

Quantitative and qualitative analyses of explicitation in translations

Stella Neumann & Silvia Hansen-Schirra
Universität des Saarlandes
st.neumann@mx.uni-saarland.de, hansen@coll.uni-sb.de

What is Explicitation?

Blum-Kulka (1986) states the **explicitation hypothesis** "which postulates an observed [rising] cohesive explicitness from SL [source language] to TL [target language] texts regardless of the increase traceable to differences between the two linguistic and textual systems involved."

Steiner's definition

"We assume "explicitation" if in a translation (or language-internally in a pair of register-related texts) meanings (not only ideational, but including interpersonal and textual) are realized in the more explicit variant which are not realized in the less explicit variant, but which are in some theoretically-motivated sense implicit in the latter. The resulting text is more "explicit" than its counterpart." (cf. Steiner 2005)

Example

English original
First, the opponents of free trade claim **to want to help** poor countries but at the same time are reluctant to invest or trade with them -- both of which are crucial to their development.

German translation
Erstens behaupten die Gegner von freiem Handel, **sie wollten** den armen Ländern **helfen** zögern aber gleichzeitig, Investitionen in diese Länder zu tätigen oder Handel mit ihnen zu treiben. Beides wäre für deren Entwicklung entscheidend.



Methodological approach

Different constellations of bottom-up and top-down approaches:

- Analysis of linguistic evidence measurable in the corpus allowing interpretation and generalisation
- Careful derivation of linguistic indicators from abstract concepts via linguistic categories

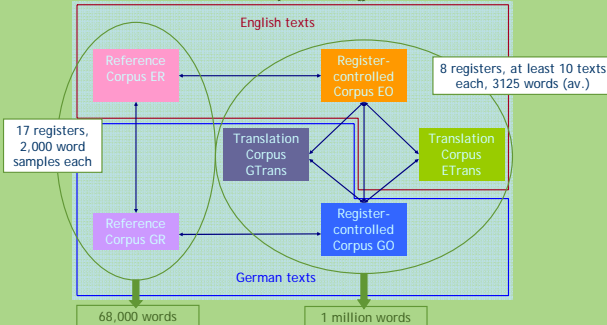
Aligned source and target texts are included in the study (differentiation between explicitness and explicitation possible by comparing comparable and parallel corpora)

Multi-purpose, theory-neutral annotation independent of linguistic research question (e.g. analysis of simplification, normalization, levelling out also possible)

Multidimensional research architecture: layering of annotation and alignment allows merging multiple constellations of data in queries and interpretations

Interpreting the output against (at least) three sources of explanation: language typology (→ reference corpora), register typology (→ register-controlled corpora) and translation process (→ parallel and comparable corpora)

Corpus design



Combining quantitative and qualitative analyses

For investigating the relation between explicitation and cohesion (on the basis of direct anaphoric relations) quantitative as well as qualitative analyses are needed:

- Part-of-speech tagging for pronominal relations (quantitative analysis)
- Manual annotation for IS-A relations (qualitative analysis)
- String-matching for total and partial recurrences (algorithm)

Quantitative analysis of explicitation and cohesion

XML annotation and alignment suggest using XQuery. This query language is particularly suited to retrieve information from different sources like for instance individual annotation and alignment files. The use for multi-layer annotation is shown in Teich et al. (2001).

Example query retrieving an explicitation of pronominal relation

a palmist, inferring the future out of his own lined flesh ein Handleser, **der** seine Zukunft aus den eigenen Linien **ableitete**

XQuery: Return all units with a PRELS part-of-speech tag which are not aligned on the token level

Explicitation + participant role + tense + mood

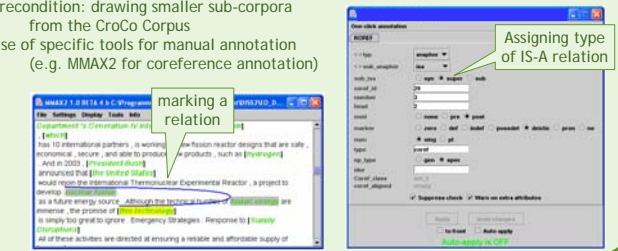
```
xquery version "1.0";
for $i in doc("e2g.wordAlign.xml")/wordAlign[@id="1" and @idref=""]
let $tokstg := doc("gt.tag.xml")/token[@id=$i and @idref=$i]
where ($tokstg/@pos = "prels")
return $tokstg
```

Qualitative analysis of explicitation and cohesion

Annotation specific to the research question, here IS-A relations (cf. Kunz 2006)

Precondition: drawing smaller sub-corpora from the CroCo Corpus

Use of specific tools for manual annotation (e.g. MMAX2 for coreference annotation)



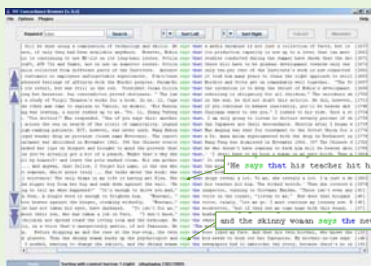
Corpus-driven research on explicitation

Concordances of lexical strings

Baker (1996) refers to explicitation as one of several translation universals, i.e. typical characteristics of translated texts irrespective of the involved languages. These characteristics should be identified in corpora of texts translated from many different languages in comparison to originals from the same domain(s) in the target language.

Operationalisation of the abstract concept 'explicitation'

Olohan & Baker (2000) discuss the use of the optional **that** in the Translational English Corpus (TEC) as compared to comparable parts of the BNC. Concordances of forms of the verbs **say** and **tell** are retrieved from the corpora and compared.



Olohan & Baker show that the optional **that** occurs significantly more frequently in translation than in originals and interpret this as an indicator of explicitation.

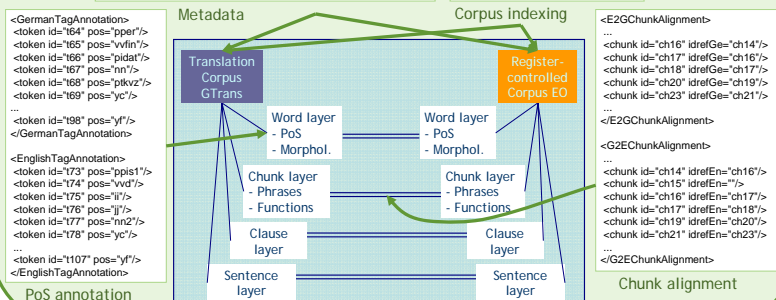
Linguistic evidence used as an indicator for a given concept has to be clearly related to the abstract concept. With string-based operationalisations like Olohan & Baker's optional **that** for explicitation, there is a wide gap between the abstract concept and the concrete indicator. Annotating a corpus with linguistic information allows more comprehensive analyses and more general statements.

Annotation and alignment of the corpus

```
<teiHeader>
<fileDesc>
<filename>GO_FICTION_001.txt</filename>
<subcorpus>FICTION GO</subcorpus>
<language>German</language>
<title>Stmnt<
<title>Mein Jahr als Mörder</title>
<author>Dellius, Friedrich Christian</author>
<title>Stmnt<
<translation></translation>
<publication>Stmnt<
<publisher>Rowohlt Berlin Verlag</publisher>
<date>2004</date>
<distributor>http://www.linix.de/mmo/priv/15719-WEB.pdf</distributor>
<availability>local</availability>
<publication>Stmnt<
<registerAnalysis>
...
</registerAnalysis>
...
</teiHeader>

<GermanTokenAnnotation>
<token id="164" strg="Ich"/>
<token id="165" strg="spalte"/>
<token id="166" strg="viele"/>
<token id="167" strg="Möglichkeiten"/>
<token id="168" strg="durch"/>
<token id="169" strg=""/>
...
<token id="198" strg=""/>
</GermanTokenAnnotation>

<EnglishTokenAnnotation>
<token id="173" strg="I"/>
<token id="174" strg="ran"/>
<token id="175" strg="through"/>
<token id="176" strg="numerous"/>
<token id="177" strg="possibilities"/>
<token id="178" strg=""/>
...
<token id="1107" strg=""/>
</EnglishTokenAnnotation>
```



Conclusion

Combination of quantitative linguistic evidence that can be re-used for different purposes in linguistics, translation studies and natural language processing with qualitative data useful for an in-depth investigation

References

Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers, ed., *Terminology, LSP and Translation Studies in Language Engineering in Honour of Juan C. Sager*. Benjamins, Amsterdam, 175-186.

Blum-Kulka, S. (1986). Shifts of cohesion and coherence in Translation. In J. House and S. Blum-Kulka, eds., *Interlingual and Intercultural Communication*. Narr, Tübingen, 17-35.

Kunz, K. (2006). Investigating nominal coreference in originals and translations. *Proceedings of SPRK Conference 2006*, Oslo.

Kunz, K. and S. Hansen-Schirra (2003). Coreference annotation of the TIGER treebank. *Proceedings of the Workshop Treebanks and Linguistic Theories 2003*. Vaxjo, 221-224.

Müller, C. and M. Strube (2003). Multi-Level Annotation in MMAX. *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*, Sapporo, Japan: 198-107.

Olohan, M. and M. Baker (2000). Reporting that in Translated English. Evidence for Subconscious Processes of Explicitation? *Across Languages and Cultures* 1(2): 141-150.

Steiner, E. (2005). Explicitation, its lexico-grammatical realization, and its determining (independent) variables - towards an empirical and corpus-based methodology. *SPRKRports* 36: 1-143.

Teich, E., S. Hansen and P. Fankhauser (2001). Representing and querying multi-layer annotated corpora. *Proceedings of the IRCS Workshop on Linguistic Databases*. Philadelphia: 228-237.