Deliverable no.4 (Arbeitspaket AP4 / Work Package WP4)

Modelling of language comparison and language contact within the research architecture proposed

Erich Steiner, December 2009

Contents

1. The contrastive perspective

Figure 1 lists research questions in a text-based contrastive perspective on cohesion English-German. For further details see Deliverable no.3 *Suggestions for a corpuslinguistic analysis of cohesion*, under 1. *A corpus-based analysis of cohesion* [1].

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?

Figure 1: Research questions

---

[1] http://fr46.uni-saarland.de/uploads/media/GECo_AP3.pdf

2. Traces of language contact in our data

2.1 TERMINOLOGICAL CLARIFICATION

In language contact research (cf. e.g. Thomason 2001: 66ff), *borrowing* usually refers to importing some structure or configuration from a contact language into a receptor language. Imperfect learning plays no role; often, though not necessarily, borrowing is enacted by native speakers of the receptor language and should be relatively permanent rather than just short-term. *Shift* refers to traces left by one's first (often native) language when using a second (often non-native foreign) language. Both borrowing and shift are forms of *interference*.

The three phenomena so far described are usually applied to language contact situations other than translation.

Finally, *shining-through* was coined in Teich (2003) for investigations of translations, referring to some kind of empirically identifiable trace left by the source language through the source text in some target text, (and potentially in the target language?). *Shining through* could be identified in the domain of individual constructions, but usually is predicated on proportional frequencies of constructions or constructional configurations. It is originally explained largely by systemic differences between the language systems (and possibly also the registers) involved.

2.2 APPROACHES TO THE ANALYSIS OF OUR DATA

a) Direct comparisons of frequencies of parallel lexicogrammatical items:

- e.g. *3rd person neuter pronouns* vs. *masculine and feminine pronouns*
- between originals (EO vs. GO) of the same register,
- between SL-TL pairs of the same register (EO vs. GTrans)

b) Frequencies in cohesive vs. non-cohesive usage for contrastively corresponding lexicogrammatical means:

- e.g. 3rd person singular neuter pronouns in cohesive vs. non-cohesive usage
- between originals (EO vs. GO) of the same register,
- between SL-TL pairs of the same register (EO vs. GTrans)

c) Comparisons of frequencies between originals and translations within the same language:

- either of 2.1 and b), but
- between originals and translations within one language and one register (EO vs. ETrans).

The independent variables would be language, register, and status (as original vs. translation). The dependent variables would be (proportional) frequencies between different lexico-grammatical structures in cohesive (possibly vs. non-cohesive) function.

d) Comparisons in terms of frequencies:

- of corresponding cohesive relationships, but realized through different lexicogrammatical realizations
- of different types of cohesion (e.g. referential vs. substitutional vs. elliptical vs. fully lexical) for corresponding textual (coherence?) relationships (e.g. of co-reference or co-classification).

If we are investigating such comparisons between originals and translations within the same language, we are looking at contact phenomena in addition to contrast.

The independent variables under d) would be language, register, and status (as original vs. translation). The dependent variables would be (proportional) frequencies between different lexico-grammatical structures realizing corresponding cohesive relationships, or else *different* type of cohesion, but the same type of *coherence relationship*.

So, we might have, as an example, a personal pronoun in a source text, corresponding to a demonstrative (article/ pronoun) phrase in the target text.

Or we might have cohesive reference in one case, vs. lexical cohesion or ellipsis of substitution in the other, in all cases establishing co-reference or co-classification.

e) Comparisons of vagueness/ ambiguity of scope associated with usage of different realizations of some cohesive relationship, or even of different cohesive relationships, again also of the different frequencies of each type

Comparisons between originals and translations within one language should open a window into contact studies.

An example would be the analysis of using a generic fully lexical phrase vs. a definite phrase vs. a phrase pre-modified through possessive vs. deixis vs. demonstrative vs. a phrase headed by a pro-form (demonstrative vs. pronoun). The interest would here not be in the phenomenon as such, but in the different kinds of ambiguity and/ or vagueness associated with each case. Contact phenomena should play a role in investigating originals and translations within one language and within one register.

f) Comparison of the breadth of variation between and within comparative registers in the two languages (cf. Neumann 2009)

Initially, this would be a clear study of contrasts. If we then could show that this variation is different between translations and originals again within one language, we would be on the trace of contact phenomena.

For example, if we can show that the breadth of variation in terms of logico-semantic relationships is smaller or larger in translations than in originals within one language (or register), or that register differences are smaller or larger between translations and originals, we would have candidates for contrast or even contact, depending on the corpora examined (and on their difference to proportionalities obtaining in the reference corpora).

g) Contrastive register-specificity of cohesive configurations, and again their behaviour under contrast vs. contact conditions

These configurations could be length of lexical or referential chains, density of chains, etc.

2.3   SOME GENERAL THOUGHTS ABOUT TRACES OF CONTACT PHENOMENA IN OUR DATA

Contrast in type or frequency of realization – highly interesting in itself – presupposes some assumed *tertium comparationis*, for example difference in frequencies in the realization of one and the same functional relationship (say, *personal reference,* or difference in the *referential status* of some co-referent in the sense of Figure 3 in Deliverable 2 *Cohesion - conceptualizations and systemic features of English and German*[2]). The functional relationship is the *tertium comparationis*. Differences in the mere frequencies of some lexicogrammatical realization as such may be interesting as possible indicators, but are not indicative of a real contrast by themselves.

In order to identify instances of *contact*, rather than of *contrast* only, we need to show that cohesive relationships in texts in a contact situation – in our case prototypically translations[3] – behave differently from these relationships in non-contact situations. In other words, we need to trace differences between originals and translations within the same language (i.e. not ST and TT) and within the same register. These differences can then be specified along the axes indicated above, and others.

There is, of course, at least one other and additional explanation for variation between originals and translations within one and the same language:

The phenomena observed in translations could be due to the *translation process*, as we have often argued within the CroCo project (e.g. Hansen-Schirra, Neumann & Steiner (2007), cf. also Alves et al. in press), rather than to interactions between source language and target language, or between source-text and target-text. And in a trivial sense, this is necessarily true as each TL-text is the outcome of a translation process. However, a careful analysis as to which properties of the product (e.g. "explicitness") is due to "shining-through" and which to the problem solving process as such should enable us to locate the source of the phenomenon observed with some accuracy. Hence our experimental side-line work on Work Package 4.2 (in progress).

---

[2] http://fr46.uni-saarland.de/uploads/media/GECo_AP2.pdf
[3] Of course, translations are by no means the only contact variety, cf. results in e.g. House (2002), but most of the existing literature on language contact altogether (Thomason & Kaufmann 1988, Thomason 2001; Haspelmath 2004; Siemund & Pietsch 2008; Stolz, Bakker, and Palomo . eds. 2008)

3. Our data and our findings in relationship to studies of language contact in general

How do we assume our findings to connect with findings about language contact in other relevant research contexts? We assume that they complement findings about contrastive grammar (and texts, where such findings exist), especially those (relatively few, as yet) which are based on empirical investigations.

- Freiburg Graduiertenkolleg 1624/1
  "Frequenzeffekte in der Sprache"

- Oslo
  (Languages in Contrast http://www.hf.uio.no/forskningsprosjekter/sprik/english/)

- Meaning and Understanding Across Languages, Centre for Advanced Study at Norwegian Academy of Sciences
  (http://www.cas.uio.no/research/1011acrosslanguages/abstract.pdf)

- Manchester translational English Corpus
  http://www.llc.manchester.ac.uk/ctis/research/english-corpus/

- Hong Kong
  http://www.engl.polyu.edu.hk/department/academicstaff/ChristianMIMMatthiessen.html

- Hamburg
  http://www.uni-hamburg.de/sfb538/

- cf. Siemund & Kintana (eds.) 2008

- Krug 2000 (esp. 18ff).

For a general discussion to which we would like to make a contribution cf. Haspelmath (2009).

A particularly interesting line of thought in the area of language contact would stipulate the following:

- Language contact is clearly sensitive to socio-cultural factors (external, such as prestige, number of speakers, intensity of contact (Thomason 2001: 69ff), power relationships of various sorts, but also degree of establishment of the genre in question) and language-internal factors (internal, systemic, such as whether the levels involved are lexicon, morphology, syntax, the typological relatedness of the languages involved, etc.).

- It happens classically in the form of either borrowing or else shift (Thomason & Kaufmann 1988, Thomason 2001: 66ff); but cf. Koller (2000: 113ff) for translation as language contact, as well as House in many places and Becher, Höder and Kranich 2009.

- The postulate of (at least structural) contact requires (Thomason 2001: 93ff) effects in more than just one local system of the target language, an assumed source language, shared structural features in both source and target language, and evidence of a state of affairs prior to the assumed contact.

- Furthermore, contact happens on a hierarchy of levels (from phonology thorough lexis and syntax to morphology). The question of which level(s) are more affected by contact processes has to do with whether we are in a borrowing or in a shift scenario, and it depends heavily on the intensity of the contact situation. Furthermore, contact-phenomena on the more "systemic levels" (syntax, morphology) may be more far-reaching in terms of impact than those on the less grammaticalized levels (lexis, text and discourse in particular), as these latter are more directly open to influence from socio-cultural norms.

- If this general assumption is valid, then cohesion would appear to be an obvious early stage of language contact, often more so than some more "systemic levels", at least in cases of "borrowing" rather than "shift" (cf. Siemund & Pietsch 2008: 370), but it would interestingly involve frequencies and proportionalities of lexical as well as grammatical and phonological means, rather than just lexis, the obvious candidates for early stages of language contact.

Our additions to more established studies would be the following:

1. In terms of borrowing scales, we would be extending the hierarchy of levels "upwards" in looking at cohesion, rather than on lexicogrammar directly (which latter of course always plays a role, as all cohesive configurations have to be realized lexicogrammatically)

2. While our operationalizations always need to make reference to lexicogrammatical and cohesive structures, we would model configurations of these in terms of *properties* of texts/ discourses (linking up with work such as House, Traugott, etc.)

3. We would extensively use empirical methodologies, studying translations between English and German as a contact variety.[4] If we are trying to situate our type of data relative to classical empirical data in linguistics (corpora, responses to questionnaires, experimental data, ontogenesis, phylogenesis, testing predictions based on ranked constraints derived from models, but cf. the comprehensive discussion of types of data in linguistics in Haspelmath (2009) and the other contributions in that same volume), then we are extensively using corpora, combined with experimental data in a restricted domain. Other types of data are close to our methodology (historical/ ontogenetic, using models with ranked constraints), but currently not part of the architecture.

---

[4] It is interesting to note that Haspelmath (2009: 158) classifies „translation" among behaviour-oriented linguistic data, yet on the next level as „meta-linguistic" as opposed to the "purely linguistic data", which latter include the bulk of our own data as "non-controlled, spontaneous" data. I think that what makes source- and target text pairs (alignments in our architecture) "meta-data" in his sense is precisely their property of being multi-functional paraphrases of each other (cf. Steiner 2001: 179f)

4. A possible case study: conjunctive relations in English, German, Chinese and Japanese

Conjunctive relations may be of particular interest in our line of research because they are an established area of interest on the borderline between discourse and grammar (cf. Matthiessen & Thompson 1988, Fabricius-Hansen 1999, 2005), language contact studies and studies of language change, and because the semantics underlying them permeates entire language systems as "fractal systems" (cf. Halliday & Matthiessen 1999: 222ff). One line of research we envisage therefore is one within which we contrast and compare the realization of conjunctive relations in cohesion and lexicogrammar, covering Chinese, Japanese, English and German as languages in and established collaboration (cf. also Matthiessen et al. 2008).

5. References

Alves, F., Pagano, A., Neumann, S., Steiner, E. and S. Hansen-Schirra. In press. "Translation Units and Grammatical Shifts: Towards an Integration of Product- and Process-based Translation Research." In: Shreve, G. & Angelone, E. (eds.). Translation and Cognition. Amsterdam: John Benjamins. 46ff.

Becher, V., Höder, S. and S. Kranich. 2009. A tentative typology of translation-induced language change. Paper given at Workshop Multilingual Discourse Production. 6–7 November 2009. University of Hamburg Research Centre on Multilingualism.

Fabricius-Hansen, C. 1999. Information packaging and translation: Aspects of translational sentence splitting (German – English/ Norwegian). In: *Studia Grammatica*, 47:175-214.

Fabricius-Hansen, C. 2005. Elusive connectives. A case study on the explicitness dimension of discourse coherence. In: *Linguistics* 43-1: 17-48.

Halliday, M.A.K. and Matthiessen, C.M.I.M. 1999. *Construing Experience through Meaning*. London: Cassell.

Haspelmath, M. 2004. How Hopeless is genealogical Linguistics and How Advanced is Areal Linguistics? Review of Dixon, R.M.W., Aikhenvald, A.Y. (eds.): Aral Diffusion and Genetic Inheritance: Case Studies in Language Change. Oxford: Oxford University Press. In: *Studies in Language*, 28, 209-223.

Haspelmath, M. 2009. Welche Fragen können wir mit herkömmlichen Daten beantworten? In: *Forum Zeitschrift für Sprachwissenschaft ZfS*. 2009. 28.1..: 157ff.

Krug, M. G. 2000. *Emerging English Modals . A corpus-Based Study of Grammaticalization. Topics in English Linguistics*. Berlin, etc.: Mouton De Gruyter.

Matthiessen, C.M.I.M. and Thompson, S.A. 1988. The structure of discourse and 'subordination'. In: Haiman, J. and Thompson, S.S. (eds.). *Clause combining in grammar and discourse*. Amsterdam: John Benjamins. 275-329.

Matthiessen, C.M.I.M., Teruya, K. and Canzhong, W. 2008. Multilingual Studies as a Multi-dimensinal space of Interconnected Language Studies. In: Webster, J. (ed.) 2008. *Meaning in Context. Implementing Intelligent Applications of Language Studies*. London: Continuum.

Siemund, P. and Kintana, N., eds. 2008. *Language contact and contact languages*. Amsterdam: John Benjamins (Hamburg Studies in Multilingualism Vol. 7).

Siemund, P. and Pietsch, L. 2008. Contact-Induced Change and Linguistic Universals: The Case of Irish English. In: Stiersdorfer, K., ed. 2008. *Anglistentag 2007 Münster*. Trier: Wissenschaftlicher Verlag. 369-384.

Steiner E. 2005. Some properties of lexicogrammatical encoding and their implications for situations of language contact and multilinguality. In: Franceschini, R., ed. 2005. *In einer anderen Sprache. Sondernummer der Zeitschrift für Literaturwissenschaft und Linguistik* 35, 139. Stuttgart: Metzler Verlag. 54-75.

Steiner, E. 2008. Empirical studies of translations as a mode of language contact - "explicitness" of lexicogrammatical encoding as a relevant dimension. In: Siemund, P. and Kintana, N., eds. 2008. *Language contact and contact languages.* Amsterdam: John Benjamins (Hamburg Studies in Multilingualism Vol. 7). 317-346.

Stolz, T., Bakker, D. and Palomo, R., eds. 2008. *Aspects of Language Contact. Methodological and Empirical Findings with Special Focus on Romanisation Processes.* Berlin: Mouton de Gruyter.

Thomason, S. G. 2001. *Language Contact. An Introduction.* Washington D.C.: Georgetown University Press.