

SABINE KIRCHMEIER

## empirische Grundlagen der offiziellen Rechtschreibung in Dänemark

### 1. Die Dänische Sprache

Ungefähr 5,8 Mio. Menschen in der Welt sprechen heute die dänische Sprache – die meisten davon in Dänemark. Dänisch hat seine Wurzeln im Germanischen und begann sich zwischen den Jahren 800 und 1000 aus dem Urnordischen als selbständige Sprache zu entwickeln. Noch heute weist Dänisch viele gemeinsame Züge mit dem Norwegischen und Schwedischen auf. In den meisten nordischen Ländern kann man sich auf Norwegisch, Schwedisch oder Dänisch verständlich machen, wenn man deutlich und langsam spricht. Untersuchungen in den letzten 20 Jahren (Delsing/Åkesson 2005) weisen jedoch darauf hin, dass das internordische Sprachverständnis besonders bei der jüngeren Generation auf immer größere Schwierigkeiten stößt, und dass die dänische Sprache im Norden nicht mehr so leicht verstanden wird wie früher. Das ist hauptsächlich den Unterschieden in der Phonetik und der Prosodie zuzuschreiben. Generell ist das Verhältnis zwischen Lautindruck und Schriftbild im Dänischen komplexer als in den beiden anderen Sprachen. Englisch ist daher immer häufiger die Sprache, in der sich junge Menschen in den nordischen Ländern zunächst bei ihren ersten Begegnungen verständigen.

Dänisch war viele Jahrhunderte eine dominierende Sprache in den nordischen Ländern. Bis Mitte des 19. Jahrhundert dominierte Dänisch in Norwegen und bis 1944 in Island. Bis 2009 war Dänisch offizielle Sprache in Grönland und ist noch heute eine der offiziellen Sprachen auf den Färöer-Inseln, die zusammen mit Grönland Teil der dänischen Reichsgemeinschaft sind. Seit 1945 hat der Einfluss der dänischen Sprache im Norden stark abgenommen und ihre Rolle als Lingua Franca ist nach und nach von der englischen Sprache übernommen worden.

Charakteristische Merkmale der Dänischen Sprache sind vor allem die unterschiedlichen Vokalqualitäten, der Einheitsakzent der dazu dient, Prozesse zu beschreiben: *læse a'vis* (Zeitung lesen), und der Stoßlaut, der semantisch distinktiv ist: *stien* (Pfad (+Stoß)) – *stigen* (Leiter (-Stoß)). Mit den anderen Zentralskandinavischen Sprachen hat Dänisch unter anderem den suffigierten bestimmten Artikel gemeinsam: *en mand* – *manden* (ein Mann – der Mann). Wie im Deutschen hat man im Dänischen die dynamische Bildung von Kom-

posita und vielen Partikelverben, aber im Gegensatz zum Deutschen oft mit teilweise lexikalisierte Bedeutung.

- *skrive af* (von jemandem abschreiben)
- *afskrive* (Einen Betrag abschreiben/jemanden abschreiben)

Im Gegensatz zum Deutschen ist das Kasussystem nur noch rudimentär bei den persönlichen Pronomina zu erkennen. Die Wortstellung im Dänischen ist daher aus topologischer Sicht relativ fixiert. Nur das Vorfeld ist auswechselbar:

Vorfeld	Mittelfeld			Nachfeld		
	v	n	a	V	N	A
<i>Jeg</i>	<i>har</i>		<i>ikke</i>	<i>hentet</i>	<i>bogen</i>	<i>i dag</i>
<i>I dag</i>	<i>har</i>	<i>jeg</i>	<i>ikke</i>	<i>hentet</i>	<i>bogen</i>	
<i>Bogen</i>	<i>har</i>	<i>jeg</i>	<i>ikke</i>	<i>hentet</i>		<i>i dag</i>
Das Buch	habe	ich	nicht	geholt		heute

Besonders für Deutsche sollte die dänische Schriftsprache eigentlich leicht verständlich sein. Durch die vielen politischen, wirtschaftlichen und kulturellen Kontakte zwischen den beiden Nachbarländern, vermittelt z.B. durch die Hanse im Ostseeraum, haben sich viele deutsche, insbesondere niederdeutsche Wörter und Redewendungen in der dänischen Sprache eingebürgert, und man kann noch heute dänische Sätze bilden, in denen alle inhaltstragende Wörter aus dem Deutschen entliehen sind (Torp 2004).

Eine Untersuchung des aktuellen Wortschatzes anhand eines mittelgroßen dänischen Wörterbuches, *Nudansk Ordbog* (2001), mit 70.000 Einträgen (Jarvad/Sandøy 2008 (Hg.)), hat gezeigt, dass der Anteil der Lehnwörter aus der deutschen Sprache mit ca. 3,6% an erster Stelle steht, dicht gefolgt von Latein (besonders im Mittelalter und bis 1850 als Sprache der Wissenschaft) mit 3,4% und Französisch (im 18. Jahrhundert) mit 2,4%.

Der Einfluss des Englischen auf den etablierten dänischen Wortschatz steht mit 1,8% erst an vierter Stelle und ist bisher gemessen an der Anzahl der Wörter, die aus beiden Sprachen in die dänische Sprache übernommen worden sind, nur ungefähr halb so groß wie der des Deutschen. Dieses Verhältnis wird sich in Zukunft jedoch mit großer Wahrscheinlichkeit ändern. Seit 1945 ist Englisch/Amerikanisch mit ca. 90% der neu etablierten Lehnwörter die wichtigste Quelle des dänischen Wortimportes. Deutsch ist die zweitgrößte Quelle, liefert jedoch nur 1,25% des Zuwachses.

Auch in der dänischen Rechtschreibung hat man sich mit dem Einfluss aus Deutschland auseinandergesetzt. Vor 1948 schrieb man noch wie im Deutschen alle Substantive groß, doch unter den Eindrücken des Zweiten Welt-

krieges hatte sich das politische Klima so weit gewandelt, dass es möglich wurde, nicht nur die Kleinschreibung einzuführen, sondern auch den im Schwedischen und Norwegischen gebräuchlichen Buchstaben *å*, der den Laut [ɔ] repräsentiert, der bis dahin *aa* geschrieben wurde. Auf diese Art konnte man von einem Tag auf den anderen sogar im Schriftbild signalisieren, dass man mit dem Deutschen möglichst wenig gemeinsam haben wollte und sich eher den nordischen Nachbarn verbunden fühlte. In der Praxis dauerte es jedoch mehrere Jahre, bis diese Änderungen überall durchgeführt waren.

## 2. **Dansk sprognavn – das offizielle nationale dänische Sprachinstitut**

*Dansk Sprognavn* – das offizielle nationale dänische Sprachinstitut – ist eng mit der Rechtschreibreform von 1948 verknüpft, da es, als es 1955 gegründet wurde, die Aufgabe bekam, das erste offizielle Rechtschreibwörterbuch nach der Reform herauszugeben. Seit 1997 ist die Arbeit des Instituts gesetzlich verankert.

Es gibt ein Rechtschreibgesetz, das bestimmt, dass alle öffentlichen Einrichtungen, wie das Parlament, die Gerichtshöfe, öffentliche und öffentlich unterstützte Schulen usw., der Rechtschreibung folgen müssen, die das Sprachinstitut durch die Herausgabe des offiziellen Rechtschreibwörterbuches festlegt.

Neben dem Rechtschreibgesetz legt ein weiteres Gesetz die Aufgaben des nationalen Sprachinstitutes fest. Das Institut muss demnach:

- die Entwicklung der dänischen Sprache beobachten,
- über Sprache und Sprachgebrauch informieren und
- die Rechtschreibung festlegen.

Die Aufgaben des Institutes fallen somit in drei Hauptbereiche: Forschen, Informieren und Normieren. Das Institut sieht den Forschungsauftrag als Grundlage für die beiden anderen Tätigkeiten, und deswegen haben 90% der 18 Mitarbeiter einen sprachwissenschaftlichen Hintergrund. Neben diesen drei Tätigkeiten hat das Institut die Verpflichtung, mit entsprechenden Institutionen in den anderen nordischen Ländern zusammenzuarbeiten, unter anderem um die Verständigung mit Hilfe der nordischen Nachbarsprachen zu fördern.

Seit 2015 betreibt das Sprachinstitut auch ein Sekretariat für die dänische Gebärdensprache.

### 3. kodifizierung der dänischen rechtschreibung

Durch das Rechtschreibgesetz und durch sprachliche Bestimmungen in anderen Gesetzen ist der Status der dänischen Sprache nur indirekt festgelegt. Es gibt keinen Abschnitt im dänischen Grundgesetz (*Grundloven*) von 1848, der die dänische Sprache als Landessprache festlegt. Es gibt auch kein Sprachgesetz, das den Status der dänischen Sprache, z.B. als Unterrichtssprache und Amtssprache, vorschreibt.

Dass Dänisch de facto Hauptsprache des Landes ist, geht z.B. daraus hervor, dass man nur mit einer Sondergenehmigung in der Grundschule und im Sekundarbereich eine andere Unterrichtssprache als Dänisch benutzen darf. Die Sprache der Gesetzgebung und der Rechtsprechung ist hingegen ausdrücklich Dänisch, während Dänisch als Amtssprache im Verwaltungsgesetz lediglich in dem Sinne festgelegt ist, dass man dänischen Behörden auferlegt, mit den Bürgern in einem klaren und verständlichen Dänisch zu kommunizieren.

Unmittelbar scheint dem nationalen Sprachinstitut im Bereich der Rechtschreibung eine bedeutende Machtposition zugewiesen zu sein. Es ist jedoch auch gesetzlich festgelegt, dass die wissenschaftlichen Mitarbeiter des Institutes die Verantwortung für die dänische Rechtschreibung nicht allein verwalten können. Dem Institut ist ein Gremium beigeordnet, das aus 43 Personen besteht, die entweder persönlich eine große Kompetenz im Bereich der dänischen Sprache besitzen oder Institutionen repräsentieren, die große Bedeutung für die Entwicklung der Sprache haben, z.B. die Staats-,<sup>1</sup> Bildungs-, Justiz- und Kulturministerien, das Fernsehen, die Zeitungen, Journalisten-, Schriftsteller-, Schauspieler- und Übersetzerverbände, Universitäten und Kulturinstitutionen.

Betrachten die Repräsentanten eine vorgeschlagene Änderung als prinzipiell, muss sie dem Kulturminister vorgelegt werden, der im Einvernehmen mit dem Bildungsminister darüber entscheidet, ob die vorgeschlagene Änderung vorgenommen werden soll. Es ist schon vorgekommen, dass vorgeschlagene Änderungen zurückgewiesen werden. So wurde z.B. 1986 der Vorschlag, neben der traditionellen Form *gymnasium* die angepasste Form *gymnasie* zuzulassen, von dem damaligen Kulturminister verworfen. Erst 2012 wurde diese Änderung eingeführt. Auch in Hinblick auf die Zeichensetzung gab es im Laufe der Jahre politischen Einfluss, sodass heute im Rahmens des Kommasystems zwei Wahlmöglichkeiten existieren.

In der dänischen Gesetzgebung gibt es zwei Prinzipien, die die Festlegung der Rechtschreibung bestimmen. Das eine nennt sich das Traditionsprinzip, nach dem die Rechtschreibung der dänischen Sprache prinzipiell festliegt.

<sup>1</sup> *Staatsministeriet* entspricht dem *Bundeskanzleramt* in Deutschland.

Das andere Prinzip ist das Sprachgebrauchsprinzip, nach dem man die Rechtschreibung der Praxis angleicht, die bei guten und sicheren Benutzern der Schriftsprache zu finden ist. (Sprognævnsbekendtgørelsen (Erlass zur Organisation und Arbeitsform des dänischen Sprachrates) 1997, § 1, stk. 3 und 4).

Dem **Traditionsprinzip** nach werden Wörter und Wortformen im Dänischen in Übereinstimmung mit der Praxis geschrieben, die seit der Verordnung No. 24 vom 27. Februar 1892 geltend ist, und die seit 1955 in den Rechtschreibwörterbüchern zum Ausdruck kommt, die das nationale Sprachinstitut herausgegeben hat. Dem Traditionsprinzip zufolge liegen die Schreibweisen des existierenden Wortschatzes prinzipiell fest mit Ausnahme der Justierungen, die durch das Sprachgebrauchsprinzip veranlasst werden.

Ein Element des Traditionsprinzips ist das Prinzip, dass Fremdwörter, die im Dänischen allgemein verbreitet sind, nach den gleichen Regeln geschrieben werden, die für ursprünglich Dänische Wörter und für ältere Lehn- und Fremdwörter gelten.

Dieses Prinzip gilt hauptsächlich Wörtern aus dem Griechischen, Lateinischen und Französischen und Wörtern, die griechische, lateinische und französische Bestandteile enthalten. Wörter aus anderen Sprachen, insbesondere neuere Wörter, werden in der Regel in Übereinstimmung mit der Schreibweise dieser Sprachen oder der internationalen Praxis geschrieben.

Dem **Sprachgebrauchsprinzip** zufolge werden Wörter und Wortformen im Dänischen in Übereinstimmung mit der Praxis geschrieben, die bei guten und sicheren Benutzern der Schriftsprache zu finden ist. Damit sind nicht Einzelpersonen gemeint, sondern eher die Praxis, die in Texten zu erkennen ist, die von Menschen geschaffen werden, die professionell mit der Sprache arbeiten.

Das offizielle dänische Rechtschreibwörterbuch (Retskrivningsordbogen 2012) enthält ca. 65.000 Wörter. Die elektronische XML-Datenbank, von der das Wörterbuch automatisch generiert und publiziert wird, ist darüber hinaus mit sämtlichen flektierten Formen und dementsprechender morphosyntaktischer Information versehen. Die Datenbank enthält insgesamt ungefähr 450.000 flektierte Formen und wird in vielen Zusammenhängen für sprachtechnologische Zwecke genutzt, z.B. für Rechtschreibkontrolle, Suchmaschinen, elektronische Wörterbücher, Computerspiele, Kontrolle von Sprachinput, Wörtertrennung, Unterrichtsprogramme, Generierung von Login-Namen, usw.

Das Sprachinstitut legt großen Wert darauf, dass die Information über die offizielle Rechtschreibung in so vielen Kontexten wie möglich zugreifbar ist.

Das Rechtschreibwörterbuch wird seit 2014 ein Mal jährlich aktualisiert. Es werden neue Wörter eingesetzt, neue Informationen in existierende Artikel eingefügt (z.B. Zusammensetzungs- und Gebrauchsbeispiele, Bedeutungen

und fehlende Flexionsinformation), und neue Beispiele, die z.B. eine Rechtschreibregel besser erläutern. So wächst die Wortmenge ständig an. Ungefähr alle 10 Jahre wird eine gründlichere Revision vorgenommen, wobei veraltete Worte und Wortformen entfernt und existierende Formen der aktuellen Schreibweise angepasst werden, natürlich erst nach vorheriger Befragung der Repräsentanten und zuständigen Ministerien. Die veralteten Worte verschwinden jedoch nicht völlig, sondern sind weiterhin in einem historischen Wörterbuchportal im Internet zugreifbar, in dem man die Entwicklung der Rechtschreibung von 1872 bis zur Gegenwart verfolgen kann. Es wird daran gearbeitet, auch die Änderungen, die seit 1777 vorgenommen worden sind, in diesem Portal zu erfassen.

Zurzeit bemerkt man folgende Entwicklungstendenzen in der dänischen Rechtschreibung:

- Zunehmende Getrenntschreibung von Komposita, insbesondere Substantivkomposita, teilweise so wie man sie auch im Deutschen antrifft, z.B. Substantiv + Adjektiv: *video overvåget* statt *videoovervåget* ('video überwacht' statt 'videoüberwacht'), am häufigsten jedoch Substantiv-Substantivkomposita: *kalve lever* statt *kalvelever* ('Kälber leben' statt 'Kalbsleber') (Heidemann Andersen 2015). Wie im Deutschen verursacht die Getrenntschreibung oftmals auch ganz andere Lesarten der Wortfolge.
- Zunehmende Reduktion der Verbalbeugung (Präsens ⇒ Infinitiv: *jeg lærer/ jeg lære* (ich lerne/ich lernen). Hier ist das Problem, dass die Form mit dem fehlenden *-r* mit dem Infinitiv identisch ist und oft nicht vor der Rechtschreibkontrolle erfasst wird. Hinzu kommen Änderungen in der Aussprache, insbesondere Abschwächungen im Auslaut. Junge Menschen können das finale *-r* kaum hören, insbesondere wenn es nach einer größeren Anzahl von Konsonanten auftritt, z.B. *ændrer* (ich ändere) (Jervelund/Schack 2016).
- Zunehmender Einfluss des Englischen auf den Wortschatz. Die steigende Anzahl englischer Lehnwörter ist insbesondere in der Wirtschaft, in der Informationstechnologie und an den Universitäten deutlich. 56% der dänischen Industrieunternehmen gaben 2015 an, Englisch als Konzernsprache zu benutzen. Dänemark liegt in Bezug auf die Anwendung des Englischen als Unterrichtssprache an den Hochschulen in Europa mit an erster Stelle (Kirchmeier-Andersen et al. 2012). Neben den neuen Lehnwörtern sieht man in zunehmendem Maße neue Lesarten existierender griechischer oder lateinischer Wörter, wie z.B. *adoptere*<sup>1</sup> (adoptieren), *adoptere*<sup>2</sup> (neu: sich aneignen – to adopt a point of view), *adressere*<sup>1</sup> (mit einer Adresse versehen), *adressere*<sup>2</sup> (neu: zur Sprache bringen – to address a problem).
- Beginnender Einfluss des Englischen auf das Flexionssystem. Im Flexionssystem zeigt sich der englische Einfluss durch die Beibehaltung des engli-

schen Pluralismorphems -s. Englische Lehnwörter werden im Plural teilweise nach dänischem Muster flektiert (*baby-er*), oder sie behalten die englische Form (*airbags*) oder beides (*apps/app-er*).

- Beginnender Einfluss des Englischen auf die Syntax. Hier bemerkt man zunächst nur kleine Veränderungen, insbesondere bei Konstruktionen mit Imperativ und Adverbium: Statt *vent venligst* (warte bitte) hört man immer häufiger *venligst vent* (bitte warte – please wait).

#### 4. empirische Grundlagen

Seit 2008 arbeitet der das Sprachinstitut intensiv mit korpuslinguistischen Methoden, d.h. schriftliche Quellen, z.B. Tageszeitungen, Blogs und soziale Medien, werden systematisch überwacht und eingesammelt. Die Sektoren, in denen die Vorgabe des Sprachinstitutes Pflicht ist, werden besonders überwacht. Es werden Texte aus dem Ausbildungsbereich von der Grundschule bis zu den ersten Semestern des Universitätsunterrichtes, und Texte aus dem öffentlichen Bereich, insbesondere öffentliche Informationen im Internet und Texte, die direkt an Bürger gerichtet sind, eingesammelt und analysiert. Die Texte aus den übrigen Bereichen dienen unter anderem dazu, Ergebnisse zu kontrastieren.

In verschiedenen Forschungsprojekten werden den gewählten Forschungsfragen entsprechend relevante Korpora zusammengestellt. Die Texte dienen auch dazu, die Norm in den unterschiedlichen Kontexten zu erschließen und die relative Verbreitung bestimmter Varianten festzustellen.

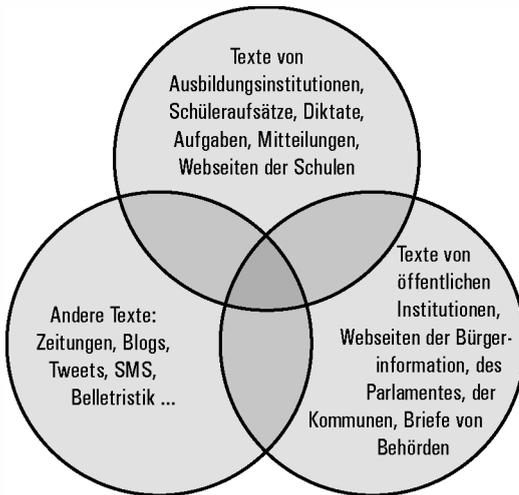


Abb. 1: Zusammensetzung der Quellen für korpuslinguistische Analysen

Mit verschiedenen Korpusprogrammen können relative und absolute Frequenzen erstellt werden, die die Entwicklung und Präferenzen in der geschriebenen Sprache verdeutlichen. Z.B. wird deutlich, dass bei neuen Lehnwörtern, wie dem Wort *App*, im Plural das englische Pluralmorphem *-s* dem dänischen *-er* vorgezogen wird. Das Verhältnis ist 1:10.



Abb. 2: Vorkommen der Form *apps* (insgesamt 930) in einem Zeitungskorpus von ca. 400 Mio. Wörtern

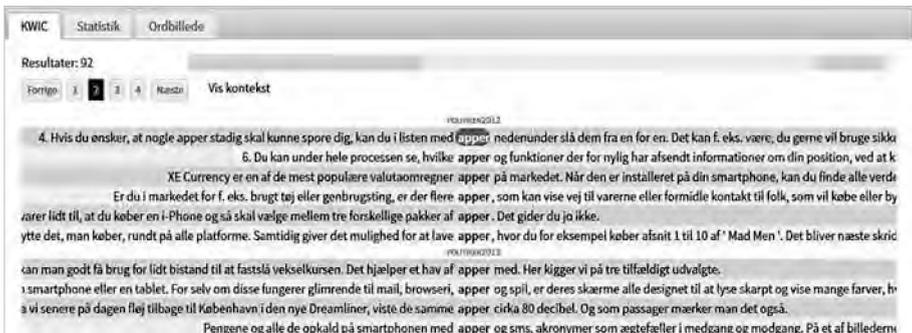


Abb. 3: Vorkommen der Form *apper* (insgesamt 92) in einem Zeitungskorpus von ca. 400 Mio. Wörtern

In diesem Fall sind im Wörterbuch beide Morpheme beibehalten, in anderen Fällen entscheidet man sich für die eine oder andere Form.

Auf welcher Grundlage eine solche Entscheidung getroffen wird, ist keine leichte Frage. Sollte die Frequenz allein ausschlaggebend sein, müssen an die Qualität und Repräsentativität der verwendeten Korpora große Anforderungen gestellt werden. Das heißt, die Texte müssten im Idealfall individuell anhand ihrer sprachlichen Qualität ausgewählt und ausgewogen zusammengesetzt werden. Das ist bei den großen Datenmengen, die mittlerweile in der Korpuslinguistik verwendet werden, oft nicht möglich. Deshalb sind die großen Textsammlungen in den meisten Wörterbuchredaktionen weiterhin nur ein Ergänzung zu den traditionellen Methoden: die manuelle Sammlung von neuen Wörtern und die jahrelange Erfahrung der Lexikografen. Das Sprach-

institut experimentiert seit einigen Jahren mit neuen Arbeitsmethoden, um dieses Problem zu lösen. Dazu mehr im folgenden Abschnitt.

Alternative Formen lassen sich nie ganz umgehen und existieren auch in offiziellen Normen anderer Sprachen, die allgemein als relativ restriktiv angesehen werden. Z.B. im französischen *Le Grand Robert* (2009) hat man einen Großteil der Änderungen, die um 1990 beschlossen wurden, eingearbeitet. Heute enthält das Buch 6.000 Wörter, bei denen sowohl eine traditionelle als auch eine alternative Schreibweise autorisiert ist, was ungefähr 6% der 100.000 Wörter ausmacht. Zum Vergleich existieren in dem aktuellen dänischen Rechtschreibwörterbuch ungefähr 1.500 Wörter von 65.000 (ca. 2%), bei denen mehrere Alternativen erlaubt sind. Das ist nicht immer ein Vorteil für die Benutzer, da man öfter nachschlagen muss. Sowohl von privater als auch von öffentlicher Seite wird oft verlangt, so wenig Alternativen wie möglich zuzulassen. Die Entscheidung für die eine oder andere Variante betrachtet das Sprachinstitut jedoch eher als eine Frage der Sprachpolitik der einzelnen Organisation.

Die Behandlung von alternativen Schreibweisen ist im Norden sehr unterschiedlich. In Norwegen gibt es neben beiden offiziellen Standardvarianten Bokmål und Nynorsk eine große Anzahl von ebenfalls autorisierten Nebenformen, die frei kombiniert werden können, was die Kombinationsmöglichkeiten der verschiedenen Varianten innerhalb eines Satzes schier explodieren lässt. In Schweden dagegen schreibt Svenska Akademiens Ordlista überwiegend nur eine Schreibweise vor und gibt bei alternativen Formen sogar an, welche zu bevorzugen ist.

## 5. Der gute sichere Benutzer der Schriftsprache

Ein Problem in der Anwendung von korpuslinguistischen Methoden bei der Festlegung der Rechtschreibnorm ist es, den Status der Korpus Texte zu bestimmen: Sind die gefundenen alternativen Schreibweisen Ausdruck für eine reelle Variation, d.h. liegt dahinter eine bewusste Überlegung des Verfassers oder eine neue Konvention, oder ist die Variation zufällig, d.h. dreht es sich lediglich um einen relativ weit verbreiteten Schreibfehler? Laut dem oben zitierten **Sprachgebrauchsprinzip** werden Wörter und Wortformen im Dänischen in Übereinstimmung mit der Praxis geschrieben, die bei guten und sicheren Benutzern der Schriftsprache zu finden ist. Die Frage ist, wie man diese identifiziert.

Das Sprachinstitut hat viele Jahre lang manuell die Norm von einer breiten Auswahl von Zeitschriften, Büchern und anderen Texten abgeleitet, die von professionellen Textverfassern produziert wurden. Infolge der immer weitergreifenden Digitalisierung und dem beschleunigten Arbeitstempo z.B. in den

Medien, werden viele Texte mit deutlich verminderter Sorgfalt produziert und nur noch mit notdürftiger Qualitätskontrolle publiziert. Manche der großen dänischen Tageszeitungen überlassen das Korrekturlesen nahezu ganz dem einzelnen Journalisten, oder setzen die Korrektur nur noch ein, wenn ein Artikel frühzeitig eingeliefert wird. Das bedeutet, dass derselbe Journalist Texte von ganz unterschiedlicher Qualität veröffentlichen kann.

Auch die Einbeziehung von Texten aus dem Internet und den sozialen Medien werfen Probleme auf. Untersuchungen in den sozialen Medien haben erwiesen, dass in manchen Zusammenhängen sehr auf korrekte Sprache geachtet wird und quasi dieselben Normen herrschen wie in anderen Bereichen. In anderen Zusammenhängen etablieren bestimmte Gruppen z.B. auf Facebook eigene Regeln und betrachten bestimmte orthographische Varianten als einen Teil ihrer Identität (Hyttel-Sørensen/Stæhr 2014). Diese Entwicklungen sind in einem großen, diversen Korpus oft schwer zu identifizieren und verwirren das Bild.

Stellen wir uns vor, dass die Redakteure des Rechtschreibwörterbuches erwägen, das Wort *autenticitet* (Authentizität) in die nächste Ausgabe des Wörterbuches aufzunehmen. In der großen dänischen Medien-Datenbank Infomedia sieht man die lange Form *autenticitet* in etwa 75% der Suchergebnisse, während die kürzere Form *autencitet* in 25% der Ergebnisse auftritt. Wie aber kann man erkennen, welche dieser Formen von den guten, sicheren Benutzern der Schriftsprache geschrieben sind, von denen im Sprachgebrauchsprinzip die Rede ist?

Das erörtern Mitarbeiter des Sprachinstitutes in einem Forschungsprojekt mit dem Ziel, bessere Methoden zu entwickeln, um die Texte zu identifizieren, die maximal mit der Norm übereinstimmen, und zwar in der Annahme, dass gute und sichere Verfasser, die qualitativ gute Texte schreiben, auch die Norm eher berücksichtigen als andere, und dass Abweichungen von der offiziellen Norm in diesen Texten daher ernster genommen werden müssen als Abweichungen, die in weniger normkonformen Texten vorkommen (Diderichsen/Christensen/Schack 2015).

Konkret will man eine Methode entwickeln, die es ermöglicht, die einzelnen Texte in einem Korpus nach dem Grad ihrer Normkonformität einzustufen – ein sogenanntes orthographisches Qualitätsbarometer: An der Spitze des Barometers findet man die Texte, die völlig mit der Norm übereinstimmen, und am Ende die Texte, die am weitesten davon entfernt sind. Jeder Text im Korpus soll also im Verhältnis zur Norm klassifiziert und im Barometer eingeordnet werden.

Zur Erstellung einer ersten Version des Qualitätsbarometers wurden zwei Elemente kombiniert: Erstens ein Korpus von ca. 370.000 Zeitungstexten aus sieben dänischen Tageszeitungen von 2004-2014, insgesamt 200 Millionen

Wörter. Zweitens eine Liste mit besonders typischen Fehlern. Schreibfehler lassen sich in zwei Kategorien aufteilen: Die eine enthält die Schreibfehler, die sogenannte Nicht-Wörter ergeben, also Wörter, die nicht der autoritativen Schreibweise entsprechen, und die prinzipiell in jedem Kontext als Fehler eingestuft werden. Z.B. *hierachisch* statt *hierarchisch*. Daneben gibt es auch Fehler, die nur im Kontext als Fehler zu erkennen sind, z.B. *Muße* oder *Muse*.



Abb. 4: Qualitätsbarometer zur Beobachtung der Normkonformität von Texten

Zunächst hat man sich auf die Nicht-Wörter konzentriert, da diese sich in jedem Kontext eindeutig identifizieren lassen und somit einen sicheren Ausgangspunkt geben. Die Liste der Fehler, die aus Nicht-Wörtern besteht, enthält 20.000 verschiedene nichtkontextuelle Fehlerformen oder Nicht-Wörter, die sich auf 90 verschiedene Fehlertypen verteilen, z.B. fehlende Kompositionsfugen, *-tion* statt *-sion* usw. Die verschiedenen Formen stammen unter anderem aus einem Projekt, in dem wir die Funktionsweise verschiedener Rechtschreibkontrollen untersucht haben.

Es stellt sich natürlich die Frage, ob Texte, die keine oder nur wenige dieser nichtkontextuellen Fehler aufweisen, auch generell von guten, sicheren Verfassern geschrieben sind und somit dem Sprachgebrauchsprinzip entsprechen. Texte können unterschiedliche strukturelle, rechtschreibmäßige, formulierungsmäßige und argumentative Qualitäten aufweisen. Bei guten Texten scheinen diese Merkmale einander zu unterstützen. Ist eins der Merkmale vertreten, sind die anderen es meist auch. Die Annahme liegt nahe, dass Texte, die viele nichtkontextuelle Fehler enthalten, auch im Hinblick auf die anderen Merkmale qualitativ zu wünschen übrig lassen.

Das wurde in einem internen Experiment untersucht (Schack/Diderichsen 2015): Anhand des Korpus wurden automatisch drei Textpaare erstellt, die thematisch, genremäßig und im Umfang einander gleich waren, z.B. zwei

Fußballreportagen, zwei Präsentationen eines neuen Autos usw. Sie waren im gleichen Jahr erschienen, die Verfasser waren ungefähr gleich alt, gleichen Geschlechtes und waren mit ungefähr der gleichen Textmenge im Korpus repräsentiert. Der einzige Unterschied zwischen den Texten in jedem Paar war die Menge der nichtkontextuellen Fehler. Der eine Text enthielt keinen dieser Fehler, der andere wies eine größere Menge nichtkontextueller Fehler auf. Diese Fehler wurden korrigiert, die Texte wurden anonymisiert, und danach wurden acht Mitarbeiter des Sprachinstitutes gebeten, die Textpaare zu beurteilen.

In den ersten beiden Paaren wurde jeweils der Text mit den wenigsten nichtkontextuellen Fehlern von den Mitarbeitern eindeutig als der beste Text klassifiziert. Obwohl die nichtkontextuellen Fehler nicht sichtbar waren, bekamen die ursprünglich fehlerhaften Texte Prädikate wie „unsicher“, „geschwollen“, „unpräzise“, „häufige Stilbrüche“ und „mangelhafte Kohärenz“. Die Bevorzugung der fehlerfreien Texte wurde generell mit Bemerkungen wie „flüssiger Stil“, „leicht zu lesen“, „gut disponiert“, „durchgehend objektiver“ und „informativ“ begründet.

Im letzten Fall dagegen wurde der Text mit den wenigsten kontextuellen Fehlern als der schlechteste Text beurteilt. Es stellte sich heraus, dass in diesem Text sehr viele kontextuelle Fehler vorkamen, die von dem automatischen Auswahlprozess nicht berücksichtigt worden waren. Das bestärkt die Annahme, dass die Anzahl der Fehler mit der erlebten Qualität der Texte zwar korreliert, dass aber die nichtkontextuellen Fehler allein nicht ausreichen, um die besten Texte zu identifizieren.

Betrachtet man die kontextuellen und anderen Fehler, z.B. Kommafehler, in den drei Texten, wird deutlich, dass nichtkontextuelle Fehler durchaus als Indikation für die Textqualität dienen können.

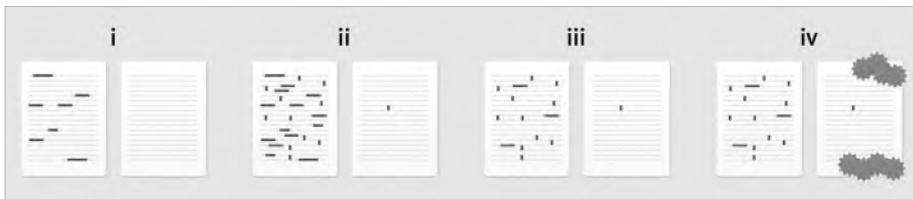


Abb. 5: Verschiedene Stadien der Texte und ihre Beurteilung

In Abbildung 5 sind die verschiedenen Stadien der Texte und ihre Beurteilung illustriert:

- i. Zwei ähnliche Texte werden ausgewählt. Der eine enthält mehrere nichtkontextuelle Fehler, der andere keine.

- ii. Beide Texte werden auf weitere Fehler untersucht. Der erste Text weist mehrere nichtkontextuelle Fehler, einen Kongruenzfehler und zahlreiche Kommafehler auf. Der zweite Text enthält nur einen Kommafehler.
- iii. Die nichtkontextuellen Fehler werden korrigiert. Die restlichen Fehler werden in beiden Texten stehengelassen.
- iv. Der Text mit den wenigsten nichtkontextuellen Fehlern wird von allen Beurteilern bevorzugt.

Unsere Untersuchungen haben, wenn sie sich weiterhin als haltbar erweisen, interessante Perspektiven. Z.B. ist es anhand der nichtkontextuellen Fehler möglich, festzustellen, welche Tageszeitungen in unserem Korpus die wenigsten Fehler zu enthalten scheinen, was tatsächlich damit korreliert, welche Zeitungen nach eigener Aussage am meisten darauf Wert legen, eine gründliche Textkorrektur vorzunehmen.

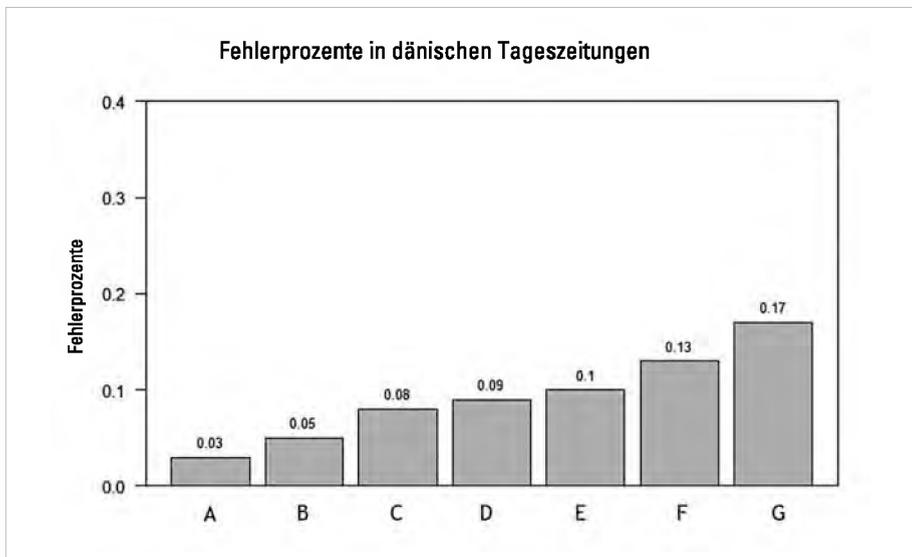


Abb. 6: Fehlerprozent in dänischen Tageszeitungen

Es lässt sich auch feststellen, dass die Anzahl der Fehler im Laufe der letzten zehn Jahre geringer geworden ist, was wahrscheinlich Verbesserungen in der automatischen Rechtschreibkontrolle zuzuschreiben ist.

Generell sind die Fehlerprozent in dem untersuchten Korpus relativ gering. Insgesamt enthalten die Texte 10 Mio. Wörter, die anhand unserer Fehlerliste potenziell falsch geschrieben werden können. Es treten aber in der Praxis nur 10.000 Fehler auf, d.h. nur 1 Promille der möglichen nichtkontextuellen Fehler werden tatsächlich gemacht und nicht im Redaktionsverfahren erfasst.

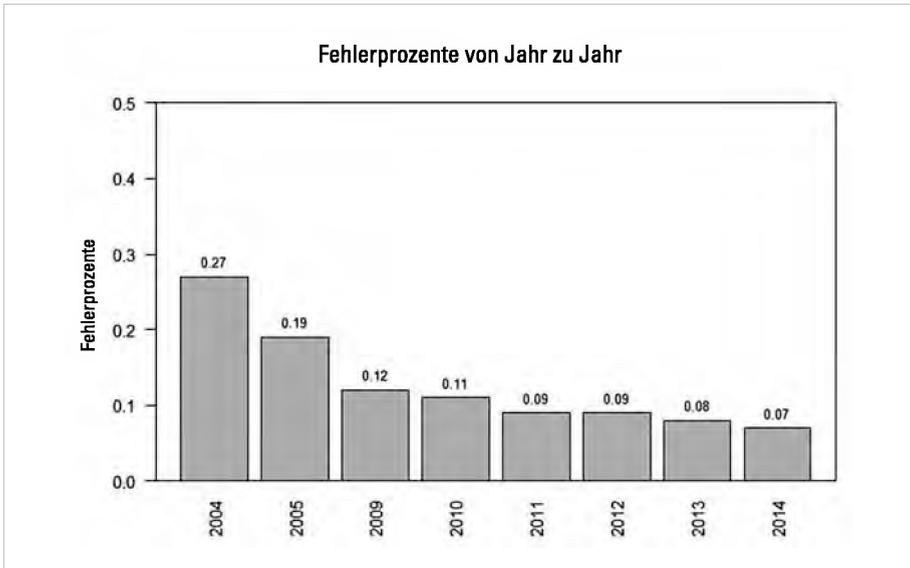


Abb. 7: Fehlerprozent von Jahr zu Jahr

Die Frage stellt sich nun, ob der Zusammenhang zwischen orthographischen Fehlern und der generellen Textqualität noch zuverlässiger nachgewiesen werden kann. In einem der Texte haben nichtkontextuelle Fehler im wesentlichen Grad die Beurteilung beeinflusst. Deshalb wird im weiteren Verlauf des Projektes daran gearbeitet, auch die kontextuellen Fehler automatisch zu erfassen. Das ist eine weitaus schwerere Aufgabe, da man zu diesem Zweck den Kontext genau analysieren muss, um z.B. morphologische oder syntaktische Fehler zu erkennen. Will man herausfinden ob *laufe* im Kontext *wir laufe* ein Fehler ist, muss man also den Numerus des Pronomens *wir* erfassen, und dazu auf das Beugungsmuster des Verbs *laufen* im Computer Zugriff haben.

Obwohl wir bisher nur vorläufige Ergebnisse zu verzeichnen haben, ist es nicht unwahrscheinlich, dass wir in absehbarer Zeit unser Qualitätsbarometer in Gebrauch nehmen können. Wenn nicht als einziges Mittel, um Veränderungen in der Rechtschreibung beurteilen zu können, dann doch als eine Methode, mit der Varianten schon erheblich präziser beurteilt werden können als mit einer rein quantitativen Betrachtungsweise.

Man kann sich auch vorstellen, dass die Methode zu anderen Zwecken verwendet werden kann. In mehreren Studien der Sprachformen in den sozialen Medien (Androutopoulos 2011; Hyttel-Sørensen/Stæhr 2014) wurde festgestellt, dass die korrekte Rechtschreibung auch in diesen Kontexten weiterhin eine wichtige Rolle spielt, und dass sich weitgehend die Konventionen der klassischen Medien etabliert haben. Die Bedenken vieler Sprachforscher, dass

sich in den sozialen Medien selbständige Konventionen entwickeln, die die Schreibenden verunsichern, und die korrekte Rechtschreibung somit an Bedeutung verliert, haben sich bis jetzt als unbegründet erwiesen. Mit der entwickelten Methode wäre es durchaus möglich, diese Ergebnisse auch quantitativ zu dokumentieren, und die Entwicklung weiterhin zu verfolgen.

## 6. Literatur

- Androutsopoulos, Jannis (2011): Language Change and Digital Media. In: Kristiansen, Tore et al. (Hg.): Standard Languages and Language Standards in a Changing Europe. Oslo, S. 145-159.
- Deising, Lars-Olof/Åkesson, Katarina L. (2005): Håller språket ihop Norden? En forskningsrapport om ungdomars förståelse av danska, svenska och norska. (= Tema Nord 573). Kopenhagen.
- Diderichsen, Philip/Christensen, Sune J./Schack, Jørgen (2015): Ranking corpus texts by spelling error rate. Poster. Conference on Corpus Based Research in the Humanities. Warschau.
- Heidemann Andersen, Margrethe (2015): Særskrivningsfejl i de nye medier og i skolen. In: Duncker, Dorthe/Jensen, Eva S./Ravnholt, Ole (Hg.): Rette ord. Festskrift til Sabine Kirchmeier-Andersen i anledning af 60-årsdagen. Frederiksberg, S. 21-30.
- Hyttel-Sørensen, Liva/Stæhr, Andreas C. (2014): Normativitet som social resource på Facebook. In: Karoli Christensen, Tanya et al. (Hg.): Sprogbrug i de nye medier. (= Nydanske sprogstudier 46). Frederiksberg, S. 67-102.
- Jarvad, Pia/Sandøy, Helge (Hg.) (2008): Stuntman og andre importord i Norden. Om udtale og bøjning. (= Moderne importord i språka i Norden 7). Oslo.
- Jervelund, Anita Ågerup/Schack, Jørgen (2016): Den der skriver d i gjort. Staveproblemer i folkeskolens ældste klasser. Frederiksberg.
- Kirchmeier-Andersen, Sabine et al. (2012): European Language Monitor (ELM). In: Stickel, Gerhard/Carrier, Michael (Hg.): Education in creating a multilingual Europe. Frankfurt a.M., S. 181-190.
- Le Grand Robert (2009): Dictionnaires Le Robert. Paris.
- Retskrivningsordbogen (2012): Dansk sprognævn. Frederiksberg. [www.dsn.dk/ro](http://www.dsn.dk/ro).
- Schack, Jørgen/Diderichsen, Philip (2015): Jagten på den gode sikre Sprogbruger. In: Nyt fra Sprognævnet 3, S. 1-8.
- Torp, Arne (2004): Nordiske sprog i fortid og nutid. In: Stampe Sletten, Iben (Hg.): Nordens sprog med rødder og fædder. Kopenhagen, S. 19-74.