Deliverable no.3 (Arbeitspaket AP3 / Work Package WP3)

Suggestions for a corpuslinguistic analysis of cohesion

Kerstin Kunz, Karin Maksymski, Erich Steiner, November 2009

Contents

After giving an overview of the resources for representing knowledge about languages[1] and presenting conceptualizations and systemic features of cohesion in English and German[2], in the following we will focus on methodological aspects. The intended analysis of cohesion will be corpus-based, thus enabling us to address the question of the use, i.e. relative frequencies, of different cohesive devices. In a first part, we outline the advantages of a **corpus-based analysis** on the basis of a few examples taken from the CroCo corpus[3]. In a second part, we will take a closer look at this corpus to determine whether it can be used for an analysis of cohesion and where **changes** (in registers, in size, in annotation levels, etc.) need to be made. The implications concerning language contact resulting out of a corpus-based analysis will be addressed in Work Package 4.

1. A corpus-based analysis of cohesion

An analysis of cohesion, it could be argued, does not necessarily require a large corpus to find out about the number of different cohesive devices and about the various contexts in which they are used (cf. Kunz 2009 for a contrastive account of co-reference). A theoretical approach (in the sense of WP2) would be sufficient to define the range of different cohesive devices provided by different language systems. Furthermore, examining text samples on an exemplary basis could give an insight about the instantiation of different cohesive devices in a text and thus complement theoretical approaches in terms of use and function. This will be illustrated on the basis of the short extract below taken from the CroCo corpus:

English original	German translation
'*WHY DO YOU want me to go?*' *I asked her the night before.* '*Because if you don't go, I'll have to go to prison.*' *She picked up the knife.* '*How many slices do you want?*' '*Two,*' *I said.* '*What's going in them?*' '*Potted beef, and be thankful.*' '*But if you go to prison you'll get out again. St Paul was always going to prison.*' '*I know that*' *(she cut the bread firmly, so that only the tiniest squirt of potted beef oozed out) ...* '*but the neighbours don't. Eat this and be quiet.*' *She pushed the plate in front of me. It looked horrible.* '*Why can't we have chips?*' '*Because I haven't time to make you chips. There's my feet to soak, your vest to iron, and I haven't touched all those requests for prayer. Besides, there's no potatoes.*' (EO_FICTION_008)	*"Warum willst du, daß ich hingehe?" fragte ich sie am Abend vorher.* *"Weil ich, wenn du nicht gehst, ins Gefängnis komme." Sie griff nach dem Messer. "Wieviel Scheiben willst du?"* *"Zwei", sagte ich. "Was machst du drauf?" "Sülze, und sei gefälligst dankbar."* *"Aber wenn du ins Gefängnis kommst, kommst du auch wieder raus. Der heilige Paulus war auch dauernd im Gefängnis."* *"Ich weiß" (sie schnitt das Sandwich mit fester Hand durch, so daß nur ein ganz kleines bißchen Sülze an den Seiten herausquoll). "Aber die Nachbarn wissen es nicht. Iß jetzt und sei still."* *Sie schob den Teller vor mich. Er sah gräßlich aus. "Wieso gibt es keine Pommes?"* *"Weil ich keine Zeit habe, dir welche zu machen. Ich muß noch ein Fußbad nehmen und deine Bluse bügeln, und dabei habe ich mit den vielen Bitten um Gebete noch nicht einmal angefangen. Außerdem sind keine Kartoffeln da."* (GTrans_FICTION_008)

[1] Work Package 1 (AP1): http://fr46.uni-saarland.de/uploads/media/AP1_final_01.pdf
[2] Work Package 2 (AP2): http://fr46.uni-saarland.de/uploads/media/AP2_final.pdf
[3] http://fr46.uni-saarland.de/croco/index_en.html

We will go through the text sentence by sentence, showing the various cohesive devices used.

In example (1) below, a cohesive tie is set up via a personal pronoun and thus by personal reference in English while in the German translation (2), cohesion is created by demonstrative reference, i.e. via the colloquial form of a pronominal adverbial:

(1) *'How many slices do you want?' 'Two,' I said. 'What's going in them?'*

(2) *"Wieviel Scheiben willst du?" "Zwei", sagte ich. "Was machst du drauf?"*

In the next example, the position and/ or type of the referring item is changed in the translation. More precisely, there is demonstrative reference in the first clause of the English text (3) where there is nominal ellipsis in German (4); furthermore the second clause of the English text exhibits lexical ellipsis whereas the German second clause makes us of a personal pronoun (reference), which is less marked than the demonstrative pronoun in the first English clause.

(3) *'I know that' (...) ... 'but the neighbours don't.*

(4) *"Ich weiß" (...). "Aber die Nachbarn wissen es nicht.*

In the next sentence, the exophoric reference in the original (5) is eliminated in the translation (6) and cohesion in the German version is established on the basis of lexical cohesion (plus the definite article as a case of reference) only (*iß* ⇔ *Teller*):

(5) *Eat this and be quiet.' She pushed the plate in front of me.*

(6) *Iß jetzt und sei still." Sie schob den Teller vor mich.*

There are also instances, where both languages make use of the same category of cohesive devices (here: personal reference), but use them to refer to different things. The ambiguity in the English sentence (7), where *it* can refer to the plate as well as to the bread with potted beef on it mentioned a few sentences before (one would assume that the last one is the intended reference), becomes unambiguous in the German sentence (8). Here, it is possible to refer only to one of these items, because of the different genders of Brot (bread) and Teller (plate) in German. The translator makes a reference to the plate by choosing *er* instead of *es*:

(7) *She pushed the plate in front of me. It looked horrible.*

(8) *Sie schob den Teller vor mich. Er sah gräßlich aus.*

Not only can a translation resolve ambiguity in the original, which then is a case of "explicitation", but it can offer alternative possibilities for creating cohesion. In the last sentences of the text, lexical cohesion in the English version is used by repeating *chips* (9). The German text (10) shows that the author could have opted for *substitution* using "welche":

(9) *'Why can't we have chips?' 'Because I haven't time to make you chips.*

(10) *"Wieso gibt es keine Pommes?" "Weil ich keine Zeit habe, dir welche zu machen.*

In our example above, we have analysed occurrences of different types of cohesion in one original text and compared these with their translational equivalents in the target text. We have interpreted the findings on the basis of the theoretical assumptions stated in WP2.

However, this hermeneutic and example-based approach limits the scope of our linguistic study to the range of possibilities that create cohesion in one text only (and, even within this one text, not all possibilities have been captured). In a sense, it presupposes that we have a good initial idea of what range and types of phenomena we are looking for, using the text to be analysed as source of illustration and possibly as source of additional findings, provided there are any in this particular text.

Below, some findings from the CroCo corpus will be presented as examples of how a corpuslinguistic analysis complements the theoretical and example-based approaches described above, thus yielding additional and above all different types of insights.

The following analyses were carried out on subcorpora of the CroCo corpus[4]. As for the first study, we investigated personal reference in the sense of Halliday & Hasan (1976:43 ff). More precisely, we were interested in the use of the neuter pronoun in English and German: *it* in English and *es* in German. Although both languages have this pronoun, assumptions about contrastive differences in frequency and use were already formulated on the basis of theoretical considerations presented in the deliverable for WP2.

Table 1 below shows the findings from the corpus query into the two corpora FICTION and ESSAY.

ESSAY	***EO***		***Etrans***		***Gtrans***		***GO***		
tokens	31248		37344		31579		31291		
it	118		177		99		130		***es***
It	40		36		42		41		***Es***
total	158	0,51 %	206	0,55 %	141	0,45 %	171	0,55 %	**total**
Cohesive *It*	19	47,9 %	21	58,3 %	6	14,29 %	5	12,20 %	**Cohesive *Es***
FICTION	***EO***		***Etrans***		***Gtrans***		***GO***		
tokens	31316		33302		31908		30767		
it	330		335		260		244		***es***
It	61		72		50		65		***Es***
total	391	1,25 %	407	1,22 %	310	0,97%	309	1,0%	**total**
Cohesive *It*	38	62,3 %	35	48,61 %	16	36,29 %	23	32 %	**Cohesive *Es***

Table 1: Query "it/es" in the CroCo registers ESSAY and FICTION

As the table illustrates, both registers comprise two original subcorpora, one in English and one in German, as well as subcorpora containing translations from these into the respective other language. Queries were made as to the number of all instances of the neuter pronoun, inside the sentence (third row ***total***), at the beginning of the sentence (fourth row), as well as

[4] for further details on corpus design see http://fr46.uni-saarland.de/croco/corpus_design.pdf

the sum of these two queries. As cohesive uses could not be distinguished from non-cohesive uses automatically, we concentrated on occurrences of the neuter pronoun at the beginning of the sentence. Here, each instance was assigned manually to one of the above-mentioned categories. The number of neuter pronouns used cohesively at the beginning of the sentence is displayed in the fifth row of the table.

First of all, the findings reveal that in most respects, there is a greater similarity between originals and translations in the same language than in texts across languages in both registers, pointing to a contrastive difference in terms of frequency of the neuter pronoun. Comparing frequencies in originals with those in translations may also give some hints about translations as possible media of language contact.

Yet, it is not the contrast in frequency alone we are interested in when considering more extensive data. The findings from the queries also permit to look into the reasons lying behind the differences in frequency, particularly when comparing translations and originals. For instance, investigating all instances of the neuter pronoun occurring at the beginning of the sentence, we see that *Es* in German more often has a non-cohesive function than a cohesive one whereas English *It* tends to have a cohesive function as often as a non-cohesive one.

Examining the cohesive occurrences more closely we detect that, in both registers of English, the neuter pronoun is often employed in cases where a personal pronoun indicating feminine or masculine gender is used in the German parallel texts. This is illustrated by an extract from an English original and its German translation below:

(11) *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. [EO_ESSAY_003]*

(12) *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. [GTrans_ESSAY_003]*

Thus, the findings corroborate the assumptions from theoretical and example-based approaches (see deliverable for WP2). Yet, the output of the queries also shows that the contrast in gender of personal pronouns is not the only reason for the contrast in frequency. The German translations from the English originals allow for an investigation of the different translational options for *it*. The comparison reveals that an equivalent translation for *it* in German may be realized via a pronominal adverbial (as in (14), which is additionally focussed by having the German demonstrative article “das” instead of the English personal pronoun “they”), a demonstrative pronoun (as in (16), yet again giving a feature of increased deictic (demonstrative fore to the German translation), and also, via cohesive ellipsis, or a full lexical noun phrase.

(13) *We work for prosperity and opportunity because they' re right. It' s the right thing to do. [EO_ESSAY_006]*

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

(15) *And he answered them courteously that they should speak on, for he had not come so far and so wearily simply in order to turn back.*

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]

(16) *Und er erwiderte ihnen artig, daß sie weitersprechen sollten, denn er habe die Mühsal und Beschwerden des weiten Weges nicht auf sich genommen, um nun kehrtzumachen. Und zudem habe sein Vater ihn mit einer Aufgabe betraut, die er an diesem Ort zu enthüllen nicht gesonnen sei. "Dies ist uns bekannt", sagten die drei Jungfrauen. [GTrans_FICTION_002]*

Hence, while in English the neuter pronoun is quite frequently used to establish cohesion, German often makes use of other cohesive devices. Thus, examining the instantiations of one particular cohesive device seems a good source for investigating the range of possible realizations for creating cohesion in the two languages.

Quantitative information from particular corpus queries such as the one obtained above cannot be obtained on the basis of an exemplary or theoretical approach in which only assumptions can be made and examples given.

However, in order to gain a more comprehensive picture of the distribution and function of cohesive devices holding for texts produced in English and German, we would have to look into more than just two different registers. Furthermore, an investigation of more linguistic devices establishing cohesion would be necessary as well together with an analysis of co-occurrences of different means of cohesion in particular texts across registers and languages. We intend to do this in the large-scale project proposed for the future[5].

Apart from the general tendencies described above, register dependent features across languages can be discerned on the basis of the findings from corpus queries. For instance, the distribution of the neuter pronoun is higher in FICTION than in ESSAY across languages; what is more, there are more cohesive instances at the beginning of sentences in FICTION than in ESSAY[6]. Examining these instances more closely reveals that many instances which are cohesive in FICTION have a wide and partially ambiguous scope; in contrast, most of the instances traced in ESSAY have a more limited and well-defined scope. Finally, we also detect that the findings for the English registers are more similar to each other than the findings for the German registers. Examining the respective instances of the neuter pronoun in the German texts at the beginning of the sentence we find that this may mainly result from the fact that in the German FICTION texts, more instances of the neuter pronoun seem to have a very wide scope than in the English FICTION texts. Moreover, while the scope of the cohesive device can be defined very easily in most cases in ESSAY, the scope of the antecedent in FICTION quite often remains rather vague.

For an illustration, consider the following extract of a German original text of the register FICTION:

(17) *Er war ein eher ängstliches Kind, sagte die Mutter.*
Er log nicht. Er war anständig. Und vor allem, er war tapfer, sagte der Vater, schon als Kind. Der tapfere Junge.

[5] Here, we have to be aware of the fact that quite often of a large-scale manual analysis is required to account for different functions of the same cohesive device.
[6] It has to be noted that in some cases it is very difficult to delineate cohesive from non-cohesive instances

So wurde er beschrieben, auch von entfernten Verwandten. Es waren wörtliche Festlegungen, und sie werden es auch für ihn gewesen sein.[GO_FICTION_008]]

Here, the neuter pronoun *es* at the beginning oft the last sentence points to a text passage spanning the six preceding sentences. In fact, the scope of the pronoun in the German text is not clearly discernible. This may have caused the translator to use a demonstrative pronoun instead in the corresponding English translation, as shown below:

(18) He was rather a timid boy, said our mother.
He didn't tell lies. He was well-behaved, and above all, said our father, he was brave even as a child. People described him as that brave boy, even distant relations. These were verbatim observations, and they will have been meant for him too. [ETrans_FICTION_008]

With our corpuslinguistic analysis, we also intend to identify new tendencies in frequency, function and use of cohesive devices dependent on register, so far largely neglected in comparative research into English and German. For instance, the literature on the subject usually makes the assumption that the use of the German demonstrative pronouns *der, die, das* is restricted to registers of spoken language (see deliverable for WP2). Indeed, tcorpuslinguistic comparison shows that more occurrences are traced in the German subcorpora FICTION and SPEECH, which approximate registers of spoken English to some extent, than in other registers of the CroCo corpus (compare the findings for the German subcorpora of the CroCo Corpus as indicated in Table 2).

	Das	**das**	**Der**	**der**	**Die**	**die**
GO_SPEECH	74	99	1	3	1	3
Gtrans_SPEECH	18	20	-	3	-	-
GO_FICTION	40	73	8	7	7	5
Gtrans_FICTION	28	72	3	7	3	4
GO_POPSCI	52	58	1	3	-	1
Gtrans_POPSCI	22	22	1	2	-	1
GO_TOU	11	20	2	7	-	2
Gtrans_TOU	6	8	-	2	1	-
GO_SHARE	23	21	2	1	1	-
Gtrans_SHARE	21	25	1	2	-	-
GO_ESSAY	47	43	-	1	1	2
Gtrans_ESSAY	26	23	-	-	-	-
GO_INSTR	11	9	-	-	-	-
Gtrans_INSTR	11	7	-	-	-	-
GO_WEB	8	23	-	1	-	2
Gtrans_WEB	11	16	-	1	-	-

Table 2: Frequencies in numbers of the demonstratives der, die, das in the German subcorpora of the CroCo corpus

The following example displays one instance taken from GO_FICTION and its English translation:

(19) *Ich lenkte mich ab, suchte Schlaf, vergaß, sank weg - prompt schoss mir das entscheidende Bild in den Kopf: mein Freund Axel am Tisch der Mensa, neben uns die Zeitung, aufgeschlagen die Seite mit einer Überschrift zum beginnenden Prozess gegen diesen Richter, der am Volksgerichtshof mindestens 230 Todesurteile gefällt hatte. Sogleich stellte sich der Ton zu diesem Bild ein, der bittere, verächtliche Satz, den Axel hatte fallen lassen und der mich erst jetzt, im Bett, wie eine böse Erleuchtung traf: Der hat das Urteil für meinen Vater fabriziert, der und der Freisler. [GO_FICTION_001]*

(20) *I distracted myself, sought sleep, forgot, drifted off - and promptly the crucial image popped into my head: my friend Axel at the cafeteria table, the newspaper next to us opened to a headline about the start of the trial of this judge who had passed at least 230 death sentences at the People's Tribunal. Immediately the soundtrack to this image kicked in, Axel's bitter, contemptuous words which hit me only now, in bed, like an evil epiphany: He fabricated my father's verdict - him and Freisler. [ETrans_FICTION_001]*

As the example shows, partial equivalence in meaning in English can be obtained by employing a personal pronoun.

Example (22) below illustrates, that another option is to use a proximate demonstrative pronoun:

(21) *Bei den Gebuehren fuer Rundfunk kann ich es mir, verehrter Herr Ministerpraesident, ganz leicht machen: die duerfen nur die deutschen Laender erheben. [GO_SPEECH_012]*

(22) *As for the licence fee issue, I have a very simple answer, for these, Mr Minister-President, are a matter purely for the Laender." [ETrans_SPEECH_012]*

However, we also detect some occurrences in other registers, e.g. TOU, SHARE and POPSCI (see Table 2). These findings either may point to the fact that there is no clear preference for *der, die, das* to be employed in registers of spoken language or that they might be indicators of language change towards making the standard more "spoken" in general.

As the CroCo corpus currently does not comprise registers of spoken language, we can neither measure the respective frequencies in these registers nor compare them to those that are already available in the existing corpus. Thus an expansion of the corpus to other registers is required in order to obtain a more comprehensive picture. As other previous studies of the CroCo corpus have shown, this seems to be particularly desirable for the investigation of instantiations of cohesive ellipsis and substitution (see Klein 2007 and Birster 2007).

The findings in Table 2 also reveal that the translations exhibit much lower values than the originals in most registers. The figures for each morphological form show that a differentiation has to be made between the neuter form of the demonstrative pronoun on the one hand, and the masculine and feminine form on the other hand, possibly not only in terms of frequency but also with respect to the pragmatic function.

The above examples of corpuslinguistic research may serve as an indication that empirical corpuslinguistic studies are an essential step towards a contrastive model of cohesion in English and German not only because they yield statistics about the frequency of

cohesive devices, but also because they allow a more comprehensive interpretation and because they show a wider range of realizational possibilities for one and the same cohesive relation than one would suspect without corpus evidence. .

In particular, the study of cohesion in a large amount of texts permits to answer research questions that regard:

a) additional possibilities not covered in purely theoretical approaches		⇨ Which devices do exist?
b) the use of cohesive devices c) the nature of the cohesive ties set up between a cohesive device and its antecedent d) the nature of cohesive chains	• concerning the actual utilization of the theoretical possibilities	⇨ Which of them are used?
	• in the sense of frequency	⇨ How often are they used? ⇨ Are there typical co-occurrences in texts of the same language?
	• in relation to their cognitive function	⇨ Which mechanisms of cognitive text processing do they reflect?
	• in relation to their pragmatic/ interpersonal function	⇨ In which contexts of situation/ registers do they occur? ⇨ Which cohesive devices do co- occur in which registers?
	• in translations	⇨ What can be said on their range, frequency and function in translations?

Table 3: Research questions in a text-based contrastive perspective on cohesion English-German

An example-based theoretical and hermeneutic approach is an important exploratory step to generate assumptions about the range and use of cohesive devices in English and German. The next step, i.e. the one from assumptions to actual linguistic evidence, though, is only possible on the basis of a text corpus for both languages. Including translations in the analysis is especially interesting here: not only do they hint at analogies between cohesive devices in the two languages, they also show areas where one-to-one equivalents are not preferred, or even non-existent[7].

2. Review of the CroCo corpus

[7] It is, of course, important here to include the work of more than one translator in the corpus; otherwise the findings could be due to an idiosyncratic translation strategy

This section provides an overview of those architectural features of the CroCo corpus that are relevant for the investigation of cohesion in the proposed project. We evaluate the existing CroCo architecture against the background of possible and necessary extensions in terms of corpus design, annotation and alignment as well as techniques for querying the information encoded in the corpus.

2.1 SIZE AND COMPOSTION, REGISTERS, REFERENCE CORPORA

As described in earlier works of the research group, the architecture of the CroCo corpus has already yielded results for a number of different investigations on various linguistic levels. Although the texts are well balanced and representative in terms of text size and number of texts per register (Neumann 2005), for the investigation of cohesion in English and German some modifications have to be made, particularly with respect to register.

The examples for a corpuslinguistic analysis above have shown that in order to trace as many different cohesive devices as possible, the existing corpus requires an expansion to registers of spoken language. For instance, the values for devices such as substitution and ellipsis do not allow for significance testing, and therefore cannot be regarded as being representative. Hence, especially those types of cohesive devices, which only occur to a very small extent in the existing corpus require further analysis in other registers, especially those containing spoken dialogue.

2.2 ANNOTATION LAYERS AND ALIGNMENT

The corpus of the CroCo project is encoded on five different levels of annotation, i.e. it is enriched with information about morphology, parts of speech, syntactic functions and types of syntactic structure (chunks), as well as about clauses[8]. A sixth layer comprising the annotation of semantic relations is currently under construction. In addition, the parallel corpora are aligned on several linguistic levels[9]. The existing annotation is sufficient for tracing the linguistic items which may function as cohesive devices in English and in German. For instance, particular cohesive devices establishing reference or substitution can be investigated on the part of speech level. Other types such as cohesive conjunctions can be identified when examining the part of speech as well the chunk level. In addition, for the investigation of ellipsis combined queries into different layers of annotation can be employed.

However, most of these items do not necessarily have a cohesive function in particular linguistic environments: they may be non-referential, e.g. *it*/ *es* employed as dummy subject, or they may point to the extralinguistic context of situation, e.g. demonstratives employed for exophoric reference, or they may rather function as clause-internal grammatical devices such as conjunctions introducing subordinate clauses, or as clause internal ellipsis. Some devices may even be assigned to different categories of cohesion depending on the structure of the linguistic context (e.g. pronominal adverbs may either function as conjunctions and/ or may establish demonstrative reference). Hence, another layer of annotation is required to distinguish between cohesive and non-cohesive use. Furthermore, only the semantic layer includes information as to the antecedent of the cohesive devices, the type of cohesive tie as well as the nature of the cohesive chain and thus will yield enough information for a sound interpretation in terms of lexical cohesion as soon as the annotation is finished. The currently

[8] for further information on annotation of the CroCo Corpus see http://fr46.uni-saarland.de/croco/deliverable4.pdf

[9] for further information on alignment of the CroCo Corpus see http://fr46.uni-saarland.de/croco/corpus_align.pdf

available layers do not permit an immediate identification of antecedents, nor do they directly encode the nature of cohesive ties and chains. Thus, the investigation of co-reference in particular would require additional encoding of antecedents and co-reference chains. For this purpose, guidelines for annotators need to be developed beforehand containing rich information as to the classification and identification of cohesive ties and chains, the delineation of cohesive and non-cohesive relations and the special nature of the cohesive relation established.

2.3 QUERIES – TECHNIQUES AND SOFTWARE

At the current stage of the CroCo project, the part of speech layer can be queried with cqp (Christ 1994) and the chunk level with cqp2 in some registers. However, some registers still require conversion into the adequate format. Furthermore, only the subcorpora SHARE, FICTION and SPEECH can be in queried in cqp in terms of alignment on sentence level, i.e. when querying particular parts of speech in one parallel corpus, the output also contains the aligned sentence in the other parallel corpus.

In addition, strings can be queried in all texts with grep and egrep and each layer can be queried text per text with the query tool of the MMAX2 annotation tool (Müller & Strube 2006).

However, different elaborate *perl* scripts are required in order to be able to query every possible piece of information about each layer of annotation and in order to do queries into combined layers of annotation. This needs to be done either by an IT specialist or a computer linguist.

In addition, an expansion of the corpus as highlighted in 2.1 and an extension of the annotation to other layers as described in section 2.2 would require implementing and developing other tools in order to query all possible aspects of cohesion.

3. Literature

Birster, L. 2007. Kohäsionsmittel im Englischen und Deutschen - ein Vergleich anhand ausgewählter Phänomene. Diploma thesis. Universität des Saarlandes, Fachrichtung 4.6. Angewandte Sprachwissenschaft sowie Übersetzen und Dolmetschen.

Christ, O. 1994. A modular and flexible architecture for an integrated corpus query system. In: Proceedings of the 3rd Conference on Computational Lexicography and Text Research. Budapest, Hungary.

Halliday, M.A.K. and R. Hasan. 1976. *Cohesion in English.* London, New York: Longman.

Klein, Y. 2007. Übersetzungsspezifische Eigenschaften - eine korpusbasierte Studie am Beispiel der Kohäsion. Diploma thesis. Universität des Saarlandes, Fachrichtung 4.6. Angewandte Sprachwissenschaft sowie Übersetzen und Dolmetschen.

Kunz, K. forthcoming. English and German nominal Coreference. A study of political essays.

Kunz, K. / Maksymski, K. / Steiner, E. 2009. Brief review of existing resources for representing knowledge about languages (English and German). Deliverable No. 1 of the GECo Project (http://fr46.uni-saarland.de/index.php?id=geco).

Kunz, K. / Maksymski, K. / Steiner, E. 2009. Cohesion - conceptualizations and systemic features of English and German. Deliverable No. 2 of the GECo Project (http://fr46.uni-saarland.de/index.php?id=geco).

Müller, C. and M. Strube. 2006. Multi-Level Annotation of Linguistic Data with MMAX2. In: Braun, S., Kohn, K. & Mukherjee, J. (eds.), *Corpus Technology and Language Pedagogy. New Resources, New Tools, New Methods.* Frankfurt: Peter Lang, 197-214.

Neumann, S. 2005. Corpus Design. Deliverable No. 1 of the CroCo Project. (http://fr46.uni-saarland.de/croco/corpus_design.pdf).