

**LMU**

LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

Zentrum  
historische  
Sprachwissenschaften

[www.sprachwiss.lmu.de](http://www.sprachwiss.lmu.de)

<http://epub.ub.uni-muenchen.de/>

# Mubahis

## Münchener Beiträge zur Allgemeinen und Historischen Sprachwissenschaft

herausgegeben von

Peter-Arnold Mumm

Institut für Vergleichende und Indogermanische Sprachwissenschaft sowie Albanologie

Wolfgang Schulze

Institut für Allgemeine und Typologische Sprachwissenschaft

Band 4

**Cornelia van Scherpenberg, Fleur Szabó,  
Stefanie Eckmann (Hgg.)**

**Tagungsband der 53. Studentischen Tagung  
Sprachwissenschaft (StuTS), 9. – 12. Mai  
2013, LMU**

2014

Tagungsband der 53.  
Studentischen Tagung  
Sprachwissenschaft  
(StuTS)

9. – 12. Mai 2013  
Ludwig-Maximilians-Universität  
München

Herausgegeben von

Cornelia van Scherpenberg, Fleur Szabó, Stefanie Eckmann

## **Danksagung**

Unser herzlicher Dank gilt allen, die direkt oder indirekt an der Veröffentlichung dieses Tagungsbandes beteiligt waren.

Allen voran Prof. Dr. Wolfgang Schulze und PD Dr. Peter-Arnold Mumm für die Gelegenheit, diesen Tagungsband in der Reihe *Münchener Beiträge zur Allgemeinen und Historischen Sprachwissenschaft* (MUBAHIS) zu veröffentlichen.

Unser Dank geht außerdem an die Autoren der hier veröffentlichten Aufsätze für ihre Beiträge und ihre Geduld. Gleichzeitig möchten wir uns bei den Peer Reviewern für ihre fachlichen Hinweise bedanken.

Cornelia van Scherpenberg, Fleur Szabó, Stefanie Eckmann

## **Vorwort**

Im vorliegenden Tagungsband sind zehn Vorträge, die auf der 53. Studentischen Tagung Sprachwissenschaft (StuTS) vom 9.-12.05.2013 an der LMU in München gehalten wurden, gesammelt. Die Autoren sind Studenten der Linguistik von verschiedenen Universitäten in Deutschland und im europäischen Ausland. Ihre Aufsätze behandeln Themen, mit denen sich die Studenten in Seminaren und in Abschlussarbeiten beschäftigt haben. Die Aufsätze wurden von anderen Teilnehmern der Tagung in einem Peer Review überprüft. Für inhaltliche Fehler übernehmen die Autoren selbst die Verantwortung. Bezüglich genderkorrekter Sprache wurden von den Herausgebern keine Vorgaben gemacht.

Cornelia van Scherpenberg, Fleur Szabó, Stefanie Eckmann

Ludwig-Maximilians-Universität, München

München, den 12.10.2014

## Inhaltsverzeichnis

Danksagung.....	
Vorwort.....	
A comparative analysis of lexis in the Bible: Old English, Middle English and the Authorized Version.....	
Justyna Beba & Olga Laskowska (Universität Warschau).....	1
Präverbien im Ungarischen - Ihre Funktion als Aspektmarker.....	
Laura Becker (Universität Mainz).....	10
Vergleichendes Verbalparadigma: Standarddeutsch – Mittelhessisch/Oberhessisch.....	
Timm Braun (Universität Stuttgart).....	20
Grammatische und semantische Kongruenz bei anaphorischem Verweis auf Diminutiv-Nominalphrasen: Überlegungen zum Einfluss syntaktischer Komplexität.....	
Henning Czech (Universität Osnabrück).....	26
Tschechisch als eine SAE-Sprache: Im Zentrum oder an der Peripherie?.....	
Martin Konvička (Freie Universität Berlin).....	36
Nothing but “typed talk”? Analysing discourse in computer-mediated communication by employing Koch and Oesterreicher’s framework model.....	
Gerrit Kotzur (Technische Universität Berlin).....	46
Implementation eines regelbasierten Pronomenresolutionsalgorithmus für das Deutsche.....	
Ludger Paschen (Ruhr-Universität Bochum).....	57
Das gegenwärtige polnische Lautsystem und seine historische Entwicklung im Lichte der Präferenzgesetze der Silbenstruktur von Theo Vennemann.....	
Katarzyna Plucińska (Universität Bamberg).....	71
Pink im Deutschen und Dänischen Eine Studie zur Semantik von Farbwortentlehnungen.....	
Cornelia van Scherpenberg (Ludwig-Maximilians-Universität München).....	81
The representation of women and men in English idioms: a corpus-based study.....	
Karin Wiecha (Universität Potsdam).....	93

# **A comparative analysis of lexis in the Bible: Old English, Middle English and the Authorized Version**

*Justyna Beba & Olga Laskowska (Universität Warschau)*

## **1. English translations of the Bible and language change**

The paper offers a diachronic account of lexical changes that took place in the English language from Old to Early Modern English, based on a corpus of the three versions of St Mark's Gospel, chapter 15. Since the Bible is a text which has been examined and translated in every crucial stage of development of the English language, its translations from different periods provide a substantial amount of data for the analysis of the language of the three periods and for drawing diachronic comparisons.

The first of the three translations is the Anglo-Saxon version (later referred to as 'AS') in Old English, which covers fragments of the text of the Bible. The biblical texts were mostly word to word glosses based on the Latin Vulgate (Shepherd 1969: 362–386).

The Middle English version, the first full text of the Bible in English, also known as the Wycliffe–Purvey translation, which for the purpose of this study is referred to as 'WP', was also based on the Vulgate. It is a 14<sup>th</sup> century translation which was most probably made by John Wycliffe, a priest, and his secretary John Purvey (Shepherd 1969: 404).

The paper also looks at the Authorized Version, which is commonly known as King James Bible. Over a span of seven years, forty seven scholars appointed by King James were working on this translation of the Bible. Completed in 1611, this version encompasses the New Testament (translated from Greek), the Old Testament (translated from a Hebrew text), and the Apocrypha (translated from Greek and Latin). In the present study the abbreviation KJ is used to refer to the Authorized Version (Weigle 1969: 361–382).

## **2. The Norman Conquest and its influence on English**

The year 1066 is a symbolic date of a transition from Old to Middle English. As Matthew Townend writes, during the first few decades only the Norman invaders spoke French, but soon the situation began to change and by the mid 12<sup>th</sup> century most of the aristocracy were bilingual. From the 13<sup>th</sup> century onwards, French began to be used also by the middle class, and thus its usage grew in the

various spheres of life. The phenomenon of the popularity of French at that time must have been due to its social prestige and status as lingua franca on the continent rather than to the fact of the armed conquest that took place over two centuries earlier (Townend 2006: 61–85).

“A borrowed word, or loanword, is one that has been introduced at some time from another language, from either a related or unrelated language.” (Millward 1996: 49). According to the compilers of *The Oxford English Dictionary*, 10 203 words were borrowed from French between 1066 and 1500. The delayed influx of French loanwords can be aptly presented on the following timeline.

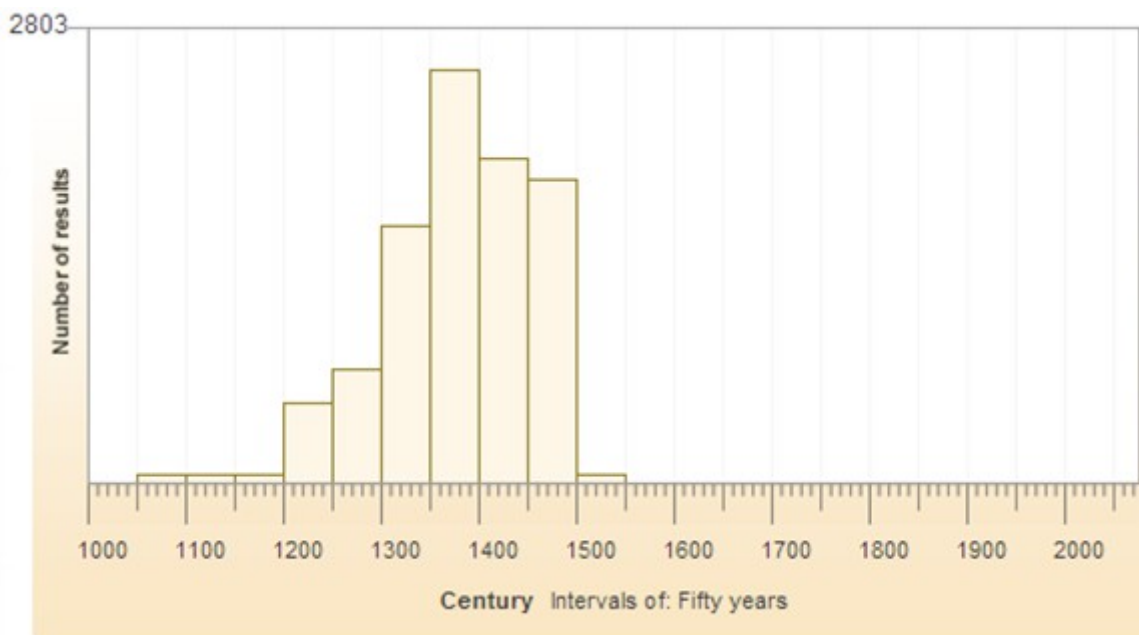


Figure 1: The influx of French borrowings from 1066 to 1500. (*The Oxford English Dictionary online*)

The analysis of lexis in our corpus has revealed 65 borrowings of French origin in the Wycliffe-Purvey translation and 95 in the Authorized Version (including repetitions). As both texts contain around 380 words, French borrowings make up respectively 17% and 24% of the whole text. The examples of such borrowings are:

WP *counsel*, *accusiden* ‘accused’, *feeste* ‘feast’, *seductioun* ‘treason, treachery’, etc.

KJ *chief*, *deliver*, *marvel*, *soldiers*. Of course, not only French words have been borrowed into English. A number of Scandinavian items (*take*, *fro*, *till* ‘to’) and Latin loanwords (*priest*, *scripture*, *scribe*, *temple*) were found in the text but their number was considerably lower.

	Number of words in the text (approx.)	Number of French borrowings and hybrids (approx.)	Number of other borrowings and hybrids (approx.)	Total number of borrowings (approx.)
Wycliffe– Purvey	380	65 ~17%	20~5%	85~22%
King James	390	95~24%	25~6%	120~31%

Table 1: The ratio of the lexical items of various origin in the WP and KJ versions

Hybrids, such as *murder* (OE *morðor* reinforced in Middle English by OF *murdre*) or *crying* (OF *crier* with the English suffix *-ing*) have also been counted as borrowings. Table 2 shows that the number of borrowings gradually increased in the course of time and French borrowings began to play a vital role in English lexis.

### 3. Displacements

The process of borrowing and words of a similar meaning coexisting at the same time in a language were the triggers of semantic change (McMahon 1994: 176). The influence of the French borrowings on the lexicon of English caused that many native words were either lost or drastically changed their meanings.

The present corpus study has revealed a number of displacements resulting in the falling out from the common usage of native English words, for example:

- (1) (a) 13 Hi eft **hrymdon** and cwædon, Hoh hine. (AS)  
13 And thei eftsoone **crieden**, Crucifie hym. (WP)  
13 And they **cried** out again, Crucify him. (KJ)
- (b) 15 Pilatus wolde ða ðam folce **gecweman**, and forgef him Barraban, and sealde him ðone Hælend, beswungenne, ðæt he ahangen wære. (AS)  
15 And Pilat, willynge to make **aseeth** to the puple, lefte to hem Barabas, and bitok to hem Jhesu, betun with scourgis, to be crucified. (WP)  
15And so Pilate, willing to **content** the people, released Barabbas unto them, and delivered Jesus, when he had scourged him, to be crucified. (KJ)
- (c) 29 And ða ðe forþ-stopon hine **gremedon**, and hyra heafod cwehton, and ðus cwædon,



Wala! se towyrpþ ðæt tempel and on þrim dagum eft getimbraþ, (AS)

29 And as thei passiden forth, thei **blasfemyden** hym, mouynge her heedis, and seiynge, Vath! thou that distriest the temple of God, and in `thre daies bildist it ayen, (WP)

29 And they that passed by **railed** on him, wagging their heads, and saying, Ah, thou that destroyest the temple, and buildest it in three days, (KJ)

Item (1a) contains the verb *hrymdon* ‘cried’, which survives in Present-Day English as *reme*, which according to the OED, is considered a rare or regional form, replaced elsewhere. by *cry*, a verb of French origin. Item (1b) exemplifies a displacement which later led to a complete disappearance of the verb *gecweman* whose continuation *queem* failed to survive beyond the 17<sup>th</sup> century. Item (1c) contains the verb *gremian* ‘to irritate, provoke’, which in later versions was replaced by OF *blasfeme/n* ‘to speak profanely, to blaspheme’ and the verb *rail* ‘to joke, to jest’.

Of course, not only loanwords caused replacements. Item (2) is an example of a noun (*sceaþan*pl. ‘one who does harm, a criminal’) fully replaced by another native word *thief* (OE *þēof*).

(2) 27 And hi ahengon mid him twegen **sceaþan**, anne on his swyðran healfe, and oðene on his wynstran. (AS)

27 And thei crucifien with hym twei **theues**, oon `at the riythalf and oon at his lefthalf. (WP)

27 And with him they crucify two **thieves**; the one on his right hand, and the other on his left. (KJ)

Old English did not use too many borrowings and, while translating the text of the Bible instead of imposing Latin loanwords, the translators would rather create their own terms, often by means of compounding (Millward 1996: 120).

<i>Line</i>	<i>vulgate</i>	<i>AS</i>	<i>WP</i>	<i>KJ</i>
6	festum	<b>symbol-dæge</b>	feeste dai	feast
7	seditiosis	<b>swic-cræft</b>	seducioun	insurrection
16	praetorii	<b>dom-ernes</b>	porche	Praetorium
26	inscriptus	<b>ofer-gewrit</b>	titil	superscription
36	centurio	<b>hundred-man</b>	centurien	centurion

*Table 2: The comparison of compounds in the Old English version*

As exemplified in Table 3, by the time when Wycliffe– Purvey translation was created loanwords were commonly accepted. Moreover, in King James Version words seem to be more complex and sophisticated.

Let us take a look at a few examples of native and foreign words illustrating semantic change:

- (3) (a) 17 And scryddon hine mid purpuran. And him onsetton þyrnenne **helm**, awundenne; (AS)  
17 and clothiden hym with purpur. And thei writhen a **coroun** of thornes, and putti den on hym (WP)  
17 And they clothed him with purple, and platted a **crow**n of thorns, and put it about his head (KJ)
- (b) 40 And ða **wif** wæron feorran behealdende, and betwux ðam wæs seo Magdalenisce Maria, and Maria Iacobes moðor [ðæs gingran, and Iosepes moder,] and Salomeæ. (AS)  
40 And ther weren also **wymmen** biholdynge fro afer, among whiche was Marie Maudeleyn, and Marie, the modir of James the lesse, and of Joseph, and of Salome. (WP)  
40 There were also **women** looking on afar off: among whom was Mary Magdalene, and Mary the mother of James the less and of Joses, and Salome. (KJ)
- (c) 8 And ða he ferde, ða ongan seo menegeu hine **bidan**, swa heo symle dyde. (AS)  
8 And whanne the puple was gon vp, he bigan to **preie**, as he euer more dide to hem. (WP)  
8 And the multitude crying aloud began to **desire** him to do as he had ever done unto them. (KJ)

- (d) 1 Ða sona on mergen, worhton ða heah-sacerdas hyra gemot mid ealdrum, and boce-  
rum, and eallum werodum, and læddon ðæne Hælend, gebundenne, and **sealdon** hine  
Pilato. (AS)  
1 And anoon in the morewtid the hiyeste prestis maden a counsel with the elder men,  
and the scribis, and with al the counsel, and bounden Jhesu and ledden, and **bitoken**  
hym to Pilat. (WP)  
1 And straightway in the morning the chief priests held a consultation with the elders  
and scribes and the whole council, and bound Jesus, and carried him away, and **deli-  
vered** him to Pilate. (KJ)
- (e) 39 Ða se hundred-man, ðe ðar stod agen, geseah ðæt se Hælend swa clypiende forþ-  
ferde, he cwæþ, **Soplice**, ðes man wæs Godes sunu.(AS)  
39 But the centurien that stood forn ayens siy, that he so crynge hadde diede, and  
seide, **Verili**, this man was Goddis sone.(WP)  
39 And when the centurion, which stood over against him, saw that he so cried out,  
and gave up the ghost, he said, **Truly** this man was the Son of God. (KJ)

OE *helm* (PDE *helm*, *helmet*) could mean ‘that part of the armour which covers the head; a helmet’ (now poet. and arch.), i.e. the meaning we all may expect, but in (3a) the noun has a different sense: ‘Christ’s crown of thorns’. Therefore it might be assumed that the meaning of the word underwent the process of narrowing and the noun was replaced by Fr. *coroun* (WP) *crown* (KJ) denoting ‘an ornamental fillet, wreath, or similar encircling ornament for the head, worn for personal adornment, or as a mark of honour or achievement; a coronal or wreath of leaves or flowers’. Although these words coexisted in the Middle English period (OED), sharing this specific meaning, it was not until Late Middle English that the meaning of *helm* narrowed down.

Another change concerns OE *wif* which, in general, meant ‘a woman’. Looking at (3b) it may be assumed that in the Anglo-Saxon Bible (fragment of Mark 15) the noun *wif* had that sense. However, in Wycliffe – Purvey and King James translations the authors used the forms *wymmen* and *women* respectively. Such a change is another example of the process of specialization.

According to *The Oxford English Dictionary* the contemporary sense of ‘female spouse’ began to emerge already in Old English. Although nowadays the meaning of ‘spouse’ is the one that directly comes to one’s mind, the earlier sense was preserved in single words, for instance *midwife*, and phrases like *old wives’ tale*. Because of the ongoing sense specification, the noun *wif* was replaced in later texts by *woman*, which preserved its general sense of an ‘adult female human being’.

According to McMahon (1994: 177) the change of OE (*ge*)*bēd* ‘prayer’ to ModE *bead* ‘small woo-

den or other ball on a string' has its source in the discontinuous nature of language transmission in which children do not receive a complete grammar from the previous generation, but build a grammar of their own on the basis of incoming information. Hence a child informed by his/her parent about 'counting beads' while praying the rosary would not realize that *bead* means 'prayer'. Rather, the child would associate beads with wooden balls on the string, moving under the praying person's fingers.

In (3d), the contrast between *sellan* (AS), *bitake* (ME) and *deliver* (KJ) is displayed. The three words refer to the same concept. However, today they would not be considered synonyms. Taking a closer look at the history of the verb *sell* it might be discovered that it did not always mean what we chiefly associate with today's 'to give up or hand over for money (or something that is reckoned as money)'. In Old English it chiefly meant 'to give' in the various senses of the word. The verb used in the Middle English version, *bitake*, which according to the OED did not exist before Middle English, had the sense 'to hand over, deliver, give up, grant, place at a person's disposal', is now considered obsolete. In the Anglo-Saxon and Middle English versions, *sellan* (AS) and *bitake* (ME) are of native origin, whereas King James Bible uses *deliver* from French.

In (3e), yet another word is worth commenting upon. In the place of OE *soplice* 'truly, really, certainly, verily', the authors of the Middle English version used *verili*, which reflected the Old French meaning of the adjective *verai* 'true'. The contrast can be easily noticed on confronting data from King James Version where the word *truly* is used. It can be assumed that by that time *very* underwent the process of grammaticalization, losing its initial meaning and becoming an intensifier.

A feature worth discussing is the way in which Jesus Christ was referred to. While analyzing the three versions of the text, one cannot fail to notice some discrepancies concerning words denoting God's son. Generally, what in WP and KJ's translations is simply *Jhesu* (WP) *Jesus* (KJ) is in the Anglo-Saxon version consistently substituted with *Hælend* 'healer'. In Old English the name of God's son was considered a taboo word, hence this euphemism coined to replace it. The names *Crist* (AS, WP) and *Christ* (KJ) are used in all the versions, the reason being that *Christ* is a borrowing from the Latin translation of the Hebrew word *Messiah* 'anointed'. Other ways of addressing Jesus in Old English are *Iudea cyning* and *Godes sunu*, which correspond to *Kinge of Jewis/Israel* (WP), *Godde sone King of the Jews, the Son of God* (KJ). Those circumlocutions are direct translations from the Vulgate.

When discussing the employment of euphemisms aimed at getting rid of taboo words, it is also worth emphasising that words which denoted an action of dying were also avoided in English (Burnley 1992: 445). This process can be epitomized by OE *forþ-ferde*, a compound consisting of *forþ* 'forth' and *feran* 'to go, make a journey, set out' which might be related to the contemporary

euphemism *passed away*. While the Wycliffe– Purvey version contains the verb *die*, KJ uses the idiomatic phrase *gave up the ghost*.

#### 4. Conclusions

The analysis of the texts from the period under investigation allows one to formulate the following conclusions:

- (1) The number of borrowings in English has been constantly increasing in the course of its history.
- (2) The majority of borrowings came from French.
- (3) The process of translating words by means of compounding seems to have lost its productivity.
- (4) Semantic changes were inevitable, especially considering the fact of the coexistence of borrowed and native words with a similar meaning.
- (5) Some native words failed to resist the linguistic consequences of the Norman Conquest and have failed to survive.

#### 5. References

##### a) Sources

*Anglo-Saxon Bible* online: <http://wordhord.org/nasb/>

*Bosworth —Toller Anglo-Saxon dictionary* online: <http://bosworth.ff.cuni.cz/>

*King James Bible* online:

<http://www.kingjamesbibleonline.org/book.php?book=Mark&chapter=15&verse=>

Marlowe, Michael D. (2001) *Bible research. Internet resources for students of Scripture.*

<http://www.bible-researcher.com/plumptre1.html>

*Middle English dictionary* online: <http://quod.lib.umich.edu/m/med/>

*Wycliff—Purvey Bible* online: [http://www.bibledbdata.org/onlinebibles/wycliffe\\_nt/41\\_015.htm](http://www.bibledbdata.org/onlinebibles/wycliffe_nt/41_015.htm)

##### b) Special studies

Burnley, Davis. 1992. “Lexis and semantics”. In: Richard M. Hogg (ed.): *The Cambridge history of the English language*. 409–499.

Hogg, Richard M. (ed.). 1992. *The Cambridge history of the English language*. Vols. 1–4. Cambridge: Cambridge University Press.

Lampe, Geoffrey. W. H. (ed.). 1969. *The Cambridge history of the Bible*. Vols. 1–2. *The West from the Fathers to the Reformation*. Cambridge: Cambridge University Press.

McMahon, April M. S. 1994. *Understanding language change*. Cambridge: Cambridge University Press.

- Millward, Celia M. 1996. *A biography of the English language*. Fort Worth: Harcourt Brace.
- Mugglestone, Lynda. 2006. *The Oxford history of English*. Oxford: Oxford University Press.
- Shepherd, Geoffrey. 1969. "English versions of the Scriptures before Wycliff". In: Geoffrey W. H. Lampe (ed.): *The Cambridge history of the Bible*. Vol. 2. 362–386.
- Townend, Matthew. 2006. "Contacts and conflicts: Latin, Norse, and French". In: Lynda Mugglestone (ed.): *The Oxford history of English*. 61–85.
- Weigle, Luther A. 1969. "English versions since 1611". In: Geoffrey W. H. Lampe (ed.): *The Cambridge history of the Bible*. Vol. 1. 361–382.

# Präverbien im Ungarischen - Ihre Funktion als Aspektmarker

Laura Becker (Universität Mainz)

## 1. Einleitung

Die Präverbien des Ungarischen stellen ein stark diskutiertes Forschungsgebiet dar, da es sich um verbale Präfixe handelt, die durch ihre atypischen Eigenschaften immer wieder Fragen aufwerfen. Besonders ihre syntaktischen Eigenschaften, da sie unter bestimmten Umständen vom Verb trennbar sind, wurden bisher untersucht.

Dieser Aufsatz jedoch betrachtet jedoch einen Bereich der Semantik der Präverbien: ihre aspektuellen Funktionen. In der Literatur wurden bereits verschiedene Versuche unternommen, die aspektuellen Eigenschaften zu klassifizieren; allerdings scheint bisher ein allgemein typologischer Ansatz, welcher das Thema z.B. zum Vergleich mit anderen Sprachen öffnet, zu fehlen.

Daher versucht dieser Aufsatz, (1) zunächst einen Überblick über die besonderen Eigenschaften der Präverbien im Ungarischen zu geben. Anschließend (2) wird mit dem typologisch orientierten Klassifizierungsmodell des Aspekts nach Plungian (2011) ein neuer Ansatz dargestellt, um die aspektuellen Funktionen der Präverbien zu untersuchen. Damit lässt sich zeigen, dass, obwohl kein ausgebautes Aspektsystem wie es etwa im Slavischen vorliegt, das Ungarische die Präverbien dennoch zum Ausdruck aspektueller Werte nutzen kann.

## 2. Vorbemerkungen: Was sind Präverbien? Was sind ihre Besonderheiten?

Präverbien sind Verbpräfixe, die jedoch nicht vollkommen an den Rest des Verbs gebunden sind und unter bestimmten Umständen noch trennbar und somit syntaktisch eigenständig sind<sup>1</sup>. Diese Eigenschaft besitzen sie, da sie gewissermaßen „untergrammatikalisierte“ Präfixe sind, die noch eine Zwischenstufe der diachronen Entwicklung eines freien Ausdrucks zu einem gebundenen Morphem darstellen, wobei sich viele der Präverbien aus ursprünglichen Adverbien entwickelten.

Die verbalen Präfixe, die im Ungarischen anzutreffen sind, müssen als Präverbien eingestuft werden, da sie vom Rest des Verbs trennbar sind. Dass sie der Grammatikalisierung unterworfen sind, lässt sich z.B. an ihrer Semantik feststellen. Zunächst hatten und haben die Präverbien vorrangig lexikalische und spezifische, nämlich räumliche, Bedeutungen, die mit der Zeit durch abstraktere und

---

<sup>1</sup> Das Deutsche etwa verfügt ebenso über Präverbien, die in der Germanistik traditionell „Verbpartikeln“ genannt werden (z.B.: *mit-*, *nach-*, *ent-*, *be-*, *zer-*, *an-*, *aus-*, *auf-*, *ab-*, ...-gehen).

stärker grammatikalisierte aspektuellen Werte ergänzt und z. T. ersetzt wurden.

Andererseits zeugt die Trennbarkeit (z.B. bei Negation<sup>2</sup>) des Präverbs vom Rest des Verbs noch davon, dass es sich um zwei ehemalige unabhängige Wortformen handelt:

- (1) a. László el-dug-t-a a cigaretta-k-at.  
 László PRV:weg-steck-PST-S<sub>3SG</sub>.O<sub>DEF</sub> ART:DEF Zigarette-PL-ACC  
 ‘László versteckte die Zigaretten.’
- b. László nem dug-t-a el a cigaretta-k-at.  
 László NEG steck-PST-S<sub>3SG</sub>.O<sub>DEF</sub> PRV:weg ART:DEF Zigarette-PL-ACC  
 ‘László versteckte die Zigaretten nicht.’

Die folgenden sechs ältesten und ebenso frequentesten Präverbien (Soltész 1959) stellen gewissermaßen den Kern des Präverb-Inventars dar:

meg	‘KOMPLETTE HANDLUNG’ <sup>3</sup>
el	‘weg’
le	‘her-hinunter’
fel/föl	‘her-/hinauf’
ki	‘her-/hinaus’
be	‘her-/hinein’

### 3. Präverbien und Aspekt

Im Folgenden wird dargestellt, nach wie die Kategorie „Aspekt“ verstanden werden kann, um zu zeigen, welche Funktionen von Präverbien übernommen werden.

#### 3.1 Was kann unter „Aspekt“ verstanden werden?

Um Themen wie „Aspekt“ zu betrachten, ist es sinnvoll, Verben in aktionale Klassen<sup>4</sup>, welche auf ihre interne zeitliche Struktur verweisen, einzuteilen.

Situationen können statisch, und somit hochgradig unabhängig von einem zeitlichen Referenzpunkt, oder dynamisch sein, wobei im letzten Fall ein Zustandswechsel stattfindet. Dieser Wandel kann kontinuierlich sein, dann handelt es sich um einen Prozess; ist er hingegen augenblicklich, so stellt die Handlung ein Ereignis dar. Prozesse können noch weiter unterteilt werden in endliche Situationen (telisch) und zyklische oder sich wiederholende Situationen (atelisch).

Aspektuelle Werte beschreiben die Situation in Bezug auf ihren „internen zeitlichen Aufbau“ (Comrie 1976: 3). In diesem Aufsatz sollen aspektuelle Funktionen nach Plungian (2011: 377–422) in pri-

2 Weitere Faktoren für die Trennung von Präverb und Verb, bzw. die Inversion beider sind die morphologische Markierung des Imperativs und ein fokussiertes Element im Satz, welches die präverbale Position besetzt und das Präverb gewissermaßen verdrängt.

3 Dieses Präverb besitzt keine lexikalische Bedeutung mehr, sondern zeigt nur noch aspektuelle Bedeutungen an und wird deshalb auf diese Art übersetzt.

4 Diese Einteilung geht auf Vendler (1957) zurück.



mären und sekundären Aspekt aufgeteilt werden. Dabei sind primäre aspektuelle Funktionen der Präverbien solche, die in Einklang mit der inhärenten Semantik des Verbs stehen (z.B. das Hinweisen auf eine bestimmte Phase der Handlung bei in sich begrenzten Handlungen oder die unbegrenzte Darstellung von inhärent atelischen Prozessen). Der sekundäre Aspekt hingegen steht für die Verschiebung des präverblosen Verbs in eine andere Aktionsklasse. Da diese Funktionen sehr spezifisch sind, übernehmen ihre Marker häufig verschiedene, ähnliche Funktionen, was zu Clusterbildungen führt. Die zwei Hauptcluster sind dabei Imperfektive und Perfektive. Durch eine solche Einteilung wird die "klassische" polaristische Zweiteilung in perfektive und imperfektive Werte, die auf das Ungarische etwas unzureichend anwendbar ist, vermieden. Denn das Ungarische verfügt über keine Mittel, Aspekt z. B. als Flexionskategorie anderweitig am Verb zu markieren. Die einzige Möglichkeit besteht in seiner derivativen Kennzeichnung mithilfe der Präverbien, wobei auch dies nicht für jedes Verb zur Verfügung steht. Da es sich um Derivation präverbloser Verben zu neuen Lexemen bestehend aus Präverb und lexikalischem Verb handelt, erfolgt die Aspektmarkierung nur eingeschränkt und hochgradig lexikalisch abhängig, sodass viele Wechselwirkungen der Verbsemantik und der (ursprünglichen und häufig noch bestehenden räumlichen) Präverbsemantik zu beachten sind. All dies würde in einem zweigeteilten System vernachlässigt, kann jedoch hier Beachtung finden. Zusätzlich fällt durch gleichwertige Betrachtung spezifischer Werte und Bündelung in Cluster die Unterscheidung zwischen Aspekt und Aktionsart<sup>5</sup>, deren Werte teilweise durch die gleichen Präverbien gekennzeichnet werden, weg.

### 3.2 Markierung des primären Aspekts

Da der primäre Aspekt eng an die Verbsemantik gebunden ist, kann er auch als „linearer Aspekt“ verstanden werden, sodass die aspektuelle Funktion des Präverbs mit den aspektuellen Werten des Verbs überlappt. Nach Lehmann (2010), Breu (2000) und Plungian (2011) kann ein Verb eine Situation mit einer vorbereitenden, resultativen (extern) und beginnenden, mittleren und finalen (intern) Phase darstellen, wobei jedoch nicht jedes Verb alle dieser Phasen besitzen muss. Die Einteilung der Situation in verschiedene Phasen erlaubt es, den inneren zeitlichen Aufbau des Verbs zu abstrahieren und für die Betrachtung der Aspektmarkierung relevant zu machen:

#### 3.2.1 Inchoative und Inzeptive

Nach Bybee (1985: 147) korrespondieren Inchoative bzw. Inzeptive mit der vorbereitenden Phase einer Situation.

Zusätzlich lassen sich die beiden Funktionen anhand der Verbklasse, bei welcher sie auftreten, un-

---

<sup>5</sup> Nach Comrie wird Aspekt als "Grammatikalisierung relevanter semantischer Unterscheidungen" gesehen, während "Aktionsart die Lexikalisierung der Unterscheidungen" darstelle (1976: 6 et passim; eigene Übersetzung).

terscheiden, sodass Inchoative die vorbereitende Phase dynamischer Verben und Inzeptive diese Phase der statischen Verben markieren.

Eine Aufteilung lässt sich im Ungarischen belegen, da es zwei Präverbien gibt, die jeweils an Inchoativen bzw. Inzeptiven beteiligt sind: *el* 'weg' und *meg* 'VOLLSTÄNDIG'.

Das Präverb *el* ist mit verschiedenen Verbklassen kompatibel, bei dynamischen atelischen Verben, die in ihrer Argumentstruktur keinen Undergoer<sup>6</sup> besitzen, scheint *el* ein „Dummy-Pronomen“ zu erfordern, da es semantisch ein Thema benötigt. Das Reflexivpronomen übernimmt diese Aufgabe:

- (2) a. Nevet-Ø-nek.  
lach-PRS-3PL  
'Sie lachen.'
- b. **El**-nevet-Ø-ik                      mag-uk-at.  
**PRV:INCH**-lach-PRS-3PL      REFL-3PL-ACC  
'Sie **fangen plötzlich** und **unerwartet an** zu lachen.'

Handelt es sich um z.B. intransitive Verben, die nur ein Undergoer- und kein Actor-Argument besitzen, so kommt der Inchoativausdruck ohne zusätzliches Pronomen aus, da das Thema bereits durch das einzige Argument gegeben ist:

- (3) a. kész-ül  
mach.fertig-DECAUS  
'fertig werden'
- b. **el**-kész-ül  
**PRV-INCT**-mach.fertig-DECAUS  
'**anfangen** fertig zu werden'

Bei Kognitions- und Emotionsverben tritt *meg* als Marker der Inzeptivität auf:

- (4) a. Szeret-t-em                      a              spenót-ot.  
mag-PST-S<sub>1SG</sub>.O<sub>DEF</sub>      ART:DEF      Spinat-ACC  
'Ich **mochte** Spinat.'
- b. Meg-szeret-t-em                      a              spenót-ot.  
**PRV:INCPT-mag**-PST-S<sub>1SG</sub>.O<sub>DEF</sub>      ART:DEF      Spinat-ACC  
'Ich fing an, Spinat zu mögen.'

### 3.2.2 Progressive

Progressive verweisen auf die mittlere Phase der Situation in Bezug auf das Beobachtungszeitfenster-

---

6 Die Termini „Actor“ und „Undergoer“ (Van Valin 2001) folgen der zweigeteilten Klassifikation semantischer Makrorollen.

ter. Im Ungarischen ist daran nicht ein bestimmtes Präverb beteiligt, sondern die syntaktische Position einer jeden semantisch geeigneten PRV-V-Kombination markiert diesen Aspekt.

Ohne einen anderen der in Abschnitt 1 erwähnten Trigger für die Inversion des Präverbs und Rest des Verbs, kann diese auftreten. In diesem Falle markiert sie Progressivität:

- (5) a. Mari vit-t-e ki a szemet-et,  
 Mari bring-PST-S<sub>3SG</sub>·O<sub>DEF</sub> PRV:hinaus ART:DEF Müll-ACC  
 a-mikor haza-jött-em.  
 REL-wenn nach.Hause-komm.PST-1SG  
 ‘Mari **war dabei**, den Müll **hinauszubringen**, als ich nach Hause kam.’
- b. Mari ki-vit-t-e a szemet-et,  
 Mari PRV:hinaus-bring-PST-S<sub>3SG</sub>·O<sub>DEF</sub> ART:DEF Müll-ACC  
 a-mikor haza-jött-em.  
 REL-wenn nach.Hause-komm.PST-1SG  
 ‘Mari **brachte** den Müll **hinaus**, als ich nach Hause kam.’

Eine solche Aspektmarkierung ist jedoch nur möglich, sofern das Verb semantisch geeignet ist. Handelt es sich um ein statisches oder auch atelisches dynamisches Verb, so kann die Inversion nicht erfolgen, um z.B. Gleichzeitigkeit zu kennzeichnen. Es muss also daraus geschlossen werden, dass statische und atelische dynamische Verben keine Markierung des progressiven Aspekts zulassen.

Auch hatten meine Informanten<sup>7</sup> Probleme, Konstruktionen wie in (5a) zu akzeptieren, wenn kein geeigneter Kontext gegeben war. Ein solches Beispiel wurde beinahe von allen befragten Muttersprachlern korrigiert, indem sie das Präverb wie in (5b) wieder vor das Verb stellten.

Auch sei erwähnt, dass eine Progressiv-Kennzeichnung nur möglich ist, wenn kein anderer Trigger die Inversion hervorruft. Bei z.B. gleichzeitiger Fokussierung eines anderen Elementes im Satz kann auf diese Weise nicht mehr die Progressivität der Handlung ausgedrückt werden, da die Inversion bereits durch den anderen Trigger unabhängig der aspektuellen Bedeutung gefordert wird:

- c. Mari a bicikli-t hoz-t-a be,  
 Mari ART:DEF Fahrrad-ACC hol-PST-S<sub>3SG</sub>·O<sub>DEF</sub> PRV:hinein  
 nem a labdá-t  
 NEG ART:DEF Ball-ACC  
 ‘Mari **holte** das Fahrrad **rein**, nicht den Ball.’ /  
 ‘Mari **war dabei**, das Fahrrad **reinzuholen**, nicht den Ball.’

7 Es handelt sich um fünf Muttersprachler (3 männlich, 2 weiblich) des Ungarischen aus Budapest, Debrecen und Hajdúszoboszló, mit denen die hier vorliegenden Beispiele elizitiert und aus der Literatur zitierte Beispiele kommentiert wurden.

### 3.2.3 Resultative

Resultative beschreiben den „Zustand, der durch eine Handlung in der Vergangenheit hervorgerufen wurde“ (Bybee et al. 1994: 63; eigene Übersetzung).

In Verbindung mit dem Präverb *meg* ‘VOLLSTÄNDIG’ weisen vor allem statische Verben eine resultative Bedeutung auf:

- (6) a. Csak kevés pénz **marad-t-Ø**.  
nur wenig Geld bleib-PST-3SG  
‘Nur wenig Geld **wurde gelassen**.’
- b. Csak kevés pénz marad-t-Ø **meg**.  
nur wenig Geld bleib-PST-3SG PRV:RES  
‘Nur wenig Geld **ist übriggeblieben**.’

Von dynamischen telischen Verben lassen sich auch Resultative bilden. Dies geschieht durch eine Konstruktion eines mit einem Präverb versehenen Konverbs und der Kopula. Zwar sieht die Konstruktion dem „europäischen Typ des Passivs“ sehr ähnlich, jedoch handelt es sich nicht um ein Passiv, da sich kein Agens mehr in die Handlung einfügen lässt (was Kriterium für ein Passiv wäre). Zusätzlich wird ausschließlich auf die resultative Phase der Handlung verwiesen:

- (7) A számítógép szét-van szed-ve.  
ART:DEF Computer PRV:auseinander-sein.PRS.3SG nem-CONV<sup>8</sup>  
‘Der Computer **ist auseinandergenommen**.’

### 3.2.4 Delimitative

Delimitative zeigen Handlungen als zeitlich limitiert, sodass sie genau in das Beobachtungsfenster passen. Im Ungarischen übernimmt das Präverb *el* ‘weg’ diese Funktion. Auffällig dabei ist, dass bei z.B. Kiefer (2006: 167) und Dékány (2008: 12) Beispiele für eine solche aspektuelle Markierung genannt werden, jedoch nicht weiter darauf eingegangen wird, dass es sich in jedem Fall um bereits abgeleitete Verben bzw. Verben diminutiver Semantik handelt:

- (8) a. Söröz-get-ünk a munka után.  
Bier.trink-DIM-1PL ART:DEF Arbeit PTPS:nach  
‘Wir **trinken ein paar Bier** nach der Arbeit.’
- b. El-söröz-get-ünk a munka után.  
PRV:DELIM-Bier.trink-DIM-1PL ART:DEF Arbeit PTPS:nach  
‘Wir **trinken für einige Zeit ein paar Bier** nach der Arbeit.’

Meinen Informanten zufolge sei eine Derivation zu einem delimitativen Verb teilweise auch von

---

8 Konverb.



- (12) **El-fogy-ott** a leves.  
**PRV:weg-verminder-PST.3SG** ART:DEF Suppe  
 ‘Die Suppe ist **alle**.’
- (13) **El-olvas-t-ad** a regény-t?  
**PRV:CMPL-les-PST-S<sub>2SG</sub>·O<sub>DEF</sub>** ART:DEF Roman-ACC  
 ‘Hast du den Roman **ausgelesen**?’

### 3.3.3 Frequentative

Frequentative stellen häufig in einer bestimmten Zeit nacheinander wiederholte Ereignisse dar (Bybee 1985: 129). Sie werden durch die Reduplikation des Präverbs ausgedrückt.

Dabei kann auch ein bereits aspektuell markiertes Verb, wie z.B. ein semelfaktiver Ausdruck als frequentative Handlung dargestellt werden:

- (14) **Meg-meg-** kavar-Ø-ja a leves-t.  
**PRV:SEMEL-PRV:FREQ-** rühr-PRS-S<sub>3SG</sub>·O<sub>DEF</sub> ART:DEF Suppe-ACC  
 ‘X rührt die Suppe **von Zeit zu Zeit einmal um**.’

Lediglich statische Verben erlauben keine Frequentativ-Bildungen (Kiefer 206: 151).

## **4. Schlussbetrachtung**

Vor allem die zwei ältesten Präverbien, *meg* und *el* (Soltész 1959), werden häufig als Perfektivierer bezeichnet (Soltész 1959, Máta 1989, Kenesei 2006, Farkas & Sadock 1989: 319).

Das perfektive Cluster umfasst die im Ungarischen ausdrückbaren Aspektbedeutungen des Delimitativs, Inchoativs/Inzeptivs, Semelfaktivs, Kompletivs und Resultativs.

Wie in den vorherigen Abschnitten gezeigt wurde, werden diese Funktionen von einem kleinen Set an Präverbien – *meg* ‘VOLLSTÄNDIG’, *el* ‘weg’, *fel* ‘her-/hinauf’, *be* ‘her-/hinein (und *ki* ‘her-/hinaus) – geclustert übernommen. Besonders die ersten beiden scheinen, wie es in der Literatur oft beschrieben wurde, viele Funktionen innerhalb des perfektiven Clusters auszudrücken. Allerdings wurde auch gezeigt, dass die Aussage, alle Präverbien seien Perfektivierer die ganze Bandbreite an verbsemantik-spezifischen Wechselwirkungen außer Betracht lässt.

Ebenso kann nicht gesagt werden, dass im Gegenzug Verben ohne Präverb imperfektive Formen wären. Vielmehr sind diese neutral in Bezug auf aspektuelle Bedeutung und können erst in Opposition zu einem perfektiven PRV-V-Pendant als imperfektivisch reanalysiert werden.

Die einzige wirkliche Markierung der Imperfektivität findet bei der Inversion zur Progressiv-Kennzeichnung statt, da der Progressiv als imperfektivischer Wert gilt. Es handelt sich in diesem Falle auch um die einzige formale Operation, die nicht als Derivation bezeichnet werden kann, denn es entsteht kein neues Lexem: aus einem anfangs präverblosen Verb, welches als imperfektivisch re-

analysiert wird, kann eine PRV-V-Kombination abgeleitet werden, die ein neues Lexem mit spezifischer, in das Perfektiv-Cluster fallenden Bedeutung entstehen lassen kann. Durch die Inversion wiederum kann von einem solchen Verb eine progressive, also imperfektivische, Form entstehen, die jedoch an der Semantik des Verbs nichts verändert. Dieses Phänomen ist dem „slavisches Aspekttyp“ auffallend ähnlich. Im Slavischen werden unpräfigierte Verben als imperfektivisch angesehen und durch lexikalisch-determinierte Präfigierung perfektiviert. Durch die sekundäre Imperfektivierung (z.B. Lehmann 1988: 173), welche durch Suffigierung erfolgt, entstehen ohne semantische Änderungen der perfektivischen Verben wieder imperfektive Formen. Anders als im Ungarischen ist jedoch dies jedoch im Slavischen ein „flächendeckendes“ System, was dadurch eine Aspektpartnerzuweisung (Perfektiv – Imperfektiv) an jedem Verb erzwingt. Nur in dem Fall darf eigentlich von Aspekt gesprochen werden, da die Kategorie erst dann als solche an jedem Verb markiert wird bzw. ausgedrückt werden muss. Das Ungarische weist kein so stabiles und durchgehendes System auf. Vielmehr werden eng an die Verbsemantik geknüpfte Markierungen einzelner aspektueller Werte durch bestimmte Präverbien übernommen, wie in Abschnitt 3 gezeigt wurde. Damit lässt sich zusammenfassend sagen, dass das Ungarische wohl kaum eine grammatische Kategorie „Aspekt“ besitzt, jedoch viele aspektuelle Funktionen durch Präverbien ausgedrückt werden können.

## 5. Literatur

- Breu, Walter. 2000. „Zur Position des Slavischen in einer Typologie des Verbalaspekts (Form, Funktion, Ebenenhierarchie und lexikalische Interaktion)“. In: Breu, Walter (Hrg.): *Probleme der Interaktion von Lexik und Aspekt (ILA)*. (Linguistische Arbeiten 412). Tübingen: Max Niemeyer Verlag. 21–54.
- Bybee, Joan. 1985. *Morphology. A Study of the Relation of Meaning and Form*. (Typological Studies in Language. Vol. 9). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Bybee, Joan, Perkins, Revere & William Pagliuca. 1994. *The Evolution of Grammar. Tense, Aspect, and Modality in the Languages of the World*. Chicago/London: The University of Chicago Press.
- Comrie, Bernard. 1976. *Aspect. An Introduction to the Study of Verbal Aspect and Related Problems*. (Cambridge textbooks in linguistics). Cambridge: Cambridge University Press.
- Dékány, Éva. 2008. „El + verb complex predicates in Hungarian“. In: Peter Svenonius & Inna Tolskaya: *Tromsø Working Papers on Language & Linguistics: Nordlyd 35*, special issue on Complex Predication. Tromsø: CASTL. 1–17.
- Farkas, Donka F. & Sadock, Jarrold M. 1989. „Preverb Climbing in Hungarian“. *Language*. Vol. 65

(2). 318–338.

Kiefer, Ferenc. 2006. *Aspektus és akcióminőség. Különös tekintettel a magyar nyelvre*. Budapest: Akadémiai Kiadó.

Lehmann, Volkmar. 1988. „Der russische Aspekt und die lexikalische Bedeutung des Verbs“. *Zeitschrift für slavische Philologie*. Vol. 48. 171–181.

Lehmann, Volkmar. 2010. „Der Aspekt in den slavischen Sprachen im Licht der kognitiven Linguistik“. In: Tanja Anstatt & Boris Norman (Hrg.): *Die slavischen Sprachen im Licht der kognitiven Linguistik / Slavjanskije jazyki v kognitivnom aspekte*. Wiesbaden: Harrasowitz. 77–99.

Mátai, Mária. 1989. *Igekötőrendszerünk történetéből*. Budapest: A Magyar Nyelvtudományi Társaság.

Plungian, Vladimir. 2011. *Vvedenie v grammatičeskiju semantiku. Grammatičeskie značenija i grammatičeskie sistemy v jazykakh mira*. Moskau: RGGU.

Soltész, Katalin J. 1959. *Az ősi magyar igekötők. Meg, el, ki, be, le, fel*. Budapest: Akadémiai Kiadó.

Van Valin, Robert. 2001. *An Introduction to Syntax*. Cambridge: Cambridge University Press.

Vendler, Zeno. 1957. „Verbs and Times“. *The Philosophical Review*. Vol. 66 (2). 143–160.



# Vergleichendes Verbalparadigma: Standarddeutsch – Mittelhessisch/Oberhessisch

*Timm Braun (Universität Stuttgart)*

## 1. Was ist Mittelhessisch/Oberhessisch?

Das Mittelhessische bzw. Oberhessische gehört zu den westmitteldeutschen Dialekten. Von den Sprechern wird die Varietät oft als „Platt“ bezeichnet. Das Mitteldeutsche liegt wie eine horizontale Sanduhr zwischen dem Niederdeutschen und dem Oberdeutschen und stellt eine Übergangszone dar, deren Grenzen schwer festzustellen sind. Mittelhessisch ist ein Ausdruck, der sich auf die Verbreitung des Dialekts im heutigen Bundesland Hessen bezieht, während Oberhessisch seinen Ursprung in der Bezeichnung der Provinz Oberhessen im Großherzogtum Hessen-Darmstadt hat. Der Begriff Zentralhessisch, wie er in der Karte angegeben ist, ist genauer auf die Region zwischen Lahn und Main, sowie zwischen Westtaunus und Vogelsberg bezogen. Die Untersuchung bezieht sich auf die Ortsmundart von Steinfurth in der Wetterau und streift die umliegenden Dörfer. Dieses Gebiet befindet sich etwa in der Mitte des Sprachraumes. Der Aufsatz stellt keine abschließenden Ergebnisse vor, sondern Besonderheiten und Phänomene, die noch genauer zu untersuchen und gegebenenfalls zu berichtigen sein werden.



Abbildung 1: <http://www.uwefrank.de/mundartUF.htm> (21.10.2013)

Die Daten stammen hauptsächlich aus meinem familiären Sprachgebrauch und sind introspektiv bzw. durch Befragung erhoben. Im Rahmen meiner Masterarbeit werde ich diese noch präzisieren und gegebenenfalls berichtigen. Ebenso die Transkription, welche nur ein ungefähres Lautbild vermitteln soll. Die Graphie orientiert sich an der (teilweise sehr uneinheitlichen) mittelhessischen Literaturschreibweise.

## 2. Lautliche Besonderheiten

Kennzeichnend für die Dialekte Mittelhessens ist das koronale *r* [r], welches obligatorisch ist. Das uvulare *r* [R] klingt dagegen stark markiert und wird von Muttersprachlern so gut wie nicht verwendet. Im Nordwesten des Sprachgebiets, im Westerwald, stellt für die hessischen Mundarten sogar das retroflexe *r* [ɹ] die Norm dar.

Während das Deutsche drei Diphthonge kennt ([aʊ],[aɪ],[ɔɪ]), verwendet das Oberhessische zusätzlich noch mindestens sechs weitere. Es handelt sich um vier fallende Diphthonge unter (1), sowie zwei zentrierende Diphthonge, die jeweils auf Schwa enden unter (2).

- (1) [ɛɪ] Bäier [bɛɪər] (Bier)  
[ɔʊ] Schou [ʃɔʊ] (Schuhe)  
[ɛɐ̯] Dreack [dɾɛɐ̯k] (Dreck)  
[ɔɛ] Borg [bɔɛk] (kastrierter Eber)
- (2) [ɪə] Gieft [gɪəft] (Gift)  
[ʊə] Duett [dʊət] (Tüte)

Was die mittelhessische Aussprache außerdem deutlich von der standarddeutschen unterscheidet, ist die Nicht-Realisierung eines Schwa am Wortauslaut auf *-er*. So wird im Standarddeutschen das Wort *bitter* wie in (3a) mit [ʌ] ausgesprochen. Im Mittelhessischen würde dies mit finalem koronalem [r] realisiert (3b).

- (3) a. bitter [bitʌ]  
b. bitter [bɪətr]

Ein Problem für die phonetische Notation stellt das kaum silbenfähige Schwa in *Blerrer* (Blätter) dar, welches gleich in beiden Silben zu finden ist. Es stellt sich noch die Frage ob man es mit einem vokalisiertem *r* zu tun hat oder ob wirklich ein [ə] realisiert wird.

### 3. Verbalflexion

#### 3.1 Infinite Verbformen

Die Suffixe für Infinitive im Standarddeutschen sind *-n* und *-en*, die sich abhängig von bestimmten Kennzeichen auf die Verben verteilen. Das Mittelhessische kennt drei unterschiedliche Morpheme, die sich teilweise regelhaft auf die Verben verteilen. Das häufigste Suffix ist ein einfaches Schwa (4a), welches auch für die meisten Neologismen verwendet wird. Ein *-n*-Suffix erhalten Verben grundsätzlich, wenn die standarddeutsche Entsprechung vor dem *-en* einen Liquid aufweist, wie man in (4b) erkennen kann. Dabei gibt es aber auch einige Ausnahmen. Zur Gruppe der Verben mit Nullmorphem (4c) gehören besonders hochfrequente Verben, sowie die Auxiliare *haben* und *sein*. Auch hier kann man dies nicht an einer bestimmten lautlichen Gestalt des Standarddeutschen festmachen.

- (4) a. glawe [gla:və] (glauben)  
fläje [flejə] (fliegen)  
fächte [fɛ:çtə] (fürchten)
- b. foin [fɔɪn] (fühlen)  
spinn [ʃpɪn] (spielen)  
foahrn [fɔɐn] (fahren)
- c. gieh [gi:] (gehen)  
seu [zɔɪ] (sein)  
troa [trɔ:] (tragen)  
hu [hʊ] (haben)

Das Partizip im Präsens ist im Oberhessischen ungebräuchlich, bis auf einige feste Wendungen, wie *kochend haas* (kochend heiß) oder *de leawiche Duud* (der lebende Tod). Das Suffix des zweiten Beispiels unterscheidet sich hier deutlich von der hochdeutschen Entsprechung. Perfektpartizipien hingegen werden produktiv gebildet. Das Standarddeutsche bietet hierfür zwei Möglichkeiten: die Zirkumfixe *ge - en* und *ge - t*. Die erstgenannte Variante wird für starke, die zweite für schwache Verben angewendet. Im Mittelhessischen gibt es auch hier eine größere Formenvielfalt. Unter (5a) finden sich Beispiele für die Partizipienbildung bei starken Verben, wobei noch abschließend zu klären ist, nach welchen Regeln welches Zirkumfix auf die verbalen Stämme angewendet wird. (5b) zeigt die Möglichkeit für schwache Verben, die sich mit der standarddeutschen Variante deckt.

- (5) a. gə - ə gefuene [gə'fʊənə] (gefunden)  
 gə - n gefoahrn [gə'fʌrn] (gefahren)  
 gə - Ø gedoo [gə'do:] (getan)  
 Ø - ə gonge ['gʊŋə] (gegangen)
- b. gə - t getraat [gə'tra:t] (getragen)

Ein weiterer Unterschied zum Hochdeutschen lässt sich beim Verben feststellen, die in Bezug auf ihre Regelmäßigkeit von der standarddeutschen Entsprechung abweichen. So gibt es einige Beispiele für Verben, deren standarddeutsche Variante stark konjugiert wird, die im Mittelhessischen aber die schwache Konjugation aufweisen, so (5b) oder (6a). Außerdem gibt es wie in zahlreichen anderen Dialekten Partizipien, die ohne *ge*-Präfix gebildet werden (6b).

- (6) a. baje gebaigt [gə'baikt] (biegen - gebogen)  
 vedearwe vedeapt [fə'dɛəpt] (verderben - verdorben)
- b. geann goann [gʌn] (gelten - gegolten)  
 krije kritt [kritt] (kriegen - gekriegt)

### 3.2 Finite Verbformen

Das auffälligste regelhafte Merkmal der Konjugation im Mittelhessischen ist der Synkretismus zwischen dem Infinitiv und den Präsensformen der 1. Person Singular und Plural und der 3. Person Plural. Dieser Synkretismus wird im Hochdeutschen nur in der 1. und 3. Person Plural realisiert. Im Gegensatz dazu herrscht im Oberhessischen auch in der 1. Person Singular eine große Regelmäßigkeit, die sich sogar auf das Auxiliar *sein* erstreckt. In (7) ist zu erkennen, dass dieser „113-Synkretismus“ für alle oben erwähnten Infinitiv-Formen anwendbar ist. Ausnahmen zu dieser Regel bilden die Modalverben, jedoch nicht *sollen*, und *wissen*.

- (7) roufe [rʊfə] (rufen): aich roufe, me roufe, däi roufe  
 bestenn [bə'ʃtɛn] (bestellen): aich bestenn, me bestenn, däi bestenn  
 seu [zʊ] (sein): aich seu, me seu, däi seu

Das Standarddeutsche kennt bei der Konjugation im Präsens und auch im Präteritum den regelhaften Ablaut bei Vokalen (8a). Im Mittelhessischen ist dieser Wechsel auch vorhanden (8b), jedoch gibt es auch einen Konsonantenwechsel, wie er beispielsweise beim Verb *fragen* angewendet wird (8c).

- (8) a. helfen: ich helfe, du hilfst, er hilft, wir helfen, ...

- b. komme [kɔmə] (kommen): aich komme, dou kimmst, der kimmt,  
me komme, ihr kommt, däi komme
- c. freje [frɛjə] (fragen): aich freje, dou fregst frɛ:kst

Dieses Phänomen lässt sich bei näherer Betrachtung als regelmäßiger Konsonantenwechsel bezeichnen, da sich bei bestimmten Verben der stammauslautende Konsonant regelhaft in der 2. und 3. Person Singular ändert. Man kann zu einem vorläufigen Ergebnis kommen, welches die Konsonantenwechsel in (9a) zeigen. In (9b) sind Beispiele aufgeführt, die unter anderem auch die Kombination von Vokal- und “Konsonantablaut” verdeutlichen. Bei der Bildung des Vergangenheitspartizips behalten schwache Verben den Konsonantablaut bei, während starke Verben vom Infinitivstamm ausgehen.

- (9) a. aje/äje → aig/äig; eje → eg; w → p; r → t
- b. baje [baɪjə] (biegen): aich baje, dou baigst [baɪkst]
- howwe [hɔvə] (heben): aich howwe, dou hobbst [hɔpst]
- ruere [rʊərə] (raten): aich ruere, dou rietst [ri:tst]

Auch für diese Regel gibt es einige Ausnahmen, so müsste *laje* [laɪjə] (liegen), eigentlich in der 2. Person Singular ein *k* einfügen, dies wird aber nicht realisiert. Ebenso gehört das Verb *haa* [ha:] (hauen) eigentlich nicht zur Gruppe mit regelhaftem Konsonantenwechsel, dennoch findet dieser statt.

### 3.3 Tempus und Modus

Für das Hochdeutsche gilt, dass es zwei synthetische Tempora gibt: Präsens und Präteritum. Das Oberhessische kennt hingegen strenggenommen nur ein Tempus, da die regelmäßige Bildung von Präteritum fehlt. Dieses existiert lediglich in einigen Formen, die hauptsächlich hochfrequente Verben sind (10).

- (10) gieh [gi:] (gehen)      aich gong [gɔŋ] (ging)
- hu [hʊ] (haben)          aich hatt [hat] (hatte)
- seu [zɔɪ] (sein)          aich woar [vɔər] (war)

Das Perfekt und das Plusquamperfekt werden wie im Standarddeutschen aus Hilfsverb und Vollverb analytisch gebildet. Das Futur I ist im Mittelhessischen eine sehr seltene Form. Häufiger ist der Gebrauch des Präsens, welches durch temporale Adverbien futurisch verwendet wird.

Der Konjunktiv kann anders als im Hochdeutschen nicht synthetisch gebildet werden. Ausnahmen bilden hier auch wieder hochfrequente Verben, wie die Hilfsverben. Regelmäßig wird die analytische Umschreibung mit *deed* (täte) oder *wir* (wäre), wie in (11).

- (11)   aich deed brenge [det brɛŋə] (ich brächte)  
      dou wirst gewease [wiəst gəvɛɪzə] (du wärest gewesen)

#### **4. Literatur**

Friebertshäuser, Hans. 1987. *Das hessische Dialektbuch*. München.

Hasselberg, Joachim. 1979. *Differenzgrammatik – Mittelhessisch: Hochsprache*. Gießen.

# **Grammatische und semantische Kongruenz bei anaphorischem Verweis auf Diminutiv-Nominalphrasen: Überlegungen zum Einfluss syntaktischer Komplexität**

*Henning Czech (Universität Osnabrück)*

## **1. Vorbemerkungen**

Im Rahmen eines Studienseminars im Wintersemester 2012/13 habe ich mich mit Referenzbezügen in Schüleraufsätzen befasst. Es hat sich in dem Zusammenhang ein interessanter Untersuchungsansatz zur semantischen Kongruenz herausgestellt, den ich in meinem Aufsatz skizzieren werde. An der Erhebung und Auswertung der im Folgenden präsentierten Daten war meine Kommilitonin Lara Muchall beteiligt.

Ich orientiere mich an folgender Gliederung: In einem ersten Schritt führe ich in die Thematik ein, indem ich den Begriff der semantischen Kongruenz bestimme und den Gegenstandsbereich der Überlegungen einschränke. Anhand einer Analyse von Schülertexten aus dem Ludwigsburger Aufsatzkorpus (Fix & Melenk 2002) gebe ich dann einen Überblick über die Wahl semantischer Kongruenz beim Rückverweis auf Diminutiv-NPs. Der Hauptteil des vorliegenden Aufsatzes behandelt die Frage, inwiefern die syntaktische Komplexität einen Einfluss auf die Wahl grammatischer bzw. semantischer Kongruenz haben könnte.

## **2. Einführung: Grammatische und semantische Kongruenz**

Wenn durch ein Pronomen oder Artikelwort bzw. „Artikelpronomen“ (vgl. Thieroff 2012: 120) auf ein Antezedens verwiesen wird, dann kongruiert das Pro-Wort mit dem Antezedens in der Regel in den Kategorien Numerus und Genus. Ist das nicht der Fall, liegt aus grammatischer Perspektive ein Kongruenzfehler vor. Ich gebe ein Beispiel aus einem Aufsatz eines Viertklässlers:

(1) *es war ein mal [ein kleines Mädchen]<sub>i</sub>. [es]<sub>i</sub> wohnte in einem kleinen dorf in einem kleinen Haus mit [ihrer]<sub>i</sub> schwester und [ihrem]<sub>i</sub> Vater [...].*

(Textsammlung Musan/Noack)

Das Pronomen *es* trägt das Genus Neutrum und kongruiert auf grammatischer Ebene mit dem Ante-

zedens *ein kleines Mädchen*. Bei den darauf folgenden Artikelwörtern *ihrer* und *ihrem* wechselt das Genus aber ins Femininum. Trotz der fehlenden grammatischen Kongruenz erscheint der Textauszug den meisten Muttersprachlern vermutlich völlig akzeptabel.<sup>1</sup> Erklären lässt sich die Akzeptabilität dadurch, dass das Antezedens in der außersprachlichen Welt auf eine weibliche Person verweist (vgl. Duden 2006: 964). Bei der Wahl des zweiten und dritten Verweiswortes richtet sich der Schreiber hier also nicht nach dem Genus des Antezedens, sondern nach dem Sexus des Referenten. Dieses Phänomen, das durch den Status der Bezugssubstantive als „lexical hybrids“ (Corbett 2006: 163) ermöglicht wird, bezeichnet man als „semantic agreement“ (Corbett 1979: 203) bzw. „semantische Kongruenz“ (Duden 2006: 964). Daneben finden sich in der Fachliteratur die weitgehend deckungsgleichen Bezeichnungen „Konstruktion nach dem Sinn“ (ebd.), „constructio ad sensum“, „Konstruktion kata synesis“ (Duden 2009: 946), „Synesis“, „Sinnkongruenz“ (Jaeger 1992: 54), „pragmatische Kongruenz“ (Köpcke 2012: 38), „notional agreement“, „logical agreement“ (Corbett 2006: 155).

Natürlich tritt semantische Kongruenz auch in Bezug auf das Merkmal Numerus auf (vgl. op.cit.: 157); als prototypisches Beispiel für das Deutsche betrachte ich im Folgenden aber ausschließlich den Genuswechsel beim Verweis auf Diminutiv-NPs.

Nachdem ich zunächst ein Textbeispiel eines vergleichsweise ungeübten Schreibers verwendet habe, stellt sich die Frage, ob semantische Kongruenz auch bei älteren Schreibern noch zu beobachten ist. Dazu verweise ich auf ein Zitat von Peschel (2006: 180). Die Autorin bezieht sich allgemein auf 'Kongruenzfehler' beim Rückverweis durch Pronomen.

(2) „Solche Phänomene findet man auch in Texten von älteren Schülern und sogar Studierenden noch, wenn Satzbau oder thematische Struktur zu komplex werden.“

Aufgrund dieser Aussage haben wir im DWDS-Korpus nach Beispielen recherchiert. Dabei hat sich bestätigt, dass auch erfahrene Schreiber beim Verweis auf Diminutiv-NPs teilweise die semantische Kongruenz bevorzugen. Das lässt sich an einem Textausschnitt aus der ZEIT veranschaulichen:

(3) *Am Abend der Wahl schluchzte [ein junges Mädchen]<sub>i</sub>; [sie]<sub>i</sub> erlebe einen Albtraum, denn Nicolas Sarkozy würde ein rassistischer und freiheitstötender Präsident werden.*

(Die Zeit, 17.05.2007, Nr. 20; DWDS-Korpus)

Auch hier kongruiert das Pronomen im Genus nicht mit dem Antezedens.

---

1   Darauf deutet eine informelle Befragung hin, die wir mit einigen Studierenden durchgeführt haben.



### 3. Semantische Kongruenz in Schüleraufsätzen (Ludwigsburger Aufsatzkorpus)

Die beschriebenen Beobachtungen legen es nahe, auch Texte älterer Schüler auf semantische Kongruenz zu untersuchen. Dazu haben wir einige Aufsätze von Hauptschülern und Gymnasiasten der 8. Klasse gewählt. Der Bildungsgrad der Schüler wurde damit neben dem Schuljahrgang als weitere Variable berücksichtigt. Zunächst haben wir alle Textstellen extrahiert, bei denen eine Diminutiv-NP als einzig mögliches Antezedens für ein nachfolgendes Pro-Wort in Frage kommt. Zwischen Antezedens und Verweiswort tritt dabei entweder grammatische (4a) oder semantische Kongruenz (4b) auf:

(4) a. *Irmela Bender schreibt in ihrer Liebesgeschichte [...] über [ein Mädchen]<sub>i</sub>, [das]<sub>i</sub> noch nie einen Freund hatte [...].*

(Fix & Melenk 2002, Schreiber: 3jb)

b. *[Ein junges, eher schüchternes Mädchen]<sub>i</sub> hat [ihren]<sub>i</sub> ersten Freund. [...] Mit ihm kann [sie]<sub>i</sub> alles machen.*

(Fix & Melenk 2002, Schreiber: 14dh)

Der Einfachheit halber spreche ich im Folgenden von *Antezedenzen* und *koreferenten Ausdrücken*. Für Relativsätze wie in (4a) gestaltet sich die Beziehung zwischen dem Kopf der NP und dem Pronomen etwas komplizierter, da es sich bei Relativpronomen in restriktiven Relativsätzen um gebundene Variablen handelt (vgl. Heim & Kratzer 1998: 87). Das soll hier aber zunächst unberücksichtigt bleiben.

Bei der Analyse haben wir, wenn möglich, jeweils die ersten beiden Verweise auf die Diminutiv-NP betrachtet. Wenn sowohl beim ersten als auch beim zweiten Verweis die semantische Kongruenz bevorzugt wurde, haben wir den zweiten Versuch als ‚Folgefehler‘ gewertet und nicht in die Analyse aufgenommen. Das ist beispielsweise beim Textauszug (4b) der Fall. Insgesamt hat sich nach diesem Auswahlverfahren eine relativ kleine Datengrundlage ergeben:

Datengrundlage Ludwigsburger Aufsatzkorpus:

Gymnasiasten: 27 Diminutiv-NPs, 34 Verweise, 20 Schreiber

Hauptschüler: 21 Diminutiv-NPs, 25 Verweise, 15 Schreiber

Abgesehen von einer Textstelle handelt es sich beim Kopf der jeweiligen Diminutiv-NP, die als Antezedens fungiert, ausschließlich um das Wort *Mädchen*. Die Auswertungsergebnisse sind in Tabelle 1 zusammengestellt.

Bevorzugung semantischer Kongruenz bei Gymnasiasten	61,76 %
... bei Hauptschülern	72,00%
... bei Hauptschülern mit Deutsch als Erstsprache	83,33%
... bei Hauptschülern mit Deutsch als Zweitsprache	68,42%
... bei Mädchen insgesamt	52,00%
... bei Jungen insgesamt	76,00%
Durchschnittlicher Wortabstand zwischen Verweiswort und nächstmög- lichem Antezedens bei Bevorzugung semantischer Kongruenz	3,92 Wörter
... bei Bevorzugung grammatischer Kongruenz	2,64 Wörter

*Tabelle 1: Auswertungsergebnisse Ludwigsburger Aufsatzkorpus*

Auch wenn die Ergebnisse bei dieser Stichprobengröße natürlich wenig repräsentativ sind, lassen sich der Tabelle einige Tendenzen entnehmen: Zunächst fällt auf, dass der Anteil der Fälle, in denen semantische Kongruenz bevorzugt wird, insgesamt deutlich höher als 50% liegt. Im Vergleich zu den Gymnasiasten bevorzugen die Hauptschüler häufiger<sup>2</sup> semantische Kongruenz. Bemerkenswert ist außerdem eine Differenzierung innerhalb der Gruppe der Hauptschüler: Die Schüler mit Deutsch als Erstsprache wählen häufiger die semantische Kongruenz als die Schüler mit Deutsch als Zweitsprache. Und schließlich ist zu beobachten, dass Jungen häufiger als Mädchen semantische Kongruenz bevorzugen. Zusätzlich zu diesen Ergebnissen haben wir den durchschnittlichen Wortabstand zwischen Verweiswort und letztgenanntem koreferenten Ausdruck berechnet. Dieser Abstand ist bei semantischer Kongruenz mit 3,92 Wörtern deutlich größer als bei grammatischer Kongruenz – dort sind es nur 2,64 Wörter.

#### **4. Semantische Kongruenz und syntaktische Komplexität**

Nachdem sich gezeigt hat, dass selbst in Texten von älteren Schülern und von Erwachsenen noch semantische Kongruenz zu finden ist, scheint diese Beobachtung bezüglich des Wortabstands die zweite Aussage Peschels zu bestätigen. Hierzu noch einmal das Zitat:

(2') „Solche Phänomene findet man auch in Texten von älteren Schülern und sogar Studierenden noch, wenn Satzbau oder thematische Struktur zu komplex werden.“

(Peschel 2006: 180)

Die Frage ist nun, ob es auch Evidenz dafür gibt, dass mit zunehmender syntaktischer Komplexität die Präferenz für semantische Kongruenz zunimmt. Um das zu testen, haben wir einen Fragebogen entwickelt, bei denen die Probanden zwischen Personalpronomen wählen sollten, die entweder

<sup>2</sup> Da es sich um eine explorative Studie handelt, wurde auf eine inferenzstatistische Auswertung der Daten verzichtet.

grammatisch oder semantisch mit dem Antezedens kongruierten. Als Indikator für syntaktische Komplexität haben wir wegen der genannten Tendenzen den Wortabstand zwischen Pronomen und Antezedens gewählt. Dieser Parameter wurde variiert, indem wir die Diminutiv-NPs links bzw. rechts des Phrasenkopfes ausgebaut haben. Als Diminutiv wurde ausschließlich das Wort *Mädchen* verwendet. Der Fragebogen enthält 8 Testsätze mit unterschiedlich stark ausgebauter Diminutiv-NP. Dazu zwei Beispielsätze:

(5) a. *Annika ist [ein Mädchen von 16 Jahren]. Sein/Ihr Geburtstag ist im August.*

b. *Brigitte ist [ein Mädchen mit viel Spaß am Fußballspielen]. Ihr/Sein Lieblingsspieler ist Thomas Müller.*

Insgesamt wurde der Fragebogen von 56 Erstsemester-Studierenden der Germanistik ausgefüllt. Weil dabei fast ausschließlich semantische Kongruenz bevorzugt wurde, hat sich der vermutete Zusammenhang zwischen erhöhtem Wortabstand und der Wahl semantischer Kongruenz nicht bestätigt. Zu beachten ist aber, dass der Fragebogen im Nachhinein eher ungünstig gestaltet war: Der eingangs genannte Eigenname könnte ein alternatives Antezedens darstellen; und außerdem fungiert die Diminutiv-NP syntaktisch als Prädikativ, was bei der Anaphernresolution möglicherweise zur Bevorzugung des Eigennamens (in diesem Fall gleichzeitig des Topiks) führt. Das liegt daran, dass ein Prädikativ semantisch gesehen kein Individuum bezeichnet, sondern eine Eigenschaft ausdrückt. Die Suche nach einem weiteren Indikator für syntaktische Komplexität hat zur Entwicklung eines neuen, zweiten Fragebogens geführt. Auch daraus zwei Beispielsätze:

(6) a. *Jedes Mädchen freut sich, wenn es/sie viele Geschenke bekommt.*

b. *Das Herrchen ist verärgert, wenn er/es im Regen mit dem Hund nach draußen muss.*

Wie hier bereits zu sehen ist, haben wir bei den neuen Testsätzen nicht nur die beiden Diminutive *Mädchen* und *Herrchen* gleichmäßig auf die Sätze verteilt, sondern auch weitere Veränderungen vorgenommen: Zum einen ist das Diminutiv natürlich das einzig mögliche Antezedens. Zum anderen befindet sich das Verweiswort in einem Nebensatz ersten Grades. Mit der Wahl der hypotaktischen Konstruktion geht in diesen Beispielen einher, dass das Pronomen von der Diminutiv-NP c-kommandiert wird. Ich erläutere kurz, was damit gemeint ist:

*C-Kommando* ist ein wesentlicher Grundbegriff der Bindungstheorie, die im Rahmen der Generativen Grammatik entwickelt wurde (vgl. Reinhart 1976, Chomsky 1986). Der Terminus beschreibt eine grammatische Relation zwischen über- und untergeordneten Konstituenten, die unter anderem folgendermaßen definiert wird:

- (7)  $\alpha$  c-kommandiert  $\beta$ ,  $\alpha \neq \beta$ , genau dann, wenn
- a) jeder verzweigende Knoten  $\gamma$ , der  $\alpha$  dominiert, auch  $\beta$  dominiert,
  - b)  $\alpha$   $\beta$  nicht dominiert.
- (Philippi & Tewes 2010: 145)

Diese Konstituentenbeziehung veranschaulicht der Strukturbaum für einen der Testsätze aus Fragebogen II:

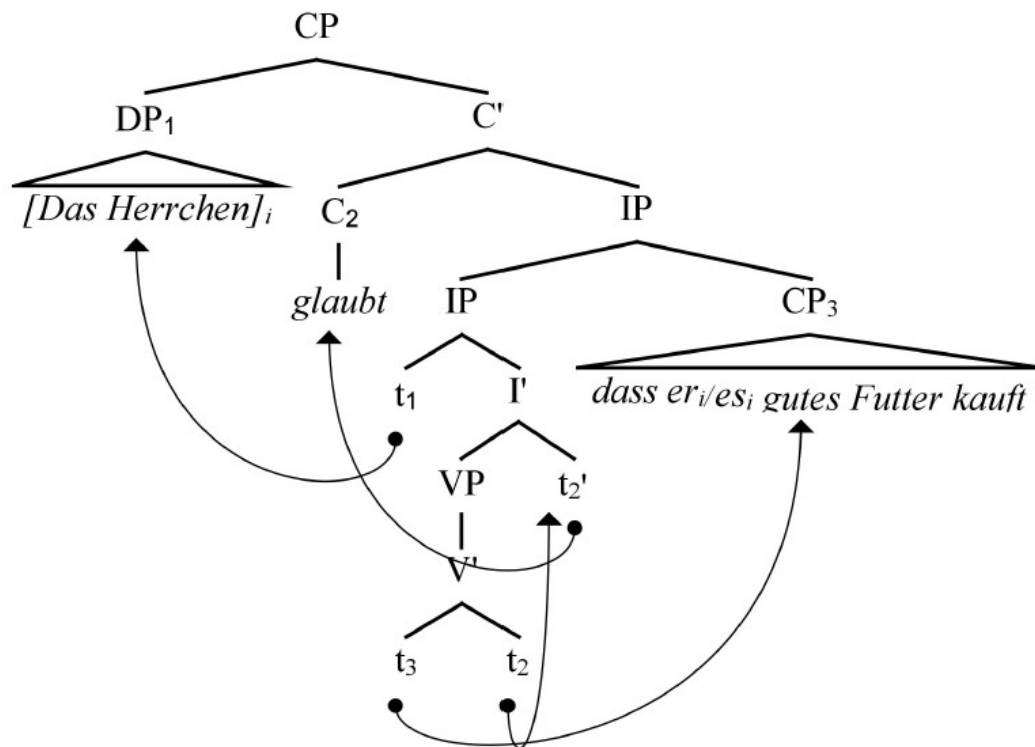


Abbildung 1: Strukturbaum zu Fragebogen II

Der nächste verzweigende Knoten, der die Phrase *das Herrchen* (in diesem Modell eine DP) dominiert, ist die CP. Diese dominiert neben der DP auch den Knoten C' sowie alle untergeordneten Konstituenten. Zwischen dem Antezedens und dem Pronomen *er* bzw. *es* liegt demnach die als C-Kommando bezeichnete syntaktische Relation vor.

Zu Beginn meiner Ausführungen habe ich mich an der Aussage von Peschel (2006) orientiert und vermutet, dass syntaktische Komplexität die Wahl semantischer Kongruenz fördert. Vielleicht ist es aber eher so, dass bestimmte syntaktische Konstellationen grammatische Kongruenz fördern. Um das zu erläutern, muss ich etwas ausholen:

Man weiß aus verschiedenen Konstruktionstypen, dass C-Kommando einen besonders engen syntaktisch-semantischen Bezug zwischen Antezedens und Pronomen ermöglicht. Erstens ist das im Falle von Reflexivpronomen zu beobachten: Nach dem Prinzip A der Bindungstheorie muss ein Re-

flexivpronomen in seiner Bindungsdomäne gebunden sein (vgl. Carnie 2013: 482). Etwas vereinfacht bedeutet das: Ein Reflexivpronomen braucht im Regelfall ein Antezedens, das im selben Satz steht und von dem das Reflexivpronomen c-kommandiert wird (vgl. op.cit.: 154f.). Dies zeigen die Beispielsätze in (8). Das Pronomen *sich* in (8b) wird von der NP *den Minister* nicht c-kommandiert und kann demnach nicht mit dieser NP koindiziert werden.

(8) a. *[Der Präsident]<sub>i</sub> hat [den Minister]<sub>j</sub> mit [sich]<sub>i/j</sub> ausgesöhnt.*

b. *[Der Präsident]<sub>i</sub> hat mit [sich]<sub>i/\*j</sub> [den Minister]<sub>j</sub> ausgesöhnt.*

Zweitens besteht zwischen einer quantifikationellen NP und einem koindizierten Pronomen eine semantisch sehr enge Beziehung der Variablenbindung (vgl. Lohnstein 2011: 81); dafür muss normalerweise zwischen der quantifikationellen NP und dem Pronomen ebenfalls C-Kommando vorliegen (vgl. Reinhart 1991: 537). Ein Vergleich der folgenden Sätze soll dies veranschaulichen:

(9) a. *Eine Frau, die [den Mann]<sub>i</sub> kennt, glaubt, dass [er]<sub>i</sub> seine Mutter liebt.*

b. *?? Eine Frau, die [jeden Mann]<sub>i</sub> kennt, glaubt, dass [er]<sub>i</sub> seine Mutter liebt.*

c. *[Jeder Mann]<sub>i</sub> sagt, dass [er]<sub>i</sub> seine Mutter liebt.*

Die Sätze (a) und (b) haben dieselbe Struktur: Das Pronomen kann vom Antezedens nicht c-kommandiert werden, weil das Antezedens *den Mann* bzw. *jeden Mann* zu tief in die Subjekt-NP des Matrixsatzes eingebettet ist. Satz (a) mit definitiver NP klingt bei Koindizierung der markierten Ausdrücke akzeptabel, Satz (b) mit quantifikationeller NP weniger. In Satz (c) dagegen liegt C-Kommando vor. Der Verweis auf die quantifikationelle NP ist wiederum völlig akzeptabel.

Die Überlegung ist nun die folgende: Wenn C-Kommando eine besonders enge Beziehung zwischen Antezedens und Pronomen schafft, dann gibt es vielleicht auch einen Zusammenhang zwischen dem Auftreten der strengeren grammatischen Kongruenz und dem Vorliegen von C-Kommando. Insbesondere könnte das der Fall sein, wenn quantifikationelle NPs als Antezedenzen auftreten. Um dies zu überprüfen und den vermuteten Effekt nicht zu verfehlen, haben wir eine weitere Variable eingebaut und sowohl definite als auch quantifikationelle Diminutiv-NPs als Antezedenzen verwendet.

Ebenso wie der erste Fragebogen enthält die überarbeitete Version 8 Testsätze. Der Fragebogen wurde von 90 anderen Erstsemester-Studierenden ausgefüllt. Die Auswertungsergebnisse habe ich

in Tabelle 2 zusammengestellt:

Bevorzugung grammatischer (!) Kongruenz bei NP mit Quantor ( <i>Mädchen</i> )	74,45%
... bei NP mit Quantor ( <i>Herrchen</i> )	35,39%
... bei NP mit def. Artikel ( <i>Mädchen</i> )	55,00%
... bei NP mit def. Artikel ( <i>Herrchen</i> )	47,78%

Tabelle 2: Auswertungsergebnisse C-Kommando

Bei den definiten NPs ist weder für grammatische noch für semantische Kongruenz eine deutliche Präferenz zu erkennen. Bemerkenswert ist aber die Verteilung bei den Verweisen auf quantifikatorische NPs: Beim Diminutiv *Mädchen* wird die grammatische Kongruenz eindeutig präferiert, wenn die NP quantifikatorisch ist. Beim Diminutiv *Herrchen* zeigt sich aber die gegenläufige Tendenz: Tritt das Diminutiv in einer quantifikatorischen NP auf, dann wird in den meisten Fällen semantische Kongruenz gewählt. Für dieses Ergebnis könnte die Tatsache eine Rolle spielen, dass die Frequenz der beiden Diminutive im alltäglichen Sprachgebrauch sehr unterschiedlich hoch ist. Konkret würde das bedeuten, dass sich mit steigender Frequenz des Diminutivs, das innerhalb einer quantifikatorischen NP auftritt, die Präferenz für grammatische Kongruenz beim Verweiswort erhöht (zum Einfluss der Frequenz vgl. z. B.: Oldfield & Wingfield 1965). Natürlich müsste das in einer gesonderten Studie untersucht werden.

## 5. Ausblick

Das ist der aktuelle Stand. Neben den genannten Aspekten ließe sich in einer größer angelegten Studie beispielsweise untersuchen, ob die Wahl von grammatischer bzw. semantischer Genuskongruenz im Deutschen in Abhängigkeit von unterschiedlichen Pro-Wort-Typen variiert. Aufgrund der Distribution von grammatischer und semantischer Kongruenz in den Daten des Ludwigsburger Aufsatzkorpus bietet es sich an, Relativpronomen (ebenfalls gebundene Variablen) und sonstige Pro-Wörter einander gegenüberzustellen. Die Vermutung wäre dabei, dass bei Relativpronomen häufiger grammatische Kongruenz gewählt wird als bei Personalpronomen. Dies wiederum würde mit den Annahmen der „Agreement Hierarchy“ (Corbett 1979, The Universals Archive: Nr. 576) übereinstimmen – einer implikationellen Hierarchie, die einen Zusammenhang zwischen der strukturellen Distanz eines kongruierenden Elementes zu seinem Bezugsausdruck einerseits und der Wahrscheinlichkeit für das Auftreten semantischer Kongruenz andererseits beschreibt.

## 6. Literatur

- Carnie, Andrew. 2013. *Syntax. A Generative Introduction*. 3. Auflage. Malden etc.: Wiley-Blackwell.
- Chomsky, Noam. 1986. *Barriers*. 3. Auflage. Cambridge, Massachusetts: The MIT Press.
- Corbett, Greville G. 1979. „The Agreement Hierarchy.“ *Journal of Linguistics* 15. 203–224.
- Corbett, Greville G. 2006. *Agreement*. New York: Cambridge University Press.
- Duden 2006. *Die Grammatik. Unentbehrlich für richtiges Deutsch*. Überarbeiteter Neudruck der 7., völlig neu erarbeiteten und erweiterten Auflage. Mannheim etc.: Dudenverlag.
- Duden 2009. *Die Grammatik. Unentbehrlich für richtiges Deutsch*. 8., überarbeitete Auflage. Mannheim, Zürich: Dudenverlag.
- DWDS-Korpus: <http://www.dwds.de/> (letzter Zugriff am 01.09.2013).
- Fix, Martin & Melenk, Hartmut 2002. *Schreiben zu Texten – Schreiben zu Bildimpulsen: das Ludwigburger Aufsatzkorpus. Mit 2300 Schülertexten, Befragungsdaten und Bewertungen*. 2., korrigierte Auflage. Baltmannsweiler: Schneider-Verlag.
- Heim, Irene & Kratzer, Angelika. 1998. *Semantics in Generative Grammar*. Malden: Blackwell.
- Jaeger, Christoph. 1992. *Probleme der syntaktischen Kongruenz. Theorie und Normvergleich im Deutschen*. Tübingen: Niemeyer.
- Köpcke, Klaus-Michael. 2012. „Konkurrenz bei der Genuskongruenz. Überlegungen zum Grammatikunterricht in der Sekundarstufe II.“ *Der Deutschunterricht* 1. 36–46.
- Lohnstein, Horst. 2011. *Formale Semantik und natürliche Sprache*. 2., durchgesehene und erweiterte Auflage. Berlin, New York: de Gruyter.
- Oldfield, Richard C. & Wingfield, Art 1965. „Response Latencies in Naming Objects.“ *The Quarterly Journal of Experimental Psychology* 17. 273–281.
- Peschel, Corinna. 2006. „Verweismittel – Anaphorik – thematische Fortführung: Ein Thema für den Grammatikunterricht?“ In: Carmen Spiegel & Rüdiger Vogt(Hrg.): *Vom Nutzen der Textlinguistik für den Unterricht*. 171–186.
- Philippi, Jule & Tewes, Michael. 2010. *Basiswissen Generative Grammatik*. Göttingen: Vandenhoeck & Ruprecht.
- Reinhart, Tanya. 1976. *The Syntactic Domain of Anaphora*. Cambridge, Massachusetts: The MIT Press.
- Reinhart, Tanya. 1991. „Pronouns.“ In: Arnim von Stechow & Dieter Wunderlich (Hrg.): *Semantik. Ein internationales Handbuch der zeitgenössischen Forschung*. 535–548.
- Musan, Renate & Noack, Christina. Textsammlung. Stand: 02.06.2012 (nicht veröffentlicht)

The Universals Archive:

<http://typo.uni-konstanz.de/archive/nav/search.php?PHPSESSID=abi85j2ak1gdj7og8fg548e-di4u3gt1g> (letzter Zugriff am 01.09.2013)

Thieroff, Rolf. 2012. „Die indeklinablen neutralen Indefinitpronomina. *Etwas, was, irgendetwas, irgendwas* und *nichts*.“ In: Björn Rothstein (Hrg.): *Nicht-flektierende Wortarten*. (Linguistik – Impulse & Tendenzen 47). Berlin etc.: de Gruyter. 117–147.



# Tschechisch als eine SAE-Sprache: Im Zentrum oder an der Peripherie?

*Martin Konvička (Freie Universität Berlin)*

## 1. Sprachbundforschung: Eine Einleitung

Wenn mehrere Sprachen gemeinsame strukturelle Merkmale aufweisen, kann es für diesen Zustand grundsätzlich drei logische Erklärungen geben: Erstens könnte es sich um in beiden Sprachen zufällig entwickelte Parallelerscheinungen handeln. Zweitens können die Merkmale vererbt sein, weil es um genetisch verwandte Sprachen geht. Drittens kann es um durch Sprachkontakt erworbene Merkmale gehen. Die letztgenannte Situation wird in der Sprachbundforschung untersucht und in Bezug auf einen potentiellen europäischen Sprachbund, als Standard Average European (SAE) bezeichnet, auf den folgenden Seiten thematisiert.

Im Vordergrund steht hier das Großprojekt EUROTYP (u. a. Haspelmath 2001; van der Auwera 1998; König 2010), das sich die systematische Suche nach den nicht durch genetische Verwandtschaft erklärbaren gemeinsamen Merkmalen der europäischen Sprachen zum Ziel gesetzt hat. Bevor die einzelnen Merkmale des SAE-Sprachbundes unter die Lupe genommen werden, sind einige Grundlagen der Sprachbundforschung zu besprechen.

Die Sprachbundforschung versteht sich als ein gewisser Gegenpol zur klassischen diachronen Linguistik, da sie sprachliche Strukturen erforscht, die nicht den früheren Sprachstadien der jeweiligen Sprachen entstammen können. So verbreiten sich in den einem Sprachbund angehörenden Sprachen strukturelle Merkmale, d. h. nicht nur einfache Entlehnungen auf der Ebene des Lexikons, die sie von anderen Sprachen, die dem Sprachbund nicht angehören, absondern. Dahl (1990) betrachtet dementsprechend die Merkmale des Standard-Durchschnitts-Europäisch (SAE) sogar als exotisch.

Damit die Merkmale, die den Sprachbund von seinen Nachbarn abgrenzen, wirklich dem gegenseitigen Sprachkontakt zugeschrieben werden können, sollen die einzelnen Sprachen des Sprachbundes verschiedenen Sprachfamilien bzw. verschiedenen Zweigen einer Sprachfamilie angehören (vgl. Mišeska Tomić 2006: 3). Diesbezüglich deutet Van Pottelberge (2001: 15) kritisch darauf hin, dass die europäischen Sprachen fast ausschließlich eine Sprachfamilie repräsentieren (vgl. Načeva-Marvanová 2010a: 40–41, Laury 1997). Die parallele Herausbildung ähnlicher oder sogar identischer Strukturen in Sprachen mit einer gemeinsamen Protosprache wäre nicht so verwunderlich, wie die Herausbildung solcher Strukturen in genetisch unverwandten Sprachen.

Die Sprachbundmerkmale sollen auch progressiv und innovativ sein. Die gemeinsamen Merkmale

sollen sich in den Sprachen des Sprachbundes oder in den angrenzenden Sprachen verbreiten. Merkmale wie Dativ für externe Possessoren im SAE-Sprachbund (s. 3.3) eignen sich deswegen nicht als gute Sprachbundmerkmale. Bloßes Feststellen gemeinsamer Merkmale in mehreren Sprachen reicht also nicht aus, um einen Sprachbund postulieren zu können.

In der Literatur (u. a. Stolz 2010: 409) wird auch die Frage der Merkmalqualität diskutiert: ob negative Sprachbundmerkmale akzeptiert werden sollen, d. h. ob die Abwesenheit eines Merkmals im Sprachbund im Vergleich zu den benachbarten Sprachen der Anwesenheit eines Merkmals gleich ist. Wird jedoch der Abwesenheit eines Merkmals derselbe Wert wie der Anwesenheit eines Merkmals beigemessen, entsteht mit jedem Sprachbund gleichzeitig ein virtueller Gegensprachbund, der nur durch Negation des ursprünglichen Sprachbundes charakterisiert wird. Darum wird an dieser Stelle auf den höheren Wert der positiven Merkmale hingewiesen, es sei denn, dass das Fehlen eines Merkmals in der Tat eine gemeinsame strukturelle Innovation darstellt.

Die Merkmalquantität, d. h. die Anzahl der Merkmale, bleibt dagegen größtenteils arbiträr. Genauso arbiträr ist auch die Zahl der Sprachen eines Sprachbundes. Ein minimaler Sprachbund könnte folglich auch aus lediglich zwei Sprachen mit nicht mehr als einem gemeinsamen Merkmal bestehen. Solch ein minimalistischer Sprachbund wird u. a. von Haspelmath (2001: 1492) akzeptiert, der aber gleichzeitig solche Sprachbünde als uninteressant ansieht.

Mit diesem Punkt ist auch die folgende Frage verbunden: Ist die Zugehörigkeit zu einem Sprachbund absolut oder graduell? Haspelmath (2001) und van der Auwera (1998) verwenden im Rahmen des europäischen Sprachbundes die sogenannten Isoplethen, um die Sprachen nach der Anzahl der geteilten Merkmale, zu gruppieren. Solch eine graduelle Herangehensweise, obwohl legitim, entkräftet jedoch den eigentlichen SAE-Sprachbund zu einem bloßen Dachbegriff, zu einer Subsumierung der minimalen Sprachbünde.

Van Pottelberge (2001) warnt noch davor, dass die Äußerungen über Sprachbünde von kulturellen, ideologischen oder politischen Ansichten der LinguistInnen unbewusst geprägt sein können. Er weist darauf hin, dass beispielsweise Jakobsons (1931) Arbeiten zum eurasischen Sprachbund durch ihre Emigrationserfahrungen beeinflusst sein könnten. Ähnliche unterbewusste außersprachliche Einflüsse befürchtet Van Pottelberge (2001: 14) auch bei dem Projekt EUROTYP zur Untersuchung des SAE-Sprachbundes. Der Sprachbund, sein Kern und die wellenartige Erweiterung seiner Merkmale korrelieren nämlich laut Van Pottelberge viel zu auffällig mit der politischen Entwicklung der Europäischen Gemeinschaft bzw. Europas.

## **2. SAE-Sprachbund**

Linguistisch gesehen wird Europa nach Haspelmath (2001: 1493) durch zwölf sog. Europäismen,

d. h. SAE-Merkmale, definiert. Neben diesen zwölf Hauptmerkmalen wurden zusätzlich auch andere Europäismen vorgeschlagen (u. a. van der Auwera 1998, 2001, 2011), die jedoch in diesem Artikel nicht berücksichtigt werden können. Die zwölf Hauptmerkmale des SAE-Sprachbundes nach Haspelmath (2001: 1494–1501) sind hier aufgelistet: i) Bestimmte und unbestimmte Artikel, ii) Relativsätze mit Relativpronomina, iii) Haben-Perfekt, iv) Nominative Experiencer, v) Passiv mit Partizip, vi) Antikausativ-Prominenz, vii) Dativ für externe Possessoren, viii) Negationspronomen und verbale Negation, ix) Vergleichspartikel, x) Äquationskonstruktionen, xi) Affixe als reine Kongruenzmarker, xii) Intensivierer verschieden von Reflexivpronomen.

Für die SAE-Hypothese, wie sie u. a. von Haspelmath vertreten wird, ist von wesentlicher Bedeutung, dass die einzelnen Merkmale allmählich aus einer Sprache (*model language*, Heine & Kuteva 2006: 49) in andere Sprachen (*replica language*, Heine & Kuteva 2006: 49) übergehen. Wie jedoch Hüning (2008: 50) am Beispiel des Niederländischen zeigt, ist es meistens problematisch, die Quelle einer innovativen Struktur oder eines Prozesses zu identifizieren. Auch aus diesem Grund scheint die Darstellung des Sprachkontakts zwischen den europäischen Sprachen als der Hauptauslöser der für den SAE-Sprachbund typischen Entwicklungen, d. h. sprachliche Innovation als Produkt der Beziehung zwischen *model* und *replica language*, nicht immer selbstverständlich.

### **3. SAE-Merkmale im Tschechischen**

Das Tschechische wird in der SAE-Literatur (z.B. Haspelmath 2001: 1505) weder zu den Kernsprachen, noch zu den *model languages* des Sprachbundes zugerechnet. Die kultur-historischen Übereinstimmungen mit den Kernsprachen (Tornow 2010: 87, Weiss 2010, Grenoble 2010) werden in diesem Artikel außer Acht gelassen und lediglich die linguistischen Übereinstimmungen in Betracht gezogen. Meistens wird behauptet (u. a. Haspelmath 2001: 1505, van der Auwera 2011), dass das Tschechische sechs der zwölf Hauptmerkmale erfüllt. Im Folgenden wird nun auf die Merkmale eingegangen, die wider Behauptungen in der einschlägigen Literatur einen höheren Grad der Europäisierung des Tschechischen bezeugen. Es werden auch einige Merkmale des Non-Standard Average European (Seiler 2011) mit dem Tschechischen verglichen, um die im Titel gestellte Frage, ob das Tschechische linguistisch im Zentrum oder an der Peripherie Europas liegt, beantworten zu können.

#### 3.1. Bestimmte und unbestimmte Artikel

Haspelmath (2001: 1494) geht davon aus, dass „[i]n large parts of Europe there are no articles at all ([...] West Slavic, Finno-Ugrian other than Hungarian [...]).“ Obwohl dies in der tschechischen Standardsprache in der Tat der Fall ist, zeugt wenigstens das gesprochene Tschechisch von der

Emergenz definitiver Artikel aus den Demonstrativpronomina *ten/ta/to* (vgl. Cummins 1998, Berger 1993: 182, 2009: 18). Eine ähnliche Entwicklung ist laut Laury (1997) in letzter Zeit auch im Finnischen zu beobachten.

Auf die Schwächung der deiktischen Funktionen der Demonstrativpronomina *ten, ta, to* im gesprochenen Tschechisch bereits am Anfang des 20. Jahrhunderts hat Mathesius (1926) hingewiesen. In deiktischen Kontexten werden die Demonstrativa durch komplexere Ausdrücke wie *tento, tato, toto* ersetzt oder aber mithilfe zusätzlicher Mittel wie lokaler Adverbien oder Relativsätzen ergänzt. Anaphorische Funktionen von *ten, ta, to* wurden in der Umgangssprache nach Mathesius zu erinnernden oder emotionalen Funktionen abgeschwächt. Mathesius konstatiert weiterhin Folgendes: obwohl die Demonstrativpronomina *ten, ta, to* artikelähnliche Züge aufweisen, handelt es sich immer noch um keine definiten Artikel.

(1) **Ta** vázaná kniha.  
 Die-F.SG.NOM gebunden-F.SG.NOM Buch-F.SG.NOM  
 ‚Das gebundene Buch.‘

(2) Co dělá **ten** druhý, **ten** mladší?  
 Was machen-3SG der-M.SG.NOM ander-M.SG.NOM der-M.SG.NOM jünger  
 ‚Was macht **der** Andere, **der** Jüngere?‘

(3) To je přitom **to** nejlepší.  
 Das-N.SG.NOM sein-3SG dabei das-N.SG.NOM beste-N.SG.NOM  
 ‚Das ist aber **das** Beste.‘

Eine Übersicht der artikelähnlichen Funktionen tschechischer Demonstrativpronomina bietet Berger (1993: 182–197). Einerseits zeigt er, dass die tschechischen Demonstrativpronomina *ten, ta, to* im Gegensatz zu ihren russischen Äquivalenten deutlich artikelähnlicher sind, andererseits wird dargelegt, dass sie die Qualität der ebenfalls aus Demonstrativpronomina entwickelten bulgarischen oder mazedonischen Artikel noch nicht erreichen. Hier einige Beispiele nach Berger (1993: 187, Hervorhebungen MK):

- (4) a. Ребѣта, что вы делаете? Сейчас же оставьте **Ø** машину!  
 b. Co to prosím děláte? Okamžitě nechte být **to** auto!  
 c. Was macht ihr da bloß? Laßt sofort **das** Auto in Ruhe!

- (5) a. Что вам нужно, чтобы выиграть **Ø** процесс?  
 b. Co potřebujete, abyste **ten** proces vyhrál?  
 c. Was brauchen Sie, um **den** Prozess zu gewinnen?

Obwohl das Tschechische über keine voll ausgebildeten Artikel (definit oder indefinit) verfügt, rückt es den SAE-Kernsprachen näher und entwickelt sich in Richtung anderer europäischer Sprachen, z. B. Baskisch, Isländisch, Irisch und Walisisch (vgl. Haspelmath 2001: 1494).

### 3.2. Haben-Perfekt

Was das Perfekt in den europäischen Sprachen anbelangt, ist die Situation im Tschechischen für die geographische Lage des tschechischen Sprachraums bezeichnend. Načeva-Marvanová (2010a, 2010b) deutet darauf hin, dass sich die europäischen Sprachen in zwei große Gruppen aufteilen lassen: Das Haben-Perfekt überwiegt im Westen und das Sein-Perfekt im Osten. Im deutschen Sprachraum, der als eine Übergangsregion gilt, wird dazu noch eine Nord-Süd Unterscheidung gemacht: *ich habe gegessen – ich bin gegessen*.

Die Ost-West-Dichotomie beschreibt die Situation jedoch nicht ganz genau, da die meisten Sprachen Europas eher einen Mischtypus darstellen (vgl. Deutsch *ich bin gekommen – ich habe gegessen*, Niederländisch *ik ben gekomen – ik heb gegeten* oder Englisch *he is come* [biblisch, nicht mehr gebräuchlich] – *he has eaten*). Im Mazedonischen, das gemeinsam mit dem Bulgarischen unter den slawischen Sprachen am Weitersten auf dem Weg zum analytischen Sprachaufbau vorangeschritten ist, konkurrieren die beiden Typen, wobei das Sein-Perfekt noch aus dem Proto-Slawischen übernommen wurde und das Haben-Perfekt durch Sprachkontakt eingeführt worden ist (vgl. Mišeska Tomić 2010: 123).

Laut Načeva-Marvanová (2010a: 40) hat sich das Haben-Perfekt im 4. und 5. Jahrhundert in den späteren romanischen Sprachen verbreitet und im 12. und 15. Jahrhundert erscheint diese Konstruktion auch in den germanischen Sprachen. Im Laufe des 20. Jahrhunderts hat dieselbe Entwicklung in den (west)slawischen Sprachen angefangen. Bemerkenswert ist eine ähnliche Ausbreitung des Sein-Perfekts aus den östlichen europäischen Sprachen in die nicht indo-europäischen Sprachen wie Finnisch, Estnisch oder Karelisch (Načeva-Marvanová 2010a: 40–41).

Anfang des 20. Jahrhunderts bemerkt Mathesius (1925), dass sich das ganze tschechische Tempus-system aufgrund der Emergenz solcher umgangssprachlichen Konstruktionen langsam verändert. Später konstatiert Gallis (1960: 186), dass das tschechische Haben-Perfekt mit dem des Bulgarischen übereinstimmt aber noch nicht dem Mazedonischen gleich ist. Gallis vergleicht diesen Prozess mit einem Ähnlichen im Norwegischen, wo unter dem Einfluss norwegischer Dialekte das

Hilfsverb *haben* in perfektiven Konstruktionen das ursprüngliche *sein* ersetzte.

Im Tschechischen stimmen die Partizipien des Haben-Perfekts mit dem Objekt im Genus und Numerus überein (6a, b). Načeva-Marvanová (2010a: 42–44, Hervorhebungen der Autorin, Glossierungen MK) zeigt eine relativ neue Tendenz, die zur Änderung der Kongruenz führt. In (6c) ist bereits eine Entwicklung in Richtung der SAE-Kernsprachen zu beobachten.

(6) a. Pro koně má **připraveno** seno.  
Für Pferd-SG hat-3SG vorbereitet-N.PST.PRCP Heu-N.SG  
,Er hat für sein Pferd das Heu vorbereitet.‘

b. Ministerstvo má **připraveny** [...] **dvě** **koncepte**.  
Ministerium-N.SG hat-3SG vorbereitet-F.PL zwei Konzepte-F.PL  
,Das Ministerium hat zwei Konzepte parat.‘

c. Ministerstvo má **připraveno** [...] dvě koncepte.  
Ministerium-N.SG hat-3SG vorbereitet-N zwei Konzepte-F.PL  
,Das Ministerium hat zwei Konzepte parat.‘

### 3.3. Dativ für externe Possessoren

In dieser Hinsicht stimmt das Tschechische mit dem Standard-Durchschnitts-Europäisch überein, denn es verfügt über Konstruktionen wie (7).

(7) Matka myje dítěti vlasy  
Mutter-F.SG wäscht-3SG Kind-N.SG.DAT Haare-PL.AKK  
,Die Mutter wäscht dem Kind die Haare.‘

Wie jedoch u. a. Van Pottelberge (2001: 18) und Hüning (2002: 149) bemerken, sind externe Possessoren im Dativ im Niederländischen nicht mehr produktiv. Es geht also nicht um durch Sprachkontakt ausgelöste Innovationen, sondern um Relikte älterer Sprachstadien. Wie die SAE-Kernsprachen ersichtlich machen, werden die synthetischen Strukturen allmählich zugunsten analytischer Konstruktionen abgebaut. Da das Proto-Indoeuropäische eine stark flektierende Sprache darstellte, bildet diese Übereinstimmung zwischen den europäischen Sprachen ein ererbtes Merkmal und keine areallinguistische Neuerung im Sinne des SAE.

### 3.4. Negationspronomen und verbale Negation

Haspelmath (2001: 1498) unterscheidet in diesem Punkt zwischen zwei Typen verbaler Negation: Negationspartikel mit nicht negiertem Verb und Negationspartikel mit negiertem Verb, wobei für die SAE-Sprachen der erste Typ charakteristisch sei. Im Tschechischen sei dagegen der letztere Typ (8) vorherrschend.

- (8)        **Nikdy**        **netancuji.**  
             Niemals        tanzen-1SG.NEG  
             Ich tanze niemals.<sup>6</sup>

In den Kernsprachen des SAE-Sprachbundes lässt sich der zweite Typ jedoch auch ausfindig machen – hauptsächlich im nicht standardsprachlichen Sprachgebrauch (10), (11) und (12), aber auch in den Standardsprachen (9) Mehr zu der Untersuchung der nicht standardsprachlichen Varietäten der europäischen Sprachen siehe Seiler (2011).

- (9) Je **ne** parles **pas** français. / Ich spreche **nicht** Französisch. *Französisch*

- (10) Des måcht **kaa** Mensch **ned**. / Das macht **keiner** Mensch. *Bairisch*

- (11) Ik **en** ga **niet** naar school. / Ich gehe **nicht** zur Schule. *Flämisch*

- (12) We **don't** need **no** education.<sup>1</sup> / Wir brauchen **keine** Bildung *Englisch*

Was die verbale Negation anbelangt, unterscheidet sich die tschechische Standardsprache offensichtlich von den SAE-Kernsprachen. Sobald aber der Vergleich mit den nicht standardsprachlichen Varietäten (10), (11), (12) oder mit den europäischen Tochtersprachen (9) gezogen wird, erscheint ein völlig anderes Bild.

## **4. Zusammenfassung**

Zu Beginn des vorliegenden Artikels wurde kurz auf die allgemeinen Fragen der Sprachbundforschung eingegangen, die der eigentlichen Suche nach den erworbenen SAE-Merkmalen in den Sprachen Europas vorangehen muss. Es wurde Folgendes gezeigt: Obwohl die Anzahl der Sprachbundmerkmale arbiträr ist, spielt deren Qualität eine wichtige Rolle, indem zwischen positiven und

<sup>1</sup> Pink Floyd: Another Brick in the Wall. Part II. <http://www.pink-floyd-lyrics.com/html/another-brick-2-wall.html> (letzter Zugriff am 7. Oktober 2013)

negativen Merkmalen unterschieden werden sollte.

Des Weiteren wurden einige SAE-Merkmale, Haben-Perfekt, Artikel, externe Possessoren im Dativ und die mehrfache Negation, im Tschechischen unter die Lupe genommen. Es zeigte sich ein dynamischeres Bild der tschechischen Sprache innerhalb des SAE-Sprachbundes als gewöhnlich in der einschlägigen Literatur präsentiert wird.

Ferner wurde auf die Wichtigkeit der Untersuchung der nicht standardsprachlichen Varietäten europäischer Sprachen (Non-Standard Average European) hingewiesen, die auch zum besseren Verständnis der areallinguistischen Entwicklungen in Europa beitragen können, da die Standardsprachen, auf denen das SAE-Konzept basiert, teilweise ein künstliches Gebilde darstellen.

Zusammenfassend lässt sich konstatieren, dass die Suche nach den neu erworbenen Merkmalen in Europa Hand in Hand mit generellen Fragen der Sprachbundforschung gehen soll, sodass die Ergebnisse aussagekräftig bleiben. Es sollte nicht nur nach einfachen Übereinstimmungen, sondern auch nach Gründen dieser Konvergenzen gesucht werden.

## 5. Literatur

- Berger, Tilman. 1993. *Das System der tschechischen Demonstrativpronomina. Textgrammatische und stilspezifische Gebrauchsbedingungen*. Habilitationsschrift: LMU München.
- Berger, Tilman. 2009. „Tschechisch-deutsche Sprachbeziehungen. Zwischen intensivem Kontakt und puristischer Gegenwehr.“ In: Christel Stolz (Hrg.): *Unsere Sprachlichen Nachbarn in Europa. Die Kontaktbeziehungen. Zwischen Deutsch und seinen Grenznachbarn*. Bochum: Universitätsverlag Dr. N. Brockmeyer. 133–156.
- Cummins, George M. 1998. „Definiteness in Czech“. *Studies in Language*. 22 (3). 567–596.
- Dahl, Östen. 1990. „Standard Average European as an Exotic Language“. In: Johannes Bechert, Giuliano Bernini & Claude Buridant (Hrg.): *Empirical Approaches to Language Typology: Toward a Typology of European Languages*. 3–8.
- Gallis, Arne. 1960. „Die neuen slavischen Perfekte vom Typus *factum habeo* und \**casus sum*, \**casum habeo*“. *Scando-Slavica*. 6. 176–188.
- Grenoble, Lenore A. 2010. „Contact and the Development of the Slavic Languages“. In: Raimond Hickey (Hrg.): *The Handbook of Language Contact*. Hoboken, NJ: Wiley Blackwell. 581–598.
- Haspelmath, Martin. 2001. „The European linguistic area: Standard Average European“. In: Martin Haspelmath, Ekkehard König & Wolfgang Raible (Hrg.): *Language Typology and Language Universals*. Vol. 2. Berlin: de Gruyter. 1492–1510.
- Heine, Bernd & Kuteva, Tania. 2006. *The Changing Languages of Europe*. Oxford: Oxford Univer-



sity Press.

- Hüning, Matthias. 2002. „Zwischen Deutsch und Englisch Aspekte der sprachlichen Identität des Niederländischen.“ In: Herbert van Uffelen (Hrg.): *Sprache und Identität. Akten des Symposiums zum 10-jährigen Bestehen der Nederlandistiek am 18. und 19. Oktober 2002 in Wien*. Wien: Edition Praesens. 143–158.
- Hüning, Matthias. 2008. „Hoe Talen Elkaar Ontmoeten: Over Taalcontact En Convergentie.“ In: Irena Barbara Kalla & Bożena Czarnecka (Hrg.): *Neerlandistische Ontmoetingen. Tretpunt Wrocław*, Wrocław: Oficyna Wydawnicza ATUT – Wrocławskie Wydawnictwo Oświatowe. 40–53.
- Jakobson, Roman. 1931. „Über die phonologischen Sprachbünde“. In: *Roman Jakobson: Selected writings 1: Phonological Studies*. 1971. Paris/Den Haag: Mouton. 137–143. (Vortrag auf der Prager Phonologischen Konferenz 1930, erstmals erschienen in: *Travaux du Circle linguistique de Prague* 4, 1931)
- König, Ekkehard. 2010. „Das Projekt EUROTYP“. In: Uwe Hinrichs (Hrg.): *Handbuch der Euro-linguistik*. Wiesbaden: Harrassowitz. 425–435.
- Kruse, Jörg. 2000. *Die Entwicklung des Begriffs Sprachbund in der Balkanlinguistik*. Magisterarbeit: Georg-August-Universität Göttingen. <http://sprachbund.de/titel.html> (letzter Zugriff am 1.10.2013).
- Laury, Ritva. 1997. *Demonstratives in Interaction: The Emergence of a Definite Article in Finnish*. Amsterdam: Benjamins.
- Mathesius, Vilém. 1925. „Slovesné časy typu perfektního v hovorové češtině“ *Naše řeč* 9 (7). 200–202.
- Mathesius, Vilém. 1926. „Přívlastkové ten, ta, to v hovorové češtině“ *Naše řeč* 10 (2). 39–41.
- Mišeska Tomić, Olga. 2006. *Balkan Sprachbund morpho-syntactic features*. Berlin: de Gruyter.
- Mišeska Tomić, Olga. 2010. „The Macedonian ‘Have’ and ‘Be’ Perfects.“ In: Motoki Nomachi (Hrg.): *Grammaticalization in Slavic Languages: From Areal to Typological Perspectives*. Sapporo: Slavic Research Center, Hokkaido University. 123–146.
- Načeva-Marvanová, Mira. 2010a. „Grammaticalization and Verbal Structures (The Case of Analytic Perfect).“ *Linguistica Pragensia* 20 (1). 37–47.
- Načeva-Marvanová, Mira. 2010b. *Perfectum v současné češtině*. Praha, Ústav Českého národního korpusu FF UK – Nakladatelství Lidové noviny.
- Seiler, Guido. 2011. *Non-Standard Average European*. Vortrag gehalten an der Freien Universität Berlin am 13. Dezember 2011.
- Stolz, Thomas. 2010. „Sprachbund Europa – Probleme und Möglichkeiten“ In: Uwe Hinrichs

- (Hrg.): *Handbuch der Eurolinguistik*. Wiesbaden: Harrassowitz. 397–425.
- Tornow, Siegfried. 2010. „Wo endet Europa im Osten? Versuch einer historischen Regionalisierung“. In: Uwe Hinrichs (Hrg.) *Handbuch der Eurolinguistik*. Wiesbaden: Harrassowitz. 69–91.
- van der Auwera, Johan. 1998. „Conclusion“. In: Johan van der Auwera (Hrg.): *Eurotyp 3: Adverbial Constructions in the Languages of Europe*. Berlin: de Gruyter. 813–836.
- van der Auwera, Johan. 2001. „Standard Average European“. In: Bernd Kortmann & Johan van der Auwera (Hrg.): *The Languages and Linguistics of Europe: A Comprehensive Guide*. Berlin: de Gruyter Mouton. 291–306.
- van der Auwera, Johan. 2011. „Standard Average European“. In: Bernd Kortmann & Johan van der Auwera (Hrg.): *The Languages and Linguistics of Europe*. Berlin: de Gruyter. 291–306.
- Van Pottelberge, Jeroen. 2001. „Sprachbünde: Beschreiben sie Sprachen oder Linguisten?“ *Linguistik Online* 8 (1). [http://www.linguistik-online.de/1\\_01/VanPottelberge.html](http://www.linguistik-online.de/1_01/VanPottelberge.html) (letzter Zugriff am 7.10.2013).
- Weiss, Daniel. 2010. „Osteuropa als eurolinguistisches Areal“. In: Uwe Hinrichs (Hrg.): *Handbuch der Eurolinguistik*. Wiesbaden: Harrassowitz. 435–457.

# Nothing but “typed talk”? Analysing discourse in computer-mediated communication by employing Koch and Oesterreicher’s framework model<sup>1</sup>

Gerrit Kotzur (Technische Universität Berlin)

## 1. Introduction: the hybrid character of ‘netspeak’

With the analysis of language use in the new media (and especially forms of computer-mediated communication<sup>2</sup>) the question has arisen whether prototypical text types of CMC (such as chat conversations) should be situated somewhere between the poles of orality and literacy, since they possess, to a certain degree, features of both dimensions (see for example Crystal 2008: 28ff.).

E-mail and SMS could be arranged somewhere between personal letters and face-to-face and telephone conversation. Their style is a mixture of the written and spoken language and has “the spontaneity and informality of speech, though it is ‘written’ and thus carries with it the written language” (Moran and Hawisher 1998: 94, cited from Frehner 2008: 35).

The different designations suggested to depict the specific quality of ‘netspeak’ reflect its hybrid character, amongst them “typed talk”, “talking in writing”, “text-talk” and “text-speak” (Frehner 2008: 35). This hybrid character already comes into effect from the speaker’s point of view who plans and writes his or her contribution as part of a conversation in a graphic medium (Hess-Lüttich & Wilde 2003). However, the question remains: which aspect has a stronger impact on language use in CMC, its written or spoken quality?

Though both are instances of CMC, the following two examples that occurred in conversations of a Nintendo Wii™ fan forum hint at how the discursive constellation, the purpose of communicative contact and other factors influence the style and linguistic cues of the statements made.<sup>3</sup>

---

1 I wish to express my gratitude to the organising team and editors for the opportunity to present and publish this paper, to Edgar Baumgärtner for his advice on additional literature relevant for the topic, to the participants of the 53rd StuTS for fruitful discussions and to Denise Kotzur for her assistance in proof-reading the article. I also thank the anonymous reviewer for detailed comments and valuable suggestions for improvement of this manuscript.

2 *Computer-mediated communication* is but one of many terms used to encompass the linguistic phenomena occurring in online discourse. Since CMC is the most common expression it is also utilised in this article—despite terminological problems and the concept’s sketchiness. For a detailed discussion concerning alternative notions see Jucker and Dürscheid (2012).

3 Judging the linguistic data available in such forums, one cannot be perfectly sure whether the appearance of certain linguistic features is the result of the conceptualised proximity of discourse or simply a lack of textual competence and awareness for grammatical errors: “It is important to note that simple errors and mistakes are *not* relevant indicators of immediacy” (Oesterreicher 1997: 202). Then again, a lot of mistakes are usually tolerated or ignored in proximate talk and writing compared to distant forms of communication, which the participants are aware of, too. Proximate talk requires less planning and more spontaneity.

(1) Speakin' of arcade games, y'all might want to grab "Ghost Squad". It's a \$10~ used game that's exactly like those on-rail shooters in arcades where ya use that fake plastic gun to shoot things on-screen. (B\_#1563\_04-30-2011\_11:18 AM\_Hayzee)<sup>4</sup>

(2) Hello everyone

I would appreciate help in resetting my wii resort golf game scores back to zero. Any help gratefully received. Yvonne (A\_#892\_05-09-2011\_07:32 AM\_yvonne)

(1) violates some of the (traditional) norms of written communication by employing typically oral features like *speakin'* and *y'all*, but also vague expressions like *things*. (2) is a completely different case – except for the elliptic sentence *Any help [is] gratefully received*. The user opens with a greeting and closes the comment with a polite phrase and states her name, structures usually included in letters and other text types, but less frequently in conversations of fan forums (however, see Härvelid 2008: 28ff., 59). Possibly, the conventions of the forum are rather new to the author of the second posting.

In order to be able to apply the following framework model to authentic material, I gathered more data from the internet forum *wiichat.com*.<sup>5</sup> The material consists of one-to-many conversations between users that own or are at least interested in the Nintendo Wii™ console(s). One user's question and the respective answers of others are accessible for an undefined number of recipients who can also actively engage in the discussions once they have created a personal account.

Below, I will outline the methodical problems that simple assumptions on the supposed mixture of oral and written features in CMC can entail (point 2). Afterwards, I will explain the components of Koch and Oesterreicher's framework model,<sup>6</sup> including the poles of proximity and distance, communicative parameters and coding strategies (point 3) and apply them to forum conversations (point 4). Point 5 deals with the methodical problems of the model itself, as issued by different researchers, and suggestions for its improvement.

## 2. Written versus spoken language: continuum or dichotomy?

The 'traditional' approach to written and spoken language is conceived of as both a continuum and a dichotomy at the same time: on the one hand, any communicative act is transmitted by a medium

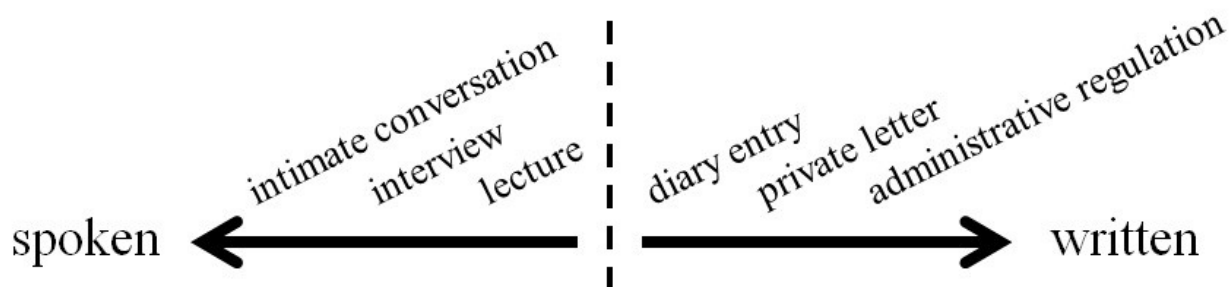
---

4 The signature includes a reference to the forum thread used (A or B, see sources), the number of the contribution in the thread, date, time and name of the user.

5 The examples used in this article solely serve as means to illustrate the theoretic assumptions. They should not be considered as representative of the text type of forum conversations in general (although it is highly possible that most of them are). Different forums can present a sheer diversity of language use (Härvelid 2008: 82ff.).

6 For another (German) overview of characteristics of 'oraliteral' text types in CMC—especially from the perspective of functional pragmatics—see Marx and Weidacher (2014: chapter 3.2.3, 107ff.).

of either graphic (e.g. a letter) or phonic (e.g. an interview) nature, so the medial aspect of language should represent a clear-cut dichotomy. On the other hand, the conception and planning of each utterance and its ultimate linguistic realisation can bear more or less resemblance with and employ features typical for spoken or written language use. This aspect is perceived as a continuous spectrum, where intermediate degrees are possible, since the number of features used is variable and depends on the communicative setting and the speaker's intention. A framework of this kind (including sample text types) might look as follows:



*Figure 1: assumed traditional framework model of written and spoken language*

Obviously, such a model is not very helpful for judging authentic data. This is why notions of ‘typed talk’ arose in the first place: medial and conceptional aspects have not always been distinguished. Therefore, it was necessary to design a model that overcomes difficulties and terminological uncertainties.

### 3. Koch and Oesterreicher’s framework model

#### 3.1 Theoretical foundations

Koch and Oesterreicher’s ([1985] 2012) more sophisticated framework for the investigation of orality and literacy was originally used in sociolinguistics and variational linguistics for the purpose of analysing certain synchronic and diachronic phenomena and discourse typology (Koch & Oesterreicher 2012: 459 f.). Since its first publication, the model has been revised and extended by several researchers from different fields (see for instance Ágel & Hennig 2007, Frehner 2008, Landert & Jucker 2011). Nonetheless, it was scarcely recognized by Anglophone academics (however, see Frehner 2008).

As I already stated, simple assumptions on the degree of orality and literacy of discourse and text types are problematic, since they do not precisely distinguish between medial and conceptional aspects. However, one can assume that what those researchers refer to when they talk about the stri-

king oral quality of certain texts is not the medial, but the conceptional aspect.

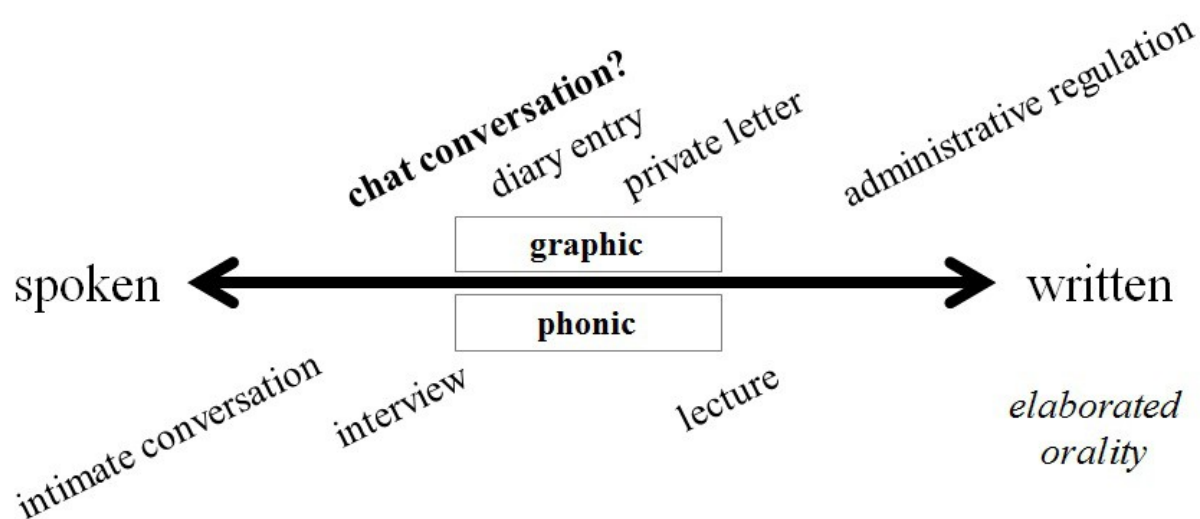


Figure 2: first step towards a refined framework model (Koch & Oesterreicher 2012: 444)

In Koch and Oesterreicher’s model, the medial aspect is designed as a dichotomy, whereas the conceptional aspect represents a continuum where text and discourse types can be positioned on, depending on the number or degree of features typical for written or spoken registers.

Though it is realised by the phonic medium of a speaker’s voice, a lecture or sermon usually makes use of many characteristics typical for written text types and in the model it is thus positioned closer to the right pole of the scale.<sup>7</sup> The authors see texts of this kind as examples for “elaborated orality” (Koch & Oesterreicher 2012: 457 f.). A diary entry or chat conversation on the other hand, though occurring in written form, can be very informal in style and employ features typical for spoken discourse, such as elliptic sentences, colloquial language and discourse markers (*well, so, you know*), so that it is plausible to position those on the left side of the scale.

Ultimately, this two-dimensional model facilitates the comparison of different text types with respect to their degree of conceptional orality and literacy. It also indicates the prospect of medium transferability: most text types can be realised by a phonic as well as a graphic medium. For instance, an interview can be conducted orally and transcribed afterwards. Yet, one has to keep in mind that for a written interview to create the same impression of proximity as the spoken version, more linguistic cues might actually have to be used as substitutes (see also point 5).

<sup>7</sup> The sample text types shown in this figure are positioned on the basis of idealised communicative events. Empirically tested investigations into a large corpus of literate and oral registers have been conducted in detail by Biber (1995), though; see point 3.3 and 5 in this paper for further information.

### 3.2 Language of proximity—language of distance

Next, Koch and Oesterreicher renamed the two conceptional poles, formerly called *spoken* and *written language*, since those labels already imply the medial aspect of language and therefore pose the risk of confusion and terminological vagueness. The poles were called *language of proximity*<sup>8</sup> and *language of distance*.

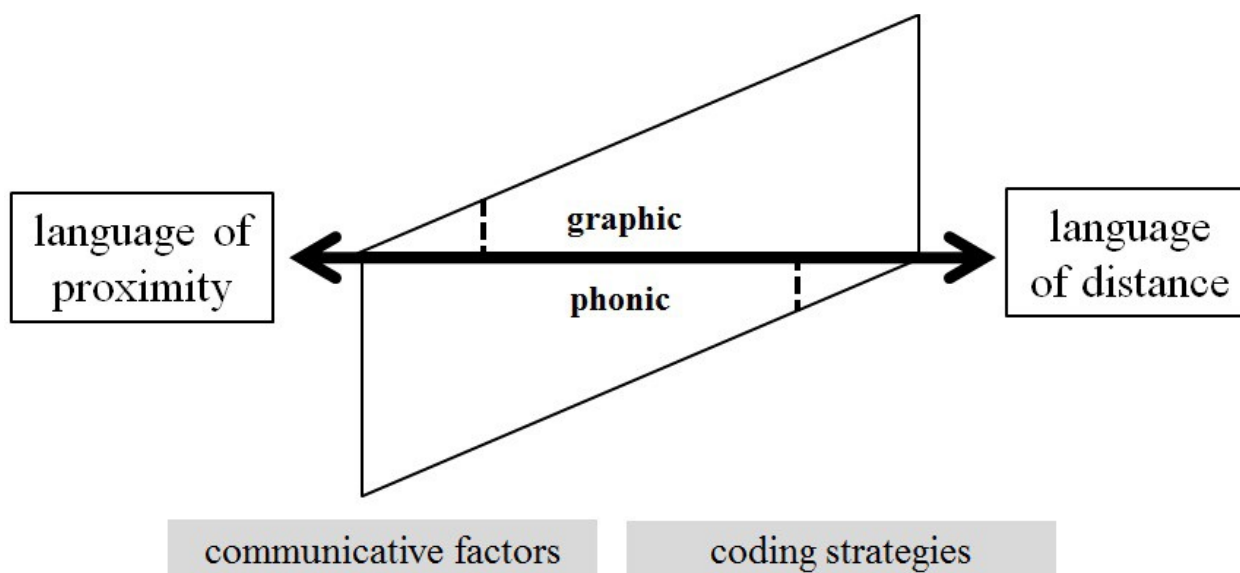


Figure 3: *language of proximity—language of distance* (Koch & Oesterreicher 2012: 450)

The triangles indicate that certain text types and poles have higher affinities towards each other: medially graphic texts are typical for the language of distance and vice versa. The dotted lines on either side signify that no medially graphic text can be fully proximate (the same applies to phonic texts). Conceptional proximity and distance are, again, decisively influenced by communicative factors and coding strategies, which themselves can be seen as scales (except for the first communicative factor of *face-to-face* versus *distant interaction of partners*).

### 3.3 Communicative factors and coding strategies

The text types placed on the conception continuum shown in figure 2 result from the interplay of several communicative parameters or factors which are (Oesterreicher 1997: 194, Koch & Oesterreicher 2012: 445, 450):

- face-to-face interaction of partners versus distance in space and time
- private/non-public versus public setting of the communicative event

8 In the translation of the original German paper the left pole is called *language of immediacy* (Koch & Oesterreicher 2012). However, I find this to be an inaccurate choice of translation, for *Nähe* in German is best translated with the word *proximity*, which is a spatial lexeme. *Proximity* and *distance* are also better choices for opposites (see also Frehner 2008: 171 f.).

- familiarity versus unfamiliarity of the partners
- context embeddedness versus contextual dissociation of a discourse
- dialogue versus monologue
- maximum versus minimum cooperation of partners
- spontaneity versus reflection
- involvement versus detachment
- free thematic development versus fixation of topics
- expressivity and affective speech versus objectivity

Conversations in online forums, such as the one examined for this article, can thus be characterized as instances of communication marked by a spatially and (more or less) temporally distant interaction, a semi-public setting, unfamiliar partners (at first), context embeddedness, (a kind of) dialogue, a high cooperation of partners (it is a forum developed for advisory discussions), (more or less) reflected reactions, involvement, fairly fixed topics (through specific threads in the forum) and (more or less) objectivity (see also Härvelid 2008: 36ff.). The communicative conditions of proximity and distance give rise to particular preferences for different coding strategies<sup>9</sup> (Oesterreicher 1997: 194, Koch & Oesterreicher 2012: 448, 450):

- process versus object orientation
- tentativeness versus finality
- information density
- integration
- compactness
- explicitness
- complexity
- elaboration
- planning

Most of these communicative factors also correspond to the dimensions gained from Biber's broadly based corpus analysis of register variation in English, Nukulaelae Tuvaluan, Korean and Somali (1995). He distinguished, amongst others, *involved* versus *informational production*, *narrative* versus *non-narrative discourse*, *situation-dependent* versus *elaborated reference*, *overt expression of argumentation* and *non-abstract* versus *abstract style*. The interplay of the communicative factors and coding strategies is quite complex: "there are multiple dimensions that distinguish between stereotypical speaking and writing; each of these dimensions has different functional underpinnings,

---

<sup>9</sup> For critical comments on the coding strategies as well as Koch and Oesterreicher's model in general see Ágel and Hennig (2007: 182ff.) and Härvelid (2008: 14ff.). Koch and Oesterreicher (2007) mostly confirmed these claims.



and each defines a different set of relations among registers” (Biber 1995: 240 f.).

I found three more studies which show how this framework model can be used to examine authentic data: Ágel and Hennig (2007) carried out a thorough analysis of chat conversations by applying 52 features of the language of proximity and contrasting the results with a radio feature. Frehner (2008) analysed 342 e-mails and 983 text messages using a refined version of Koch and Oesterreicher’s framework model and reached conclusions similar to Ágel and Hennig. Last but not least, Härvelid (2008) built up a text corpus of 290 contributions in three different news boards. He analysed situational and contextual factors, the discourse- and text-related structure of the data, as well as their syntax, lexis, morphology, typography and sociolinguistic variation.

#### 4. Fan forum conversations: conceptional orality and beyond

Taking a closer look at the contributions to the discussions examined here, one can infer that most of the statements are neither very complex nor elaborated (compared to most graphic text types) and result from a spontaneous rather than planned writing process – however, there are exceptions. Consequently, most of the utterances possess certain features typical of conceptionally proximate oral texts, which I highlighted in bold.

(3) Since setting up the 42" TV the text is **wayyyyyy** blurry and **pretty much** impossible to read. (A\_#895\_06-01-2011\_04:04 AM\_Carlito)

(4) **Erm....right....well** that would be **erm** whatever it was that came with the console Same connection **thingy** that we used to the 32", I'm sorry I'm just so **un-techy** that typing 80085 on a calculator still impresses me. **Pretty sure** it's not a scart...if that helps...**ah ha**, my other half just texted me and says it is the connector that comes with the console and splits in to red, yellow and white end **thingies** that go in to the TV

**\*trudges off to the corner in pure non-techyness shame\*\*** (A\_#897\_06-01-2011\_04:37 AM\_Carlito)

(5) **Yikes**, that certainly isn't right. (A\_#921\_12-21-2011\_09:43 PM\_Splash\_King)

(6) comes from being a gamer my whole life and tearing open two other game systems. **ouch. oh well.** thanks though. (A\_#924\_12-27-2011\_03:52 AM\_brynjadiss)

(7) **Oooooooo... ummm....** I got my awesome black wii a month ago and **im like freakin awesome** at super smash bros. brawl... (: <-- mario is so awesome! so is Samus! (B\_#1551\_12-08-2010\_06:38 AM\_\*Angel\*)

Here we find specific linguistic means typical for speech, such as hesitation phenomena (fillers like *erm* (4), *um* (7)), spontaneously coined terms like *un-techy* (4), discourse markers, interjections and

exclamations like *yikes* (5), *ouch* (6) and *ooo* (7). In most cases, there does not seem to be a real need to use those features. Still, they reveal the conception of orality characteristic for this type of discourse and the spontaneous production of text without need for revision. Fillers can also function to support the construction of a user's identity, e.g. Carlito's "non-techyness shame" in (4) and his/her hesitation towards the topic discussed. Moreover, forum conversations are associated with a certain style that developed over time. Users even found graphic means that act as substitutes for the missing vocal quality, like iterative characters to signify emphasis (*wayyyyyyy* (3)). In spite of the linguistic cues of proximity, the graphic code strongly influences the discourse form of chat conversations and thus they can never be fully proximate (Ágel & Hennig 2007: 202). It is also important to note that forum conversations are generally even less proximate in style compared to genuine chat conversations (Schwarz-Friesel & Consten 2014: 46ff.). Of course, examples of purely (conceptionally) literate language use can be found on the internet, too. Hence, one should not assume that rather oral forms of CMC reflect the whole variety of text types occurring in cyberspace (Marx & Weidacher 2014: 113), as over-generalisations by so-called language purists sometimes suggest. Interestingly, there are features which cannot be explained by the high degree of conceptual proximity: in (4) the description of one's own actions from a third-person perspective functions as an ironic comment. Such forms are quite typical for CMC. They are found less often in phonic and graphic texts (except for comics). In German, those complex "Inflektive" or verb stems pose a grammatical innovation (Teuber 1998).

To conclude, text types "reflect strong tendencies for speakers and writers to construct texts around a limited set of functions and co-occurring linguistic forms. These typologies thus give structure to the multidimensional space of textual variation, even though they do not negate the continuous nature of that space" (Biber 1995: 358).

## 5. Methodical adjustments

Firstly, the model discussed in this paper seems to promote a downgrade of medial influence, which is reduced to a matter of medial realisation (Androutsopoulos 2007: 80). Härvelid (2008: 15 f., 87 f.) underlined the influence of technical conditions, restrictions and possibilities on computer-mediated text production. On the other hand, one has to be careful not to place too much emphasis on the medial influence either, and thereby neglect the conceptual features. This consideration is also connected to the medium transferability of texts. As Oesterreicher states: simple "medium-transcoding does not affect the conceptual component of discourse" (1997: 196). But even by only transferring a text from its medially graphic realisation to a phonic medium, would this process not inevitably contribute to an increasingly proximate conception of the respective text? Oesterrei-

cher also explicates that “texts of a strict immediacy type cannot exist” (1997: 199). As a result, the possibilities of medium-transcoding are, in fact, limited. The communicative factors “produce a filter that causes immediate forms of expression to be altered or eliminated from texts in the process of writing” (Oesterreicher 1997: 199).

Secondly, positioning text types on the scale is somewhat problematic, since they are always relative to other texts (compare formal to informal e-mails and chats). This is why Ágel and Hennig (2007) suggest using a prototypical text from either side of the continuum to compare with. Different text types can also vary significantly in their amount of conceptual proximity and distance: “some registers consistently show a greater range of internal variation, while other registers are more or less well defined on particular dimensions” (Biber 1995: 314). Moreover, the interdependence and possible lack of discriminatory power of (some of) the communicative factors has to be considered. Therefore, the linguistic features under examination have to be well-defined and need to correspond to the different communicative factors.

Thirdly, the one-sided relation of text and context is questionable. A certain language use is not only the product of communicative factors and context, proximity can also be actively created through language (Androutsopoulos 2007: 80), as can be the case in CMC. After all, text types are the results of social agreements between participants and the degree of conceptual orality/literacy also depends on the communicative requirements and how discourse is perceived (Schwarz-Friesel & Consten 2014: 48, Härvelid 2008: 88, Biber 1995: 314).

Furthermore, language change needs to be considered. The positions of text types on the scales are in need of regular updates. Concerning language change on the internet, however, Elspass (2002: 27 f.) argues that what has actually changed are the attitudes and the now more conscious use of conceptual orality in a medially graphic medium.

Finally, the model reaches its limitations when features of CMC emerge that neither speech nor writing originally possessed (Ágel & Hennig 2007: 210, Frehner 2008: 176 f.). To be fair, when first publishing this model, Koch and Oesterreicher did not have in mind CMC as a field for analysis. Yet, it can be adapted to meet new demands as Frehner (2008) and others have shown by adding an additional level: *conceptionally analogue* versus *conceptionally digital* (Frehner 2008: 177).<sup>10</sup>

Chats are “quasi-synchronous” acts of communication which is why they show features typical for language planned offline. Many characteristics are not genuinely oral features, but products of compensation techniques (Ágel & Hennig 2007: 209).

---

10 Landert and Jucker (2011) added another two dimensions to the model: *accessibility*, which defines the context as public or non-public, and the *privacy* of topics discussed.

## 6. Summary and outlook

Despite refinements of Koch and Oesterreicher's framework model that need to be considered in order to analyse linguistic data of CMC more satisfyingly, the model is already well-suited to differentiate between conceptional and medial aspects of orality and literacy (Ágel & Hennig 2007, Frehner 2008, Härvelid 2008). Biber (1995) used a similar approach to examine a wide variety of text and discourse types cross-linguistically, such as press reportage, fiction, letters, speeches, face-to-face conversations and interviews. His study can also function as a starting point for more detailed investigations into CMC and comparisons with other registers.

Of course, the analysis of sample texts presented in this article is just a tentative first step. The assumptions will have to be confirmed through precise linguistic criteria. Even so, I hope to have been able to show one possible and promising direction for upcoming research in this field.

## 7. References

- Ágel, Vilmos & Hennig, Mathilde. 2007. "Überlegungen zur Theorie und Praxis des Nähe- und Distanzsprechens." In: Vilmos Ágel & Mathilde Hennig (eds.): *Zugänge zur Grammatik der gesprochenen Sprache*. Niemeyer: Tübingen (= Reihe Germanistische Linguistik 269). 179–214.
- Androutsopoulos, Jannis. 2007. "Neue Medien – neue Schriftlichkeit?". *Mitteilungen des deutschen Germanistenverbandes* 54 (8). 72–97.
- Biber, Douglas. 1995. *Dimensions of Register Variation. A Cross-linguistic Comparison*. Cambridge: Cambridge University Press.
- Crystal, David. 2008. *Language and the Internet*. Cambridge: Cambridge University Press.
- Elspass, Stephan. 2002. "Alter Wein und neue Schläuche? Briefe der Wende zum 20. Jahrhundert und Texte der neuen Medien – ein Vergleich." *Osnabrücker Beiträge zur Sprachtheorie* 64. 7–31.
- Härvelid, Frederic. 2008. "'Wusste gar nicht das man schriftlich labern kann.' Die Sprache in Deutschschweizer Newsboards zwischen Mündlichkeit und Schriftlichkeit". In: *Net.worx* 51. <http://www.mediensprache.net/networx/networx-51.pdf>.
- Hess-Lüttich, Ernst W. B. & Wilde, Eva. 2003. "Der Chat als Textsorte und/oder als Dialogsorte?" *Linguistik online* 13 (1). [http://www.linguistik-online.de/13\\_01/hessLuettichWilde.html](http://www.linguistik-online.de/13_01/hessLuettichWilde.html).
- Jucker, Andreas H. & Dürscheid, Christa. 2012. "The Linguistics of Keyboard-to-Screen Communication. A New Terminological Framework". *Linguistik Online* 56 (6). 39–64. [http://www.linguistik-online.de/56\\_12/juckerDuerscheid.pdf](http://www.linguistik-online.de/56_12/juckerDuerscheid.pdf).

- Koch, Peter & Oesterreicher, Wulf. 2007. "Schriftlichkeit und kommunikative Distanz". *Zeitschrift für germanistische Linguistik* 35 (3). 346–375.
- Koch, Peter & Oesterreicher, Wulf. 2012. "Language of Immediacy – Language of Distance. Orality and Literacy from the Perspective of Language Theory and Linguistic History". In: Claudia Lange, Beatrix Weber & Göran Wolf (eds.): *Communicative Spaces. Variation, Contact, and Change. Papers in Honour of Ursula Schaefer*. Frankfurt a. M.: Lang. 441–473. (German original: 1985. "Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte". *Romanistisches Jahrbuch* 36. 15–43.)
- Landert, Daniela & Jucker, Andreas H. 2011. "Private and public in mass media communication: From letters to the editor to online commentaries". *Journal of Pragmatics* 43. 1422–1434.
- Marx, Konstanze & Weidacher, Georg. 2014. *Internetlinguistik. Ein Lehr- und Arbeitsbuch*. Tübingen: Narr (= Narr Studienbücher).
- Moran, Charles & Hawisher, Gail. 1998. "The Rhetorics and Languages of Electronic Mail". In: Ilana Snyder (ed.): *Page to Screen*. London: Routledge. 80–101.
- Oesterreicher, Wulf. 1997. "Types of Orality in Text". In: Egbert Bakker & Ahuvia Kahane (eds.): *Written Voices, Spoken, Signs. Tradition, Performance, and the Epic Text*. Cambridge/Massachusetts, London: Harvard University Press. 190–214.
- Schwarz-Friesel, Monika & Consten, Manfred. 2014. *Einführung in die Textlinguistik*. Darmstadt: WBG.
- Teuber, Oliver. 1998. "fasel beschreib erwähn – Der Inflektiv als Wortform des Deutschen". *Germanistische Linguistik* 141/142. 7–26.

## 8. Sources

A: *Wii Troubleshooting*, started by i0n, 11-19-2006, 11:26 AM. 939 replies.

<http://www.wiichat.com/forum/nintendo-wii-chat/5269-wii-troubleshooting.html>.

B: Official 'I got a Wii' thread, started by MrKlown, 12-07-2006, 12:07 AM. 1,574 replies.

<http://www.wiichat.com/forum/nintendo-wii-chat/8136-official-i-got-wii-thread.html>.

# Implementation eines regelbasierten Pronomenresolutionsalgorithmus für das Deutsche

*Ludger Paschen (Ruhr-Universität Bochum)*

## 1. Einleitung

Die automatische Auflösung anaphorischer und koreferenter Beziehungen in Texten stand in den letzten Jahrzehnten durchgängig im Fokus der Aufmerksamkeit der Computerlinguistik (Poesio et al. 2010), da sie in vielen Bereichen der NLP (*natural language processing*) Relevanz besitzt. Strube (2010) nennt in diesem Zusammenhang u.a. Informationsmanagement, natürlichsprachige Dialogsysteme, maschinelle Übersetzung und Textgenerierungssysteme. So sollte ein System, welches den Satz (1) als Input verarbeitet, in der Lage sein, die anaphorische Beziehung zwischen „a block which is taller than the one you are holding“ und „it“ zu erkennen, damit nicht der falsche Block in die Kiste gelegt wird.

(1) Find [<sub>i</sub> a block which is taller than [<sub>j</sub> the one you are holding]] and put *it*<sub>i</sub> into the box. (nach Winograd 1972: 3)

Im Folgenden wird die Implementation eines Pronomenresolutionsprogramms für das Deutsche vorgestellt. Nach einer kurzen Darstellung des in diesem Artikel zu Grunde gelegten Anaphorikbegriffs (Kapitel 2) werden in Kapitel 3 einige automatische Resolutionsverfahren, auf denen der eigene Algorithmus aufbauen wird, besprochen. Der Implementation des eigenen Algorithmus und dem verwendeten Korpus des Deutschen, der *Tübinger Baumbank des Deutschen/Zeitungskorpus* (TüBa-D/Z) ist das vierte Kapitel gewidmet. Daran anschließend werden in Kapitel 5 die Ergebnisse der automatischen Evaluation des Algorithmus diskutiert und ein Vergleich mit den Ergebnissen anderer Algorithmen angestellt (Kapitel 6).

## 2. Anaphern und Koreferenz: Begriffsklärung

Das gemeinsame Konzept hinter Anaphorik und Koreferenz ist Referenz, also das Verweisen auf eine Entität im Diskursmodell (DM) des Produzenten bzw. des Rezipienten eines Textes (sie kann generell aber auch Teil der realen oder einer Konzeptwelt sein, vgl. Bußmann 2008: 573). Bezüglich der Beziehung der Entitäten untereinander sprechen Krifka & Musan (2012: 27) in Anlehnung an

Heim (1982) und Reinhart (1982) von einem „file card system“: Neue Informationen werden dem *common ground* so hinzugefügt, dass sie mit verwandten Entitäten assoziiert werden. Wie also die einzelnen Elemente einer Proposition zueinander und zu außerpropositionalen Dingen in Beziehung gesetzt werden (Informationsstruktur), hat Einfluss auf die sprachlichen Mittel, mit denen auf bestimmte Entitäten referiert wird. Eine referierende NP (auch: *mention*) führt entweder eine neue Diskursentität ein oder verweist auf einen bereits vorhandenen, erschließbaren (*bridging*, vgl. Clark 1975) oder unikalen Referenten (z.B. „der Papst“). Beispiele für nicht-referierende NPen sind generische Kontexte („*Störche* verbringen den Winter im Süden“), Expletiva („*Es* gibt keinen Osterhasen“) und Prädikativa („Peter ist *Student*“).

In (1) referieren „block“ und „it“ auf dieselbe Entität, was die beiden Wörter *koreferent* macht. Gleichzeitig ist „it“ die *Anapher* zu „block“, da sich die Semantik eines Personalpronomens immer erst im Kontext, der hier durch den Antezedens „block“ gegeben ist, ergibt. Formal lassen sich Koreferenz (2) und Anaphorik (3) also folgendermaßen definieren (nach van Deemter & Kibble 2000: 629, Kamp & Reyle 1993)<sup>1</sup>:

(2) Koreferenz

$\alpha_1$  und  $\alpha_2$  koreferieren genau dann, wenn  $\text{Referent}(\alpha_1) = \text{Referent}(\alpha_2)$ .

(3) Anaphorik

$\alpha_1$  ist anaphorischer Antezedens zu  $\alpha_2$  genau dann, wenn die Interpretation von  $\alpha_2$  von  $\alpha_1$  abhängt. Verhält es sich genau umgekehrt, so ist  $\alpha_1$  die Katapher zu  $\alpha_2$ .

Es ist wichtig, hierbei zwischen der *textuellen* und der (*diskurs-*)*semantischen* Ebene zu unterscheiden: Anaphorik ist eine *textuelle* Relation, bei der die Semantik der Anapher von der des Antezedens abhängt, Koreferenz hingegen eine *semantische*, die durch gleiche Referenz auf eine Entität gekennzeichnet ist. Anders ausgedrückt: „block“ und „it“ sind nicht per se koreferent, sie werden vielmehr in (1) durch ihre anaphorische Beziehung koreferent.

### 3. Related Work: Verschiedene Implementationsansätze von Pronomenresolutoren

Bevor das eigene Programm vorgestellt wird, soll ein Überblick über den Stand der Forschung im Bereich der maschinellen Anapherauflösung gegeben werden. Dabei beschränke ich mich auf regelbasierte Ansätze, die sich auf Pronomenresolution konzentrieren. Datenbasierte bzw. *machine-learning*-Verfahren und Systeme, die auch definite NPen auflösen, wie Soon et al. (2001), werde ich

<sup>1</sup> Aus dieser Definition von Anaphorik folgt, dass Anaphern auf einen schon im Diskurs eingeführten Referenten verweisen; dies allein genügt jedoch nicht zur Definition (aber vgl. Strube 2009: 399).

nicht behandeln; eine gute Übersicht bietet hier Wunsch (2010: 129ff).

### 3.1. Frühe Ansätze und MUC

Frühe Implementationen von Koreferenzauflösungsmechanismen waren oft ein Teil von größeren Programmen, die auf das *Verstehen* von natürlicher Sprache ausgerichtet waren und mit KI-Datenbanken operierten. Beispiele sind STUDENT, in dem als einziges Kriterium für Koreferenzauflösung die Nähe zum Pronomen zum Einsatz kommt (Bobrow 1964: 20), und SHRDLU (Winograd 1972), welches mit einer komplexeren Heuristik operiert.

Während der beiden letzten der sieben zwischen 1987 und 1997 abgehaltenen MUCs (*Message Understanding Conference*), auf denen nach besseren Technologien zur Informationsextraktion gesucht wurde, konnten Standards für die automatische Auflösung von Koreferenz, Evaluation und Annotation entwickelt werden (zu MUC-6 s. Grishman & Sundheim 1996, zu MUC-7 s. Baldwin et al. 1998). Bei den MUCs gehörte auch die Auflösung aller definiten NPen zum Aufgabenspektrum.

### 3.2. Implementationen von Centering

Brennan et al. (1987) stellt eine Implementation der Centeringtheorie (Grosz, Joshi & Weinstein 1983, 1995) vor. Die Centeringtheorie stellt einen Zusammenhang zwischen Salienz, Informationsstruktur und Kohärenzrelationen her und bietet einen Formalismus, der mit verschiedenen Arten von „Aufmerksamkeitszentren“ (*centers*) operiert, um bestimmte Abfolgemuster textstrukturell zu werten. Brennan et al. machen sich dieses Verfahren umgekehrt zu Nutze, um automatisch die Antezedenten von Pronomen zu ermitteln.

### 3.3 Hobbs

Der filterbasierte Ansatz für die Pronomenauflösung im Englischen von Hobbs (1978) enthält einen „naive[n]“ (ibid.: 339) Algorithmus, der die syntaktische Struktur eines zuvor geparsten Satzes nach möglichen Antezedenten durchsucht. Dazu werden vom Pronomen ausgehend bei allen NP- und S-Knoten auf dem Weg zum initialen S-Knoten die NPen links des Weges abhängig von ihrer Einbettung auf Kongruenz und selektionale Beschränkungen hin abgeglichen. Komplementiert wird der syntaktische Algorithmus durch einen semantischen Ansatz, der hauptsächlich auf lexikalische Inferenzen und Kausalketten zurückgreift.

### 3.4. Der scorebasierte Ansatz für das Englische von Lappin und Leass

Der Pronomenresolutionsalgorithmus von Lappin und Leass (1994) operiert mit Salienzparametern,



die möglichen Antezedenten auf Grund von grammatischen und textuellen Kriterien Werte zuweisen, die dann über die Wahl des Antezedenten entscheiden.

Der RAP (*Resolution of Anaphora Procedure*) genannte Algorithmus beinhaltet einen syntaktischen Filter (einschließlich Bindungsprinzipien, s. Chomsky 1981), einen morphologischen Filter, ein wortbasiertes Verfahren zur Identifizierung von Expletiva, das charakteristische Scoringsystem (s.u.), daran anschließend eine Prozedur zur Berechnung von Salienz und ein finales Entscheidungssystem.

Das Scoringsystem vergibt Werte entsprechend der grammatischen Rolle einer NP und baut dabei auf der Verfügbarkeitshierarchie von NPen (*NP accessibility hierarchy*, Keenan & Comrie 1977) auf, nach der die grammatische Funktion einer NP die Wahrscheinlichkeit, dass auf sie mit einem Relativsatz oder einer Anapher referiert wird, beeinflusst. In der durch Lappin & Leass leicht modifizierten Hierarchie (4) heißen die Salienzgewichte *emphasis* (vgl. Tab. 1); zudem finden auch existentielle Konstruktionen („Es gibt einen Hund“) Berücksichtigung.

(4) Salienzhierarchie für NPen

*SUBJ > EXIST > DO > IO > PP*

In das Scoringsystem sind zusätzlich noch fünf weitere Faktoren integriert: *recency* belohnt das Auftreten im gerade verarbeiteten Satz und wird de facto auf jeden Referenten angewandt; *head noun emphasis* ist gegeben, wenn sich die NP des Referenten nicht in einer anderen NP befindet; *non-adverbial emphasis* liegt vor, wenn sich die NP nicht in einer adverbialen PP befindet. Die letzten zwei Faktoren, *parallel roles reward* (wenn Anapher und Antezedent in gleicher syntaktischer Funktion auftreten) und *cataphora penalty* (wenn die Anapher vor dem Antezedenten steht), sind nicht-permanent, d.h. ihre Werte werden nicht im DM gespeichert, sondern nur für einzelne Pronomen addiert. Tabelle 1 bietet einen Überblick über die Salienzfaktoren (1. Spalte) und ihre Gewichtung bei Lappin und Leass (2. Spalte).

Faktor	Gewichtung				
	unconditioned	L&L	K&B	Wunsch	Eig.
<i>recency</i>	100	100	20	15	
<b>emphases</b>					
<i>subject emphasis</i>	80		170	190	
<i>existential emphasis</i>	70		---	5	
<i>possessive emphasis</i>	---	65	---		
<i>accusative emphasis</i>	50		70	50	
<i>indirect object emphasis</i>		40			
<i>preposition/oblique complement</i>	40	30	50		
<i>head noun emphasis</i>			80		
<i>non-adverbial emphasis</i>	50		---	5	
<b>temporary boosts</b>					
<i>parallelism</i>			35		

Tabelle 1: Salienzfunktionen und ihre Gewichtung bei Lappin & Leass (1994), Kenney & Boguraev (1996) und Wunsch (2010). In der letzten Spalte sind die selbst gewählten Werte (s. Abs. 4.2) wiedergegeben.

Das genaue Funktionieren des Algorithmus ergibt sich aus den einzelnen Bestandteilen der RAP (s.o.) und ist in Lappin und Leass (1994: 542ff) beschrieben. RAP bei lexikalischen Anaphern (reziproke und reflexive Pronomen) umfasst weniger Schritte als bei Personalpronomen, da sie lokal gebunden sein müssen und die komplexe Numerusunterscheidung, die Lappin und Leass bei Personalpronomen vornehmen, ebenfalls entfällt.

Kennedy und Boguraev (1996) nehmen das Bewertungssystem von Lappin und Leass auf, modifizieren das Vorgehen aber dahingehend, dass ihr System nicht auf tiefes syntaktisches Parsing angewiesen ist, sondern nur einfache morphosyntaktische und Abhängigkeitsfunktionen abfragt. Da Kennedy und Boguraev auf einem Korpus von 27 Texten mit 306 anaphorischen Pronomen der 3. Person eine *accuracy* von 75% erreichen, schließen sie, dass sich auch wissensarme RAPs, die dann zum Einsatz kommen müssen, wenn keine verlässlichen Parsingressourcen vorhanden sind, für scorebasierte RAPs eignen. Die Gewichtung von Kennedy und Boguraev ist der dritten Spalte von Tab. 1 zu entnehmen.

#### **4. Implementation eines Pronomenauflösungsalgorithmus für das Deutsche**

In diesem Abschnitt wird die Implementation eines eigenen RAP vorgestellt, die die Grundidee von Lappin und Leass (1994) für das Deutsche anpasst.

##### 4.1. Das Korpus: Die TüBa-D/Z

Die *Tübinger Baubank des Deutschen/Zeitungskorpus* (TüBa-D/Z) wurde von der Sektion Allgemeine Sprachwissenschaft & Computerlinguistik der Universität Tübingen entwickelt und beinhaltet in dem hier verwendeten Release 8 (Beck 2013) etwa 75.000 morphosyntaktisch sowie text- und diskurssemantisch annotierte Sätze aus der deutschen Zeitung „die tageszeitung“, was ca. 1,4 Mio. Tokens bzw. 3.000 Texten entspricht. Mit mehr als 100.000 Anaphorik- und Koreferenzrelationen bietet die TüBa-D/Z das zur Zeit umfangreichste deutschsprachige Korpus mit derartigen Annotationsebenen.

Das Annotationsschema der TüBa-D/Z ist ausführlich in Telljohan et al. (2012) beschrieben. Tab. 2 listet die für diese Arbeit relevanten Tags auf<sup>2</sup>; die Annotation der Wortarten entspricht dem STTS

---

<sup>2</sup> Mein Algorithmus extrahiert auch morphologische Informationen, allerdings nur zum Kongruenzabgleich, sodass

(Schiller et al. 1999).

<b>Ebene</b>	<b>Typ</b>	<b>Tag</b>	<b>Beschreibung</b>
word	<i>pos</i>	PPER PRF PPOSAT	irreflexives Personalpronomen reflexives Personalpronomen attribuierendes Personalpronomen
	<i>deprel</i>	SUBJ OBJA OBJD OBJG DET PN EXPL	Subjekt direktes/Akkusativobjekt indirektes/Dativobjekt adverbiale Verwendung Determinierer Komplement einer PP Expletivum: Vorfeld-es
relation	<i>type</i>	anaphoric cataphoric inherent_reflexive expletive	Anapher Katapher inhärent reflexiv Expletivum: Alle Typen
node	<i>cat</i>	SIMPX	Simplex-Satz

Tabelle 2: Tags für Wortarten (*pos*), Dependenzrelationen (*deprel*), Referenz (*type*) und Knoten (*node*), die von dem Programm abgefragt werden.

Die TüBa-D/Z unterscheidet drei Typen von expletiven Relationen für „es“ (vgl. Telljohan et al. 2012: 51ff), die klar von referierenden Personalpronomen („das Haus<sub>i</sub> ... es<sub>i</sub> ...“) abgegrenzt werden: Das formale Subjekt/Objekt („es regnet“), das Korrelat eines extrapolierten Phrasenkomplements („es<sub>i</sub> ist übertrieben zu sagen, dass X<sub>i</sub>“) sowie das Vorfeld-es („Es steht hier viel Kram herum“). Nur letzteres wird explizit in *deprel* als Tag markiert; alle expletiven Verwendungen von „es“ werden allerdings in der Referenzrelation *relation* als expletiv getaggt.

Unter den Relationstypen dominieren die Koreferenz- und Anaphorikrelationen, die zusammen etwa 80% der gesamten Relationen ausmachen (Beck 2013). Die Verteilung ist wortartenabhängig komplementär: Alle koreferenten Nicht-Pronomen sind mit „coreferential“ getaggt, alle Pronomen mit einer anderen Relation (für eine vollständige Liste der Relationstags s. *ibid.*).

Die TüBa-D/Z steht in mehreren Exportformaten wie Negra oder als Penn Treebank zur Verfügung; das Pronomenresolutionssystem arbeitet auf der Export-XML. Um eine reibungslose Verarbeitung des Korpus zu gewährleisten, sind vorab am Korpus<sup>3</sup> folgende Modifikationen vorgenommen worden:

- Ersetzen aller nicht ASCII-kompatiblen Zeichen durch ASCII-Zeichenfolgen: ü > ue, ß > ss usw. (vermeidet Kodierungsprobleme bei der Weiterverarbeitung)

---

sie an dieser Stelle nicht zusätzlich aufgeführt werden.

3 Offizieller Dateiname der verwendeten Version: *8final.mitKonn.exml.xml*.

- Ersetzen des xml-Attributs „xml:id” durch „xml\_id” (wird benötigt, damit das Parsen mit *Element-Tree* richtig funktioniert)
- Erstellen eines Subkorpus (Texte 1 bis 251 ohne den in der Version fehlerhaften Text 81)

#### 4.2. Die Implementation

Ein in Python 2.7 geschriebenes Programm erhält die angepasste Korpusdatei (ein Fragment der TüBa-D/Z vom Umfang von 250 Texten, s.o.) als Input, berechnet für alle referierenden Pronomen Äquivalenzklassen (s.u.), extrahiert die Gold-Äquivalenzklassen aus den Koreferenzannotationen der TüBa-D/Z und evaluiert anschließend die Güte des Resolutionsergebnisses (s. Abschnitt 5).

Der Algorithmus umfasst folgende Schritte:

- Definiere *Score*-Werte, Markables und Pronomentypen, erstelle leere Äquivalenzregister.
- Erstelle am Anfang eines Textes jeweils ein leeres DM.
- Gehe nun satzweise vor. Sammle alle Anaphern (die oben definierten Pronomen mit Ausnahme jeglicher Expletiva) und alle potentiellen Antezedenten (Markables) und merke ihre morphosyntaktischen Werte. Merke gleichzeitig die Anaphorikannotationen der TüBa-D/Z.
- Aufbau des DM für referierende Markables und Pronomen: Bestimme die grammatische Funktion und weise den Markables die zuvor definierten *Score*-Werte (s. Tab 1, letzte Spalte) zu und speise sie in das DM ein. Ein OBJA wird als EXIST-Konstruktion gewertet, wenn folgende Bedingungen erfüllt sind:
  - ON ist der Wert von *func* im *node* des „es”
  - Im Satz, in dem das OBJA steht, gibt es mind. eine *expletive*-Relation und eine Wortform „es”
  - Es steht kein SIMPX-Knoten zwischen dem „es” und dem OBJA (s.u.)
- Gleiche alle Pronomen der Reihe nach mit allen Referenten im DM ab. Wende als harte Constraints an: Genuskongruenz, Numeruskongruenz (nicht bei PPOSAT) und Bindungsprinzipien. Wende als weiche Constraints an: Kataphernrelation und Parallelismuskonstruktion (für die *Score*-Werte s. 1).
- Ermittle Referenten mit höchstem *Score* (bei Gleichstand wähle den dem Pronomen nächsten) und füge das koreferente Paar dem Äquivalenzregister hinzu.
- Lösche Referenten, die vier oder mehr Sätze zurück liegen und teile alle Werte der verbleibenden Referenten durch 2.

Ein SIMPX-Knoten zwischen einem expletiven „es” und einem direkten Objekt wie in (5)a macht es sehr unwahrscheinlich, dass eine existentielle Konstruktion vorliegt, während ein Fehlen eines SIMPX-Knotens zwischen „es” und OBJA wie in (5)b zusammen mit den Kriterien in iv. hingegen

ein verlässlicher Hinweis auf existentielle Konstruktionen ist.

- (5) a. ... hält [<sub>OS-MOD</sub> es] für selbstverständlich, [<sub>SIMPX</sub> dass sie [...] ein dienstliches Handy<sub>OBJA</sub> hat.] (s25)  
 b. [<sub>SIMPX</sub> [<sub>ON</sub> Es] gibt so eine Buchung<sub>OBJA</sub>.] (s12)

## 5. Evaluation

Es existieren eine Reihe unterschiedlicher Verfahren zur Evaluation von RAPs, denen gemein ist, dass sie einen Gold-Standard (z.B. manuelle Annotation) mit dem zu bewertenden System abgleichen (vgl. Versley 2011: 64ff). Das in dieser Arbeit verwendete *entity/partition*-Verfahren (Vilain et al. 1995) ist ein partitionsbasiertes Evaluationsverfahren, das ab MUC-6 Verwendung gefunden hat. Es unterscheidet sich von linkbasierten Ansätzen dadurch, dass anstelle von einzelnen Links (Referentenpaaren) Mengen von Äquivalenzklassen miteinander verglichen werden. Die Vorteile dieses Ansatzes liegen darin, dass die Ergebnisse unabhängig von den Linkpaaren im System und im Goldstandard werden, sodass ein in vielen Fällen höherer und realistischerer F-Score errechnet wird. Die Formeln für den für *entity/partition* relevanten Partitionsschnitt sowie für die Berechnung von Precision (P), Recall (R) und F-Score (F) lauten:

(6) a.  $S1 \cap S2 := \{A1 \cap A2 | A1 \in S1 \wedge A2 \in S2 \wedge A1 \cap A2 \neq \emptyset\}$

b.  $P = \frac{I(Gold \cap System)}{I(System)}$

c.  $R = \frac{I(Gold \cap System)}{I(Gold)}$

d.  $F = \frac{2 \cdot P \cdot R}{P + R}$

	Deutsch				Englisch	
	L&L	Wunsch	Eig.	Baseline	K&B	L&L
<b>P</b>	63.92	64.88	65.22	42.77	75	<b>86</b>
<b>R</b>	72.18	73.27	<b>73.65</b>	43.38	---	---
<b>F</b>	67.8	68.82	<b>69.18</b>	45.4	---	---

Tabelle 1. Evaluationsergebnisse des Programms mit unterschiedlichen Scoreverteilungen im Deutschen und die Evaluationen von Systemen für das Englische zum Vergleich.

Die Evaluation des in Kap. 4 vorgestellten Algorithmus (Tab. 3) ergab, dass R in allen Bedingungen etwas höher als P lag. Der F-Score der Baseline, für die immer der am nächsten gelegene potentiell kongruierende referierende Ausdruck als Antezedens ausgewählt wurde, lag bei 45,40%. Durch Integration von syntaktischen Constraints und den Salienzwerten von Lappin und Leass (1994) verbesserte sich F auf 67,80%. Eine weitere Modifikation der Salienzwerte konnte F auf 69,18% ansteigen lassen (vgl. Tab. 1, letzte Spalte). Die für das deutsche Korpus angepassten Werte unterscheiden sich teilweise frappierend von den Werten für das Englische. Im nächsten Kapitel werden

diese Unterschiede und die Evaluationsergebnisse weiter erläutert und in einen größeren Kontext gestellt.

## 6. Diskussion

Um die Ergebnisse deuten zu können, muss bedacht werden, dass die Grundvoraussetzungen bei dem hier vorgestellten TüBa-D/Z-Ansatz andere sind als bei der ursprünglichen Implementation von Lappin und Leass. Poesio et al. (2010) nennen drei grundlegende Kriterien für die Vergleichbarkeit von Koreferenzauflösungssystemen ((7)a-c), denen ich einen vierten Punkt hinzufügen möchte:

- (7) a. Ist Pre-Processing ein Teil der zu evaluierenden Funktionalität?
- b. Werden Problemfälle wie Pronomen der 1. Person und Kataphern abgedeckt?
- c. Auf welcher Textgrundlage (Register) wird evaluiert?
- d. Auf welcher Sprache sind die Texte verfasst?

Die Voraussetzungen bei Lappin und Leass und der eigenen Implementation sind im Hinblick auf (7)a und (7)b vergleichbar; hinsichtlich (7)c-d sind die Unterschiede jedoch offensichtlich: Die TüBa-D/Z besteht aus deutschen Zeitungstexten verschiedener Register, während Lappin und Leass auf einem kleinen Korpus technischer Gebrauchstexte trainiert und evaluiert haben. Da zwischen einzelnen Registern erhebliche Unterschiede hinsichtlich Form, Funktion und Häufigkeit von linguistischen Einheiten und Konstruktionen bestehen können (Biber 1995), relativiert dies die Überlegenheit der Programme in den rechten beiden Spalten von Tabelle 3, da zu erwarten ist, dass regelbasierte Ansätze schwerer auf heterogene Korpora anzuwenden sind.

Die Evaluation zeigt eine mehr als 50%-ige Überlegenheit des Algorithmus mit syntaktischem Filter und *Score*-System gegenüber der Baseline. Die wirksamsten Modifikationen – Abwertung von *re-cency*, *existential emphasis*, *non-adverbial emphasis* und Aufwertung von *subject emphasis* – führten bereits bei Wunsch (2010) auf der Tüpp-D/Z, einem Korpus der Wissenschaftsausgabe der deutschen Zeitung „die tageszeitung“, zu erheblichen Verbesserungen. Dass eine weitere Aufwertung von *subject emphasis* und eine Abwertung von *accusative emphasis* sowie des Kataphernmalus den F-Score weiter verbessert, zeigt Beispiel Fehler: Referenz nicht gefunden, in dem Äquivalenzklassen in eckigen Klammern <...> angegeben sind. Die RAP nach Wunsch hat der Äquivalenzklasse, die mit s3617\_10 beginnt, im Gegensatz zu meiner RAP auch das Element s3621\_2 hinzugefügt, welches allerdings eine Katapher zu s3621\_12 („Regiseur [sic!] Kirk Jones“) und nicht koreferent mit s3617\_10 („Ned Divine“) ist. In Fehler: Referenz nicht gefunden waren also die höhere Ge-

wichtung von *subject emphasis* und die Abschwächung des Kataphernmalus ausschlaggebend für die richtige Entscheidung.

(8) a. Korpusfragment:

„In einem kleinen Dorf im Sueden Irlands stirbt Ned<sub>s3617\_10</sub> Divine [...] . Doch ist das ein Grund, dass er<sub>s3618\_8</sub> seinen<sub>s3618\_9</sub> Gewinn nicht bekommt? Seine<sub>s3619\_1</sub> Nachbarn [...] Mit seinen<sub>s3621\_2</sub> skurrilen Gestalten [...] erzahlt Regiseur<sub>s3621\_12</sub> Kirk Jones eine Geschichte aus dem Leben [...] .

b. Gold:

<s3617\_10, s3618\_8, s3618\_9, s3619\_1>; <s3621\_12, s3621\_2>

c. Meine Gewichtung:

<s3617\_10, s3618\_8, s3618\_9, s3619\_1>; <s3621\_12, s3621\_2>

d. Gewichtung Wunsch:

<s3617\_10, s3618\_8, s3618\_9, s3619\_1, s3621\_2>

Bemerkenswert ist, dass sich die eigene Gewichtung im Vergleich zu der von Lappin und Leass erheblich von der *NP accessibility hierarchy* entfernt hat: Die Unterschiede zwischen direktem, indirektem und PP-Objekt sind annulliert, während das Subjekt alle anderen Funktionen weit überragt; existentielle Konstruktionen fallen beinahe vollkommen aus der Hierarchie heraus bzw. hinter PP-Komplementen an das untere Ende. Die Dominanz der Subjektrolle im Deutschen theoretisch zu deuten, liegt außerhalb des Rahmens dieser Studie. Ein Modell, welches ein solches leistet, muss aber in jedem Fall die Subjektrolle mit der kohäsiven Funktion von Anaphern und den informationsstrukturellen Eigenschaften syntaktischer Konstituenten in Verbindung setzen. Speyer (2007) weist auf die kohäsive Funktion des Vorfelds im Deutschen hin, seine Daten zeigen jedoch auch, dass Vorfeldbesetzung allein kein sicheres Indiz für Topikstatus ist, da das Vorfeld in seiner Studie nur in ca. 50% der Fälle mit einem Topik (dem Cb nach Grosz & Weinstein 1995) besetzt war. Dass in seiner Studie fast 70% aller untersuchten Topiks gleichzeitig Subjekte waren, legt allerdings den Schluss nahe, dass die Subjektrolle im Deutschen (neben der Salienz) auch für Topikalität prädestiniert ist, sodass dem Subjekt eine informationsstrukturelle Sonderposition zukommen scheint, die stärker ausgeprägt ist als im Englischen.

Von einem typologischen Blickwinkel her betrachtet ist es in den Sprachen der Welt nicht unüblich, die Unterscheidung von Subjekt und Topik overt zu markieren. So schließen sich im Koreanischen der Subjektmarker *i-ka* und der Topikmarker *un-nun* gegenseitig aus, was ambige Strukturen verhindert (vgl. No 1991), während Topikalität im Deutschen syntaktisch (und prosodisch) nicht markiert werden muss, wie die Gegenüberstellung von (9) und (10) zeigt. Das Deutsche scheint sich auf einer Topikalitätsskala an einem extremen Pol, an dem Subjektrolle und Topikalität häufig korreliert

ren, zu befinden; an dem gegenüberliegenden Pol stehen „Topiksprachen“ wie das Koreanische. Innerhalb von „Subjektsprachen“ wie Deutsch oder Englisch lässt die schwächere Gewichtung der Subjektkontrolle für Saliens im Englischen die Vermutung aufkommen, dass das Englische auf der Skala dem Deutschen zwar nahe, aber etwas weiter vom Subjekt-Pol entfernt steht.

- |     |                                                                 |  |                                                                  |
|-----|-----------------------------------------------------------------|--|------------------------------------------------------------------|
| (9) | a. Koreanisch                                                   |  | b. Koreanisch                                                    |
|     | <i>p<sup>hit</sup>a-ka t<sup>eksi-und</sup>ansu-ro ilhan-da</i> |  | <i>p<sup>hit</sup>a-nun t<sup>eksi-und</sup>ansu-ro ilhan-da</i> |
|     | Peter-SUBJ Taxi-Fahrer-als arbeiten-FIN                         |  | Peter-TOP Taxi-Fahrer-als arbeiten-FIN                           |
|     | ‘[ <sub>SUBJ</sub> Peter] fährt Taxi.’                          |  | ‘[ <sub>TOP</sub> Peter] fährt Taxi.’                            |

- |      |                                           |                                                            |                                                                    |
|------|-------------------------------------------|------------------------------------------------------------|--------------------------------------------------------------------|
| (10) | a. Deutsch                                | b. Deutsch                                                 | c. Deutsch                                                         |
|      | [ <sub>SUBJ/ TOP</sub> Peter] fährt Taxi. | [ <sub>TOP</sub> Peter], [ <sub>SUBJ</sub> er] fährt Taxi. | [ <sub>TOP</sub> Was Peter angeht], [ <sub>SUBJ</sub> er] fährt T. |

Das in dieser Studie implementierte System besitzt klar benennbare Grenzen. Dazu zählen die Beschränkung auf Pronomen der 3. Person, das Nicht-Ausschöpfen der Ressourcen im Hinblick auf Relativpronomen, das Außenvorlassen semantischer Information (s. dazu aber Wunsch 2010: 206) und vor allem die schon bei Lappin und Leass zu Tage tretende Begrenztheit manueller *Scoring*-Systeme. Dass trotz dessen und trotz des breit gestreuten Korpus (vgl. (7)c) solide (d.h. weit über der Baseline liegende) Resultate erzielt werden konnten, zeigt aber, dass der gewählte Ansatz bis zu einem bestimmten Grad gut funktioniert. Idealerweise wäre die Methodik dieses regelbasierten Ansatzes in ein System zu integrieren, welches zusätzlich auf semantisches Wissen und maschinelle Lernverfahren zurückgreift.

## 7. Fazit und Ausblick

Generell hat es sich als sinnvoll herausgestellt, bei einem regelbasierten Pronomenresolutionsalgorithmus neben morphosyntaktischen „harten“ Constraints auch von durch die Hierarchie grammatischer Rollen motivierten „weichen“ Constraints Gebrauch zu machen.

Der von Lappin und Leass (1994) entwickelte scorebasierte Algorithmus für das Englische ließ sich mit minimalen Modifikationen auf das Deutsche übertragen. Die besten Ergebnisse wurden erzielt, wenn der Salienswert für die Subjektkontrolle für das Deutsche wesentlich höher gesetzt war als für das Englische. Es wurde herausgestellt, dass diese computerlinguistische Erkenntnis eine theoretische Modellierung benötigt, und zukünftige Forschungen können dort anknüpfen, indem sie bspw. der Frage nachgehen, welche Unterschiede sich im Bezug auf Salienswerte von grammatischen Rollen in verschiedenen Einzelsprachen feststellen lassen.



## 8. Literatur

- Baldwin, Breck, Morton, Tom, Bagga, Amit, Baldrige, Jason & Chandraseker, Raman. 1998. „Description of the UPENN CAMP system as used for coreference”. *Proceedings of the seventh Message Understanding Conference (MUC-7)*.
- Beck, Kathrin. 2013. *TüBa-D/Z Release 8*. [www.sfs.uni-tuebingen.de/ascl/ressourcen/corpora/tuba-dz.html](http://www.sfs.uni-tuebingen.de/ascl/ressourcen/corpora/tuba-dz.html) (letzter Zugriff am 22.11.2012).
- Biber, Douglas. 1995. *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Univ. Press.
- Bobrow, Daniel Gureasko. 1964. „Natural Language Input for a Computer Problem Solving System”. *AI Technical Reports* 66.
- Brennan, Susan. E., Friedman, Marilyn. W. & Pollard, Carl J. 1987. „A centering approach to pronouns”. *Proceedings of the 25th annual meeting on ACL*. 155–162.
- Bußmann, Hadumod. 2008. *Lexikon der Sprachwissenschaft* (4. Aufl.). Stuttgart: Kröner.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Clark, Herbert. 1975. „Bridging”. In: Roger C. Shank & Bonnie L. Nash-Webber (Hrg.): *Theoretical issues in natural language processing*. 169–174.
- Grishman, Ralph & Sundheim, Beth. 1996. „Message Understanding Conference – 6: A brief history”. *Proceedings of the 16. COLING Copenhagen*. 466–471.
- Grosz, Barbara J., Joshi, Aravind K. & Weinstein, Scott. 1983. „Providing a unified account of definite noun phrases in discourse”. *Proceedings of the 21st Annual Meeting of the Association of Computational Linguistics*. 44–50.
- Grosz, Barbara J., Joshi, Aravind K. & Weinstein, Scott. 1995. „Centering: A Framework for Modeling the Local Coherence of Discourse”. *Computational linguistics* 21 (2). 203–225.
- Heim, Irene Roswitha. 1982. *The semantics of definite and indefinite noun phrases*. PhD dissertation, University of Massachusetts.
- Hobbs, Jerry R. 1978. „Resolving Pronoun References”. *Lingua* 44. 311–338.
- Kamp, Hans & Reyle, Uwe. 1993. *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory*. Dordrecht: Kluwer.
- Keenan, Edward L. & Comrie, Bernard. 1977. „Noun Phrase Accessibility and Universal Grammar”. *Linguistic Inquiry* 8 (1). 63–99.
- Kennedy, Christopher & Boguraev, Branimir. 1996. „Anaphora for Everyone: Pronominal Anaphora Resolution without a Parser”. *Proceedings of COLING-96*. 113–118.

- Krifka, Manfred & Musan, Renate. 2012. „Information structure: Overview and linguistic issues”. In: Manfred Krifka & Renate Musan (Hrsg.): *The expression of information structure*. Berlin: de Gruyter. 1–43.
- Lappin, Shalom & Leass, Herbert J. 1994. „An Algorithm for Pronominal Anaphora Resolution”. *Computational linguistics* 20 (4). 535–561.
- No, Yongkyoon. 1991. „A centering approach to the \*[CASE][TOPIC] restriction in Korean”. *Linguistics* 29. 653–668.
- Poesio, Massimo, Ponzetto, Simone Paolo & Versley, Yannick. 2010. „Computational Models of Anaphora Resolution: A Survey“. *Linguistic Issues in Language Technology*.
- Reinhart, Tanja. 1982. *Pragmatics and Linguistics: An analysis of sentence topics*. Bloomington: Indiana University Linguistics Club.
- Schiller, Anne, Teufel, Simone, Stöckert, Christine & Thiele, Christine. 1999. *Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset)*. <http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf> (letzter Zugriff am 22.11.2012).
- Soon, Wee Meng, Ng, Hwee Tou & Lim, Daniel Chung Yong. 2001. „A Machine Learning Approach to Coreference Resolution of Noun Phrases”. *Computational linguistics* 27 (4). 521–544.
- Speyer, Augustin. (2007). „Die Bedeutung der Centering Theory für Fragen der Vorfeldbesetzung im Deutschen”. *Zeitschrift für Sprachwissenschaft* 26. 83–115.
- Strube, Michael. 2010. „Anaphernresolution”. In: Kai-Uwe Carstensen, Christian Ebert, Cornelia Ebert & Ralf Klabunde (Hrsg.): *Computerlinguistik und Sprachtechnologie. Eine Einführung* (3. Aufl.). Heidelberg: Spektrum Akademischer Verlag. 399–409.
- Telljohann, Heike, Hinrichs, Erhard W., Kübler, Sandra, Zinsmeister, Heike & Beck, Kathrin. 2012. *Stylebook for the Tübingen Treebank of Written German (TüBa-D/Z)*. <http://www.sfs.uni-tuebingen.de/resources/tuebadz-stylebook-0911.pdf> (letzter Zugriff am 22.11.2012).
- Tetreault, Joel R. 2001. „A Corpus-Based Evaluation of Centering and Pronoun Resolution”. *Computational linguistics* 27 (4). 507–520.
- van Deemter, Kess & Kibble, Rodger. 2000. „On coreferring: Coreference in MUC and related annotation schemes”. *Computational linguistics* 26 (4). 629–637.
- Versley, Yannick. 2011. *Resolving Coreferent Bridging in German Newspaper Text*. PhD thesis.
- Vieira, Renata & Poesio, Massimo. 2001. „An Empirically Based System for Processing Definite Descriptions”. *Computational linguistics* 26 (4). 539–593.
- Vilain, Marc B., Burger, John D., Aberdeen, John S., Connolly, Dennis & Hirschman, Lynette. 1995. „A model-theoretic coreference scoring scheme”. *MUC 1995*. 45–52.

- Winograd, Terry. 1972. „Understanding Natural Language”. *Cognitive Psychology* 3 (1). 1–191.
- Wunsch, Holger. 2010. *Rule-based and Memory-based Pronoun Resolution for German: A Comparison and Assessment of Data Sources*. PhD thesis.

# **Das gegenwärtige polnische Lautsystem und seine historische Entwicklung im Lichte der Präferenzgesetze der Silbenstruktur von Theo Vennemann**

*Katarzyna Plucińska (Universität Bamberg)*

## **1. Einleitung**

Der vorliegende Beitrag behandelt historische Sprachverwandlungen des Polnischen, die auf ihre Silbenstruktur hin untersucht werden. Das hier angesprochene Thema bezieht sich auf Vennemanns Präferenzgesetze, die dem Leser im ersten Teil näher gebracht werden sollen. Den Ausgangspunkt der folgenden Überlegungen bildet Vennemanns Voraussetzung, dass Sprache ein System ist, das sich ständig verbessert. Im zweiten Teil werden phonologische Eigenschaften des Polnischen erläutert, indem zuerst das polnische Lautinventar dargestellt wird, und danach wird ein grober Umriss von für die polnische Sprache üblichen Lautzusammensetzungen präsentiert. Im dritten Teil wird der Leser mit Mechanismen der erfolgten sprachlichen Veränderungen bekannt gemacht. Der Aufsatz fokussiert sich auf die Kontraste zwischen dem Altpolnischen und dem Neupolnischen. Die Schlussfolgerungen am Ende des Aufsatzes fassen die wichtigsten Ergebnisse zusammen. Der hinzugefügte Anhang soll dabei helfen, sich notwendige Informationen bezüglich des polnischen Lautbestands anzueignen.

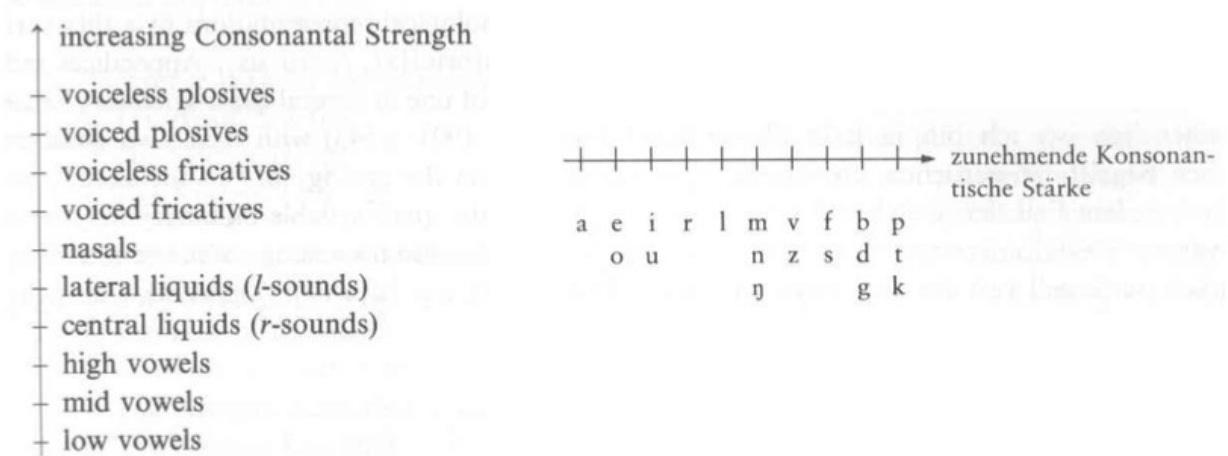
## **2. Das Modell der Präferenzgesetze**

Vennemann setzt mit seinen Präferenzgesetzen der Silbenstruktur voraus, dass Veränderungen von Silben die Silbenstruktur nur verbessern können. Seiner Meinung nach ist die Verschlechterung nicht möglich, weil die Silbe sich immer einem Parameter annähert. Innerhalb dieses Parameters wird die Silbenstruktur also verbessert (vgl. Vennemann 1986: 32–33). Außerdem greift Vennemann auf das Sonoritätskonzept zurück, welches annimmt, dass man die Laute jeder Sprache nach Sonoritätsklassen ordnen und auf dieser Basis den Silbenbau für alle Sprachen bestimmen kann (vgl. Eisenberg 2006: 104). Die Sonoritätsskala bzw. Konsonantische Stärke<sup>1</sup> soll den Linguisten helfen, Verhältnisse von Einzellauten zueinander innerhalb von Silben besser zu verstehen. Sie soll auch ein möglichst allgemeines Instrumentarium bieten, universell präferierte Abfolgen von Lauten beschreiben zu können (vgl. Haspelmath et al. 2001: 1311). Jedem Vokal und jedem Konsonanten

<sup>1</sup> Sonoritätsskala vs. Konsonantenstärke bzw. Konsonantische Stärke werden als synonyme Begriffe verwendet. In der Sonoritätstheorie haben Vokale den maximalen Wert, Plosive den minimalen. Bei Vennemann sieht das umgekehrt aus.

wird ein Wert auf der Skala zugeschrieben. Je weiter weg ein Laut von den tiefen Vokalen entfernt ist, desto kleiner ist die Chance, dass er den „erwünschten“ Silbenkern bilden wird.

Vennemanns Konsonantische Stärke präsentiert sich wie folgt:



Vennemann schlägt acht Präferenzgesetze vor, die die Jakobsonschen (Un-)Markiertheitsauffassung relativieren sollen. Vennemann operiert mit solchen Ausdrücken wie: *um so stärker bevorzugt, je...* anstatt wie Jakobson *markiert vs. unmarkiert* (vgl. Haspelmath et al. 2001: 1312). Drei von seinen acht Gesetzen thematisieren Einzelsilbe, die übrigen fünf beschreiben Silbenfolgen.

Die ersten drei Gesetze des Silbenbaus lauten (Vennemann 1986: 38–39):

Anfangsrandgesetz (The Head Law) - „Ein Anfangsrand ist um so stärker bevorzugt, (a) je näher die Anzahl seiner Sprachlaute bei Eins liegt, (b) je größer die Konsonantische Stärke seines ersten Sprachlauts ist und (c) je schärfer die Konsonantische Stärke vom ersten Sprachlaut auf die Stärke folgenden Nukleus zu abfällt.“

Endrandgesetz (The Coda Law) - „Ein Endrand ist um so stärker bevorzugt, (a) je kleiner die Anzahl seiner Sprachlaute ist, (b) je geringer die Konsonantische Stärke seines letzten Sprachlauts ist und (c) je schärfer die Konsonantische Stärke vom letzten Sprachlaut auf die Stärke des vorangehenden Nukleus zu abfällt.“

Nukleusgesetz (The Nucleus Law) - „Ein Nukleus ist um so stärker bevorzugt, je geringer seine Konsonantische Stärke ist.“

Aus diesen Präferenzen ergibt sich die CV-Silbenstruktur als ideale Struktur. Wenn wir diese Struktur mithilfe der Konsonantenstärke abbilden, dann bekommen wir eine Kurve, die vom Kopf zum Nukleus abfällt. Je größer der Abstand zwischen dem Konsonanten im Kopf und dem Vokal im Nukleus ist, desto bevorzugter ist die Silbe. Da jedoch auch geschlossene Silben in den Sprachen der Welt vorzufinden sind, bekommt man auf dem Abbild einen „Knick“. Das ganze Bild ähnelt dann der V-Form. Die Kurve fällt vom Kopf zum Nukleus ab und danach steigt sie zur Coda (die sog. Trogform der Silbe). Silben, die eine große Anzahl von Konsonanten im Onset bzw. im Offset auf-

weisen, können zugleich keine besondere Schärfe zwischen dem Kopf und dem Nukleus bzw. zwischen dem Nukleus und der Coda haben. Das hat zur Folge, dass solche Silben weniger der „idealen“ Gestalt entsprechen, und laut Vennemann weniger bevorzugt seien.

Ziel der Präferenzgesetze für Silbenfolgen ist, den bestmöglichen Silbenkontakt zu definieren. Vennemann formulierte sie folgendermaßen (Vennemann 1986: 39, 44):

Prokoschs Gesetz (The Weight Law) - „Eine (dynamische) akzentuierte Silbe ist um so stärker bevorzugt, je näher ihr Gewicht bei zwei Moren liegt (Die optimale akzentuierte Silbe ist zweimorig).“

Initialgesetz (The Law of Initials) - „Wortmediale Anfangsränder sind um so stärker bevorzugt, je weniger sie sich von möglichen wortinitialen Anfangsrändern des Sprachsystems unterscheiden.“

Finalgesetz (The Law of Finals) - „Wortmediale Endränder sind um so stärker bevorzugt, je weniger sie sich von möglichen wortfinalen Endrändern unterscheiden.“

Stärkeassimilationsgesetz (The Strength Assimilation Law) – “Wird in einem Silbenkontakt die Konsonantische Stärke assimiliert, so verringert sich die Konsonantische Stärke des stärkeren Sprachlauts.“

Silbenkontaktgesetz (The Contact Law ) - „Ein Silbenkontakt ist um so stärker bevorzugt, je größer die Differenz in der Konsonantischen Stärke zwischen dem zweiten und dem ersten Sprachlaut ist.“

An dieser Stelle konzentriere ich mich auf das letztgenannte Gesetz, denn es ist für die weitere Analyse besonders relevant. Das Silbenkontaktgesetz folgt aus dem Anfangsrand- und aus dem Endrandgesetz. Anhand der Konsonantenstärke kann man feststellen, welche Verteilungsart von Lauten von dem optimalen Silbenkontakt gefragt ist. Vennemann vergleicht das Wort *Wartha* mit dem Wort *Tatra*. Er beruft sich auf Werte der Konsonantischen Stärke, um zu veranschaulichen, dass der Silbenkontakt /r:t/ bevorzugt, und der /t.r/ nicht bevorzugt ist (vgl. Vennemann 1986: 41).

## 2. Umriss des gegenwärtigen polnischen Lautsystems

### 2.1. Phoneminventar<sup>2</sup>

Im Laufe der Zeit verschwand aus dem Polnischen die Vokalquantität, was dazu beitrug, dass das heutige Vokalsystem sehr vereinfacht ist. Im Gegensatz dazu wurde der Konsonantismus stark ausgebaut, z.B. durch Palatalisierungsprozesse. Ähnliche Abhängigkeiten zwischen dem reduzierten Vokalsystem und dem ausgeweiteten Konsonantensystem finden wir beispielsweise auch im Russischen oder in einem ostbulgarischen Dialekt (vgl. Rospond 1971: 132). Die phonologische Struktur des Polnischen entwickelte sich harmonisch, indem die gut ausgebaute Opposition zwischen weichen und harten Konsonanten entstand. Harmonisch war die Veränderung auch deswegen, weil die

<sup>2</sup> Da *Graphem-Phonem-Entsprechungen* im Polnischen recht komplex und oft nicht eindeutig sind, wird hier darauf verzichtet zahlreiche orthographische Regeln zu schildern. Polnische Wörter werden in dem Aufsatz transkribiert, so dass der deutschsprachige Leser jedes einzelne Beispiel selbst analysieren kann.

Palatalisierung von Konsonanten die Depalatalisierung von Vokalen auslöste, so dass es zu keinem einseitigen phonologischen Ausbau kam (vgl. Rospond 1971: 131–132).

Heutzutage weist das Polnische sechs Grundvokale (zudem zwei nasalierte Grundvokale) und 35 Konsonanten auf (vgl. Wójtowicz 1975: 86). Die Diskrepanz zwischen der Häufigkeit des Vorkommens der Konsonanten und der der Vokale beträgt 59,9% für die Konsonanten und 40,1% für die Vokale (vgl. Wójtowicz 1975: 128–129). Obwohl das polnische Vokalinventar nicht so groß wie das deutsche ist, finden wir Nasalvokale und den [i] Laut, die der deutschen Sprache fremd sind. Interessant kann für den Leser sein, dass unter allen slawischen Sprachen nur die polnische die Nasalvokale beibehält. Es scheint auch erwähnenswert zu sein, dass der [i] Laut in der polnischen Gegenwartssprache nicht, wie im Russischen, den tiefen Vokalen angehört, sondern den hohen. Dadurch ist die Differenz zwischen [i] und [i] nicht besonders prägnant, was noch einmal beweist, dass das polnische Vokalsystem relativ einfach gebaut ist (vgl. Rospond 1971: 133, Bartnicka et al. 2004: 66).

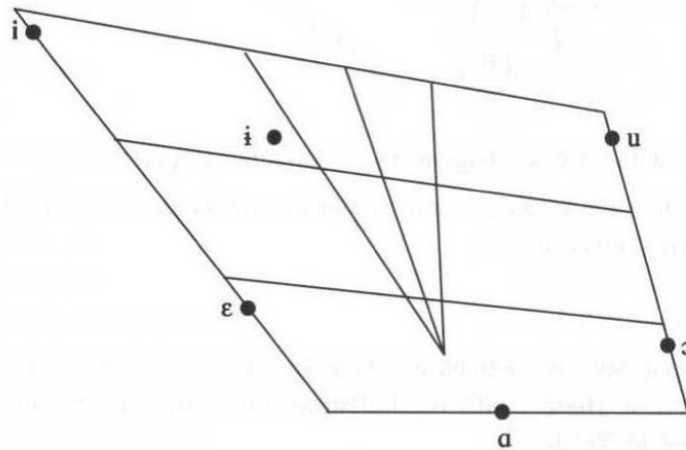


Abbildung 2: Vokalviereck (Bartnicka et al. 2004: 30)

Bei der Einführung in das polnische Konsonantensystem konzentriere ich mich auf weich realisierte Konsonanten, weil sie dem deutschen System unbekannt sind. „Die Palatalisierung eines Lautes, seine Erweichung, kann durch die Hebung der Zungenmitte gegen das Palatum bewirkt werden“ (Wójtowicz 1975: 54). Nicht alle palatalen Laute sind aber hinsichtlich ihrer Artikulation gleich. Man muss sie am besten in zwei separate Gruppen einordnen, denn für die Laute wie [ɕ], [ʐ], [tɕ] oder [dʑ] bildet das Palatum (Prä- oder Postpalatum) die Hauptartikulationsstelle.

„Man kann nicht behaupten, daß sie aus den dem Konsonanten s eigentümlichen Bewegungen sowie aus der palatalisierenden Bewegung der Zunge gegen das Palatum bestehen“ (Wójtowicz 1975: 54).

Laute, wie [p,], [b,], [f,] oder [v,] bestehen dagegen aus zwei Artikulationsbewegungen; aus einem Verschluss bzw. einer Enge und aus einer Hebung der Zunge gegen das Palatum. Die Laute aus der

zuerst besprochenen Gruppe werden als Palatale bezeichnet, die übrigen, die durch eine zusätzliche Bewegung der Zunge erweicht werden, nennt man palatalisierte Laute (vgl. Wójtowicz 1975: 54–55).

Obwohl das Polnische über einen so reichen Konsonantenbestand verfügt, fehlen ihm der Hauchlaut [h] und der Knacklaut [ʔ], die dagegen das Deutsche besitzt (vgl. Bartnicka et al. 2004: 33).

## 2.2. Phonotaktik

„Die Vokale [ɑ], [ɔ], [u], [ɛ], [ɨ] und [i] können vor allen Konsonanten auftreten. Beschränkt ist jedoch ihr Vorkommen nach Konsonanten“ (Bartnicka et al. 2004: 70). Die Nasalvokale [ẽ] und [õ], deren Häufigkeit des Vorkommens unter 1% liegt, stehen nie im Onset<sup>3</sup>. Am seltensten tauchen im Polnischen die hohen Vokale auf, und am häufigsten die Vokale der mittleren Reihe (vgl. Wójtowicz 1975: 132). Die Vokalanhäufung ist für die polnische Sprache auch nicht üblich. Noch seltener stoßen wir in einheimischen Wörtern auf Vokalverbindungen mit [i]. Man könnte vereinfacht sagen, dass die Hauptfunktion von [i] darin besteht, einen davor stehenden Konsonanten zu palatalisieren. Das Polnische verfügt auch über keine Diphthonge und nebeneinander stehende Vokale (maximal drei Vokale) werden bei der Aussprache „recht klar voneinander getrennt“, wobei sie auch gedehnt gesprochen sein können (vgl. Bartnicka et al. 2004:70).

Das spezifischste Merkmal des polnischen Konsonantismus, welches Wójtowicz zur Sprache bringt, ist die Konsonantenkorrelation, die einerseits „auf der Stimmhaftigkeit und der Stimmlosigkeit“ beruht, andererseits auf der Palatalität und die Härte (Wójtowicz 1975: 128). Typisch für das Polnische ist außerdem nicht nur eine große Anzahl von Konsonanten, sondern auch ausgebaute Konsonantenzusammensetzungen, die bis fünf Laute im Inlaut zählen können. Die konsonantischen Gruppen im An- oder im Auslaut sind geringer und zählen bis vier Elemente (vgl. Bartnicka et al. 2004: 73). Konsonanten, die angehäuft werden, müssen nicht die gleiche Artikulationsstelle bzw. Artikulationsart vertreten. „Die Konsonantengruppen aus Obstruenten müssen [aber K.P.] homogen bezüglich der Stimmhaftigkeit sein“ (Bartnicka et al. 2004: 73). Die Obstruenten tauchen in jeder Position auf, aber „am absoluten Auslaut“ verlieren sie ihre Stimmhaftigkeit (Auslautverhärtung). Konsonantische Geminaten werden im Polnischen wie ein langer Konsonant oder wie zwei kurze Konsonanten ausgesprochen (vgl. Wójtowicz 1975: 114).

---

3 Im Polnischen kommen zusätzlich folgende nasale Vokalvarianten vor : [ĩ], [ĩ̃], [ã] und [ũ] (vgl. Bartnicka et al. 2004: 30, Wójtowicz 1975: 90). Da sie aber eine relativ marginale Rolle im Polnischen spielen, werden sie in der Arbeit nicht angesprochen.



### 3. Übergang vom Altpolnischen zum Neupolnischen<sup>4</sup>

#### 3.1. Kontraktion

Die Zusammenziehung von Lauten lässt sich mit Beispielen von Adjektiven mit Deklinationsendungen *-aja* und *-oje* präsent machen. Zwei Vokale, die durch den Halbvokal (nicht-silbenbildend) geteilt waren, wurden zusammengezogen (vgl. Rospond 1971: 93).

*-aja* > *ā*      Bsp.: (1) *dobraja* [dɔ.bra.ja] > *dobrā* ['dɔ.bra:] (dt.: eine gute)

*-oje* > *ē*      (2) *dobroje* [dɔ.brɔ.jɛ] > *dobrē* ['dɔ.brɛ:] (dt.: ein gutes bzw. die guten)

In Konsequenz entstand ein langes *a* bzw. *e*, das später seine Quantität [+lang] verlor. Beide Wörter, (1) und (2), bewiesen von Anfang an eine hohe Konsonantische Stärke im Onset der ersten und zweiten Silbe. Alle Silben waren und blieben offen, und auch der Silbenkontakt brauchte keine Verbesserung. Die Wortverkürzung war aber trotzdem aus sprachökonomischen Gründen positiv.

#### 3.2. Metathese

Das Slawische, also auch das Polnische, tendiert seit der balto-slawischen Zeit zu Sprachen mit offenen Silben. Anstelle der VC-Kombination setzte sich die Folge CV durch und infolgedessen wurde die Silbe geöffnet (vgl. Kempgen 2005, Mazur 1993)<sup>5</sup>. „Das Slawische entscheidet sich [...] für eine Realisierung des typologisch grundlegenden Modells, nämlich der Abfolge C(C)V– als Grundform der Silbe“ (Kempgen 2005: 122). Wie Kempgen weiter argumentiert, ist dieses Silbennmuster „natürlicher“, denn aus artikulatorischer Sicht ist diese Struktur weniger aufwendig als VC-Struktur. Obwohl Prozesse, die zur Silbenöffnung führten, charakteristisch für die früheste Epoche aus der slawischen Geschichte waren, kann man aufgrund des Altpolnischen belegen, dass auch später Silben mithilfe der Metathese „verbessert“ wurden.

Bsp.: (3) *ordlo* [ɔr.dlɔ] > *radło* ['ra.dwɔ] (dt.: Hakenpflug)

Die Reihenfolge V + C im Wortonset *ordlo* wurde so verändert, dass eine neue offene Silbe entstand<sup>6</sup>. Nach der Konsonantischen Stärke gilt die *or*-Silbe als nicht meliorativ (vgl. Rospond 1971: 98).

#### 3.3. Assimilation

Nur diejenigen Laute wurden assimiliert, die einen ähnlichen Artikulationsort hatten. Im Altpolnischen

---

4 Alle Beispiele im untersuchten Material wurden von Rospond (1971) übernommen.

5 In der Fachliteratur wird das Prozess der Silbenöffnung als *Gesetz der offenen Silbe* oder als *Tendenz zur steigenden Sonoritätswelle* bezeichnet.

6 Hier kam es auch zum Austausch von *o* gegen *a*.

schen bezog sich der Assimilationsprozess auf die dentale und alveolare Nachbarschaft von Lauten. Das veranschaulichen folgende Beispiele:

*cs > c* Bsp.: (4) *csō* [ʦsɔ] > *co* [ʦɔ] (dt.: was)

(5) *csnota* [ʦsnɔ.ta] > *cnota* [ʦnɔ.ta] (dt.: Tugend)

*czs > c* (6) *swiadeczstwo* [sfja.dɛʦʦ.stfɔ] > *świadectwo* [ɛfja.'dɛʦʦ.tfɔ] (dt.: Zeugnis)

Im 14. Jh. waren noch die alten nicht-assimilierten Formen gebräuchlich, ein Jahrhundert später schon die assimilierten (vgl. Rospond 1971: 122).

Bei (4) und (5) wurden die Anfangsränder optimiert, indem die früher „ähnlich klingenden“ Elemente im Onset neutralisiert wurden. Demzufolge wurde das Wort aus auditiver Sicht verbessert. Im Wort (6) wurde der markierte Silbenkontakt zugunsten des unmarkierten abgebaut; eine Affrikate + ein Frikativ → eine Affrikate + ein stimmloser Plosiv.

### 3.4. Vereinfachung von Phonemgruppen

Unten werden Beispiele angeführt, die die altpolnische Vereinfachung innerhalb von dreiteiligen Phonemgruppen darstellen. Alle Phoneme in den angesprochenen konsonantischen Anhäufungen gehören zu Lauten, die durch die Vorderzunge realisiert werden (vgl. Rospond 1971: 127):

*zdn > zn* Bsp.: (7) *Gniezdno* [gnɛ.zdnɔ] > *Gniezno* [ˈgnɛ.znɔ] (Stadtname, dt: Gnesen)

(8) *pozdno* [pɔ.zdnɔ] > *póź.no* [ˈpu.znɔ] oder [ˈpu.znɔ] (dt.: spät)

*ldn > ln* (9) *żołdnierz* [ʒɔw.dnɛʃ] > *żołnierz* [ˈʒɔw.nɛʃ] (dt.: Soldat)

*stb > zb* (10) *istba* [is.tba] oder [ist.ba] > *izba* [ˈiz.ba] (dt.: Stube)

Die sprachlichen Vereinfachungen waren nicht willkürlich, sondern hingen vom Charakter der phonetischen Verbindung ab. Zwar wurden in (7) bis (10) die Silben nicht geöffnet, aber die Wörter wurden trotzdem verkürzt. Als Gegenbeispiele könnte man *miłosierdzie* [miwɔɕɛrdzɛ] (dt.: *Barmherzigkeit*) und *żold* geben. Das Wort *miłosierdzie*<sup>7</sup> enthält das nicht reduzierte Element *sierd*<sup>8</sup>, und kann heute als Sprachdenkmal der vergangenen altpolnischen Periode gelten. Einerseits fiel von der Silbe *żold* (9) das *d* weg, andererseits bewahrte das Polnische das Wort *żold*. Dieses wurde aus dem Deutschen entlehnt; *żold* – dt.: *Sold*.

Dank der oben geschilderten Vereinfachungen nahmen die Wortcodas ab und näherten sich den Standards des Endrandgesetzes an. Der Silbenkontakt wurde bezüglich der Konsonantenstärke nicht verbessert, eine Ausnahme ist das Beispiel (9)<sup>9</sup>. Bei (10) kann man nicht eindeutig sagen, ob der Kontakt im Ausgangspunkt tatsächlich „schlecht“ war. Bei dem Wort *istba* lässt sich die Silbengren-

7 Ein Beispiel für die Optimierung des Silbenkontakts über die morphologische Grenze hinweg: *mi.ło.sier.dzie*

8 Das Element kam im Wort: *sierdce > serce* (‘Herz’) vor.

9 Für die Untersuchung bezüglich der Konsonantischen Stärke wird [w] als Halbvokal zwischen Liquiden und hohen Vokalen platziert.

ze verschieden setzen und keine Lösung berücksichtigt zugleich die Anfangsrand- und die Endrandparameter.

### 3.5. Zusammenfassung des untersuchten Materials

Das Slawische, wie in dem Aufsatz schon angedeutet wurde, reduziert den Auslaut stark und strebt offene Silben an. Die Präferenz der slawischen Sprachen stimmt also mit dem Präferenzgesetz des Silbenauslauts überein. Die polnische Sprache passt sich auch dem Nukleusgesetz gut an, weil sie nur Vokale, und keine silbischen Konsonanten als Silbenkern erlaubt. Ein besonders breites und interessantes Thema ist der Silbenkontakt. Auch hier gelten silbenbezogene Restriktionen, die ihren Bezug auf Werte der Konsonantischen Stärke nehmen. Im Fall eines einzigen intervokalischen Konsonanten (VCV) steht er immer als erstes Element im Anlaut, und nicht als letztes im Auslaut, d.h. V.CV, und nicht VC.V. Die genannte Art der Silbentrennung untermauert Vennemanns Anfangsrand- und Endrandgesetz. Man kann noch ein Argument finden, das eine solche Zerlegung von Lautketten im Polnischen, aber auch im Deutschen, klärt. Obstruenten im Offset unterliegen nämlich der Auslautverhärtung. Aufgrund der Aussprache können wir also folgende Schlussfolgerung ziehen: die V.CV-Schilderung gibt die einzige korrekte Silbentrennung wieder. Zu der Frage, wie man Konsonantengruppen in Silben teilen kann, gibt es mehrere Meinungen. Bartnicka et al. sagen zum Beispiel, dass Konsonantengruppen zwischen den Vokalen nach Belieben geteilt werden könnten (vgl. Bartnicka et al. 2004: 89). Bethin zitiert dementsprechend eine interessante Anmerkung von Szober, der erstens für die Beibehaltung der offenen Silben, zweitens für die Markierung der Silbengrenzen zwischen Sonoranten und Obstruenten plädiert:

Szober (1963: 20–21) claims that open syllables prevail in Polish and that medial consonant clusters all belong to the beginning of the following syllable and not the preceding one: ma.tka, prze.cha.dzka, kre.wny etc. Exceptions are only in the case of sonorants, e.g., war.tki, wal.ka ...” (Bethin 1992: 25).

Unter dem Gesichtspunkt der steigenden Sonorität betrachtet sind Szobers Annahmen „silbenkontaktgesetzfreundlich“.

## **4. Schlussfolgerungen**

Die Sonoritätshierarchie, die in der Form der Sonoritätsskala zum Ausdruck gebracht wurde, soll als universell gelten. Der universelle Charakter der Skala könnte dadurch bestätigt werden, dass die Skala nicht für Einzellaute, sondern für Lautklassen aufgebaut wurde. Sie soll eine übergreifende Ordnung von Lauten aller Sprachen ermöglichen, was aber nicht immer gelingt.

Nun zu Vennemanns Gesetzen: Da sie auf der nicht einwandfreien Hierarchie der Konsonantischen Stärke aufgestellt wurden, können sie logischerweise auch keine universellen Regeln vertreten. Die

polnische Sprache lässt sich den Gesetzen auch nicht in jeder Hinsicht unterordnen, z.B. durch ihre komplexen Konsonanzzusammensetzungen. Vennemann betont, dass in einem Sprachsystem „niemals alle Parameter optimiert werden können“, sondern es lässt sich nur eine bestimmte Gesamtqualität erreichen (vgl. Vennemann 1986: 34). Da die Tendenz zur unmarkierten Serialisierung oft durchbrochen wird, kann man die Gesetze, was übrigens Vennemann selbst unterstreicht, nur als Präferenzgesetze ansehen. Alle Abweichungen kann man dann mit Köhlers Zitat begründen, der sagt, dass Sprache sich zu keinem Zeitpunkt in einem Zustand befinde, der sich als Sollzustand interpretieren lasse (Kempgen 2005: 119).

Vennemanns Model lässt viele Fragen offen. Seine Konzeption bietet eine strukturelle Verallgemeinerung an, womit man Relationen zwischen Lauten innerhalb einer Silbe und zwischen Silben darstellen kann. Sie erklärt jedoch nicht, warum in manchen Sprachen komplexe Konsonantencluster so beliebt sind, dass sie seit Jahrhunderten nicht abgebaut wurden. Einerseits könnte man als Befürworter der Präferenzgesetze die Frage mithilfe von nur zwei Argumenten antworten. Erstens werden „fehlerhafte“ Strukturen in Zukunft noch verbessert. Zweitens ist eine Silbengestalt, die gemäß einem Parameter nicht optimal zu sein scheint, nach einem anderen Parameter doch angemessen. Andererseits kann aber insbesondere die Sonoritätsskala einen Anstoß dazu geben, dass man für einzelne Sprachen ähnliche, aber adäquatere Systeme entwirft, die individuelle sprachliche Einzelheiten besser umfassen können. Das heißt jedoch nicht, dass man hier den universellen Aspekt der Präferenzgesetze ablehnen sollte. Die Präferenzgesetze der Silbenstruktur geben den Linguisten dennoch die Möglichkeit, sich mit verschiedenen Sprachen auseinanderzusetzen und zu zeigen, welche Strukturen in den jeweils untersuchten Sprachen in der Welt bevorzugt werden.

## 5. Literaturverzeichnis

- Bartnicka, Barbara, Hansen, Björn, Klemm, Wojtek, Lehmann, Volkmar & Satkiewicz, Halina. 2004. *Grammatik des Polnischen*. München: Otto Sagner.
- Bethin, Christina Y. 1992. *Polish syllables. The Role of Prosody in Phonology and Morphology*. Ohio: Slavica Publishers, Inc.
- Eisenberg, Peter. 2006. *Grundriss der deutschen Grammatik. Bd.1.: Das Wort*. Stuttgart/ Weimar: Metzler.
- Haspelmath, Martin, König, Ekkehard, Oesterreicher, Wulf & Raible, Wolfgang. 2001. *Language Typology and Language Universals/ Sprachtypologie und sprachliche Universalien/ La typologie des langues et les universaux linguistiques*. Bd. 2. Berlin: de Gruyter.
- Kempgen, Sebastian (Hrg.). 2005. *Slavistische Linguistik 2003. Referate des XXIX. Konstanzer Slavistischen Arbeitstreffens Bamberg, 15. – 19.9.2003*. München: Otto Sagner.

- Mazur, Jan. 1993. *Geschichte der polnischen Sprache*. Frankfurt a.M.: Peter Lang.
- Rospond, Stanisław .1971. *Gramatyka historyczna języka polskiego*. Warszawa: PWN.
- Wójtowicz, Janina. 1975. *Phonetik der polnischen Sprache*. Bochum: Studienverlag.
- Vennemann, Theo. 1986. *Neuere Entwicklungen in der Phonologie*. Berlin: Mouton de Gruyter.
- Vennemann, Theo. 1988. *Preference Laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin*. Berlin: de Gruyter.

# ***Pink* im Deutschen und Dänischen**

## **Eine Studie zur Semantik von Farbwortentlehnungen**

*Cornelia van Scherpenberg (Ludwig-Maximilians-Universität München)*

### **1. Einleitung**

Wer heutzutage in Deutschland Töchter aufzieht, wird sich früher oder später mit einer Welt konfrontiert sehen, die meist nicht aus der eigenen Kindheit oder Jugend bekannt ist: Die Welt des Pinkfarbigen – von Kleidung über Lebensmitteln bis zum Haare färben. Mädchen wollen sich nicht mehr nur mit dem typischen Baby-Rosa umgeben, sondern fordern ein neues pinkes Kleid oder pinken Nagellack.

Die Farbe **Rosa** wird schon seit den 1920er Jahren klischeehaft mit allem Mädchenhaften assoziiert (Heller 1989: 116–118). **Pink** als Farbwort des Deutschen ist dagegen erst in der zweiten Hälfte des 20. Jahrhunderts aufgetreten. Im Jahr 1970 erscheint es als Modefarbwort neben *Bleu*, *Rouge* und *Phosphor*<sup>1</sup>, die alle drei heute zumindest im alltäglichen Sprachgebrauch nicht mehr vorkommen. 1997 wird *Pink* in einer Studie des Marktforschungsinstituts iconkids & youth<sup>2</sup> schon von 9% der Mädchen im Alter von 6-8 Jahren und 4% der Mädchen im Alter von 9-11 Jahren als Lieblingsfarbe angegeben. Ungefähr gleich häufig werden z.B. *Lila/Violett* und sogar *Gelb* und *Grün* genannt. 2002 nimmt es der Verein für Deutsche Sprache (VDS) in seinen Anglizismenindex auf – beschrieben als „verdrängend“<sup>3</sup>. Das aus dem Englischen entlehnte Farbwort *pink* hat sich neben dem autochthonen *rosa* eingereiht und sowohl eine eigene Denotation als auch eine eigene Semantik übernommen, die sich zu gewissen Teilen von der des Englischen gelöst hat.

Folgender Satz, den eine Siebenjährige aus Deutschland heute problemlos sagen könnte, soll exemplarisch die grundlegenden Unterschiede zeigen: „Mama, ich will heute nicht das **pinke** Kleid anziehen, sondern das **rosane**!“. Dieser Satz ist in seiner wörtlichen Übersetzung im Englischen nicht möglich: \*„I want to wear the **pink** dress today, not the **pink** one!““. Im Dänischen bestand bei privat befragten Muttersprachlern noch Uneinigkeit, ob „Jeg vil have den **pink** kjole på i dag, ikke den

---

1 Verkaufswerbung Ausgabe 31/70 der Kaufhof AG, über [http://www.korpora.org/cgi-bin/Limas/quellen.pl?nummer=046&wortid=046\\_0763#087](http://www.korpora.org/cgi-bin/Limas/quellen.pl?nummer=046&wortid=046_0763#087). Dies ist der früheste Beleg, den ich bei einer Korpusrecherche gefunden habe. Carstensen (2001: 1062) gibt 1967 als ersten Beleg an.

2 Studie „Lieblingsfarben der Kids“, [http://www.iconkids.com/deutsch/04publikationen/04\\_1/bis1999/1999.html#10](http://www.iconkids.com/deutsch/04publikationen/04_1/bis1999/1999.html#10).

3 Für diese Information danke ich Reiner Pogarell vom IFB-Verlag.

**lyserøde**<sup>4</sup> ein „guter“ Satz ist.

Diese Arbeit beschäftigt sich mit der Art und Weise des Eintretens von *pink* in das Deutsche und Dänische. Es sollen Prozesse und Folgen der Entlehnung des englischen Farbworts aufgezeigt, seine semantische und morphologische Integration in das Farbwortsystem der jeweiligen Sprachen analysiert und sein Verwendungsspektrum beschrieben werden. Im Deutschen wird dies vor allem in Abgrenzung zum autochthonen Farbwort *rosa* getan, im Dänischen zu *lyserød*. Letztendlich soll dabei die Frage geklärt werden, inwieweit *pink* im Deutschen und Dänischen eine eigene Farbe bezeichnet, bzw. ob es „eine Art von“ *rosa/lyserød* oder ein Hyponym dazu ist (siehe auch unten 1.2.). Insgesamt wird anhand des Beispiels *pink* untersucht, wie sich Farbwörter im Sprachkontakt verhalten können.

### 1.1. Hintergrund der Fragestellung: Das EoSS-Projekt

Das Farbwort *pink* im Deutschen war schon öfter Gegenstand sprachwissenschaftlicher Untersuchungen (siehe 1.2.). Zuletzt wurde seine Präsenz sowohl im Deutschen als auch im Dänischen in folgendem Zusammenhang eindeutig belegt:

Das Forschungsprojekt „Evolution of Semantic Systems“ (EoSS) am Max-Planck-Institut für Psycholinguistik in Nijmegen, Holland, untersucht sprachliche Kategorisierungen in vier kognitiven Domänen in den indogermanischen Sprachen<sup>5</sup>. Eine dieser Domänen ist Farbe. In jeder Sprache wurden je ca. 20 Studierende in einstündigen Interviews zu den verschiedenen Bereichen befragt. Das Durchschnittsalter der deutschen Teilnehmer betrug 21,1, das der dänischen 26,5. Die Farbstudie bestand insgesamt aus drei Aufgaben, von denen eine hier präsentiert und besprochen werden soll. Den Teilnehmern wurden dabei 84 Stimuli in Form von neutralen, grauen Chips gezeigt, auf denen je eine andere Farbe abgebildet war, die durch das Munsell Color Systems gekennzeichnet waren<sup>6</sup>. Dabei wurde die Frage gestellt „Was für eine Farbe ist das?“. Dieselbe Methode wandten Kay et al. (2009: 13–14) in ihrer „World Color Survey“<sup>7</sup> an, jedoch mit 330 Farbchips. Bei einer Zusammenstellung der Studienergebnisse aus den teilnehmenden germanischen Sprachen (neben Deutsch noch Schwedisch, Norwegisch, Dänisch, Isländisch, Englisch, Luxemburgisch, Niederländisch und Friesisch) zeigte sich aber eine Auffälligkeit: *pink* kam als Antwort nur im Dä-

---

4 Übersetzung von Nicoline Munck Vinther (Kopenhagen).

5 Majid, Asifa, Jordan, Fiona & Dunn, Michael. 2011. *Evolution of semantic systems procedures manual*. Nijmegen: Max Planck Institute for Psycholinguistics.

6 Das Munsell Color System teilt Farben gemäß Farbton, Sättigung und Helligkeit ein: <http://munsell.com/about-munsell-color/how-color-notation-works/> (aufgerufen am 9.2.13).

7 <http://www1.icsi.berkeley.edu/wcs/> (aufgerufen am 13.2.13).

nischen und Deutschen vor<sup>8</sup>, nicht in den übrigen nord- und westgermanischen Sprachen, mit Ausnahme natürlich des Englischen. Diese auffällige Verteilung veranlasste uns dazu, ausgehend von der EoSS-Studie die Farbwortentlehnung genauer zu untersuchen.

### 1.2. Bisherige Literatur zu *pink*

Die Tatsache, dass *pink* als Lehnwort aus dem Englischen ins Deutsche eingetreten ist, wurde schon in zwei Arbeiten behandelt, welche bezüglich des Status seiner semantischen Integration zu unterschiedlichen Schlüssen kommen. Zum einen wird das Farbadjektiv aus semantischer Perspektive ausführlich untersucht in einer Dissertation von 2006 zur Semantik von *rosa*, *pink* und *rot* (Kaufmann 2006). Diese korpusbasierte Arbeit stellt letztendlich die Hypothese auf, dass *pink* ein Hyponym zu *rosa* darstellt (Kaufmann 2006: 38). Der zweite Beitrag, eine Magisterarbeit, wertet in einem Fragebogen sowohl Farbwörter generell als auch Assoziationen zu den Farbtönen *rosa* und *pink* und Angaben über den Gebrauch der jeweiligen Begriffe aus (Frenzel 2006)<sup>9</sup>. Obwohl Claudia Frenzel ihre Arbeit im gleichen Jahr wie Caroline Kaufmann veröffentlichte, kommt sie zu dem Schluss, dass *pink* und *rosa* im Deutschen zwei verschiedene Farben sind (Frenzel 2006: 77 und Frenzel-Biamonti 2011). Beide Arbeiten sind wichtige Stützpunkte jeder weiteren Untersuchung des Themas, so auch vorliegender Darstellung. Soweit bekannt, ist dagegen das Phänomen im Dänischen noch nicht erforscht worden.<sup>10</sup>

## **2. *Pink* in EoSS**

Im Folgenden werden die Ergebnisse der EoSS-Studie hinsichtlich des Farbworts *pink* möglichst ergiebig beschrieben und analysiert.

### 2.1. Verteilung

In der deutschen Studie des EoSS-Projekts wurde *pink* von allen 20 Teilnehmern genannt. Es wurde auf insgesamt 11 Stimuli in 71 Fällen angewandt. Daraus ergibt sich folgende Verteilung, hier schematisch dargestellt<sup>11</sup>:

---

8 Und im Schweizerdeutschen, was für diese Zwecke nicht extra behandelt werden muss.

9 Mein herzlicher Dank gilt Claudia Frenzel-Biamonti für das Zusenden ihrer Arbeit, die unter der Internetadresse [http://www.tu-chemnitz.de/phil/english/ling/research\\_student\\_projects.php](http://www.tu-chemnitz.de/phil/english/ling/research_student_projects.php) leider nur unvollständig abrufbar ist.

10 Für muttersprachliche Einschätzungen danke ich deshalb befreundeten dänischen Studenten und Studentinnen aus Aarhus und Kopenhagen; für linguistische Unterstützung außerdem meinem Kollegen Carsten Levisen, Universität Aarhus.

11 Alle Farbtabellen (= Abbildungen 1-7) sind modifizierte Versionen der Tabelle der World Color Survey (über University of California at Berkeley, <http://www1.icsi.berkeley.edu/wcs/images/jrus-20100531/wcs-chart-4x.png> (aufgerufen am 9.2.13)). Sie stammen aus der unveröffentlichten EoSS-Datenbank.





Zwar sind es die gleichen Chips wie in der deutschen Studie, die am häufigsten unter den 84 als *pink* bezeichnet wurden, aber die Anzahl der Antworten ist mit 6 - 11 deutlich geringer. Die Verteilung des Farbworts im Dänischen hat jedoch im Vergleich zum Deutschen klarere Grenzen, da bei anderen Farbchips *pink* nicht besonders häufig genannt wurde, während es im Deutschen auch bei Chips wie A19, B1 und C18 noch relative hohe Frequenzen aufzeigt. Insgesamt ist *pink* in der dänischen Studie deutlich weniger repräsentiert als in der deutschen.

## 2.2. Modifikation der Antworten

Als weiterer Punkt aus der Analyse der EoSS-Daten soll die Häufigkeit und Art der Modifikation des Farbworts in den Antworten der Teilnehmer herangezogen werden. In beiden Sprachen kam es bei der Benennung der Chips häufig vor, dass Teilnehmer nicht nur eine Farbe nannten, sondern diese auf verschiedene Arten genauer beschrieben. Damit kamen sie dem Gefühl nach, besonders genau, aber auch abwechslungsreich bei der Benennung der großen Anzahl von unterschiedlichen Farbchips vorgehen zu müssen. Auch *pink* wurde in einigen Fällen modifiziert, jedoch wiederum deutlich mehr von den deutschen Teilnehmern als den dänischen, nämlich in 34% aller Antworten (= 24 von 71<sup>12</sup>) in der deutschen Studie und 11% (= 4 von 38) in der dänischen. Vor allem die Art und Weise, wie *pink* von den Teilnehmern der beiden Studien genauer beschrieben wurde, sagt einiges darüber aus, wie sehr das Farbwort schon in den Farbwortschatz der Sprecher integriert ist.

Im Deutschen können verschiedene Eigenschaften<sup>13</sup> des Farbtons modifiziert werden. So beziehen sich Modifikationen wie *helles Pink* und *dunkles Pink* auf die Helligkeit des Farbtons, *mattes Pink* und *grelpink* auf seinen Glanz, und die Komposita *pinklila*, *pinkviolett*, *rosapink* etc. bezeichnen eine Mischung aus zwei Farbtönen zu gleichen Teilen. In einigen Fällen wurden auch Farbwörter koordiniert, wie etwa *rot und pink* oder *Rosa-Pink-Gemisch*, wobei eine gleiche Mischung beider Farbtöne beschrieben wird. Bei den Beispielen aus der Studie *pink mit orange* oder *lila mit Pinkstich* dagegen überwiegt der erstgenannte Farbton in der Mischung, der zweite ist zu geringeren Anteilen enthalten.

In der dänischen Studie wurde die Modifikation *meget pink* genannt, die sich wohl auf die Sättigung bezieht. Als einfacher *intensifier* ist *meget* 'sehr, viel' aber diesbezüglich nicht eindeutig. *Mørk pink*, 'dunkles Pink', dagegen beschreibt eindeutig die Helligkeit des Farbtons. Bei *hudfarvet hen mod pink* 'hautfarben in Richtung pink' dient *Pink* als modifizierender Farbton von *hudfarvet* 'hautfarben'

Die Unterschiede zwischen den beiden Sprachen sind deutlich. Im Deutschen ist es möglich, die

---

12 Gezählt wurden jeweils die Antworten mit [=entspricht *pink*] gegenüber [=enthält *pink*].

13 Für eine genaue Einteilung der Parameter siehe Biggam 2012: 3-5.

Qualität des Farbtons *Pink* auf praktisch jede Art zu beschreiben, auf welche auch die von Grundfarbwörtern wie *rot*, *grün* und *blau* beschrieben werden. Im Dänischen ist dies nur minimal möglich, vor allem aber jeweils nur in einem Bereich der Skala: Die Helligkeit kann nur als dunkel beschrieben werden, die Sättigung nur als intensiv (*meget*). Im Deutschen dagegen kann *pink* nicht nur dunkel oder grell sein, sondern auch hell und matt. Besonders auffällig ist aber, dass *pink* im Deutschen mit mehreren anderen Farbtönen kombiniert bzw. gemischt werden kann: Lila, Violett, Rot, Rosa und Orange. Das zeigt zum einen natürlich die Nähe dieser Farbtöne im Spektrum auf, da Mischfarben meist aus ähnlichen Farbtönen generiert werden (z.B. gibt es *blaugrün*, aber nicht *rotgrün* als Farbbeschreibung). Zum anderen macht es sehr deutlich, dass *pink* im Deutschen den Status eines in Denotation und morphologischer Freiheit unabhängigen Farbworts hat – gleich Grundfarbwörtern wie *rot* und sekundären Farbwörtern wie *türkis* oder *orange*. Besonders die Antwort *rosapink* einer Teilnehmerin macht aber auch die Abgrenzung zu *rosa* deutlich: Wäre *pink* ein Hyponym zu bzw. eine Unterkategorie von *rosa* (so wie es noch Kaufmann in ihrer Studie anführt (2006: 38)), könnte es nicht in einem Kopulativkompositum an zweiter Stelle neben *rosa* stehen, da diese immer aus der Zusammenstellung zweier semantisch gleichwertiger Konstituenten bestehen. So nennt z.B. Motsch als Bedingung für Kopulativkomposita: „Die mit dem Erstglied verbundene Bedeutung ist kein Modifikator der Bedeutung des Zweitglieds“ (2004: 270)<sup>14</sup>.

So konnten anhand der EoSS-Daten einige wichtige Unterschiede zwischen dem Dänischen und Deutschen bezüglich Verteilung und morphologischer Integration von *pink* aufgefunden werden, obwohl die Menge an verwertbaren Daten nur zu einem bestimmten Grad repräsentativ sein kann.

Ausgehend von der reinen Datenanalyse sollen nun einige semantische Aspekte des „neuen Pinks“ untersucht werden.

### 3. Die Entlehnung aus dem Englischen

Höchstwahrscheinlich geschah die Entlehnung von *pink* aus dem englischen Farbwortschatz in den Deutschen und Dänischen schon nach dem 2. Weltkrieg (siehe Koller 2008: 420 En<sup>15</sup> 11), als der Einfluss vor allem der USA wirtschaftlich sowohl kulturell in der Form einsetzte, wie wir ihn auch heute kennen. Damit einher ging natürlich auch der sprachliche Einfluss des Englischen auf die europäischen Sprachen. So Muhr (2002: 9–10): „Seit Beginn der 60er Jahre des 20. Jhds. lässt sich die verstärkte Zunahme von Anglizismen im Deutschen feststellen, die sich schließlich in den 90er Jah-

---

14 *Magentarot* z.B. wäre dagegen ein Determinativkompositum, da *rot* ein Grundfarbwort und *magenta* ein sekundäres Farbwort ist, die Konstituenten also semantisch nicht gleichwertig sind. *Magenta* modifiziert die Bedeutung von *rot*.

15 En = Endnote.

ren enorm verstärkt“. Denn nicht nur die Anzahl an Englischsprechern sei gestiegen, sondern Englisch dominiere auch „hinsichtlich des kulturellen Einflusses und der Stellung als internationale Sprache gegenüber anderen Sprachen“. Es ist anzunehmen, dass dieser Prozess auf das Dänische übertragbar ist, so dass in diesem Zusammenhang auch die Entlehnung von *pink* in die beiden Sprachen gesehen werden kann.

Aufgrund seiner phonologischen Struktur ist *pink* im Deutschen und Dänischen sowohl phonetisch als auch orthographisch vollständig integriert. Die morphologische Integration wurde schon unter 2.2. angesprochen. Hinzuzufügen ist noch, dass *pink* mittlerweile wie ein deutsches Adjektiv flektiert werden kann (z.B. *das pink-e Kleid, eines pink-en Fahrrads*). Lediglich derivative Beschränkungen gibt es, z.B. ist *\*pink-lich, \*pink-ig* nicht möglich, genauso wie auch *\*orange-ig* oder *\*rosa-lich* nicht gebildet werden können. Im Dänischen kann *pink* standardsprachlich noch nicht flektiert werden, umgangssprachlich kommt z.B. *den pink-e kjole* durchaus vor.

Im Englischen ist *pink* ein basic colour term (u.a. Berlin & Kay 1991: 94). Die Konnotationen von *pink* mit allem Mädchen- und Frauenhaften sind auch im Englischen vorhanden (Koller 2008). Allerdings umfasst das Farbwort im Englischen einen deutlich größeren Bereich des Spektrums als im Deutschen und Dänischen. Dies ist gut anhand der EoSS-Daten erkennbar<sup>16</sup>:

#### Modal response

white	A	pink	peach	peach	yellow	yellow	yellow	green	green	green	green	green	green	blue	blue	blue	blue	purple	pink/purple	pink	pink	pink
grey	B	pink	orange	orange	yellow	yellow	green	green	green	green	green	green	green	blue	blue	blue	blue	purple	purple	pink	pink	pink
grey	C	red	orange	brown	brown	green	green	green	green	green	green	green	green	blue	blue	blue	blue	purple	purple	purple	pink	pink
grey	D	maroon/red	brown	brown	brown	green	green	green	green	green	green	green	green	blue	blue	blue	blue	purple	purple	purple	purple	purple
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Abbildung 5: Häufigste Antworten für die jeweilige Farbchips im Englischen

Die Anzahl der Farbtöne und Helligkeiten, die im Englischen mit *pink* bezeichnet werden können, ist im Vergleich zum Deutschen und Dänischen viel größer und es ist kein Fokus auf eine besondere Schattierung des Farbbereichs erkennbar. Als eines der fünf Farbörter unter den 11 basic colour terms des Englischen, die sogenannte „gemischte Farben“ bezeichnen (d.h. *orange, pink, purple, grey, brown*) lässt sich *pink* also relativ simpel als „weiß + rot“ beschreiben (s. Wierzbicka 1990: 136). Alle Farbtöne und –schattierungen, die im Englischen mit *pink* bezeichnet werden, haben unterschiedliche Anteile von Weiß und Rot. Bei der Entlehnung des Farbwortes ins Deutsche und Dänische ist nun seine Extension eingeschränkt worden: *pink* bezeichnet dort nur diejenigen Schattierungen, die einen hohen Rot- und geringen Weißanteil haben, wie in den Abbildungen 1-4 oben erkennbar ist. Im Vergleich zum Englischen haben das deutsche und das dänische *pink* demnach einen beschränkten Bedeutungsumfang. Der restliche Teil des durch von den englischen Muttersprachlern mit *pink* abgedeckten Farbbereichs wird im Deutschen und Dänischen mit eigenen, älteren Farb-

<sup>16</sup> Es handelt sich hierbei um Daten des Britischen Englisch, die Interviews wurden in London durchgeführt.

wörtern bezeichnet, die im Folgenden untersucht werden sollen.

#### 4. Der Farbwortschatz des Dänischen und Deutschen nach dem Eintritt von *pink*

Bevor *pink* in die beiden Sprachen eintrat, deckten autochthone Farbwörtern den gesamten chromatischen Bereich ab, der aus „rot + weiß“ zusammengemischt ist. Im Deutschen ist dies *rosa*, im Dänischen *lyserød*.

*Rosa* ist ein relativ junges Farbwort des Deutschen: seine Verwendung als Adjektiv etablierte sich ca. im 18. Jahrhundert (Grimm 1854-1954: 1161). Es hat jedoch den Status eines basic colour terms des Deutschen und war bis zur Entlehnung von *pink* das einzige Grundfarbwort für den Farbbereich zwischen rot und weiß. Um ein etwas kräftigeres Rosa zu bezeichnen, wurde teilweise das Determinativkompositum *rosarot* verwendet. Heute ist dieses kaum mehr gebräuchlich, aber es ist z.B. in Phraseologismen wie „durch die rosarote Brille sehen“ erhalten (Heller 1989: 121). Ein berühmter Fall zeigt außerdem, dass *rosarot* zur Übersetzung von dem englischen *pink* diente: Die Zeichentrickfigur aus der Serie *The Pink Panther* heißt auf Deutsch *Der rosarote Panther*.

Seit dem Eintreten von *pink* in den deutschen Farbwortschatz wurde *rosa* marginalisiert: Es bezeichnet nun meist nur noch helle Farbtöne, die entstehen, wenn man Rot mit viel Weiß mischt. *Pink* dagegen bezeichnet die dunkleren Farbtöne mit wenig Weiß- aber eventuell auch Blauanteil. Nach Wierzbicka (1990) könnte man also *rosa* als „rot + **weiß**“ und *pink* als „**rot** + weiß“ beschreiben. Um diesen Zustand zu verdeutlichen, sollen noch einmal die Farbpaletten aus dem EoSS-Projekt angeführt werden:

##### **rosa**

	A	12	3	1													3	12	16	18	
	B	8															1	4	7	16	
	C																	1		2	
	D																				
0		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Abbildung 6: Antworten mit *rosa* im Deutschen

Im Vergleich zu der Farbpalette mit *pink* sieht man deutlich die Konzentration von *rosa* auf hellere Farbtöne in den oberen Zeilen der Tabelle. Chip C19 z.B. erhielt gar keine Antworten mit *rosa*, aber 15 mit *pink* (siehe oben 2.1.). Auch bei B19 und C20 sieht man, dass *pink* *rosa* ergänzt (insgesamt ergeben die Zahlen 20, was der Anzahl der Teilnehmer entspricht), die beiden Farben aber in den Antworten nicht vermischt werden.

Auch die Konnotation von *rosa* wurde eingeschränkt. Stand früher *rosa* an sich für alles Weibliche (Heller 1989: 115, 118), kann jetzt schon unterschieden werden zwischen *rosa* als Farbe für alles Kindlich-Mädchenhafte und *pink* als mutigeres, selbstbewussteres Äquivalent für Mädchen im Tee-

nageralter. Junge Frauen färben sich die Haare als „Statement“ pink, nicht rosa, ebenso lackieren sich modische Frauen die Nägel pink. Wird ein neues elektronisches Gerät oder ein Einrichtungsmöbel im extravaganten Design entworfen, könnte es pink sein – rosa wahrscheinlich eher nicht.

Doch die Farbe der Homosexuellen ist Rosa (siehe z.B. die Wählerinitiative „Rosa Liste“ der Homosexuellen in München) nicht Pink (anders als im Dänischen, siehe unten). Neugeborene Mädchen werden in rosa gekleidet. *Pink* ist knallig und laut, *Rosa* ist zart und leise<sup>17</sup>.

Die autochthone dänische Entsprechung ist das Kompositum *lyserød*<sup>18</sup>. Wörtlich genommen ist *lyserød* das Übersetzungsäquivalent zu Deutsch *hellrot* (*lys* ‘hell’, *rød* ‘rot’). Während *hellrot* aber ein modifiziertes Rot ist, scheint *lyserød* nicht mehr kompositionell verstanden zu werden sondern die Funktion eines eigenständigen Farbworts zur Repräsentation des Farbbereichs zu haben, den im Deutschen *rosa* abdeckt.<sup>19</sup>

Die EoSS-Daten sollen auch hier zur Verdeutlichung dienen<sup>20</sup>:

A	12	2														6	16	18	17	lyserød	
B	10																5	11	19	lyserød	
C																	1	6	6	lyserød	
D																				lyserød	
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	lyserød

Abbildung 7: Antworten mit *lyserød* im Dänischen

Es ist klar erkennbar, dass *lyserød* in der Darstellung zu Großteilen mit dem deutschen *rosa* übereinstimmt, was seine Ausbreitung auf dem Farbspektrum betrifft. Wie zu vermuten, ist aber auch seine Abgrenzung zu *pink* nicht so deutlich. Wo *pink* am häufigsten genannt wurde (Chip C20), ist *lyserød* trotzdem ähnlich häufig.

Was die Semantik des neuen Farbworts angeht, so muss betont werden, dass die visuelle Auffälligkeit (*visual conspicuousness*, Vejdemo et al. 2013) eines knalligen Pinktons ausschlaggebend ist. So lässt sich z.B. beobachten, dass ein Fahrrad, das farblich unter anderen Fahrrädern heraussticht, eher *pink* als *lyserød* genannt wird, unabhängig von seinem genauen Farbton (Carsten Levisen, persönliche Mitteilung).

Auch ist die Verwendung von *pink* noch mehr kontextabhängig auf Kleidung und Kosmetik etc. bezogen. Laut dem Wörterbucheintrag des *Den Danske Ordbrog* ist *pink* „af en kraftig og klar viole-

17 Siehe Kaufmann 2006 für eine ausführliche Analyse dazu.

18 Das Farbwort *rosa* existiert auch im Dänischen, kam jedoch in den EoSS-Daten nicht vor.

19 Dagegen ist z.B. das schwedische Kognat zu *lyserød*, *ljusröd*, kompositionell wie Deutsch *hellrot*.

20 Da beim Eintragen der Interviewdaten der dänischen Studie *lyserød* als Unterkategorie von *rød* behandelt werden musste (die genannten Erkenntnisse ergaben sich erst im Nachhinein), wurden *lyserød* und *rød* nicht mit eigenen Tabellen wie den oben abgebildeten versehen. Die hier gezeigte ist eine eigene Zusammenstellung von Carsten Levisen.

trød eller lyserød farve især om farve på tøj“ ‘von einer kräftigen und klaren violett-roten oder rosa-nen Farbe, besonders als Farbe für Kleidung‘<sup>21</sup>.

Eine Assoziation von *pink*, die seit der Entlehnung aufgekommen ist, ist die zur homosexuellen Szene. Ein Beispiel von Levisen (2012b): „I lørdags åbnede Danmarks første popup gaydisco, PINK“. ‘Am Samstag eröffnete Dänemarks erste Schwulendisko, PINK.’

Da es noch keine Publikationen zu *pink* im Dänischen gibt, muss es bei dieser knappen Darstellung zur Semantik bleiben.

## 5. Zusammenfassung und Ausblick

In dieser Arbeit wurde das Farbwort *pink* im Deutschen, Dänischen und Englischen aus mehreren Perspektiven betrachtet. Zum einen wurde, ausgehend von den Daten des EoSS-Projekts, festgestellt, welche Verteilung *pink* in den drei Sprachen hat. Dabei wurde eine dreiteilige Abstufung festgestellt: Die beschränkteste Verteilung hat *pink* im Dänischen, die weiteste im Englischen. Demnach ist *pink* im Englischen auf eine viel größere Zahl von Farben anwendbar als im Deutschen und Dänischen, wo die Extension und Semantik des Farbworts bei seiner Entlehnung eingeschränkt wurde. Des Weiteren wurde *pink* sowohl phonetisch als auch morphologisch an die jeweiligen Systeme angepasst, wobei dies im Deutschen auch standardsprachlich, im Dänischen bisweilen nur umgangssprachlich durchgesetzt ist. Auch die morphologische Varianz bei den Modifikationen von *pink* als Farbadjektiv ist im Deutschen deutlich größer, wie die Analyse der EoSS-Daten ergab.

Diese Punkte lassen darauf schließen, dass *pink* besser in den deutschen (Farb-) Wortschatz integriert ist als in den dänischen. Da leider für das Dänische keine entsprechende Zeitangabe zu finden war, kann nur spekuliert werden, ob diese Tatsache auf eine spätere Entlehnung von *pink* aus dem Englischen zurückzuführen ist.

Im Gegensatz zum Englischen haben Deutsch und Dänisch nun zwei Wörter, um auf den Farbbereich „rot + weiß“ zu referieren: Das neue *pink* und die jeweiligen autochthonen Begriffe. Beide, *rosa* und *lyserød*, wurden in ihrer Funktion, alleine diesen Farbbereich abzudecken, eingeschränkt und beschreiben nun meist nur noch hellere Farbtöne. Dazu kommt, dass in beiden Sprachen die neuen Farbwörter auch je eine eigene Semantik entwickelt haben. Die visuelle Auffälligkeit von *Pink* als ein knalliger, kräftiger Farbton wird auf seine Verwendung als Farbe der rebellischen Jugend, des kreativen und innovativen Designs und des modernen Lifestyles (nach dem Motto „Think Pink!“) übertragen. Es hat demnach nur einen Teil der ursprünglichen Semantik des englischen *pinks* behalten. Dies ist typisch für Lehnwörter, deren Semantik meist nur zu Teilen der in der Aus-

---

21 <http://ordnet.dk/ddo/ordbog?query=pink&tab=for> (aufgerufen am 23.05.13); eigene Übersetzung.

gangssprache entspricht (Schmidlin 2003: 145<sup>22</sup>).

*Pink* gegenüber stehen *rosa* bzw. *lyserød*, die wiederum ihre eigene Semantik haben, welche beim Eintreten von *pink* eingeschränkt wurde. Zumindest für das Deutsche kann festgestellt werden, dass sich die Zartheit und Blassheit von *rosa* in seiner Bedeutung als sanfte, mädchenhafte, leise Farbe widerspiegelt. Da der Gegensatz zwischen *pink* und *rosa* im Deutschen nicht nur auf der Ebene der Zusammensetzung der Farben, sondern auch in ihrer jeweiligen Semantik so deutlich ist, komme ich zu dem Schluss, dass *pink* und *rosa* im Deutschen zwei unterschiedliche Farben bezeichnen, nicht jedoch in einem Verhältnis von Hyponymie stehen, womit ich der Hypothese von Frenzel (2006) und Frenzel-Biamonti (2011) zustimme. Ausgehend davon ist auch der in der Einleitung angeführte Beispielsatz plausibel, in dem ein siebenjähriges Mädchen zwischen den zwei Farben ihrer Kleider unterscheidet. Im Dänischen lassen die geringe Salienz und stärkere Kontextgebundenheit von *pink* an dieser Einteilung zweifeln. Es bleibt abzuwarten, ob der Wandel diesbezüglich noch fortschreiten wird.

Diese Arbeit war auf die zwei Sprachen Deutsch und Dänisch im Vergleich mit dem Englischen beschränkt. Auf eurolinguistischer Ebene bliebe noch zu untersuchen, warum *pink* nur ins Dänische und Deutsche entlehnt wurde, aber nicht in andere nahverwandte Sprachen wie das Norwegische, Schwedische oder Niederländische. Hier spielen sicher sowohl linguistische als auch kulturelle Gründe eine Rolle<sup>23</sup>. Mit Hilfe der EoSS-Daten könnte auch ein Überblick über den gesamten europäischen Sprachraum bezüglich der Entlehnung von *pink* gewonnen werden.

Um den Blick noch weiter in die Ferne zu richten, wäre auch ein weltweiter Vergleich möglich, als Ergänzung zu den Untersuchungen von Stanlaw (1997), der *pink* schon im Japanischen entdeckt hat. Denn es besteht Grund zu Annahme, dass *pink* als ein Kernbegriff der Anglisierung und Amerikanisierung mehr als nur die zwei hier untersuchten Sprachen beeinflusst hat.

## 6. Literatur

- Berlin, Brent & Kay, Paul. 1991. *Basic Color Terms. Their Universality and Evolution*. Berkeley and Los Angeles: University of California Press.
- Biggam, Carole. 2012. *The semantics of colour: A historical approach*. Cambridge: Cambridge University Press.
- Carstensen, Broder. 2001. *Anglizismen-Wörterbuch: Der Einfluss des Englischen auf den deut-*

---

22 Schmidlin sieht in der semantischen Eigenständigkeit sogar das „deutlichste Zeichen der Integration“ (ebd.).

23 Zum Beispiel hat Schwedisch ein Farbwort *ceris*, das einen ähnlichen Bereich des Farbspektrums abdeckt wie deutsch/dänisch *pink*, und deshalb evtl. einer Entlehnung des englischen Begriffs im Wege stand. Isländisch dagegen ist vielleicht zu sprachkonservativ, um ein englisches Farbwort zu integrieren.



- schen Wortschatz nach 1945*. Band 3 (P-Z). Berlin/New York: de Gruyter.
- Frenzel, Claudia. 2006. *Is 'pink' the death of 'rosa'? A critical assessment of the endpoint of colour-term evolution*. Magisterarbeit, Technical University of Chemnitz. [http://www.tu-chemnitz.de/phil/english/ling/research\\_student\\_projects.php](http://www.tu-chemnitz.de/phil/english/ling/research_student_projects.php)
- Frenzel-Biamonti, Claudia. 2011. „Rosa Schätze – Pink zum kaufen: Stylistic confusion, subjective perception and semantic uncertainty of a loaned colour term.“ In: Carole P. Biggam et al. (Hrg.): *New directions in colour studies*. 91–103.
- Grimm, Jacob & Wilhelm. 1854–1954. *Deutsches Wörterbuch*. 16 Bde. Leipzig. Nachdruck in 33 Teilbänden: München 1984. [Teilbd. 14]
- Heller, Eva. 1989. *Wie Farben wirken*. Hamburg: Rowohlt.
- Kay, Paul et al. 2009. *The world color survey*. Stanford: CSLI Publications.
- Kaufmann, Caroline. 2006. *Zur Semantik der Farbadjektive 'rosa', 'pink' und 'rot'*. München: Herbert Utz.
- Koller, Veronika. 2008. „‘Not just a colour’: pink as a gender and sexuality marker in visual communication“. *Visual Communication* 7 (4). 365–423.
- Majid, Asifa, Jordan, Fiona & Dunn, Michael. 2011. *Evolution of semantic systems procedures manual*. Nijmegen: Max Planck Institute for Psycholinguistics.
- Motsch, Wolfgang. 2004. *Deutsche Wortbildung in Grundzügen*. Berlin/New York: Walther de Gruyter. 2. Aufl.
- Muhr, Rudolf. 2002. „Anglizismen als Problem der Linguistik und Sprachpflege in Österreich und Deutschland zu Beginn des 21. Jahrhunderts“. In: Rudolf Muhr & Bernhard Kettermann (Hrg.): *EUROSPEAK – Der Einfluss des Englischen auf europäische Sprachen zur Jahrtausendwende*. Wien: Peter Lang. 9–54. 2. Aufl.
- Schmidlin, Regula. 2003. „Vergleichende Charakteristik der Anglizismen in den standardsprachlichen Varietäten des Deutschen“. In: Beat Dittli et al. (Hrg.): *Gömmers MiGro? Veränderungen und Entwicklungen im heutigen Schweizer Deutschen*. 141–160.
- Stanlaw, James. 1997. „Two observations on culture contact and the Japanese color nomenclature system“. In: Clyde L Hardin & Luisa Maffin (Hrg.): *Color categories in thought and language*. 240–260.
- Vejdemo, Susanne, Levisen, Carsten, van Scherpenberg, Cornelia, Guðmundsdóttir Beck, Þorhalla, Næss, Áshild, Zimmermann, Martina, Stockall, Linnaea, Whelpton, Matthew. 2014. „Two Kinds of Pink: Development and Difference in Germanic Colour Semantics“. Im Druck.

# The representation of women and men in English idioms: a corpus-based study

Karin Wiecha (Universität Potsdam)

## 1. Introduction

Much of the language used daily consists of relatively fixed, prefabricated idiomatic expressions. Boers and Lindstromberg (2009: 2) estimate about half of English written discourse to be composed of fixed expressions with very likely an even higher percentage in spoken discourse. As Cacciari and Tabossi (1993: xi) point out, idiomatic expressions usually originate from and are based on conventions, shared beliefs and cultural models of a certain speech community. A closer look at the idioms used by a speech community may, therefore, reveal underlying biases and stereotypes shared by its speakers and conveyed by fixed expressions such as idioms.

A very productive field for such implicit biases is the representation of women and men<sup>1</sup> in idiomatic expressions. This specific aspect of idiomaticity is the main focus of this paper. The intention of this study is to reveal the way women and men are represented in English idioms on the basis of frequency and connotations of idioms related to women, men or both genders. The database used for obtaining the information for this study consists of three idiom dictionaries and the *British National Corpus*.

## 2. Definition

Before going into greater detail about the nature and use of gender-related idioms a definition of the term *idiom*<sup>2</sup> as it is used here is required. The prototypical example usually mentioned when describing idioms is to *kick the bucket* (e.g. Everaert et al. 1995: 2, Gibbs 1995: 97, Tabossi & Zardon 1995: 273). There are, however, as Prodromou (2008: 46f.) suggests, a wide range of fixed expressions also qualifying as idioms, ranging from simple every-day phrases like *I don't know* or *you know* to longer fixed expressions like proverbs, song texts, or limericks. He distinguishes between four main types of idioms: (A) formulae, (B) collocations, (C) phrasal verbs, and (D) cultural idioms.

This study focuses on type D (*cultural idioms*)<sup>3</sup>. As Prodromou (2008: 48) remarks, cultural idioms

1 I am aware that the gender-binary is to be viewed very critically at the least. For the sake of this paper I will refer to two genders – female and male – nonetheless. This is not to discriminate against individuals who might not fall into either of these categories, but due to the linguistic nature of my research and the fact that the English grammar distinguishes between these two genders only, as do the idioms that are the focus of this study.

2 For the purpose of this paper, the terms *idiom* and *idiomatic expression/phrase* are used synonymously.

3 Under this heading Prodromou (2008: 47) subsumes 22 different kinds of idioms: (1) colourful binomials; (2) co-

are “deeply embedded in the socio-cultural norms of a particular speech community” and are “in various degrees, culturally shaped”. This kind of idiom, hence, qualifies well for an analysis of gender-related biases in a given speech community.

### 3. Frequency of female- and male-related idioms

Many idioms are built around the word *man*: *man of means*, *man enough to*, *make a man (out) of*, or *a man of few words*, just to name a few. In many of these phrases, the word *man* can, theoretically, be replaced by *woman*. The question is whether such options are suggested in dictionaries of idioms and how frequent they are in actual language use. This is the focus of the following sections. Section 3.1 takes a look at three dictionaries of idioms, whereas Section 3.2 observes the frequency of a sample of gender-related idioms in the *British National Corpus* (BNC).

#### 3.1 Frequency in idiom dictionaries

All idioms, referring to either males, females, or both, listed in the *Dictionary of Idioms* (Fowler 1986), *English Idioms* (Seidl & McMordie 1989), and the *Oxford Dictionary of English Idioms* (Cowie, Mackin & McCaig 1996) were counted. Not only idioms using the words *man/men* and *woman/women*, but all possible terms referring to gender such as *boy/girl*, *lady/lord*, *king/queen*, *mother/father*, *Mr/Mrs*, and so on were included. Idioms not containing any of such terms but clearly gender-related in meaning were also counted, as in the saying *the hand that rocks the cradle (rules the world)* which is meant to refer to mothers, as Cowie et al. (1996: 249) specify in the dictionary entry.

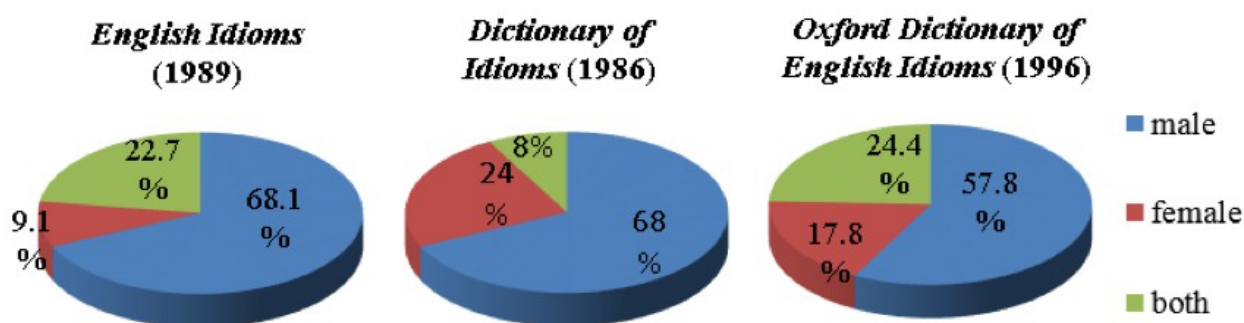


Figure 1: Frequency of idioms referring to males, females, and both genders in three different dictionaries of idiom

Figure 1 shows the frequency of gender-related idioms in the examined dictionaries. The average

- 
- lourful trinomials; (3) turns of phrase/verbal idiom; (4) metaphors, figurative; (5) similes; (6) proverbs and sayings; (7) quotations; (8) cultural allusions; (9) catchphrases, slogans; (10) slang phrases; (11) taboo expressions;
  - (12) mottoes; (13) understatement; (14) hyperbole; (15) pop-song lyrics; (16) nursery rhymes; (17) rhyming slang; (18) limericks; (19) lexicalised clauses; (20) dyadic discourse routines; (21) variable idioms; and (22) creative idiomaticity, puns.

percentages of all relevant dictionary entries in these dictionaries of idioms combined are: 59.6% male-related, 17.7% female-related, and 22.7% related to both genders. In all of the three examined idiom dictionaries entries referring to males are dominant. In the *Dictionary of Idioms and English Idioms*, the percentages for this category are almost equal. In the *Oxford Dictionary of English Idioms*, which is the most recent one, the percentage of male-related idioms is slightly lower, but still higher than 50%.

It is important to note, that, compared to the other two dictionaries, the *Oxford Dictionary of English Idioms* is considerably larger in volume as well as in number of gender-related idioms. The *Dictionary of Idioms and English Idioms* only list a total amount of 25 and 22 gender-related idioms, respectively, whereas the *Oxford Dictionary of English Idioms* lists 216 in total. These counts cannot be seen as representative for the gender ratio of all gender-related idioms of the English language (since the analysed dictionaries of idioms cannot and do not make the claim to be exhaustive). Nevertheless, the male bias is obvious in all three dictionaries.

The question arises as to why the authors of the respective idiom dictionaries list some of the gender-related idioms in a form that only refers to one gender, while other idioms, however, appear in forms possible to apply to both genders. Why, for instance, does the *Oxford Dictionary of English Idioms* list the idiom *one of the boys* only in the male form (Cowie et al. 1996: 438), whereas for the idiom, *jobs for the boys*, the female equivalent *jobs for the girls* is also mentioned (Cowie et al. 1996: 325)? As Cowie et al. (1996: viii) point out, their *Oxford Dictionary of English Idioms* is based on a self-compiled corpus. This suggests that they based such entries on their occurrence or non-occurrence and respective frequencies in their corpus data. This is tested in Section 3.2 with the help of the *British National Corpus*.

### 3.2 Frequency in the *British National Corpus*

In the following, the results of a corpus query of some examples of gender-related idioms that can refer to both genders are presented. All occurrences of the queried lemmas were manually filtered so that only the occurrences in which it can be derived from the context that the phrases are used idiomatically are included in Table 1. Sentences like (1) or (2)<sup>4</sup> refer to the literal and not the figurative, idiomatic meaning of the phrase queried and were therefore disregarded as were those hits for which the context gave no hints as to whether the phrase was used idiomatically or literally.

(1) We had better torment our **own man** next.

(2) There's the **girl next door**, she's just had a baby.

---

4 Both sentences are taken from the *British National Corpus*; emphasis added.

#	Idioms	hits		
		total	female	male
1	family man/woman	92	2	90
2	own man/woman	86	9	77
3	man/woman of letters	60	3	57
4	man/woman of the world	53	13	40
5	man/woman of action	37	-	37
	self-made man/woman	37	2	35
6	marked man/woman	30	2	28
7	dirty old man/woman	20	-	20
8	whipping boy/girl	18	-	18
9	girl/boy next door	17	15	2
	man/woman of few words	17	1	16
10	man/woman enough to	15	1	14
11	Renaissance man/woman	10	-	10
	kept woman/man	10	7	3
	man/woman of means	10	1	9
12	earth mother/father	8	8	-
	man/woman of his/her word	8	1	7
13	make an honest woman/man	6	5	1
<b>sum of hits</b>		534	70	464
<b>percentage</b>		100	13.11	86.89

Table 1: Frequency of selected gender-related idioms in the British National Corpus

Table 1 shows the frequencies for both gender variants of each idiom of a sample of 18 randomly selected gender-related idioms in the *British National Corpus* (BNC). The idioms are ranked by their frequency starting from the most frequent example. The respective rank for each idiom is numbered in the very left column (#). The frequency according to which they are ranked refers to the total number of occurrences (female and male combined) for each idiom. Some idioms produced the same number of total occurrences and, therefore, share one rank. Each idiom is listed in the female and male form, the dominant gender form (with the most occurrences) of each idiom is mentioned first.

The results support the dictionary findings: the ratio of gender-related idioms seems to be clearly in favour of the male variants. 4 out of the 18 queried expressions (*man of action*, *dirty old man*, *whipping boy*, and *Renaissance man*) do not occur in the BNC in the female form (i.e. *\*woman of action*, *\*dirty old woman* etc.). However, only one of the idiomatic expressions (*earth mother*) produced no BNC hits in the male form. It is true that many idioms can be applied to both genders. However, the corpus results in Table 1 show that the majority of idioms are considerably more frequent in the male form than in the female. Only 4 out of the 18 queried idioms occur more frequently in the fe-

male variant. This result is also supported by Mutlovà's (2009: 59) research who found that "the majority of *man* based idioms have evolved a *woman* based alternative. Yet, the frequency of such alternatives is very low."

The four idiomatic expressions which seem to primarily refer to women are *the girl next door*, *a kept woman*, *earth mother*, and *to make an honest woman (out) of somebody*. The overall frequency of these idioms is, however, quite low compared to the idioms that are mainly used in the male form. Two of the four female-dominated idioms (*earth mother* and *make an honest woman*) are the least frequent of all the queried examples and the most frequent primarily female form (*girl next door*) ranks only in the 9<sup>th</sup> of 13 ranks. Other idioms potentially referring to women, like *to marry her way out of something* (not included in Table 1), did not produce any hits in the *BNC*. The overall percentage of the female variants of all queried idioms is only 13.11%, in comparison to 86.89% for the male equivalents.

The queried idioms constitute only a very small sample from all possible gender-related idioms and, therefore, cannot be seen as representative. Nevertheless, the results clearly imply that there is a twofold imbalance in gender-related idioms, in terms of their frequency. Firstly, there are not as many idioms referring to females as there are for males and, secondly, the relatively few idioms primarily referring to females are less frequently used as idioms referring to males.

#### **4. Connotations of selected gender-related idioms**

In addition to mere frequency, equally relevant, or even more so, is the context in which these idiomatic phrases are used. Concerning the representation of gender in language, the context becomes particularly important, since, as Mills (2008: Foreword) suggests,

there are two forms of sexism – overt and indirect. Overt sexism is clear and unambiguous, while indirect sexism can only be understood contextually in relation to the interpretation of surrounding utterances.

In order to detect possible contextual differences between male-related idioms on the one hand and female-related ones on the other hand, a semantic analysis of the connotations of the corpus results presented in Section 3.2 was conducted. Crystal (2003: 97; emphasis in original) defines *connotation* as

part of a classification of types of MEANING [...] Its main application is with reference to the emotional associations (personal or communal) which are suggested by, or are part of the meaning of, a LINGUISTIC UNIT.

For the purposes of identifying connotations of the samples, the context for each occurrence of every idiom variant queried was examined and each occurrence was manually sorted and designated to one of the four categories: *positive*, *negative*, *neutral*, or *unclear*. The category *unclear* was used for the occurrences where a connotation could not be clearly derived and is, therefore, disregarded in the interpretation of the results.

#	Idioms	connotations							
		positive		negative		neutral		unclear	
		♀	♂	♀	♂	♀	♂	♀	♂
1	family man/woman	2	42	1	12	-	21	-	14
2	own man/woman	7	40	-	10	-	6	2	20
3	man/woman of letters	2	25	-	5	1	16	-	11
4	man/woman of the world	8	29	2	4	-	3	3	4
5	man/woman of action	-	17	-	5	-	6	-	9
	self-made man/woman	1	20	1	5	-	7	-	3
6	marked man/woman	-	3	2	20	-	1	-	4
7	dirty old man/woman	-	1	-	16	-	-	-	3
8	whipping boy/girl	-	3	-	11	-	3	-	1
9	girl/boy next door	8	1	2	1	-	1	2	2
	man/woman of few words	-	4	-	-	1	11	-	1
10	man/woman enough to	-	8	1	6	-	-	-	-
11	Renaissance man/woman	-	6	-	2	-	-	-	2
	kept woman/man	-	-	5	3	-	-	2	-
	man/woman of means	-	5	1	-	-	4	-	-
12	earth mother/father	3	-	3	-	-	-	3	-
	man/woman of his/her word	1	6	-	-	-	-	-	1
13	make an honest woman/man	1	-	2	1	1	-	1	-
<b>sum of hits</b>		33	210	20	101	3	79	13	75

Table 2: Connotation of selected gender-related idioms in the British National Corpus

In Table 2, the idioms are again listed according to frequency. The 4 idioms which occurred more often in the female form are highlighted, the remaining were more frequent in the male form. In addition to showing the frequency of each idiom, the connotations based on the context they occur in is given. Each of the four connotational categories mentioned above is in turn divided into two columns, one for each gender (labeled with ♀ for female and ♂ for male). In each line, at least two boxes are highlighted in a darker grey indicating the most common connotation for each gender variant of the individual queried idioms. Where there was an equal distribution of one form in two categories, both boxes are highlighted (e.g. *self-made woman* which occurred twice in the *BNC* with one hit positively and the other negatively connotated); in these cases three boxes are highlighted.

According to Čermák (2001: 13), “idioms are a primary means for the expression of positive and negative attitudes”. The data presented in Table 2 supports this claim. Only one of the queried idioms (*man/woman of few words*) was primarily used in a neutral context. The other 17 queried expressions, both male and female forms, were predominantly used either with positive or negative connotations. In the analysed samples, the female and male form of each idiom mainly agree in their connotation – the idioms which have positive connotations in the male form are also positive in the female form and vice versa (*self-made woman* and *boy next door* both having the same distribution of negative and positive connotations, though). The only idiom which does not fit this pattern is *man/woman enough to*, which is primarily positively connotated in the male form, whereas the only female occurrence is used in a negative context. The overall picture is that the respective idioms are consistent in their connotations, no matter whether they are used in the female or male form.

Differences in the representation of the genders in the queried idioms can, however, be found when comparing the connotations of each idiom with their frequency in the *BNC*. The 6 most frequent expressions (# 1-5, see Table 2) are primarily positively connotated for both genders. Significantly, all of these idioms are all more frequent in the male form. In comparison, 5 of the 18 queried idioms took predominantly negative connotations, these being: *marked man/woman* (# 6), *dirty old man* (# 7), *whipping boy* (# 8), *kept woman/man* (# 11), and *make an honest woman/man* (# 13). As the ranks in brackets indicate these idioms belong to the less frequent half of the samples. Two of these negatively connotated idioms do not occur in the female form (*dirty old man* and *whipping boy*), one (*marked man/woman*) occurs predominantly in the male form (with 27 occurrences in the male and only 2 in the female form), and two (*kept woman/man* and *make an honest woman/man*) occur more often in the female variant (*kept woman/man* produced 7 hits in the female and 2 in the male form, *make an honest woman/man* produced 5 hits in the female and 1 in the male form). Only 4 out of all the queried idioms produce more hits in the female than the male form. Two of these 4 female-dominated idioms have predominantly negative connotations (*make an honest woman* and *kept woman*). So in the sample analysed in this study half of the female-related idioms are negatively connotated. From the two other female-dominated expressions, *girl next door* had mainly positive connotations and *earth mother* had the same distribution of positive and negative connotations. From the data it can be seen that the female-dominated idioms are divided quite evenly into positively and negatively connotated examples, with a slight tendency to the negative, though. Out of the 14 male-dominated idioms 3 have negative connotations (*marked man*, *dirty old man*, and *whipping boy*), one (*man of few words*) has a neutral connotation and the remaining 10 have predominantly positive connotations. The ratio of positive to negative connotations is thus much more weigh-



hed towards positive connotations for the male-dominated expressions than it is for the mainly female-related idioms.

Another interesting observation can be made about the neutrally connotated occurrences. Only 3 female-related occurrences out of all queried female forms have a neutral connotation. In contrast, 79 of the male forms can be found in this category (see the bottom line of Table 2: *sum of hits*). It seems that the female-related terms tend to be situated on either the positive or the negative end of the connotational cline, whereas many of the references to men are uttered in a neutral context.

These results correspond with Romaine's (2000: 110) corpus findings suggesting that words with negative overtones collocate more often with female than male referents. She concludes that "[w]omen occupy what might be called a problematic or negative semantic space" (Romaine 2000: 103). This view is also supported by male/female word pairs like *bachelor/spinster*, *wizard (warlock)/witch*, or *master/mistress* which, according to Romaine (2000: 111), "may start out on an equal footing, but...become devalued over time". Henley (1987: 4) even speaks of the deprecation of women through the connotations of female-related terms.

## 5. Conclusion

The corpus study conducted in this paper has yielded some pretty clear, unambiguous results concerning the different ways women and men are represented in English idioms. There is a variety of different forms of idioms referring to either women or men or both genders as a count of gender-related idioms in three dictionaries of idioms has shown. The count indicates that there are considerably more male-related than female-related idioms or idioms referring to both. These findings were also supported by the corpus query. The corpus findings not only suggest that there are fewer female- than male-related idioms but also that these occur much less frequently. The semantic analysis of the corpus hits' individual use in context suggested that most gender-related idioms tend to have either positive or negative connotations. The minority of neutrally connotated occurrences are, however, almost exclusively male-related. The majority of the predominantly male-related idioms have positive connotations, whereas the female-dominated idioms are almost evenly divided into positively and negatively connotated examples, with a slight tendency towards the negative.

Even though this study is limited and more research in this area has to be conducted in order to gain more representative results, what can be clearly derived from these findings, nonetheless, is the impression that women and men are represented differently in English idiomatic language. Comparing the representations of both genders, it becomes apparent that women are on the losing end: they are less frequently talked about, in idiomatic terms at least, and if they are the connotations are more likely to be negative compared to male-related idioms.

These findings match what others have discovered about English in general (not only in terms of idiomatic language use). Henley (1987: 5) argues that “[w]omen and girls are [...] ignored in the language simply by not being the topics of discourse”. Furthermore, Graham (1975) found in a study that children’s schoolbooks had a ratio of four to one male versus female references. Moreover, Kjellmer’s (1986) investigation of the Brown and Lancaster-Oslo-Bergen corpora revealed that male pronouns as well as the words *man/men* were used considerably more often than the female equivalents. In a more recent study, using the *BNC*, Pearce (2008) also found that the words *man/men* are used approximately one and a half times more often than the words *woman/women*. For the study conducted in this paper it is, however, important to note that in the *British National Corpus* the ratio of written to spoken text is about 90% written to 10% spoken transcripts. The findings from this corpus, therefore, mainly reflect the usage of the gender-related idioms queried in written discourse. It is possible that the distribution is different in spoken discourse. Future research could examine if and in how far gender-related idioms behave differently in spoken-only corpora.

## 6. References

- Aarts, Jan & Meijs, Willem (eds.). 1986. *New Studies in the Analysis and Exploitation of Computer Corpora*. Amsterdam: Rodopi.
- Boers, Frank & Lindstromberg, Seth. 2009. *Optimizing a Lexical Approach to Instructed Second Language Acquisition*. Basingstoke: Palgrave Macmillan.
- Cacciari, Cristina & Tabossi, Patrizia (eds.). 1993. *Idioms: Processing, Structure, and Interpretation*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Čermák, František. 2001. “Substance of Idioms: Perennial Problems, Lack of Data or Theory?”. *International Journal of Lexicography* 14 (1). 1–20.
- Cowie, Anthony P., Mackin, Ronald & McCaig, Isabel R. 1996. *Oxford Dictionary of English Idioms*. Oxford: Oxford University Press.
- Crystal, David. 2003. *A Dictionary of Linguistics and Phonetics* (5<sup>th</sup> edition). Malden, MA: Blackwell.
- Everaert, Martin, Van der Linden, Erik-Jan, Schenk, André & Schreuder, Rob (eds.). 1995. *Idioms: Structural and Psychological Perspectives*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Fowler, William S. 1972. *Dictionary of Idioms* (2<sup>nd</sup> edition). Surrey: Nelson.
- Gibbs, Raymond W. 1995. “Idiomaticity and Human Cognition”. In: Martin Everaert, Erik-Jan van der Linden, André Schenk & Rob Schreuder (eds.): *Idioms: Structural and Psychological Perspectives*. 97–116.

- Graham, Alma. 1975. "The Making of a Nonsexist Dictionary". In: Barrie Thorne & Nancy Henley (eds.): *Language and Sex: Difference and Dominance*. 57-63.
- Henley, Nancy M. 1987. "This New Species That Seeks a New Language: On Sexism in Language and Language Change". In: Joyce Penfield (ed.): *Women and Language in Transition*. 3-27. Albany, NY: State University of New York Press.
- Kjellmer, Göran. 1986. "'The Lesser Man': Observations on the Role of Women in Modern English Writings". In: Jan Aarts & Willem Meijs (eds.): *New Studies in the Analysis and Exploitation of Computer Corpora*. 163-176.
- Mills, Sara. 2008. *Language and Sexism*. Cambridge: Cambridge University Press.
- Mutlová, Renata. 2009. *Semantic Reference to a Woman: Linguistic Determinism and Male Dominance: Master Thesis*. Brno: Masaryk University, Faculty of Education, Department of English Language and Literature.
- Pearce, Michael. 2008. "Investigating the Collocational Behavior of MAN and WOMAN in the BNC Using Sketch Engine". *Corpora* 3 (1). 1-29.
- Penfield, Joyce (ed.). 1987. *Women and Language in Transition*. Albany, NY: State University of New York Press.
- Prodromou, Luke. 2008. *English as a Lingua Franca: A Corpus-based Analysis*. London: Continuum.
- Romaine, Suzanne. 2000. *Language in Society: An Introduction to Sociolinguistics* (2<sup>nd</sup> edition). Oxford: Oxford University Press.
- Tabossi, Patrizia & Zardon, Francesco. 1995. "The Activation of Idiomatic Meaning". In: Martin Everaert, Erik-Jan van der Linden, André Schenk & Rob Schreuder (eds.): *Idioms: Structural and Psychological Perspectives*. 273-282.
- Thorne, Barrie & Henley, Nancy (eds.). 1975. *Language and Sex: Difference and Dominance*. Rowley, MA: Newbury House Publishers Inc.