

# Zur korpusbasierten Analyse von Pronomina in DaF-Lernertexten: Eine Pilotstudie

---

Margit Breckle

Institut für Deutsche Sprache und Literatur

Universität Vaasa

*Although learner corpus research has become much more frequent during the last years, learner corpora for German as a foreign language are still not very common. Against this background, the scientific network Kobalt-DaF has compiled a small multi-layered learner corpus for German as a foreign language in order to conduct corpus-based analyses of learner texts. This article presents a pilot study on the use of pronouns in the Swedish and German Kobalt sub-corpora. The results of the Contrastive Interlanguage Analysis show that – compared to the German control group – the Swedish learners of German strongly overuse non-reflexive personal pronouns (e.g. wir ‘we’ and ich ‘I’) and substituting indefinite pronouns (e.g. man ‘one’), whereas they strongly underuse attributive demonstrative pronouns (e.g. dies ‘this’). The error analysis reveals that pronouns in the Swedish sub-corpus are much more often corrected than in the German control corpus. The exemplary error analysis of personal pronouns suggests that pronoun errors can often be attributed to problems with gender, case or orthography; however, “real” pronoun errors associated with reference and information structure do appear as well.*

**Schlüsselwörter:** Pronomen, Lernerkorpus, Lernersprache, Kobalt-DaF, Deutsch als Fremdsprache

## 1 Einführung

Die vorliegende Pilotstudie wurde im Rahmen des Wissenschaftlichen Netzwerks *Kobalt-DaF*<sup>1</sup> durchgeführt. Im Netzwerk geht es um die korpusbasierte Analyse von Lernertexten für Deutsch als Fremdsprache, wobei Sprachlehr- und -lernforschung und Lernerkorpusforschung miteinander verbunden werden. Im Netzwerk wurden die so genannten Kobalt-Korpora erstellt; es handelt sich hierbei sowohl um L2-Korpora als auch um ein L1-Vergleichskorpus in Form elektronischer Textsammlungen, für die eine gemeinsame Vorverarbeitung vorgenommen wurde. Die Analysen innerhalb des Netzwerks bewegen sich im Rahmen der Kontrastiven *Interlanguage*-Analyse (CIA, vgl. z. B. Granger 1996; Granger 2008), bei der die Untersuchung von *Overuse* bzw.

---

<sup>1</sup> Ein Überblick über das von der Deutschen Forschungsgemeinschaft (DFG) geförderte Wissenschaftliche Netzwerk findet sich unter [www.kobalt-daf.de](http://www.kobalt-daf.de).

*Underuse* bestimmter sprachlicher Phänomene im Vordergrund steht, und der Fehleranalyse (EA, vgl. z. B. Corder 1981; Granger 2008; Lüdeling, Doolittle, Hirschmann, Schmidt & Walter 2008), die den *Misuse* bestimmter sprachlicher Phänomene fokussiert. Das abschließende Ziel des Netzwerks ist es, die im Netzwerk entstehenden exemplarischen Einzelanalysen miteinander zu vergleichen und in Beziehung zu setzen.

Der vorliegende Beitrag ist wie folgt aufgebaut: In Abschnitt 2 werden die Kobalt-Korpora vorgestellt, die als Datengrundlage der Pilotstudie fungieren, bevor in Abschnitt 3 die Pilotstudie an sich präsentiert wird. Dabei gehe ich nicht nur auf die Datengrundlage (3.1) sowie die Fragestellungen und Methoden (3.2) ein, sondern auch auf die Analyse und erste vorläufige Ergebnisse (3.3). Eine Zusammenfassung mit Ausblick in Abschnitt 4 rundet den Beitrag ab.

## **2 Kobalt-Korpora**

Die Kobalt-Korpora bestehen aus dem L1-Vergleichskorpus DEU mit Texten deutscher MuttersprachlerInnen sowie den drei L2-Subkorpora SWE, BEL und CMN, wobei anzumerken ist, dass die L1 der DaF-Lernenden – Schwedisch, Weißrussisch bzw. Chinesisch – zu unterschiedlichen Sprachfamilien (germanische V2-Sprache, slawische Aspektsprache bzw. sino-tibetische Topiksprache) gehören.

Die Texte für die Kobalt-Korpora, die von fortgeschrittenen DaF-Lernenden mit hohem B2-Niveau geschrieben wurden,<sup>2</sup> wurden stark kontrolliert erhoben. Die Subkorpora umfassen jeweils etwa 20 Texte pro Sprache, wobei der Richtwert für die Textlänge bei >500 Wörtern liegt. Die Aufgabe war das Schreiben eines argumentativen Textes zum Thema *Geht es der Jugend heute besser als früher?* Ergänzt werden die Textkorpora durch eine ausführliche Dokumentation der Lerner-Metadaten, der Sprachbiographie sowie der Erhebungsumstände.

---

<sup>2</sup> Der Sprachstand der Lernenden wurde mit dem OnDaF-Test erhoben (vgl. OnDaF-Test [online]); die Punktespanne für das hohe B2-Niveau umfasste 114–130 Punkte von max. 160 Punkten.

Die Korpora sind so aufgebaut, dass sie mehrere Ebenen umfassen. Neben der originalen Lerneräußerung (der sog. ctok-Ebene) gibt es eine Ebene mit minimaler grammatischer Korrektur (die sog. Zielhypothese 1 / ZH1-Ebene), bei der es vor allem darum geht, ähnliche Fälle wiederfinden und abrufen zu können, und eine Ebene mit zusätzlicher semantisch-pragmatischer Korrektur (die sog. Zielhypothese 2 / ZH2-Ebene). Die Besonderheit der Korpora liegt dabei in der expliziten Korrektur – es wird also sichtbar gemacht, was auf welche Weise korrigiert wurde. Die häufigsten Korrektur-Labels sind dabei CHANGE für die Veränderung eines sprachlichen Ausdrucks, INSERT für das Einfügen eines sprachlichen Ausdrucks und MOVE für das Verschieben eines sprachlichen Ausdrucks.

Die folgenden Beispiele (1), (2) und (3) sollen einen Eindruck der ctok-Ebene (Lerneräußerung), der ZH1-Ebene (Normalisierung in Form einer minimalen grammatischen Korrektur) und der ZH2-Ebene (semantisch-pragmatische Korrektur) vermitteln:<sup>3</sup>

- (1) **ctok-Ebene:** *Aber heute wird das meiste Essen von Industrie dargestellt, von der immer schlechte Skandal einem wegen lebensdrohender oder lebensharmvoller Qualität verwundert.*
- (2) **ZH1-Ebene:** *Aber heute wird das meiste Essen von **der** Industrie dargestellt, von der **einen ein** immer schlechterer Skandal wegen lebensbedrohender oder **schädlicher** Qualität verwundert.*
- (3) **ZH2-Ebene:** *Aber heute wird das meiste Essen von der Industrie **hergestellt**, **von der** einen immer **wieder schlimme** Skandale wegen lebensbedrohender oder **schädlicher** Qualität verwundern.*

Alle Texte in den Kobalt-Korpora wurden auf die gleiche Art und Weise transkribiert und vorverarbeitet und enthalten Informationen zu (i) Tokenisierung, Wortarten und Lemmata (vgl. Schiller, Teufel & Stöckert 1999), (ii) Zielhypothesen (vgl. Reznicek, Lüdeling, Krummes, Schwantuschke, Walter, Schmidt, Hirschmann & Andreas 2012; Reznicek, Lüdeling & Hirschmann 2013) mit ZH1 und ZH2, (iii) Grammatischen Funktionen, (iv) Topologischen Feldern, (v) Abhängigkeitsannotationen und (vi) (Ko-)Referenz. Diese Informationen können in den Korpora gesucht und abgerufen werden.

---

<sup>3</sup> Die Lerneräußerung ist dem Lernertext CH\_009\_2011\_03 entnommen.

### 3 Pilotstudie

#### 3.1 Datengrundlage

Für die vorliegende Pilotstudie wurden das schwedische L2-Subkorpus SWE und das deutsche L1-Vergleichskorpus DEU als Datengrundlage herangezogen.<sup>4</sup> Die Zahl der Token (ohne Satz- und Leerzeichen) in den beiden Subkorpora ist auf den einzelnen Ebenen die folgende:

**Tabelle 1.** Tokenzahl in den Subkorpora DEU und SWE auf ctok-, ZH1- und ZH2-Ebene

Subkorpus	ctok (Lernertext)	ZH1	ZH2
DEU	11052	11021	11063
SWE	7854	7879	7906

In meiner Analyse gehe ich von der ctok-Ebene, d. h. dem originalen Lernertext, und der ZH2-Ebene aus, auf der neben der grammatischen Korrektur auch eine semantisch-pragmatische Korrektur vorgenommen wurde. Einschränkend muss darauf hingewiesen werden, dass das Wortarten-Tagging auf der ctok-Ebene in der verwendeten Korpus-Version 1.6 ggf. in einigen Fällen Fehler aufweisen kann.

#### 3.2 Fragestellungen und Methode

Der Pilotstudie liegen die folgenden Fragestellungen zugrunde:

- Fragestellung 1: Wie unterscheidet sich die Häufigkeit der untersuchten Pronomen-Typen zwischen den beiden Subkorpora DEU und SWE? ⇒ Kontrastive *Interlanguage*-Analyse (CIA) ⇒ *overuse* / *underuse*
- Fragestellung 2: Welche der untersuchten Pronomen sind am fehleranfälligsten (d. h. welche Pronomen-Typen weisen die häufigsten Korrekturen auf?), und

<sup>4</sup> Die für die Pilotstudie verwendeten Subkorpora SWE und DEU stammen aus dem Kobalt-Korpus (Version 1.6).

welche Fehlertypen lassen sich exemplarisch ausmachen? ⇒ Fehleranalyse  
(EA) ⇒ *misuse*

Bei den Fragestellungen geht es sowohl um eine quantitative als auch um eine qualitative Perspektive.

### 3.3 Analyse und erste vorläufige Ergebnisse

Um einen ersten Eindruck von der Datengrundlage hinsichtlich der Verwendung von Pronomen zu bekommen, wurde, ausgehend von DEU, die Rangfolge der Frequenzen der verschiedenen Pronomen-Typen in DEU und SWE auf der ctok-Ebene aufgelistet:

**Tabelle 2.** Rangfolge der Frequenzen der verschiedenen Pronomentypen<sup>5</sup> in DEU und SWE (auf ctok-Ebene)

Pronomentypen	DEU		SWE		
	N	Rang	N	Rang	Diff.
<b>PPER</b>	460	<b>1</b>	477	<b>1</b>	⇔
<b>PIAT</b>	203	<b>2</b>	159	<b>3</b>	↓1
<b>PIS</b>	139	<b>3</b>	210	<b>2</b>	↑1
<b>PRF</b>	114	<b>4</b>	75	<b>5</b>	↓1
<b>PPOSAT</b>	111	<b>5</b>	91	<b>4</b>	↑1
<b>PRELS</b>	97	<b>6</b>	65	<b>6</b>	⇔
<b>PDS</b>	86	<b>7</b>	51	<b>8</b>	↓1
<b>PAV</b>	84	<b>8</b>	62	<b>7</b>	↑1
<b>PDAT</b>	77	<b>9</b>	33	<b>9</b>	⇔
<b>PWAV</b>	22	<b>10</b>	18	<b>11</b>	↓1
<b>PWS</b>	11	<b>11</b>	21	<b>10</b>	↑1
<b>PWAT</b>	1	<b>12</b>	0	<b>13</b>	↓1
<b>PPOSS</b>	1	<b>13</b>	0	<b>13</b>	⇔
<b>PRELAT</b>	0	<b>14</b>	1	<b>12</b>	↑2

<sup>5</sup> PPER – Personalpronomen; PIAT – attribuierendes Indefinitpronomen; PIS – substituierendes Indefinitpronomen; PRF – Reflexivpronomen; PPOSAT – attribuierendes Possessivpronomen; PRELS – Relativpronomen; PDS – substituierendes Demonstrativpronomen; PAV – Pronominaladverb; PDAT – attribuierendes Demonstrativpronomen; PWAV – adverbiales Interrogativ- oder Relativpronomen; PWS – substituierendes Interrogativpronomen; PWAT – attribuierendes Interrogativpronomen; PPOSS – substituierendes Possessivpronomen; PRELAT – attribuierendes Relativpronomen

Wie Tabelle 2 zeigt, belegen Personalpronomen in beiden Subkorpora Rang 1. Auf den Rängen 2 bzw. 3 finden sich attribuerende und substituierende Indefinitpronomen, worin sich sicherlich das Thema *Geht es der Jugend heute besser als früher?* der Argumentation widerspiegelt, das zum Schreiben eines eher allgemein gehaltenen Textes anregt, während Reflexivpronomen und attribuerende Possessivpronomen auf den Rängen 4 bzw. 5 liegen. In beiden Subkorpora belegen substituierende Relativpronomen Rang 6 und attribuerende Demonstrativpronomen Rang 9, während die Ränge 7 bzw. 8 substituierende Demonstrativpronomen und Pronominaladverbien umfassen. Alle weiteren Pronomen-Typen (Ränge 10–14) weisen jeweils weniger als 25 Vorkommen in den Subkorpora auf. Festzustellen ist, dass sich zwischen den Subkorpora DEU und SWE keine großen Verschiebungen bezüglich der Ränge der einzelnen Pronomen finden, sondern dass ein maximaler Unterschied von zwei Rängen auszumachen ist.

### 3.3.1 Overuse / Underuse

Bezüglich der Frequenzen der neun häufigsten Pronomen-Typen in DEU und SWE ergibt sich folgendes Bild, wobei der besseren Vergleichbarkeit halber eine Normalisierung auf 10.000 Token vorgenommen wurde und für SWE der Over- bzw.

Underuse der Pronomen-Typen in Prozent dargestellt ist:

**Tabelle 3.** Frequenzen verschiedener Pronomen-Typen in DEU und SWE (auf ctok-Ebene; Normalisierung auf 10.000 Token) sowie Over- bzw. Underuse in SWE

Pronomentyp	DEU (11052 Token)	SWE (7854 Token)	Over- bzw. Underuse in %
<b>PIS</b>	125,8	267,4	+112,6 %
<b>PPER</b>	416,2	607,3	+45,9 %
<b>PPOSAT</b>	100,4	115,9	+15,4 %
<b>PIAT</b>	183,7	202,4	+10,2 %
<b>PAV</b>	76	78,9	+3,8 %
<b>PRELS</b>	87,8	82,8	-5,7 %
<b>PRF</b>	103,1	95,5	-7,4 %
<b>PDS</b>	77,8	64,9	-16,8 %
<b>PDAT</b>	69,7	42	-39,7 %

Tabelle 3 zeigt den Over- bzw. Underuse der verschiedenen Pronomen im schwedischen Subkorpus SWE im Vergleich mit dem deutschen Vergleichskorpus DEU. Je dunkler die Zellen sind, desto größer ist der Overuse; je stärker gestreift die Zellen sind, desto größer ist der Underuse. Das heißt, dass die Verwendung zwar nicht eigentlich falsch im Sinne eines sprachlichen Fehlers ist, jedoch ist die *Anwendungshäufigkeit* nicht zielsprachlich.

In SWE zeigt sich im Vergleich zum L1-Vergleichskorpus DEU ein (starker) Overuse von substituierenden Indefinitpronomen (+112,6 %), Personalpronomen (+45,9 %), attribuierenden Possessivpronomen (+15,4 %) sowie von attribuierenden Indefinitpronomen (+10,2 %). Einen schwachen Underuse verzeichnen substituierende Relativpronomen (-5,7 %) und Reflexivpronomen (-7,4 %), während substituierende Demonstrativpronomen (-16,8 %) einen Underuse und attribuierende Demonstrativpronomen (-39,7%) einen starken Underuse aufweisen.

Im Folgenden soll anhand der Pronomen-Typen, die den größten Over- bzw. Underuse verzeichnen, dargestellt werden, durch welche Pronomen (Lemmata) konkret der Over- bzw. Underuse in SWE im Vergleich zu DEU zustande kommt. Bei PIS (+112,6 %) lässt sich der Overuse vor allem durch *man* erklären (SWE: 125/7854 vs. DEU: 70/11052), vgl. Bsp. (4):

- (4) Mit jeder Generation kommen Sachen, die **man** gelernt hat, die **man** gekriegt hat und alles, was **man** der nachfolgenden Generation geben kann. (SWE\_002)

Der Overuse bei PPER (+45,9 %) ist vor allem auf *ich* (SWE: 91/7854 vs. DEU: 87/11052) und insbesondere auf *wir* (SWE: 102/7854 vs. DEU: 60/11052) zurückzuführen, vgl. Bsp. (5) und (6):

- (5) **Ich** weiß nicht, aber **ich** glaube, dass **ich** nicht so glücklich sein würde, wenn mein Leben so aussähe. (SWE\_003)
- (6) **Wir** haben jetzt Facebook – aber reden nicht mehr miteinander, weil **wir** zu „busy“ mit unseren Smartphones sind. (SWE\_007)

Der Underuse bei PDAT (-39,7 %) basiert vor allem auf *dies* (SWE: 29/7854 vs. DEU: 72/11052), vgl. Bsp. (7) und (8):

- (7) Es ist heute einfacher, sich mit Freizeitaktivitäten zu beschäftigen, weil es zur Zeit mehr davon gibt. Auf **die** (ZH2: **diese**) Weise ist es einfacher, sich bei etwas zu engagieren als früher. (SWE\_006)
- (8) Denn heute hat fast jeder Jugendliche ein Smartphone Notebook Fernseher und viele bereits ein eigenes Auto dank des Führerscheins ab 17. **Dieser** verbesserte Lebensstandard spiegelt sich auch in der Kindheit wider [...] (DEU\_003)

Während im aus SWE stammenden Bsp. (7) in dem auf den vorigen Satz referierenden Ausdruck *Auf die Weise* mit *die* ein Artikel verwendet wird, der auf der ZH2-Ebene in *diese* korrigiert wird, finden sich in DEU Ausdrücke wie *Dieser verbesserte Lebensstandard* in Bsp. (8), in denen bereits auf der ctok-Ebene mit einem attribuierenden Demonstrativpronomen auf den vorherigen Satz referiert wird.

### 3.3.2 Fehleranalyse

Bei der Fehleranalyse geht es um die Frage, welche der untersuchten Pronomen am fehleranfälligen sind – im Sinne von: weisen die häufigsten Korrekturen auf – und welche Fehlertypen sich exemplarisch ausmachen lassen. In Tabelle 4 geht es um die Korrekturen, die in den Subkorpora DEU und SWE auf der ZH2-Ebene vorgenommen wurden:

**Tabelle 4.** Korrekturen in den Subkorpora DEU und SWE (auf ZH2)

Pronomentyp	DEU		SWE	
	Korrekturen (Fehler)		Korrekturen (Fehler)	
<b>PPER</b>	33/466	7,1 %	<b>50/482</b>	<b>10,4 %</b>
<b>PIAT</b>	<b>11/199</b>	<b>5,5 %</b>	29/150	19,3 %
<b>PIS</b>	12/137	8,8 %	32/206	15,3 %
<b>PRELS</b>	11/120	9,2 %	12/73	16,4 %
<b>PRF</b>	12/116	10,3 %	31/84	36,9 %
<b>PPOSAT</b>	11/112	9,8 %	<b>44/100</b>	<b>44,0 %</b>
<b>PAV</b>	10/88	11,4 %	26/71	36,6 %
<b>PDAT</b>	9/78	11,5 %	9/36	16,7 %
<b>PDS</b>	<b>9/75</b>	<b>12,0 %</b>	17/52	32,7 %

Wie in Tabelle 4 zu sehen, finden sich auch im deutschen Vergleichskorpus Korrekturen. Je nach Pronomentyp liegt dabei der Korrekturanteil zwischen 5,5 % für attribuie-

rende Indefinitpronomen und 12,0 % für substituierende Demonstrativpronomen. Im SWE-Subkorpus liegen die Korrekturanteile zwischen 10,4 % (Personalpronomen) und 44,0 % (attribuierende Possessivpronomen). In SWE finden sich also deutlich mehr Korrekturen im Vergleich zu DEU, und beispielsweise der Unterschied bei den attribuierenden Possessivpronomen ist hoch signifikant.

Im Folgenden möchte ich mich exemplarisch auf die Fehleranalyse der Personalpronomen konzentrieren, um zu zeigen, welche Korrekturtypen in DEU und SWE sowie welche Fehlertypen in SWE diesbezüglich vorkommen. Wie in Tabelle 4 zu sehen, handelt es sich dabei um einen Pronomen-Typ, bei dem es – was die Korrekturen angeht – zwischen DEU und SWE keine großen Unterschiede bezüglich der relativen Häufigkeit gibt (7,1 % DEU vs. 10,4 % SWE).

Um zu eruieren, welche Typen von Korrekturen in den beiden Subkorpora bei den Personalpronomen auf der ZH2-Ebene vorgenommen wurden, wurde für die einzelnen Korrektur-Typen jeweils die absolute und relative Häufigkeit errechnet, bezogen auf die Häufigkeit von PPER in den jeweiligen Subkorpora (siehe Tabelle 5).

**Tabelle 5.** Absolute und relative Häufigkeit der Korrektur-Typen für PPER in DEU und SWE (auf ZH2-Ebene)

Pronomen	Korrekturtyp	DEU		SWE	
PPER	CHANGE <sup>6</sup>	11/466	2,4 %	20/482	4,1 %
	INSERT <sup>7</sup>	10/466	2,1 %	16/482	3,3 %
	MOVE <sup>8</sup>	12/466	2,6 %	14/482	2,9 %

Wie in Tabelle 5 ersichtlich, kommt bei DEU die Verschiebung (MOVE) des sprachlichen Ausdrucks (hier: Personalpronomen) innerhalb des Satzes häufiger (2,6 %) als die Veränderung des vorhandenen sprachlichen Ausdrucks (CHANGE, 2,4 %) und die Hinzufügung des sprachlichen Ausdrucks (INSERT, 2,1 %) vor. Bei SWE hingegen findet sich am häufigsten die Veränderung des sprachlichen Ausdrucks (CHANGE, 4,1%),

<sup>6</sup> Veränderung des vorhandenen sprachlichen Ausdrucks

<sup>7</sup> Hinzufügung des sprachlichen Ausdrucks

<sup>8</sup> Verschiebung des vorhandenen sprachlichen Ausdrucks innerhalb des Satzes

während die Hinzufügung des sprachlichen Ausdrucks (INSERT, 3,3 %) und die Verschiebung des sprachlichen Ausdrucks innerhalb des Satzes (MOVE, 2,9 %) weniger häufig vorkommen; relativ gesehen, sind diese Korrekturtypen in SWE dennoch häufiger als in DEU zu finden. Andere Korrekturtypen wie SPLIT (Trennung eines sprachlichen Ausdrucks in zwei Teile) und MERGE (Zusammenfügung zweier sprachlicher Ausdrücke zu einem) kommen in den beiden Subkorpora bei Personalpronomen nicht vor.

Nun lässt sich fragen, welche Fehlertypen mit den Korrekturtypen CHANGE, INSERT und MOVE für Personalpronomen angegeben werden. Dabei muss man sich vor Augen halten, dass auch die Art und Weise, wie in den Kobalt-Korpora für die Zielhypothesen Korrekturen vorgenommen werden, Einfluss auf die Korrektur der Pronomen hat, nämlich, dass das finite Verb stehen bleibt und die anderen Elemente entsprechend korrigiert werden.

**Tabelle 6.** Typische Fehler für Personalpronomen in SWE auf Basis der Korrektur-Typen (auf ZH2-Ebene)

Korrektur-Typ	SWE
CHANGE	* Referenz: <i>die</i> ⇔ <i>sie</i>
	* Valenz
	* Stil/Rechtschreibung
INSERT	* Auslassung des expletiven <i>es</i>
	* Stil
	* Stil/Valenz
MOVE	* Pronominale Wortstellung im Mittelfeld
	* Stellung des finiten Verbs in Nebensätzen

Wie in Tabelle 6 zu sehen, handelt es sich um recht unterschiedliche Fehlertypen, die durch dasselbe Korrektur-Label markiert sind. Bei mit CHANGE markierten Korrekturen finden sich vor allem Referenzfehler (vgl. Bsp. 9), Valenzfehler (vgl. Bsp. 10) sowie Stil-/Rechtschreibungsfehler (vgl. Bsp. 11).<sup>9</sup>

- (9) Heute sitzen die Jugendlichen eher vor dem Computer oder dem Fernseher – **sie** (ctok: **die**) sind die Gefangenen der Technik. (SWE\_009)

<sup>9</sup> Die Beispiele (9) bis (16) stammen von der ZH2-Ebene – in Klammern angegeben ist jeweils auch die originale Lerneräußerung auf der ctok-Ebene, um die Korrektur besser nachvollziehen zu können.

- (10) Sie bekommen jeden Tag Essen in der Schule und nach dem Abitur haben sie ihre Eltern, die **ihnen** (ctok: **sie**) mit Geld aushelfen können. (SWE\_019)
- (11) Bäume sind stabil, **wir** (ctok: ; **Wir**) nicht. (SWE\_007)

Mit INSERT markierte Korrekturen weisen insbesondere auf die fehlerhafte Auslassung des expletiven *es* (vgl. Bsp. 12), auf Stilfehler (vgl. Bsp. 13) sowie auf Stil-/Valenzfehler (vgl. Bsp. 14) hin.

- (12) ..., weil **es** (ctok:  $\emptyset$ ) zur Zeit mehr davon gibt. (SWE\_006)
- (13) Heute **gibt es** (ctok: **sind**  $\emptyset$ ) viel mehr Möglichkeiten im Leben ... (SWE\_018)
- (14) Daher müssen sie nie schrecklichen Hunger haben und nach Essen suchen, wie **es** (ctok:  $\emptyset$ ) frühere Generationen während der Kriege erlebt haben. (SWE\_019)

Mit MOVE markierte Korrekturen schließlich beziehen sich insbesondere auf Fehler, die die fehlerhafte pronominale Wortstellung im Mittelfeld (vgl. Bsp. 15) und Fehler in Bezug auf die Stellung des finiten Verbs in Nebensätzen (vgl. Bsp. 16) betreffen.

- (15) Junge Menschen haben **es** (ctok:  $\emptyset$ ) heutzutage  $\emptyset$  (ctok: **es**) also sowohl besser als auch schlechter als frühere Generationen. (SWE\_011)
- (16) Ich finde, dass **wir heute** mehr zu erledigen **haben**. (SWE\_015) (ctok: dass **heute haben wir**...)

Sowohl Tabelle 6 als auch Bsp. (9) bis (16) verdeutlichen, dass viele der typischen Fehler, bei denen eine – wie auch immer geartete – Korrektur des Personalpronomens vorgenommen wurde, nicht „wirkliche“ Pronomenfehler sind, sondern auf andere Fehlertypen zurückzuführen sind, z. B. Rechtschreib-, Genus- oder Valenzfehler. Bei der zukünftigen Analyse der Daten im Rahmen des Netzwerks soll jedoch der Schwerpunkt auf „wirkliche“ Pronomenfehler gelegt werden, die – wie Bsp. (9) – mit Referenz, Informationsstruktur und Salienz zu tun haben.

#### **4 Zusammenfassung & Ausblick**

Die vorliegende Pilotstudie mit den Subkorpora DEU und SWE der Kobalt-Korpora hat hinsichtlich der Untersuchung von Over- bzw. Underuse, für die die Kontrastive *Interlanguage*-Analyse angewendet wurde, Folgendes ergeben:

Im SWE-Subkorpus findet sich ein starker Overuse von substituierenden Indefinitpronomen (v. a. durch *man*) sowie von Personalpronomen (v. a. durch *wir* und *ich*). Ein starker Underuse lässt sich vor allem bei den attribuierenden Demonstrativpronomen ausmachen, der in erster Linie durch *dies* verursacht ist.

Bei der Fehleranalyse, mit der der Misuse untersucht wurde, zeigte sich, dass sich im Subkorpus SWE deutlich mehr Pronomen-Korrekturen finden als im Subkorpus DEU; auf der ZH2-Ebene liegt der Anteil der korrigierten Pronomen in SWE je nach Pronomen-Typ bei zwischen 10,4 % (Personalpronomen) und 44,0 % (attribuierende Possessivpronomen). Dabei lassen sich in SWE vor allem CHANGE, aber auch INSERT und MOVE als Korrekturtypen ausmachen. Die exemplarische Analyse der Personalpronomen-Korrekturen in SWE hat offenbart, dass die Korrekturen häufig auf Genus- und/oder Valenz- sowie Rechtschreibfehler zurückzuführen sind, jedoch finden sich auch „wirkliche“ Pronomen-Fehler (wie in Bsp. 9), die mit Referenz, Informationsstruktur und Salienz zu tun haben.

Für die weitere korpusbasierte Untersuchung von Pronomina in DaF-Lernertexten auf Basis der Kobalt-Korpora ist als Vervollständigung der Fehleranalyse für die anderen Pronomen-Typen, wobei der Schwerpunkt auf den „wirklichen“ Pronomen-Fehlern liegen soll, die Analyse der Referenzannotation sowie die Analyse mit „Kandidaten für Pronomina“ als Vergleichsgröße geplant. Zudem soll die Untersuchung auch auf die in der vorliegenden Pilotstudie unberücksichtigten Subkorpora BEL und CMN ausgeweitet werden.

## Literatur

- Corder, Stephen Pit (1981). *Error Analysis and Interlanguage*. Oxford: Oxford University Press.
- Granger, Sylviane (1996). From CA to CIA and back: An Integrated Approach to Computerized Bilingual and Learner Corpora. In: *Languages in Contrast*, 37–51. Hrsg. Karin Aijmer/Mats Johansson/Bengt Altenberg. Lund: Lund University Press.
- Granger, Sylviane (2008). Learner corpora. In: *Corpus Linguistics: An International Handbook*, 259–275. Hrsg. Merja Kytö/Anke Lüdeling. Berlin u. a.: de Gruyter.
- Lüdeling, Anke/Seanna Doolittle/Hagen Hirschmann/Karin Schmidt/Maik Walter (2008). Das Lernerkorpus Falko. *Deutsch als Fremdsprache* 2/2008, 67–73.

OnDaF-Test. [online]. [zitiert 5.5.2014]. Abrufbar unter: [www.ondaf.de](http://www.ondaf.de)

Reznicek, Marc/Anke Lüdeling/Cedric Krummes/Franziska Schwantuschke/Maik Walter/Karin Schmidt/Hagen Hirschmann/Torsten Andreas (2012). *Das Falko-Handbuch. Korpusaufbau und Annotationen Version 2.01*. Tech. Rep. Humboldt-Universität zu Berlin.

Reznicek, Marc/Anke Lüdeling/Hagen Hirschmann (2013). Competing Target Hypotheses in the Falko Corpus. A Flexible Multi-Layer Corpus Architecture. In: *Automatic Treatment and Analysis of Learner Corpus Data*, 101–124. Hrsg. Ana Díaz-Negrillo/Nicolas Ballier/Paul Thompson. Amsterdam: John Benjamins.

Schiller, Anne/Simone Teufel/Christine Stöckert (1999). *Guidelines für das Tagging deutscher Textcorpora mit STTS*. Tech. Rep. Institut für maschinelle Sprachverarbeitung. Universität Stuttgart.