

GECCO

GERMAN ENGLISH CONTRASTS IN COHESION

Title Guidelines for the correction of automatically annotated cohesive devices

Author Kerstin Kunz

Deliverable No.
Status
Availability *local*
Date 21.11.12

DFG project

Richtlinien für die Korrektur von kohäsiven Referenzmitteln

Formatierungskonventionen

kursiv kennzeichnet objektsprachliche Beispiele

fett kennzeichnet wichtige Begriffe

unterstrichen kennzeichnet in Beispielen das fokussierte Element

Allgemeine Erläuterung

Die GECCo-Korrekturrichtlinien umfassen Regeln für die manuelle Korrektur automatisch annotierter Kohäsionsmittel der Kategorie kohäsive Referenz im Englischen und Deutschen mit dem Annotationstool MMAX2.

Automatischen Vorannotation

Die automatische Annotation erfolgte auf Grundlage der bereits vorhandenen Annotation auf verschiedenen lexikogrammatischen Ebenen. Zunächst werden auf Basis dieser Annotation Regeln für die Automatische Extraktion von kohäsiven Formen im Corpus Query Processor (CQP) erstellt. Diese Regeln disambiguieren kohäsive und nicht-kohäsive Vorkommen und weisen die kohäsiven Vorkommen bestimmten Typen von Kohäsion zu. Es werden komplexe Regeln formuliert, in dem Restriktionen auf der String-Ebene, sowie auf der Lemma, Parts-of-Speech und auf der Satzglied-Ebene miteinander kombiniert werden. Diese komplexen Extraktionen werden in sogenannten Macros zusammengefasst.

In einem nächsten Schritt greift ein Perl Skript auf diese Macros zurück und definiert Kategorien von Kohäsionsmitteln, mit denen die englischen und deutschen Texte annotiert werden.

Diese Annotationen sollen mit dem manuellen Tool MMAX2 verbessert werden.

Manuelle Korrektur

Die manuelle Korrektur der automatischen Annotation beinhaltet vor allem vier Arbeitsschritte:

- 1) Überprüfung: Wurden alle Referenzmittel annotiert? Gegebenenfalls neue Markierung von bisher nicht erkannten Referenzmitteln
- 2) Überprüfung: Wurden Formen annotiert, die keine kohäsive Referenz herstellen? Die Annotation wird dann NICHT gelöscht, aber mit dem Label `correction=to_delete` versehen (siehe 4.a) weiter unten)
- 3) Überprüfung: Wurden die Referenzmittel der richtigen Kategorie zugeordnet?
- 4) Gegebenenfalls neue Zuordnung zur richtigen Kategorie

Korrektur von kohäsiver Referenz

Im Projekt GECCo geht es allgemein um die Annotation von **Kohäsionsmitteln**. Deshalb zunächst eine Definition von **Kohäsion**:

Ein Text ist dann **kohäsiv**, wenn die darin enthaltenen sprachlichen Elemente miteinander in einer Bedeutungsbeziehung stehen. Diese Verbindung besteht zwischen Elementen in unterschiedlichen Einzelsätzen (z.B. Elemente in Nebensatz und Hauptsatz), Sätzen oder Textabschnitten. Die Bedeutungsbeziehung wird von bestimmten sprachlichen Elementen, den Kohäsionsmitteln angezeigt.

Die vorliegenden Richtlinien beschränken sich auf die manuelle Korrektur **kohäsiver** Referenzmittel, also einer bestimmten Art von Kohäsionsmitteln.

Definition kohäsive Referenzmittel

Nach Halliday & Hasan wird durch die Verwendung von **kohäsiven Referenzmitteln** meist **Koreferenz** erstellt. Bei Koreferenz beziehen sich verschiedene sprachliche Ausdrücke im Text auf den gleichen außersprachlichen Referenten. Die kohäsiven Referenzmittel zeigen dabei durch ihre sprachliche Form und Struktur an, dass der Referent entweder schon einmal an einer vorhergehenden Textstelle genannt wurde (bei **anaphorischer** Referenz) oder im nachfolgenden Text genannt wird (bei **kataphorischer** Referenz). Es handelt sich bei den kohäsiven Referenzmitteln um **Pronomen**, **Nominalphrasen** oder **Adverbien**, die in ihrer äußeren Form reduziert sind und über sehr wenig semantischen Eigengehalt verfügen. Dadurch deuten sie an, dass mehr Information zum Referenten woanders im Text zu finden ist – sie verweisen also nicht nur auf den außersprachlichen Referenten sondern auch sprachlich auf andere Textstellen.

Hierzu ein Beispiel:

1. *[President Bush] recognized the growing strains on energy systems as he took office.*
2. *Recognizing the growing strains on energy systems as he took office, [President Bush] sought to develop a comprehensive and balanced energy policy.*

In dem Beispiel informiert das Personalpronomen selbst nur über Genus und Numerus des Referenten. Deshalb löst es eine Suche zu einem anderen Ausdruck im Text (dem Antezedent) aus, der diese Bedingungen ebenso erfüllt und gleichzeitig mehr Informationen zum Referenten enthält. In dem Beispiel ist der Antezedent ein Eigennamen, auf den in Satz 1 durch das Pronomen rückverwiesen wird (anaphorische Referenz), in Satz 2 erfolgt ein Vorverweis (kataphorische Referenz). Antezedent und kohäsives Referenzmittel verweisen auf den gleichen außersprachlichen Referenten.

ACHTUNG: Bei **Comparative Reference** wird meist KEINE Koreferenz hergestellt, stattdessen werden unterschiedliche Referenten verglichen, die zum selben Typ gehören.

Beispiele:

3. *They've given us [special] places in the front row. Would you prefer the other seats.*
4. *Gerald Middleton was a man [of mildly but persistent depressive temperament]. Such men are not at their best at breakfast.*

Arten von Antezedenten:

Bei den Antezedenten kann es sich um sprachliche Elemente handeln, die in ihrer Form und syntaktischen Funktion sehr unterschiedlich sind:

- Einzelne lexikalische Wörter, z.B. Nomen, Adjektiv (vor allem bei Comparative Reference)
- Nominalphrasen, die einen Eigennamen oder Gattungsnamen als Kopfnomen haben:

- Einzelsätze, mit finitem oder nicht finitem Prädikat:
- Komplexe Sätze: Hauptsatz oder Hauptsatz und Nebensatz/ Nebensätze
- Größerer Textabschnitt:

Arten von kohäsiven Referenzmitteln

Man unterscheidet verschiedene Unterkategorien von kohäsiven Referenzmitteln, **personal reference** und **demonstrative reference**.

1. Personal Reference

Kohäsionsmittel: Personalpronomen, Possessivpronomen und Possessivbegleiter

Formen

	Englisch	Deutsch
Personalpronomen (pers=person- endophoric)	<i>I, you, we</i> (nicht beim ersten Vorkommen) <i>he, she, they</i> unterschiedliche Kasusformen	<i>Ich, du, wir, ihr</i> (nicht beim ersten Vorkommen) <i>er, sie</i> (3. Pers. Singular und Plural), unterschiedliche Kasusformen
Personalpronomen <i>it/es</i> (pers=it-endophoric)	<i>it</i>	<i>es</i>
Possessivpronomen (pers=possessive- endophoric)	<i>His, hers, theirs, mine, yours</i>	<i>Sein(e,r,s,m,n), Ihr(e,r,s)</i>
Possessivbegleiter (pers=possessive- endophoric)	<i>His, her, its, their, my, your</i>	<i>Sein(e,r,s,m,n), Ihr(e,r,s,m,n)</i>

HINWEIS: Vorkommen von erster und zweiter Person werden erst ab dem 2. Vorkommen annotiert!

ACHTUNG: Vor allem *it* und *es* müssen nicht bei jedem Vorkommen kohäsiv sein!

Beispiele:

5. *It is raining.*
6. *It's is clear what our citizens want from Europe.*
7. *Gewerkschaften gibt es in vielen Ländern.*
8. *Zwar ist das neue Jahr erst einige Wochen alt , aber 2005 hat es in sich.*
9. *So unheimlich war es, wenn Blaubarts Frau nach dessen Abreise in das verschlossene Zimmer eindringen will.*

Diese Vorkommen sollen aber trotzdem annotiert werden, und zwar als pers=it-exophoric!

Arten von Antezedenten

a) Antezedent=Eigenname:

10. [Tad Williams] is one of the most famous writers of modern times. In addition to *Memory, Sorrow and Thorn* he has written the acclaimed *Otherland* series.
11. [Tad Williams]? I just read on of his novels.
12. [Tad Williams]? I just read a novel of his.

b) Antezedent=lexikalische Nominalphrase:

13. *The UK has always been a strong supporter of European enlargement and I am very pleased to mark this latest accession of ten new members on 1 May. We welcome **it** as another important and historic step towards sealing over the artificial divisions created by the Cold War.*
14. *Nobody can be sure what the 21st century holds. But one characteristic is already clear. It will be dominated by the scale and pace of technological and economic change.*
15. *Großbritannien hat sich immer schon für die europäische Erweiterung stark gemacht und deshalb begrüße ich den Beitritt von zehn neuen Mitgliedstaaten am 1. Mai von ganzem Herzen. **Er** ist ein historischer Schritt auf dem Weg, die künstlichen Risse zu kitten, die der Kalte Krieg hinterlassen hat.*
16. *[Eine verantwortungsbewusste Politik] kann [diesen Prozess, der zudem von objektiven Faktoren determiniert wird,] nicht nur flankieren. Sie muss ihn vielmehr formen.*
17. *[Eine verantwortungsbewusste Politik] kann [diesen Prozess, der zudem von objektiven Faktoren determiniert wird,] nicht nur flankieren. Sie muss ihn vielmehr formen. Und sie muss außerdem ...*

Ausnahme Deutsch: Referenz auf Adjektive - im Englischen nicht möglich

18. *Ist die Suppe heiß? – Ja, sie ist es/ das ist sie.*

c) Antezedent=Einzelatz, Satz, Textabschnitt:

Im Englischen nur bei kohäsivem *it*, im Deutschen nur bei *es*.

19. *We work for prosperity and opportunity because they're right. It's the right thing to do.*
20. *Moreover he was charged by his father with a mission, which he might not reveal in that place. 'It is known to us already,' said the three damsels. [EO_FICTION_002]*

2. Demonstrative Reference

Kohäsionsmittel: Demonstrativpronomen (head), Demonstrativbegleiter (modifier), definiter Artikel (modifier), Pronominaladverbien (head) und ausgewählte Adverbien (head)

Formen

	Englisch	Deutsch
Demonstrativpronomen (dem=head)	<i>This, that, these, those</i>	<i>Diese(r,s,m,n), jene(r,s,m,n), der, die, das</i>
Demonstrativbegleiter (dem=modifier)	<i>This, that, these, those</i>	<i>Diese(r,s,m,n), jene(r,s,m,n)</i>
Definiter Artikel (dem=article)	<i>the</i>	<i>De(r,m,s,n), die</i>
Pronominaladverbien (dem=pronadv)	<i>Here + Präposition, there + Präposition</i>	<i>Hier + Präposition Da + Präposition Wo + Präposition</i>
Temporale Adverbien (dem=temporal)	<i>now, then</i>	<i>jetzt, nun, damals</i>

Lokale Adverbien
(dem=local)

Here, there

Hier, da, dort

ACHTUNG: nicht alle Vorkommen sind kohäsiv:

- Alle Formen können **exophorisch** sein, d.h. direkt auf den Situationskontext verweisen:

21. *Siehst Du den da drüben?*

22. *That man over there, he is strange.*

23. *Die Sonne steht immer höher am Himmel.*

- Viele Vorkommen von *that*, und *der, die, das* als Pronomen haben die Funktion eines Relativpronomens und sind dann nicht kohäsiv.

24. *The fact that he could not go to work, bothered him.* => nicht kohäsiv

- Manche Vorkommen von *those, jene* und *der, die, das* als Pronomen haben die Funktion eines Indefinitpronomens und sind dann nicht kohäsiv:

25. *Those, who hesitate, are lost.* => nicht kohäsiv

26. *längst hat in diesem Feld die Wirklichkeit jene, die es besser wussten, eingeholt* => nicht kohäsiv

27. *Für die, die lieber von einer festen Unterkunft Tagestouren unternehmen wollen, bietet die Insel Sylt z.B. rund 200 km Radwege*

- Der definite Artikel wird nur dann annotiert, wenn er Koreferenz erzeugt:

28. *This past spring, the U.S. Department of Education issued [a report, The Condition of Education 2000]. The report found that the benefits of attending college are greater today than ever before. [...]* => Koreferenz

29. *Insgesamt stellen sich die ökonomischen Voraussetzungen zu Beginn des neuen Jahrhunderts für Deutschland recht gut dar. Die Teilhabe an den Weltmärkten hat sich verbessert - dies belegt die positive Dynamik der Exporte und Dienstleistungen.* => keine Koreferenz

- Pronominaladverbien sind im Deutschen als **Korrelat** nicht kohäsiv:

30. *Ich habe Angst davor, alleine im Wald zu joggen* => nicht kohäsiv.

31. *Joggst Du denn auch alleine im Wald? Also ich habe Angst davor.* => kohäsiv.

- Manche Pronominaladverbien im Deutschen können je nach Kontext sowohl als kohäsives Referenzmittel als auch als **kohäsive Konjunktion** fungieren:

32. *In den vergangenen beiden Jahren haben die 200 größten deutschen Firmen insgesamt weit über 50 000 Jobs abgebaut. Mittelständische Betriebe dagegen haben alleine im Jahr 2000 unterm Strich 350 000 Jobs zusätzlich geschaffen.* => Konjunktion

33. *dann werden diese speziellen Aspekte der Komplexe deutlich miterlebt, und die Jugendlichen müssen dagegen anarbeiten.* => Referenz

Arten von Antezedenten

a) Antezedent=Eigenname (bei Demonstrativpronomen nur im Deutschen)

34. Hast Du Martin gesehen? Der war heute nicht im Büro.

b) Antezedent=NP

35. *This past spring, the U.S. Department of Education issued [a report, The Condition of Education 2000]. The report found that the benefits of attending college are greater today than ever before. [...]*

36. *The United States will be taking [three very important messages] to Johannesburg. America's vision for Johannesburg is to build on these three messages [...].*

37. *During the November/December ministerial, we created consensus in [some very important areas]. These include [...]*

38. *At present, [the only sure way to circumvent the problem of immune rejection would be to create an ES cell line using a patient's own genetic material through nuclear transfer or cloning]. This technique [...]*

39. *Bei den Gebühren für Rundfunk kann ich es mir, verehrter Herr Ministerpräsident, ganz leicht machen: **die** dürfen nur die deutschen Länder erheben. [GO_SPEECH_012]*

40. *As for the licence fee issue, I have a very simple answer, for **these**, Mr Minister-President, are a matter purely for the Länder." [ETrans_FICTION_012]*

c) Antezedent=Adverbialphrase:

41. *[In my young days] we took these things more seriously. We had different ideas then.*

42. *The plane touched down [at last]. Now, we could breathe freely again.*

43. *Is Lina going [to the doctor] today? – No, she went there yesterday.*

44. *How do you like [your new home]? - Oh, it's really wonderful, here.*

45. *Annette ist endlich fertig. Sie ist ein bisschen bummelig und unordentlich, wie ich [als Kind] gewesen sein muss. Damals hätte ich nie geglaubt, dass ich meine Kinder zurechtweisen würde, wie meine Eltern mich zurechtwiesen .*

d) Antezedent=Einzelatz, komplexer Satz, Textabschnitt:

46. *Wir [arbeiten für Wohlstand und Chancen], weil das richtig ist. Wir tun damit das Richtige. [GTrans_ESSAY_006]*

47. *[Und zudem habe sein Vater ihn mit einer Aufgabe betreut, die er an diesem Ort zu enthüllen nicht gesonnen sei]. "Dies ist uns bekannt", sagten die drei Jungfrauen.*

48. *Das ist eine schwierige Frage, weil es natürlich immer darum geht, [wer ist denn hier eigentlich der wichtigere, die die das umsetzen, technisch, oder die, die das journalistisch erdenken]. Ich möchte da keinen Unterschied machen.*

49. *Mit Ausgaben für Forschung und Entwicklung von etwa 3,5 Milliarden Euro gehört die Branche zwar nach wie vor zu den besonders forschungsintensiven Industrien der Bundesrepublik. [Tendenziell ist der Anteil der deutschen Pharmabranche an den globalen Forschungsausgaben der Branche, ebenso wie der Anteil an der Zahl neuer Wirkstoffe, aber rückläufig]. Das bereitet den Experten Sorgen.*

50. *The German pharmaceutical sector still spends more on research than most other industries in the Federal Republic - R&D expenditure totals about 3.5 billion euros. [Even so, its share of global expenditure on pharmaceutical research, as well as its share of new active-substance discoveries, is declining]. This is a cause for concern among the experts.*

3. Comparative Reference

Kohäsionsmittel: Vergleichende Adjektive oder Adverbien, Adjektive im Komparativ

Formen:

	Bedeutung	Englisch	Deutsch
Comparison general (comp =general)	Identität	<i>Same, equal, identical, identically,...</i>	<i>Derselbe (+Nomen), gleich, identisch, + Kasusmarkierung bei Adjektiven</i>
	Ähnlichkeit	<i>Similar, such, so, corresponding, similarly, likewise, ...</i>	<i>Ähnlich (adj+adv), genauso, gleichermaßen, ...</i>
	Kontrast	<i>Other, different, else, differently, otherwise</i>	<i>Anders, unterschiedlich, gegensätzlich, andersartig, ...</i>
Comparison particular (comp =particular)	Quantität	<i>More, fewer, less, further, additional, so/as/equally+Quantifier, z.B. so many</i>	Weniger, mehr,
	Qualität	<i>Adjektive im Komparativ, z.B. better, faster; so/as/equally + vergleichendes Adjektiv, z.B. equally good; vergleichende Adverbien</i>	

ACHTUNG: nicht alle Vorkommen sind kohäsiv:

- Bei allen Vorkommen, die einen Vergleich innerhalb eines Einzelsatzes herstellen, ist der Vergleich grammatikalisiert und deshalb nicht kohäsiv:

51. *It's the same cat as the one we saw yesterday.* = nicht kohäsiv

52. *Paul is bigger than Jim.*

53. *Ihr Apartment ist fast so groß wie meines.*

- Alle Vorkommen können auch allgemein referieren und keinen Bezug zu anderen Textelementen herstellen. Sie sind dann nicht kohäsiv:

54. *Most people have the same breakfast everyday.* = nicht kohäsiv

55. *Alle Befragten beschrieben ähnliche Situationen.* = nicht kohäsiv

Arten von Antezedenten

... bei Comparative Reference nicht weiter klassifizierbar. Rein strukturell kann auf alle lexikalischen Wortarten verwiesen werden, aber auch auf Phrasen und Sätze.

Unterscheidung der Kategorien general - particular

General: Die verwendeten Sprachmittel drücken Ähnlichkeit, Identität oder Kontrast aus, ohne zu spezifizieren worin diese bestehen

Particular: Ähnlichkeit, Identität oder Kontrast wird durch spezifische Eigenschaften ausgedrückt.

4. Kategorien zur Evaluierung der Annotation

Problematic

Mit dieser Kategorie wird die Sicherheit / Unsicherheit der Autorin bei der Bewertung der Zuordnung der Kategorien definiert.

Kategorie	Erklärung
No	Es besteht kein Zweifel an der Richtigkeit der Annotation.
Yes	Es bestehen Zweifel an der Richtigkeit der Annotation.

Neue Regeln (hinzugefügt bei Tabea Buth und Natalija Grbavac)

1. Konjunktionen (z.B. *wobei*, *darüber hinaus*) nicht annotieren, wie im folgenden Beispiel:

Die Container-Theorie der Gesellschaft Die "moderne" Soziologie versteht sich, wenn man eines [ihrer] Textbücher aufschlägt, als "moderne" Wissenschaft von der "modernem" Gesellschaft. [Wobei] ein Gliederungsschema des sozialen Raumes ebenso verborgen wie konsenssicher mitgedacht wird, das man die Container-Theorie der Gesellschaft nennen kann: Erstens setzen danach Gesellschaften - politisch und theoretisch - die "staatliche Beherrschung des Raumes" (J. Agnew und S. Corbridge) voraus.

Hinweis: Meistens sind es Verbletztkonstruktionen = Verb an letzter Stelle des Satzes (allerdings nicht immer).

2. Korrelate, z.B. *darin*, *dass* / *daraus*, *dass* / *dadurch*, *dass* als „correlat-pronadv“ annotieren (sie sind nicht kohäsiv, müssen aber markiert werden), wie im folgenden Beispiel:

[Dies] drückt sich [darin] aus, daß Gesellschaften Staaten (definatorisch) untergeordnet werden; Gesellschaften sind Staatsgesellschaften, Gesellschaftsordnung meint Staatsordnung.