* als typische Länderbezeichnung (bzw. *Deutschland* und *Dänemark* als typisches Land) auftaucht. In diesem Fall ist es interessant, ob es eine eindeutige Präferenz in eine Richtung gibt, d. h. ob für die jeweilige Nationalitätsbezeichnung bevorzugte Kookkurrenzen nachweisbar sind. Mit dem Verfahren des Kontextvergleichs von Dalmas et al. (2015) ist es möglich, einen Überblick über die Kookkurrenzen zu gewinnen, die zunächst von grammatischen und lexikalischen Selektionsbestimmun-

gen der einzelnen Wortformen unabhängig sind und das Morphem {dän} bzw. {tysk} als Basis für die Kookkurrenzberechnung haben. Die Wörter in den Tabellen 2 und 3 wurden so ausgewählt, dass **gleiche Stärke** eine Abweichung der Signifikanzen maximal um den Faktor zwei erlaubt, größere Abweichungen sind gekennzeichnet als stärkeres Auftreten in einer Richtung. Beispielsweise ist die Assoziationsstärke zwischen *dänisch** und *Geheimdienst* im deutschen SMiK-Korpus ähnlich stark wie zwischen *tysk** (,deutsch‘) und *efterretningstjeneste* (,Geheimdienst‘) im dänischen SMiK-Korpus. Anders ist im Deutschen *dänisch** und *Karikaturist* stark assoziiert, nicht aber im Dänischen *tysk** (,deutsch‘) und *Karikaturist*.

Die unten stehenden Tabellen 2 und 3 zeigen eine Auswahl der Kookkurrenzpartner in alphabetischer Reihenfolge. Tabelle 2 enthält die Daten für Satzkookkurrenzen und Tabelle 3 enthält die rechten Kookkurrenzpartner für *tysk*- und *dän*-Wörter. Die dänischen Wörter wurden von den Verfassern ins Deutsche übersetzt. Berufsbezeichnungen u. Ä. werden im Dänischen grundsätzlich nicht moviert, in der Übersetzung werden aber beide Formen angegeben (dänisch: *kansler* = deutsch: *Kanzler/-in*). Wörter, von denen mehrere Wortformen als Satzkookkurrenzen vorkommen, werden in einer Form zusammengefasst und die weiteren Formen mit Angaben in Klammern angezeigt (*Insel*, *Inseln* = *Insel(n)*).

Tabelle 2 zeigt die Verteilung der rechten Satzkookkurrenzen der *tysk*- und *dän*-Wörter in den beiden Sprachen, d. h. welche Kookkurrenzen nur oder bevorzugt im Dänischen (*tysk*-Wörter), nur oder bevorzugt im Deutschen (*dän*-Wörter) sowie gleich häufig im Deutschen und Dänischen vorkommen. Die Anordnung erfolgt alphabetisch:

Nur Dänisch (Auswahl)	mehr Dänisch	gleich häufig	mehr Deutsch	nur Deutsch (Auswahl)
<i>Angriff</i>	<i>Behörden</i>	<i>Außenminister/-in</i>	<i>Botschaft</i>	<i>Erstligisten</i>
<i>Arzt/Ärztin</i>	<i>Meisterschaft</i>	<i>Club/Klub</i>	<i>Hauptstadt</i>	<i>Fernsehen</i>
<i>Ausgabe</i>	<i>Soldaten</i>	<i>Geheimdienst</i>	<i>Künstler/-in</i>	<i>Firma</i>
<i>Autobahnen</i>	<i>Spitzenclub</i>	<i>Gesellschaft</i>	<i>Meister/-in</i>	<i>Flagge</i>
<i>Besatzung</i>	<i>Stadt/Städte</i>	<i>Grenze</i>	<i>Minderheit</i>	<i>Forscher</i>
<i>Besatzungs- macht</i>		<i>Grenzgebiet</i>	<i>Nationalmann- schaft</i>	<i>Herrschaft</i>
<i>Bevölkerung</i>		<i>Grenzregion</i>	<i>Nationalspie- ler/-in</i>	<i>Hersteller</i>
<i>Bundesliga</i>		<i>Krieg</i>	<i>Regierung</i>	<i>Insel(n)</i>
<i>DAX</i>		<i>Polizei</i>	<i>Sprache</i>	<i>Karikaturist(en)</i>
<i>Dienst</i>		<i>Seite</i>	<i>Zeitung</i>	<i>Kollege(n)</i>
<i>Finanzminister/- in</i>		<i>Staat(en)</i>		<i>König/Königin</i>
<i>Flüchtlinge</i>		<i>Unternehmen(s)</i>		<i>Konzern</i>
<i>Flugzeug</i>				<i>Küste</i>

<i>Gericht</i>				<i>Liga</i>
<i>Großgruppe</i>				<i>Marine</i>
<i>Gruppe</i>				<i>Medien</i>
<i>Heer</i>				<i>Ministerpräsi- dent(en)</i>
<i>Kaiser</i>				<i>Modell</i>
<i>Kaiserreich</i>				<i>Parlament</i>
<i>Konzentrations- lager</i>				<i>Prinze(n)/ Prin- zessin</i>
<i>Kriegsdienst</i>				<i>Produkte</i>
<i>Kultur</i>				<i>Radprofi</i>
<i>Magazin</i>				<i>Recht</i>
<i>Markt</i>				<i>Reederei</i>
<i>Mehrwertsteuer</i>				<i>Regierungschef</i>
<i>Nachrichten- agentur</i>				<i>Regisseur(in)</i>
<i>Offizier(e)</i>				<i>Skipper</i>
<i>Ökonomie</i>				<i>Staatsbürger</i>
<i>Politiker/-in</i>				<i>Studie</i>
<i>Reich</i>				<i>Stürmer</i>
<i>Soldat/-in</i>				<i>Südsee</i>
<i>Sozialdemokra- tie</i>				<i>Tageszeitung</i>
<i>Staatsbürger/-in</i>				<i>Team</i>
<i>Stärken</i>				<i>Universität</i>
<i>Tourist(en)</i>				<i>Vertreter</i>
<i>Truppen</i>				<i>Vorbild</i>
<i>TV</i>				<i>Zeichner</i>
<i>Ursprung</i>				<i>Zeitungen</i>
<i>Volk</i>				
<i>Wehrmacht</i>				
<i>Wein</i>				
<i>Wettbewerbsbe- hörde</i>				

Tabelle 2: Satzkookkurrenzen zu *tysk*- und *dän*-Wörtern

Tabelle 3 zeigt die Verteilung der rechten Nachbarschaftskookkurrenzen der *tysk*- und *dän*-Wörter in den beiden Sprachen, d. h. welche Kookkurrenzen nur oder bevorzugt im Dänischen (*tysk*-Wörter), nur oder bevorzugt im Deutschen (*dän*-Wörter) sowie gleich häufig im Deutschen und Dänischen vorkommen. Die Anordnung erfolgt alphabetisch:

Nur Dänisch (Auswahl)	mehr Dänisch	Gleich häufig	mehr Deutsch	nur Deutsch (Auswahl)
<i>Abstammung</i>	<i>Behörden</i>	<i>Außenminister/-in</i>	<i>Außenministerium</i>	<i>Amtskollegen</i>
<i>Autobahn(en)</i>	<i>Meisterschaft</i>	<i>Club/Klub</i>	<i>Botschaft</i>	<i>Astronom(en)</i>
<i>Bank(en)</i>	<i>Philosoph/-in</i>	<i>Fernsehsender</i>	<i>Flagge(n)</i>	<i>Botschafter</i>
<i>Besatzung</i>	<i>Soldat(en)</i>	<i>Firma</i>	<i>Flotte/Marine</i>	<i>Designer</i>
<i>Besatzungsmacht</i>	<i>Staat</i>	<i>Journalist/-in</i>	<i>Grenze</i>	<i>Fernsehen</i>
<i>besetzt</i>	<i>Truppen/Streitkräfte</i>	<i>Meister</i>	<i>Hauptstadt</i>	<i>Filmregisseur</i>
<i>Bevölkerung</i>		<i>Nationalstürmer/-in</i>	<i>Herrschaft</i>	<i>Forscher</i>
<i>Bundeskanzler/-in</i>		<i>Physiker/-in</i>	<i>Hersteller(s)</i>	<i>Gastgeber</i>
<i>Bundesland/-länder</i>		<i>Polizei</i>	<i>Kollege(n)</i>	<i>Geheimdienst</i>
<i>DAX</i>		<i>Seite</i>	<i>Künstler(s)</i>	<i>Gesamtstaat</i>
<i>Dichter/-in</i>		<i>Spitzenclub</i>	<i>Minderheit</i>	<i>Insel</i>
<i>Dienst</i>		<i>Sprache</i>	<i>Nationalmannschaft</i>	<i>Kaminofen</i>
<i>Expressionismus</i>		<i>Staatsbürger/-in</i>	<i>Nationalspieler(s)</i>	<i>Karikaturist</i>
<i>Fahrradfahrer/-in</i>		<i>Staatsbürgerschaft</i>	<i>Parlament(s)</i>	<i>Klimaministerin</i>
<i>Flüchtlinge</i>		<i>Wissenschaftler/-in</i>	<i>Reederei</i>	<i>König</i>
<i>Flug</i>			<i>Regierung</i>	<i>Königsfamilie</i>
<i>Fußball</i>			<i>Unternehmen</i>	<i>Königshaus</i>
<i>Gäste</i>			<i>Vorbild</i>	<i>Königshof</i>
<i>Gericht</i>			<i>Zeitung(en)</i>	<i>Konzern</i>
<i>Gesellschaft</i>				<i>Kronprinz/ Kronprinzessin</i>

<i>Gründlichkeit</i>				<i>Künstlergruppe</i>
<i>Gruppe</i>				<i>Küste</i>
<i>Handwerker/-in</i>				<i>Liga</i>
<i>Heer(es)</i>				<i>Märchendichter(s)</i>
<i>Kaiserreich</i>				<i>Minderheitspartei</i>
<i>Kanzler/-in</i>				<i>Ministerpräsident</i>
<i>Kapitulation</i>				<i>Modell</i>
<i>Kommandant/-in</i>				<i>Nationaltrainer</i>
<i>Konzentrationslager</i>				<i>Nordseeküste</i>
<i>Kriegsdienst/ Wehrdienst</i>				<i>Ostseeinsel</i>
<i>Kultur</i>				<i>Pavillon</i>
<i>Literatur</i>				<i>Präsidentschaft</i>
<i>Luftwaffe</i>				<i>Prinz/Prinzessin</i>
<i>Militär</i>				<i>Produkte</i>
<i>Nazis</i>				<i>Recht</i>
<i>Oberkommando</i>				<i>Regierungschef</i>
<i>Offizier(e)</i>				<i>Regisseur</i>
<i>Panzer</i>				<i>Skipper</i>
<i>Politiker/-in</i>				<i>Studie</i>
<i>Presse</i>				<i>Südsee</i>
<i>Reich(s)</i>				<i>Team</i>
<i>Reichstag</i>				<i>Thron</i>
<i>Richter/-in</i>				<i>Thronfolge</i>
<i>Schlachtschiff</i>				<i>Verband</i>
<i>Text</i>				<i>Vertreter</i>
<i>TV</i>				<i>Walforscher</i>
<i>Ursprung</i>				<i>Waren</i>
<i>Verbund</i>				<i>Wikinger</i>
<i>Veteran</i>				<i>Zeichner</i>

<i>Volk</i>				
<i>Wehrmacht</i>				
<i>Wiedervereinigung</i>				
<i>Wirtschaft</i>				

Tabelle 3: Rechte Nachbarschaftskookkurrenzen von *tysk*- und *dän*-Wörtern

Es zeigt sich, dass sowohl als Satzkookkurrenzen als auch als rechte Nachbarschaftskookkurrenzen Substantive überwiegen. Als bevorzugte linke Kookkurrenzpartner der *tysk*- und *dän*-Wörter treten insbesondere grammatische Wörter auf. Diese werden daher hier nicht aufgelistet.

5 Themenbereiche der Kookkurrenzpartner zu *dän*- und *tysk*-Wörtern

Um den Gebrauch der *tysk*- und *dän*-Wörter inhaltlich zu bestimmen, wurde eine Einordnung der 100 signifikantesten rechten Kookkurrenzen zu *tysk*- und *dän*-Wörtern nach Themenbereichen (bzw. thematischen Domänen, cf. Dalmas et al. 2015) vorgenommen. Die Wahl der Themenbereiche erfolgte datengeleitet, wobei nur wenige Wörter nicht zuordenbar waren, einige Wörter aber ggf. verschiedenen Bereichen zugeordnet werden könnten. Die Einordnung ergibt folgende Themenbereiche, in denen sich eine deutliche Übereinstimmung mit den in den deutschen und dänischen Medien üblichen Ressorts zeigt. Die fettgedruckten Wörter kommen in beiden Sprachen unter den 100 signifikantesten rechten Kookkurrenzpartnern zu *dän*- und *tysk*-Wörtern vor.

Im **Themenbereich Geographie** (geographische Phänomene) finden wir in beiden Sprachen Wortformen wie *Grenze*, *Grenzland*, *Grenzregion* und *Grenzgebiet* sowie *Stadt* und *Hauptstadt*. Die Wortform *Grenze* weist in den dänischen Kookkurrenzen die höchste und in den deutschen Kookkurrenzen die nächsthöchste Signifikanz auf. In den weiteren Kookkurrenzen handelt es sich um übergeordnete Bezeichnungen für Länder und Städte. Als nationalspezifisch können hier *Autobahnen* und *Bundesland/Bundesländer* in den dänischen Sprachdaten und *Insel*, *Inseln*, *Küste* und *Südsee* (die dänische Südsee ist ein Segelparadies südlich der Insel Fünen) sowie einzelne Städtenamen wie *Aarhus* und *Roskilde* im deutschen Korpus eingeordnet werden.

Nachbarwort zu *tysk*	Deutsche Übersetzung	Frequenz	Signifikanz	Nachbarwort zu *dän*	Frequenz	Signifikanz
<i>motorveje</i>	<i>Autobahnen</i>	127	615	<i>Aarhus</i>	873	1487
<i>delstat</i>	<i>Bundesland</i>	270	1379	<i>Grenze</i>	1665	7817
<i>delstater</i>	<i>Bundesländer</i>	223	807	<i>Grenzgebiet</i>	93	597
<i>grænse</i>	<i>Grenze</i>	1362	11004	<i>Grenzregion</i>	114	807
<i>grænseland</i>	<i>Grenzland</i>	96	1101	<i>Hauptstadt</i>	2033	8508
<i>grænseregion</i>	<i>Grenzregion</i>	59	782	<i>Herning</i>	252	935

<i>hovedstad</i>	Hauptstadt	436	1036	<i>Insel</i>	2036	4122
<i>lande</i>	<i>Länder</i>	3430	1537	<i>Inseln</i>	714	1101
<i>by</i>	Stadt	1146	2362	<i>Küste</i>	742	1098
<i>byer</i>	Städte	573	1092	<i>Roskilde</i>	525	825
				Stadt	1433	618
				<i>Südsee</i>	277	932

Tabelle 4: Rechte Nachbarschaftskookkurrenzen von *tysk*- und *dän*-Wörtern, die zum Themenbereich Geographie gehören

Im Themenbereich Sport überwiegen im Dänischen Wortformen, die sich auf Fußball und Strukturen in sportlichen Wettkämpfen beziehen. Im Deutschen kommt noch Radsport hinzu. In beiden Sprachen kommen die Wortformen *Club/Klub*, *Spitzenclub*, *Meister*, *Meisterschaft*, *Nationalmannschaft* und *Nationalspieler* vor. Als nationalspezifisch können hier nur *Bundesliga* (dänisch) und *CSC-Team* (deutsch) eingeordnet werden.

Nachbarwort zu *tysk*	Deutsche Übersetzung	Frequenz	Signifikanz	Nachbarwort zu *dän*	Frequenz	Signifikanz
<i>Bundesliga</i>	<i>Bundesliga</i>	505	3090	Club	351	567
<i>bundesliga</i>	<i>Bundesliga</i>	72	452	<i>CSC-Team</i>	315	2122
<i>klub</i>	Club	937	3627	<i>Erstligisten</i>	571	3554
<i>fodbold</i>	<i>Fußball</i>	531	1190	Klub	313	788
<i>mester</i>	Meister	211	537	<i>Liga</i>	439	1146
<i>mestre</i>	Meister (Pl.)	279	1097	Meister	898	2377
<i>mesterskab</i>	Meisterschaft	441	1709	Meisterschaft	271	643
<i>landshold</i>	Nationalmannschaft	309	835	Nationalmannschaft	950	1746
<i>landsholdsspieler</i>	Nationalspieler/-in	199	556	Nationalspieler	993	4617
<i>pokalturnering</i>	<i>Pokalwettbewerb</i>	95	718	<i>Radprofi</i>	255	839
<i>storklub</i>	Spitzenclub	218	1455	<i>Rennstall</i>	183	692
<i>storhold</i>	<i>Spitzenmannschaft</i>	103	542	Spitzenclub	73	594
				<i>Stürmer</i>	444	832
				<i>Team</i>	2488	1105
				<i>Vizemeister</i>	157	581

Tabelle 5: Rechte Nachbarschaftskookkurrenzen von *tysk*- und *dän*-Wörtern, die zum Themenbereich Sport gehören

Im **Themenbereich Politik** werden Wörter eingeordnet, die politische Funktionen und Status bezeichnen. Hier finden wir einige Bezeichnungen, die es in beiden Ländern gibt, wie z. B. *Minister, Regierung, Staat* und *Botschaft*, und auch länderspezifische Bezeichnungen wie *Kaiser, Kaiserreich, Reich* und *Bundeskanzler/in* im Dänischen. Im Deutschen sind insbesondere Bezeichnungen, die sich auf das Königshaus beziehen, signifikant, wie z. B. *König, Königin, Königshaus, Prinzen* und *Kronprinzenpaar*.

Nachbarwort zu *tysk*	Deutsche Übersetzung	Frequenz	Signifikanz	Nachbarwort zu *dän*	Frequenz	Signifikanz
<i>udenrigsminister</i>	Außenminister/-in	339	742	Außenminister	457	1413
<i>ambassade</i>	Botschaft	171	640	<i>Außenminister</i> <i>Per Stig</i>	164	803
<i>forbundskansler</i>	<i>Bundeskanzler/-in</i>	564	1361	Botschaft	906	3433
<i>finansminister</i>	<i>Finanzminister/-in</i>	178	488	<i>Herrschaft</i>	313	905
<i>kejser</i>	<i>Kaiser</i>	455	2469	<i>König</i>	2347	4333
<i>kansler</i>	<i>Kanzler/-in</i>	821	2267	<i>König Christian</i>	323	778
<i>kejserrige</i>	<i>Keiserreich</i>	116	486	<i>Könige</i>	186	658
<i>politikere</i>	<i>Politiker/-in</i>	317	552	<i>Königin</i>	1215	1806
<i>regering</i>	Regierung	1109	2305	<i>Königs</i>	502	2491
<i>rige</i>	<i>Reich</i>	507	1807	<i>Königshaus</i>	257	1135
<i>socialdemokrati</i>	<i>Sozialdemokratie</i>	81	508	<i>Kronprinzenpaar</i>	124	629
<i>stat</i>	Staat	453	943	<i>Ministerpräsident</i>	1203	3896
<i>stater</i>	Staaten	288	566	<i>Ministerpräsidenten</i>	367	2317
				<i>Parlament</i>	543	1263
				<i>Prinzen</i>	211	592
				<i>Prinzessin</i>	548	587
				Regierung	2224	6177
				<i>Regierungschef</i>	482	1774
				Staat	484	990

Tabelle 6: Rechte Nachbarschaftskookkurrenzen von *tysk*- und *dän*-Wörtern, die zum Themenbereich Politik gehören

Im Themenbereich **öffentliche Institutionen** kommen in beiden Sprachen die Wörter *Gesellschaft*, *Behörden*, *Polizei*, *Staatsbürger* und *Minderheit* vor. Als nationalspezifisch sehen wir hier nur die Kookkurrenz *Flagge* im Deutschen.

Nachbarwort zu *tysk*	Deutsche Übersetzung	Frequenz	Signifikanz	Nachbarwort zu *dän*	Frequenz	Signifikanz
<i>myndigheder</i>	Behörden	865	3990	Behörden	572	1786
<i>befolkning</i>	<i>Bevölkerung</i>	421	740	<i>Flagge</i>	318	979
<i>forbund</i>	<i>Bund</i>	243	926	Gesellschaft	507	548
<i>domstol</i>	<i>Gericht</i>	171	497	Minderheit	1023	4727
<i>selskab</i>	Gesellschaft	407	823	Polizei	1497	4475
<i>mindretal</i>	Minderheit	573	1866	<i>Recht</i>	395	690
<i>politi</i>	Polizei	895	2412	Staatsbürger	146	728
<i>statsborger</i>	Statsbürger/-in	160	644			
<i>folk</i>	<i>Volk</i>	907	817			
<i>konkurrence-myndigheder</i>	<i>Wettbewerbsbehörden</i>	98	715			

Tabelle 7: Rechte Nachbarschaftskookkurrenzen von *tysk*- und *dän*-Wörtern, die zum Themenbereich öffentliche Institutionen gehören

Im Themenbereich **Kultur** als Oberbegriff für Kunst, Wissenschaft und Presse kommen in beiden Sprachen die Wörter *Medien*, *TV/Fernsehen*, *Zeitung/Tageszeitung*, *Sprache* und *Künstler* vor.

Nachbarwort zu *tysk*	Deutsche Übersetzung	Frequenz	Signifikanz	Nachbarwort zu *dän*	Frequenz	Signifikanz
<i>udgave</i>	<i>Ausgabe</i>	580	889	<i>Architekten</i>	177	769
<i>komponist</i>	<i>Komponist/-in</i>	161	479	Fernsehen	329	1130
<i>kultur</i>	<i>Kultur</i>	457	649	<i>Forscher</i>	420	930
<i>kunstner</i>	Künstler/-in	155	482	<i>Karikaturist</i>	102	565
<i>magasin</i>	<i>Magazin</i>	178	973	<i>Karikaturisten</i>	156	665
<i>medier</i>	Medien	448	1354	Künstler	476	1461
<i>nyhedsbureau</i>	<i>Nachrichtenagentur</i>	261	1927	Künstlers	113	904
<i>filosof</i>	<i>Philosoph/-in</i>	256	1286	Medien	430	765
<i>sprog</i>	Sprache	1442	642	<i>Modell</i>	370	717
<i>text</i>	<i>Text</i>	50	471	<i>Mohammed-Karikaturen</i>	1272	1122
<i>tv</i>	TV	162	481	<i>Regisseur</i>	380	1574

<i>forfatter</i>	<i>Verfasser/-in</i>	333	834	<i>Regisseurin</i>	131	739
<i>ord</i>	<i>Wort</i>	573	507	<i>Regisseurs</i>	93	645
<i>avis</i>	<i>Zeitung</i>	472	1940	<i>Sprache</i>	711	1465
				<i>Studie</i>	659	673
				<i>Tageszeitung</i>	312	1266
				<i>Universität</i>	611	681
				<i>Zeichner</i>	244	543
				<i>Zeitung</i>	1886	6369
				<i>Zeitungen</i>	522	1685

Tabelle 8: Rechte Nachbarschaftskookkurrenzen von *tysk*- und *dän*-Wörtern, die zum Themenbereich Kultur gehören

Im Themenbereich **Militär und Krieg** gibt es Übereinstimmungen in den Wörtern *Geheimdienst*, *Krieg*, *Korps/Truppen* und *Soldaten*. Im Deutschen weist in diesem Bereich zusätzlich nur das Wort *Marine* signifikante Kookkurrenzen zu *dän*-Wörtern auf. Die dänischen Kookkurrenzen sind geprägt durch den Zweiten Weltkrieg und die Besatzungszeit. Hier finden wir Wörter wie: *Angriff*, *Besatzung*, *Besatzungsmacht*, *Wehrmacht*, *Heer*, *Offizier/Offiziere*, *Kriegsdienst*, *Dienst*, *Flüchtlinge* und *Konzentrationslager*.

Nachbarwort zu *tysk*	Deutsche Übersetzung	Frequenz	Signifikanz	Nachbarwort zu *dän*	Frequenz	Signifikanz
<i>angreb</i>	<i>Angriff</i>	870	474	<i>Geheimdienstes</i>	98	757
<i>besættelse</i>	<i>Besatzung</i>	839	4561	<i>Geheimdienst</i>	154	708
<i>besættelsesmagt</i>	<i>Besatzungsmacht</i>	326	2460	<i>Korps</i>	120	590
<i>tjeneste</i>	<i>Dienst</i>	257	670	<i>Krieg</i>	1158	1060
<i>flygtninge</i>	<i>Flüchtlinge</i>	503	1617	<i>Marine</i>	254	923
<i>efterretningstjeneste</i>	<i>Geheimdienst</i>	132	462	<i>Soldaten</i>	919	1657
<i>hær</i>	<i>Heer</i>	716	1790			
<i>koncentrationslejr</i>	<i>Konzentrationslager</i>	188	746			
<i>krig</i>	<i>Krieg</i>	976	670			
<i>krigstjeneste</i>	<i>Kriegsdienst/Wehrdienst</i>	90	668			
<i>officer</i>	<i>Offizier/-in</i>	138	465			
<i>officerer</i>	<i>Offiziere</i>	184	456			
<i>soldat</i>	<i>Soldat/-in</i>	213	718			
<i>soldater</i>	<i>Soldaten/-innen</i>	1314	3683			
<i>tropper</i>	<i>Truppen/Korps</i>	844	2494			
<i>værnemagt</i>	<i>Wehrmacht</i>	122	914			

Tabelle 9: Rechte Nachbarschaftskookkurrenzen von *tysk*- und *dän*-Wörtern, die zum Themenbereich Militär und Krieg gehören

Der Themenbereich **Wirtschaft** weist große Unterschiede auf, denn die einzigen gemeinsamen Wörter sind *Firma* und *Unternehmen*. An nationalspezifischen Wörtern fallen im dänischen Korpus *DAX* und *DAX-Index* und im deutschen Korpus *Krone* und *Kronen* auf. Hier

findet man auch Wörter, die an sich nicht nationalspezifisch sind, es aber im deutsch-dänischen Kontext durch ihre wirtschaftliche Relevanz für Dänemark bzw. Deutschland werden: *Flug*, *Markt*, *Wirtschaft*, *Mehrwertsteuer*, *Touristen* und *Wein* haben sind signifikante Kookkurrenzen zu *tysk*-Wörtern. Die Wörter *Brauereikonzern*, *Energiekonzern*, *Konzern*, *Fährreederei* und *Rederei* sowie *Hersteller*, *Produkte* und *Waren* zeigen dagegen im Deutschen signifikante Kookkurrenzen mit *dän*-Wörtern.

Nachbarwort zu *tysk*	Deutsche Übersetzung	Frequenz	Signifikanz	Nachbarwort zu *dän*	Frequenz	Signifikanz
<i>DAX</i>	<i>DAX</i>	204	1145	<i>Brauereikonzern</i>	119	924
<i>DAX-indeks</i>	<i>DAX-Index</i>	182	1116	<i>Energiekonzern</i>	168	925
<i>fly</i>	<i>Flug</i>	435	464	<i>Fährreederei</i>	75	619
<i>marked</i>	<i>Markt</i>	1098	3450	<i>Firma</i>	776	1798
<i>moms</i>	<i>Mehrwertsteuer</i>	287	800	<i>Hersteller</i>	404	951
<i>turist</i>	<i>Tourist</i>	136	849	<i>Herstellers</i>	99	675
<i>turister</i>	<i>Touristen</i>	464	1379	<i>Konzern</i>	336	806
<i>firma</i>	<i>Firma</i>	580	1925	<i>Krone</i>	626	1576
<i>virksomheder</i>	<i>Unternehmen</i> (Pl.)	678	464	<i>Kronen</i>	1069	2483
<i>vin</i>	<i>Wein</i>	276	657	<i>Produkte</i>	350	807
<i>økonomi</i>	<i>Wirtschaft</i>	601	998	<i>Reederei</i>	421	1648
				<i>Unternehmen</i>	1686	2012
				<i>Unternehmens</i>	271	681
				<i>Waren</i>	211	590

Tabelle 10: Rechte Nachbarschaftskookkurrenzen von *tysk*- und *dän*-Wörtern, die zum Themenbereich Wirtschaft gehören

Keinem Themenbereich zugeordnet wurden die Satzzeichen linke Klammer, Komma, Punkt und Doppelpunkt, grammatische Morpheme und folgende Wörter im Dänischen: *Arzt*, *Frau*, *Gruppe*, *Seite*, *Stärken*, *Ursprung* und Deutschen: *Kollege*, *Kollegen*, *Seite*, *Skipper*, *Verband*, *Vertreter*, *Vorbild*, die zu keinem der anderen Themengebiete gehören und auch keine eigene Themengebiete bilden.

Nachbarwort zu *tysk*	Deutsche Übersetzung	Frequenz	Signifikanz	Nachbarwort zu *dän*	Frequenz	Signifikanz
)		13888	674	<i>Kollege</i>	135	737
,		216035	3341	<i>Kollegen</i>	638	786
.		159261	4672	<i>Seite</i>	1188	1555
:		8382	622	<i>Skipper</i>	184	780

<i>læge</i>	<i>Arzt/Ärztin</i>	234	607	<i>Verband</i>	331	685
<i>kvinde</i>	<i>Frau</i>	365	907	<i>Vertreter</i>	682	925
<i>hold</i>	<i>Gruppe</i>	1235	1049	<i>Vorbild</i>	464	1563
<i>havde</i>	<i>hatte</i>	8393	866			
<i>eller</i>	<i>oder</i>	6569	1153			
<i>side</i>	<i>Seite</i>	1650	1965			
<i>styrker</i>	<i>Stärken</i>	515	752			
<i>og</i>	<i>und</i>	110411	3047			
<i>under</i>	<i>unter</i>	5769	467			
<i>oprindelse</i>	<i>Ursprung</i>	210	523			

Tabelle 11: Rechte Nachbarschaftskookkurrenzen von *tysk*- und *dän*-Wörtern, die keinem Themenbereich zugeordnet werden konnten

7 Diskussion und Ausblick

Für diesen Beitrag haben wir uns die Frage gestellt, ob und wie nationale Stereotype als Muster in großen Korpora automatisch und semi-automatisch erfasst werden können. Für die Beantwortung der Frage haben wir aus den Korpora der Leipzig Corpora Collection zwei Spezialkorpora extrahiert, die ausschließlich aus Belegen bestehen, die Wörter mit der Buchstabenkombination *tysk* (dänisches Korpus) und *dän* (deutsches Korpus) enthalten. Mit dieser Herangehensweise haben wir den Gebrauch der Wörter aus zwei Sprachen untersuchen können, die jeweils die Nationalität der anderen bezeichnen (cf. die Verteilung der Wörter auf Wortformen in Tabelle 1). Um die eingangs aufgestellte Frage zu beantworten, ob assoziative semantische Stereotype durch ihre Assoziationsstärke sichtbar werden, haben wir die assoziationsstärksten Nachbarwörter der in den Korpora vorhandenen *tysk*- und *dän*-Wörtern durch die Berechnung von Kookkurrenzen ermittelt. Die Berechnung wurde wortformen- und morphembasiert für alle *tysk*- und *dän*-Wörter durchgeführt. Zusätzlich haben wir ein Verfahren zur Beschreibung der Kontextualisierung der *tysk*- und *dän*-Wörter eingesetzt. Mit diesem Verfahren konnten wir auf der Grundlage der Kookkurrenzanalyse die lexikalischen Präferenzen der Nationalitätsbezeichnungen in der jeweils anderen Sprache erfassen und vergleichen. Diese Präferenzen können als Indizien für Stereotype dienen. Die einzelnen Lexeme werden als hypothetische Stereotypenkandidaten bestimmt, die jedoch erst durch die Beleganalyse als Stereotype bestätigt oder als Nicht-Stereotype verworfen werden können (cf. ausführlich dazu sowie Beispielanalysen in Hallsteinsdóttir 2016). Als ein Gesamtergebnis ist daher festzuhalten, dass auch die automatisierte Auswertung großer digitaler Korpora nicht aus sich heraus die Frage beantworten kann, wann ein Satz bzw. eine Äußerung mit einer Nationalitätenbezeichnung ein Stereotyp zum Ausdruck bringt und wann nicht.

Abschließend haben wir im Abschnitt 5 die signifikantesten Kookkurrenzen in beiden Sprachen Themenbereichen zugeordnet und interlinguale Unterschiede und Ähnlichkeiten im Vorkommen ermittelt. Die Zuordnung der Kookkurrenzpartner der *tysk*- und *dän*-Wörter zu Themenbereichen zeigt eine starke Übereinstimmung zwischen beiden Sprachen in The-

menbereichen, die die Ressorts von Medien widerspiegeln. Innerhalb dieser Bereiche kommen nationalspezifische Artefakte und Phänomene vor, die jeweils im deutsch-dänischen Kontext für eine Nation relevant sind und sich in den nationalsprachlichen Medien widerspiegeln. Es sind die Phänomene, über die in Medientexten über die andere Nationalität berichtet wird, d. h. die über diese Texte mit der Nationalität verbunden werden.

Die Zuordnung der Kookkurrenzpartner der *tysk*- und *dän*-Wörter zu Themenbereichen zeigt eine starke Übereinstimmung zwischen beiden Sprachen in Themenbereichen, die die Ressorts von Medien widerspiegeln. Innerhalb dieser Bereiche kommen nationalspezifische Artefakte und Phänomene vor, die jeweils im deutsch-dänischen Kontext für eine Nation relevant sind und sich in den nationalsprachlichen Medien widerspiegeln. Es sind die Phänomene, über die in Medientexten über die andere Nationalität berichtet wird, d. h. die über diese Texte mit der Nationalität verbunden werden. Diese Nationalspezifika zeigen eine teilweise Übereinstimmung mit den Antworten zur ersten Frage des SMiK-Fragebogens *Typisch deutsch – typisch dänisch: Schreiben Sie bitte die ersten drei Wörter auf, die Ihnen zu Dänemark/Deutschland einfallen* (cf. Abschnitt 2 sowie Hallsteinsdóttir 2015: 44, 53; Hofmann/Hallsteinsdóttir in diesem Heft). Solche Assoziationen werden in der Stereotypenforschung als ein Bestandteil des kulturellen Wissensvorrats eingeordnet, der stereotype Denkmuster charakterisiert (cf. Bolten 2006, 2007). Wenn solche Attribute als Bestandteil der Bedeutung kollektiv bei den Sprechern vorhanden sind, dann können sie aus einer linguistischen Perspektive ebenfalls als assoziative semantische Stereotype aufgefasst werden: Wenn den Sprechern beim Stimulus „Schreiben Sie bitte die ersten drei Wörter auf, die Ihnen zu Deutschland einfallen“ als Reaktion *Anden Verdenskrig* („Zweiter Weltkrieg“) nennen, so aktivieren sie ihre (positive oder negative) Bewertung von ‚Zweiter Weltkrieg‘ und übertragen es als typisches Attribut in Form eines assoziativen semantischen Stereotyps auf ‚Deutschland‘.

Die Sprecher kennen, das zeigen die Fragebogenergebnisse, auch z. T. die in den Korpora als signifikant aufgedeckten Kulturspezifika bzw. die im jeweiligen binationalen Kontext relevanten Phänomene (wie z. B. *Kopenhagen, Berlin, Anden Verdenskrig* ‚Zweiter Weltkrieg‘, *Hitler, Angela Merkel*), d. h. sie sind bei den Sprechern als assoziatives Wissen über Deutschland und Dänemark bzw. als eine Art Faktenwissen vorhanden. Ob solches Wissen als assoziative semantische Stereotype auch immer im Sprachgebrauch aktiviert wird, wird vom Thema und Kontext des jeweiligen Textes sowie dem Vorwissen und den Erfahrungen der Leser stark abhängen. Es sollte jedoch keinesfalls unterschätzt werden, dass solches assoziatives (Fakten-)Wissen (hier z. B. der Bezug zu *Hitler* und zum *Zweiten Weltkrieg* in den dänischen Korpusdaten, der nur in den Antworten zur ersten Fragebogenfrage vorkommt) in den Köpfen der Sprecher als Potenzial jederzeit für die Wiederbelebung traditioneller Stereotype aktiviert werden kann. Die Antworten zu den anderen Fragen in der SMiK-Fragebogenerhebung legen nahe, dass die in den Korpora ermittelten Kookkurrenzen nur einen Teil der nationalen Stereotype über ein anderes Land aufdecken (cf. Hallsteinsdóttir 2015; Hallsteinsdóttir 2016; Hofmann/Hallsteinsdóttir in diesem Heft; Müller/Hallsteinsdóttir 2016).

Es wurden in der bisherigen Forschung bereits mehrmals Medien-/Textanalysen zwecks Ermittlung von Nationalimages bzw. nationalen Stereotypen durchgeführt (deutsch-dänisch cf. z. B. Langer 2003; Müller 2005). Als Ausgangshypothese gilt dabei häufig, dass man die

Konzeptualisierung von nationalen Stereotypen (bzw. auch Images, Nationenbilder o. Ä.) durch die Analyse dessen, wie Medien der einen Nationalsprache die jeweils zu untersuchende andere Nation kontextualisieren, ermitteln könne (cf. z. B. Vogel 2010 und das Projekt CDI 2015). Die hier vorgestellte Analyse der Korpusdaten zu *dän*- und *tysk*-Wörtern zeigt jedoch, dass die Kontextualisierungen „des Deutschen“ und „des Dänischen“ in den untersuchten dänischen und deutschen Medientexten stark von den inhaltlich-thematischen Strukturen der Medien selbst geprägt sind, die nicht immer nationalspezifisch sind (cf. dazu die übereinstimmende Befunde zu Ungarn und der Türkei, die in Vogel [2010] beschrieben werden). Nationalspezifische Unterschiede zeigen sich in der Behandlung von gesellschaftlichen und kulturellen Phänomenen der anderen Nation innerhalb dieser Medienstrukturen. Korpuslinguistische Studien wie diese, die sich automatischer und semi-automatischer Analysen bedienen, geben daher zunächst allgemein Aufschluss darüber, welche Inhalte in Verbindung mit anderen Nationen medial thematisiert werden. Über diese Inhalte, ggf. zusätzlich über ihre positive oder negative Prägung, können eventuell Rückschlüsse auf nationale Stereotype (auch im Sinne von Bild/Image) über eine andere Nation gezogen werden. Korpusdaten sind somit ein wichtiger Baustein in der Stereotypenforschung. Die fehlende Manifestation der aus anderen Untersuchungen bekannten, aktuellen deutschen und dänischen Heterostereotype in den untersuchten Kookkurrenzen zeigt jedoch, dass Kookkurrenzanalysen nicht alleine ausreichen, um daraus die aktuellen nationalen Stereotype zu Deutschen und Dänen (bzw. das Bild oder Image von Deutschland/Dänemark) zu erfassen und beschreiben, so wie in der bisherigen Forschung z. T. suggeriert wird.

Hier stellt sich dann die Frage, was automatische und semi-automatische Analysen in großen Korpora bei der Suche nach der sprachlichen Manifestation von Stereotypen in großen Mengen von Sprachdaten leisten können? Wir sehen insbesondere folgende Ansätze als vielversprechend an:

I) Kookkurrenzen widerspiegeln musterhafte lexikalisch-semantische und morphosyntaktische Strukturen, die für jede Wortform eines lexikalischen Wortes unterschiedlich ausgeprägt sein können. Über die Beschreibung dieser Strukturen kann ein Bild gezeichnet werden, das die Inhalte der medialen Berichterstattung zu einer Nation, Nationalität und einem Land skizziert. Nicht alle solche Strukturen entsprechen jedoch (expliziten oder impliziten) nationalen Stereotypen. Signifikante Kookkurrenzpartner können aber Stereotypenindikatoren sein, d. h. sie können als Indizien für sprachliche Stereotype angesehen werden. Ob es sich in den Texten aber wirklich um ein in der Sprache manifestiertes Stereotyp handelt, kann nur durch eine inhaltliche Analyse der einzelnen Belege festgestellt werden (cf. Hallsteinsdóttir 2016).

II) Die Ermittlung von Themenbereichen gibt Aufschluss darüber, welche Inhalte mit der anderen Nation/Nationalität verbunden werden und wie über sie berichtet wird, d. h. wie sie in der anderen Sprache und Kultur kontextualisiert werden (cf. Dalmas et al. 2015). Ob hier auch der – positive und negative – Ruf einer Nation zuverlässig ermittelt werden kann, müssen weitere Untersuchungen zeigen (cf. z. B. CDI 2015).

III) Spezialkorpora mit hohem Recall und hoher Präzision, wie wir sie für die *dän*- und *tysk*-Wörter als Nationalbezeichnungen erreicht haben, bieten eine Möglichkeit, nationale Stereotype über die Suche nach den bekannten sprachlichen Markierungen von expliziten sprachlichen Stereotypen zu finden. So ist es z. B. möglich, nach Mustern und Markern wie

typisch, **stereotyp**, **an sich** und **Vorurteil** zu suchen, die zu folgenden Belegen führen¹¹ (cf. zu stereotypen syntaktischen Mustern Kilian in diesem Heft):

- (1) Auch sei von „typisch dänischem Rassismus“ die Rede gewesen.
- (2) „[...] Diese unglaubliche Gastfreundschaft ist wohl typisch dänisch, ich glaube nicht, dass eine deutsche Schule das erlaubt hätte“, meinte der Trainer Sidiki Cisse.
- (3) Und dabei setzen die Veranstalter ganz besonders auf eine angenehme, typisch dänische Atmosphäre und viel Nähe.
- (4) Meine kopfeigenen Vorurteile, die den Dänen als piefig, spießig und begeisterungsunfähig, dafür aber als versoffen einstufen, wollte ich behalten.
- (5) Menuen var bestilt via e-mail fra Danmark med et krav om, at den var typisk tysk - altså meget med kål, kartofler og kød. [„Das Menü wurde per E-Mail aus Dänemark bestellt mit einer Forderung, dass es typisch deutsch sein sollte – also viel Kohl, Kartoffeln und Fleisch.“]
- (6) Alt er stille og roligt, folk og dyr er i balance, og man hører aldrig den typiske tyske „her kommer jeg-stil“, som ellers er så almindelig hernede. [„Alles ist still und ruhig, Menschen und Tiere in Balance und man hört nie den typischen „hier komme ich-Stil“, der sonst so üblich ist hier unten.“]
- (7) 8 timer senere ankommer vi og mærker straks den typiske tyske systematik. [„8 Stunden später kommen wir an und merken sofort die typische deutsche Systematik.“]

IV) Weiterhin kann in den Korpora gezielt nach dem Vorkommen von bekannten Stereotypen gesucht werden. So ergibt die Suche nach **gemütlich** für *den gemütlichen Dänen* die folgenden Wortformen (Frequenz in eckigen Klammern):

- gemütlich [43], gemütlichen [40], gemütliche [34], gemütliches [11], gemütlicher [7], urgemütlichen [4], gemütlichem [2], ungemütlich [2], ungemütlicher [2], Hochgemütlichkeitstrakt [1], Ungemütlichkeit [1], Urgemütliche [1], gemütlich- [1], gemütlich-dänischen [1], gemütlich-familiären [1], gemütlich-netten [1], hyggelig/gemütlich [1], stilvoll-gemütliches [1], urgemütlich [1]

Bei einer inhaltliche Analyse der dazugehörigen Belege zeigt sich, ob sich bei *gemütlich* um eine sprachliche Manifestation des Stereotyps handelt:

- (8) Im Kro (Gasthof), Dänemarks Hochgemütlichkeitstrakt, blubbert die berühmte Hyggelighed leise vor sich hin, etwa beim Nachmittagskaffee mit einem Gläschen Gammel-Dansk-Bitterlikör.

Das Vorkommen des Wortes *Hyggelighed* veranlasst uns dazu, im deutschsprachigen Korpus ebenfalls nach dem dänischen Wort zu suchen. Es zeigt sich, dass folgende Wortformen von **hygge** und **Hygge** vorkommen:

¹¹ Die Beispiele 1–14 stammen aus den SMiK-Korpora [20.12.2015], die Übersetzung der dänischen Belege ins Deutsche erfolgte durch die Verfasser.

- hyggelig [12], Hyggelig [9], hygge [6], Hygge [5], hyggelige [5], hyggeligen [3], hyggelig/gemütlich [1], hyggeligsten [1], omhyggelig [1], Hygge-Kaminofen [1], Hyggekultur [1], Hyggelig :Der [1], Hyggeliges [1], Hyggelighed [1], Hyggelik [1]

Auch hier entscheidet die inhaltliche Analyse der einzelnen Belege auf der Basis des individuellen Wissens darüber, ob es sich um eine sprachliche Manifestation des Stereotyps handelt oder nicht:

- (9) „Hygge“ ist eine spezielle dänische Beschreibung von Gemütlichkeit und Wohlbefinden und der „Hygge“-Kaminofen war somit der behagliche Ofen zum Wohlfühlen.
- (10) Gemütlichkeit, Hygge, wird in Dänemark großgeschrieben und drückt sich vor allem im Essen aus.
- (11) Hygge ist ein dänisches Wort, von dem es heißt, man könne es nicht übersetzen.

Auf diese Weise kann z. B. die sprachliche Manifestation von Stereotypen, die mit anderen Instrumenten (z. B. Fragebogen, Interviews) erhoben werden, erforscht werden, wie z. B. *Ordnung* oder *Höflichkeit* (cf. ausführlich Hallsteinsdóttir 2015, 2016; Hofmann/Hallsteinsdóttir in diesem Heft; Müller 2016; Müller/Hallsteinsdóttir 2016, in diesem Heft):

- (12) „Ordnung muss sein!“ hedder det på tysk - der skal være styr på tingene, i alle livets forhold, også hvad angår sproget. [„Ordnung muss sein!“ heißt es auf Deutsch – es muss alles unter Kontrolle sein, in allen Bereichen des Lebens, auch bezüglich der Sprache.‘]
- (13) Stævnet var lagt an i stor stil, arrangeret med sædvanlig tysk sans for „ordning“ og høflig takt og tone. [Der Wettbewerb war groß angelegt, organisieret mit dem üblichen deutschen Sinn für „Ordnung“ und höfliche Umgangsformen.‘]
- (14) „Herr Nielsen,“ begyndte han på normal tysk høflig vis, når jeg havde ringet ham op. [„Herr Nielsen,“ fng er an auf die übliche deutsche höfliche Art, als ich ihn angerufen hatte.‘]

Im Hinblick auf unsere Ergebnisse¹² plädieren wir dafür, den vielfach angenommenen, starken Einfluss der Massenmedien auf die Bildung und Tradierung von (nationalen) Stereotypen ggf. zu revidieren und die Erforschung von sprachlichen Stereotypen um andere textuelle und mediale Genres zu erweitern. Wichtig wäre dabei, die Untersuchungen von (gedruckten) Texten um andere Medien wie Filme, Karikaturen, literarische Werke, Spezialmagazine (cf. Gorbahn 2016 und in diesem Heft) und – insbesondere – Bildungsmedien zu berücksichtigen, wobei zusätzlich andere Zeichensysteme als Sprache einbezogen werden sollten. Unbestritten, dies möchten wir abschließend betonen, bleibt dabei die Tatsache, dass Korpusdaten unentbehrliche Sprachdaten sind, um die Manifestation von Stereotypen im Sprachgebrauch zu untersuchen.

¹² Weiterhin ist anzumerken, dass nur wenige der Befragten in der SMiK-Fragebogenerhebung angeben, sie würden ihr Wissen über Dänemark bzw. Deutschland aus den dänischen bzw. deutschen Medien haben (cf. Hallsteinsdóttir 2015: 22, 37).

Literatur

- Agger, Gunhild/Bentikow, Barbara/Hedetoft, Ulf (eds.) (1990): *Stereotyper i Europa*. Aarhus: Aarhus Universitetsforlag.
- Bewanger, Katrin/Kosta, Peter (eds.) (2005): *Stereotyp und Geschichtsmythos in Kunst und Sprache: Die Kultur Ostmitteleuropas in Beiträgen zur Potsdamer Tagung, 16.–18. Januar 2003*. Frankfurt a. M.: Lang.
- Bolten, Jürgen (2006): „Die Entwicklung von Nationalstereotypen im Globalisierungsprozess. Hypothesen zum Auftakt einer international durchgeführten Langzeituntersuchung zu Veränderungen des Deutschlandbildes bei Studierenden“. *Zeitschrift für interkulturellen Fremdsprachenunterricht* 11/2: <http://tujournals.ulb.tu-darmstadt.de/index.php/zif/article/view/393/381> [28.12.2015].
- Bolten, Jürgen (2007): *Interkulturelle Kompetenz*. 4. vollständig überarbeitete Auflage. Erfurt: Landeszentrale für politische Bildung: www.db-thueringen.de/servlets/DerivateServlet/Derivate-20394/interkulturellekompetenz.pdf [28.12.2015].
- CDI = *Chinesisch-Deutscher Imagereport* (2015). www.cdi.korpuspragmatik.de/index.php/de/projektbeschreibung [20.12.2015].
- D-SMiK = *Deutsches SMiK-Korpus* (2013). http://wortschatz.uni-leipzig.de/ws_spezial/?dict=deu_mixed_2011_dan [10.12.2015].
- Dąbrowska, Jarochna (1999): *Stereotype und ihr sprachlicher Ausdruck im Polenbild der deutschen Presse: eine textlinguistische Untersuchung*. Tübingen: Narr.
- DÄ-SMiK = *Dänisches SMiK-Korpus* (2013). http://wortschatz.uni-leipzig.de/ws_spezial/?dict=dan_mixed_2012_deu [10.12.2015].
- Dalmas, Martine et al (2015): „Bewertung durch Adjektive. Ansätze einer korpusgestützten Untersuchung zur Synonymie“. *Zeitschrift für Literaturwissenschaft und Linguistik*: 12–29.
- Demleitner, Elisabeth (2010): *Gentlemen und Nazis? Nationale Stereotype in deutschen und britischen Printmedien*. Würzburg: Universität Würzburg. <http://opus.bibliothek.uni-wuerzburg.de/files/3961/DissDemleitnerWespa8.pdf> [21.12.2014].
- Desmarais, Fabrice/Bruce, Toni (2009): “The Power of Stereotypes: Anchoring Images Through Language in Live Sports Broadcasts”. *Journal of Language and Social Psychology* 29/3: 338–263.
- Dunning, Ted (1993): “Accurate methods for the statistics of surprise and coincidence”. *Computational Linguistics* 19/1: 61–74.
- Fleischer, Wolfgang (1997): *Phraseologie der deutschen Gegenwartssprache*. Tübingen: Niemeyer.
- Farris, Nicole D./Davis, Mary Ann/Compton, D’Lane R. (eds.) (2014): *Illuminating How Identities, Stereotypes and Inequalities Matter through Gender Studies*. New York: Springer.
- Goldhahn, Dirk/Eckhart, Thomas/Quasthoff, Uwe (2012): “Building large monolingual dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages“. In: Calzolari, Nicoletta et al. (eds.): *LREC 2014, Ninth International Conference on Language Resources and Evaluation*. www.lrec-conf.org/proceedings/lrec2012/pdf/327_Paper.pdf [28.12.2015].

- Gorbahn, Katja (2016): „Böse Nazis, gute Deutsche und singende Soldaten: Deutschlandbezogene Stereotype in dänischen Spielfilmen über die Besatzungszeit“. In: Hallsteinsdóttir, Erla et al. (eds.): *Perspektiven der Stereotypenforschung*. Frankfurt a. M. etc., Lang: 101–122.
- Gottburgsen, Anja (1995): „Zur sprachlichen Inszenierung von Geschlecht – doing gender in Kontaktanzeigen“. *Zeitschrift für Germanistische Linguistik* 23: 257–283.
- Grzybek, Peter (1990): „Kulturelle Stereotype und stereotype Texte“. In: Koch, Walter A. (ed.): *Natürlichkeit der Sprache und der Kultur*. Bochum, Brockmeyer: 300–327.
- Hallsteinsdóttir, Erla (2015): *Ergebnisse aus der SMiK-Fragebogenuntersuchung typisch deutsch – typisch dänisch? || Resultater fra SMiK-spørgeskemaundersøgelsen typisk dansk – typisk tysk?* Odense: SMiK-Projekt. www.stereotypenprojekt.eu/projektresultate-r-1/fragebogen-typisch-deutsch-typisch-dänisch-spørgeskema-typisk-dansk-typisk-tysk/ [20.12.2015].
- Hallsteinsdóttir, Erla (2016): „Deutsch-dänische Nationalstereotype in der Sprache“. In: Hallsteinsdóttir, Erla et al. (eds.): *Perspektiven der Stereotypenforschung*. Frankfurt a. M. etc., Lang: 35–58.
- Harrington, Kate et al (eds.) (2008): *Gender and Language Research Methodologies*. Basingstoke: Palgrave Macmillan.
- Heier, Anke (2016): „Dänischer Rechtspopulismus und Tyskertøser – Zu nationalen Stereotypen in ein- und zweisprachigen Wörterbüchern“. In: Hallsteinsdóttir, Erla et al. (eds.): *Perspektiven der Stereotypenforschung*. Frankfurt a. M. etc., Lang: 13–34.
- Hellinger, Marlis/Bußmann, Hadumod (eds.) (2003): *Gender Across Languages. The linguistic representation of women and men*. Band 3. Amsterdam: John Benjamins.
- Heinemann, Margot (ed.) (1998): *Sprachliche und soziale Stereotype*. Frankfurt a. M.: Lang.
- Heyer, Gerhard/Quasthoff, Uwe/Wittig, Thomas (2005): *Text Mining: Wissensrohstoff Text. Konzepte, Algorithmen, Ergebnisse*. Herdecke/Bochum: W3L-Verlag.
- Kashima, Yoshihisa/Fiedler, Klaus/Freytag, Peter (eds.) (2008): *Stereotype Dynamics. Language-Based Approaches to the Formation, Maintenance, and Transformation of Stereotypes*. New York: Lawrence Erlbaum.
- Kendall, Shari/Tannen, Deborah (1997): „Gender and Language in the Workplace“. In: Wodak, Ruth (ed.): *Gender and Discourse*. London, Sage Publications: 81–105.
- Kilian, Jörg (2001): „Kritische Semantik. Für eine wissenschaftliche Sprachkritik im Spannungsfeld von Sprachtheorie, Sprachnorm, Sprachpraxis“. *Zeitschrift für Germanistische Linguistik* 29: 293–318.
- Kilian, Jörg (2003): „Wörter im Zweifel. Ansätze einer linguistisch begründeten kritischen Semantik“. *Linguistik online* 16: 159–170.
- Kilian, Jörg (2015): „Von *blonden Däninnen* aus deutscher Sicht. Nationale Stereotype und didaktische Sprachkritik – ein Zwischenbericht aus einem laufenden Forschungsprojekt“. In: Peschel, Corinna/Runschke, Kerstin (eds.): *Sprachvariation und Sprachreflexion in interkulturellen Kontexten*. Frankfurt a. M., Lang: 155–182.
- Lakoff, Robin T. (1975/2004): *Language and the Woman's Place. Text and commentaries*. Rev. Ausgabe (ed. by M. Bucholtz). New York: Oxford University Press.
- Langer, Roy (2003): *Die Darstellung Deutschlands in dänischen Medien. Eine Mediendiskursanalyse*. Wiesbaden: Deutscher Universitätsverlag.

- LCC = *Leipzig Corpora Collection* (s. a.). <http://wortschatz.uni-leipzig.de/> [18.12.2015].
- LImA = *Linguistische Imageanalyse* (2015).
<http://friedemann-vogel.de/index.php/forschung/abgeschlossene-projekte/20-linguistische-imageanalyse-lima> [28.12.2015].
- Motschenbacher, Heiko (2010): *Language, Gender and Sexual Identity: Poststructuralist Perspectives*. Amsterdam/Philadelphia: John Benjamins.
- Müller, Katarina Le (2016): „Thematisierung von Sprache in der deutsch-dänischen interkulturellen Kommunikation“. In: Hallsteinsdóttir, Erla et al. (eds.): *Perspektiven der Stereotypenforschung*. Frankfurt a. M., Lang: 59–76.
- Müller, Katarina Le/Hallsteinsdóttir, Erla (2016): „Stereotype im Fremdsprachenunterricht“. In: Hallsteinsdóttir, Erla et al. (eds.): *Perspektiven der Stereotypenforschung*. Frankfurt a. M., Lang: 233–256.
- Müller, Kathrin (2005): *Dänemarks Selbst- und Fremdbild im Kontext der europäischen Integration. Eine Medieninhaltsanalyse*. Kiel: Christian-Albrechts-Universität zu Kiel
http://macau.uni-kiel.de/receive/dissertation_diss_00001611 [28.12.2015].
- Nübling, Damaris (2009): „Zur lexikografischen Inszenierung von Geschlecht. Ein Streifzug durch die Einträge von Frau und Mann in neueren Wörterbüchern“. *Zeitschrift für Germanistische Linguistik*: 593–633.
- Prechter, Sabine (1999): “Women’s rights – children’s games: Sexism in learners’ dictionaries of English”. *Multilingua* 18/1: 47–68.
- Pümpel-Mader, Maria (2010): *Personenstereotype. Eine linguistische Untersuchung zu Form und Funktion von Stereotypen*. Heidelberg: Winter.
- Putnam, Hilary (1975): “The meaning of ‘meaning’”. In: Gunderson, Keith (ed.): *Language, Mind and Knowledge. Minnesota Studies in the Philosophy of Science*, VII. Minnesota, University of Minnesota Press: 131–191.
- Putnam, Hilary (1985): “The meaning of ‘meaning’”. *Mind, Language and Reality*. Philosophical Papers 2: 125–271.
- Quasthoff, Uta (1973): *Soziales Vorurteil und Kommunikation. Eine sprachwissenschaftliche Analyse des Stereotyps*. Frankfurt a. M.: Fischer-Athenäum-Taschenbuch.
- Quasthoff, Uwe (2009): „Korpusbasierte Wörterbucharbeit mit den Daten des Projekts *Deutscher Wortschatz*“. *Linguistik online* 39/3: 151–162.
- Quasthoff, Uwe/Fiedler, Sabine/Hallsteinsdóttir, Erla (eds.) (2011): *Frequency Dictionary German*. Leipzig: Leipziger Universitätsverlag.
- Rapp, Reinhard (2002): “The Computation of Word associations: Comparing syntagmatic and Paradigmatic Approaches”. *Proceedings of COLING-02, Taipei, Taiwan*. Online Publikation: http://delivery.acm.org/10.1145/1080000/1072235/p7-rapp.pdf?ip=84.165.98.36&id=1072235&acc=OPEN&key=4D4702B0C3E38B35.4D4702B0C3E38B35.4D4702B0C3E38B35.6D218144511F3437&CFID=764253162&CFTOKEN=34836274&__acm__=1458748047_4799d4eb2834e9c10a31e44dd9dbf6f2 [23.03.2016].
- Steyer, Kathrin (2013): *Usuelle Wortverbindungen. Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht*. Tübingen: Narr.
- Stocker, Christa (2005): *Sprachgeprägte Frauenbilder: Soziale Stereotype im Mädchenbuch des 19. Jahrhunderts und ihre diskursive Konstituierung*. Berlin: de Gruyter.

- Süssmuth, Hans (ed.) (1996): *Deutschlandbilder in Dänemark und England, in Frankreich und den Niederlanden*. Baden Baden: Nomos.
- Vogel, Friedemann (2010): „Ungarn – das Tor zum Westen. Das Bild Ungarns in deutschsprachigen Medien 1999–2009. Eine linguistische Imageanalyse“. *Jahrbuch der ungarischen Germanistik*: 87–124.
http://friedemann-vogel.de/media/texte/Vogel_ImageanalyseUngarn_2010.pdf
[20.12.2015].
- Wenzel, Angelika (1978): *Stereotype in gesprochener Sprache: Form, Vorkommen und Funktion in Dialogen*. Mannheim: Institut für deutsche Sprache.
- Wien, Charlotte (2005): *Ældrebilledet i medierne gennem 50 år: en undersøgelse af ældrestereotyper i dagbladene fra 1953 til 2003*. Odense: Syddansk Universitetsforlag.