Ellipses as a cohesive device - a case study within the DFG project “German-English contrasts in cohesion – Towards an empirically-based comparison (GECCo)”

GECCo project context and objectives: identifying contrasts in the realization of cohesion
- across languages (English vs. German)
- across registers (different text types and communication scenarios along the written-spoken continuum)
- across production types (originals vs. translated texts)

Cohesive devices


EIIILLIPSS AS A CASE STUDY – METHODOLOGY

i) Developing systematic and fine-grained conceptualization of ellipsis as a cohesive device aiming at cross-linguistic applicability of annotation scheme categories (Menzel 2014a/b/c)

- ellipsis has been described as a ‘chameleon concept’ as linguists have produced rather heterogeneous definitions so that various constructions and discourse phenomena have been subsumed under this category
- main categories of (potentially cohesive) ellipsis relevant for this study:
  - Nominal ellipsis: omission of specific element of noun phrase (head noun) - e.g. “There are many reasons why Britain is good for Europe. Let me choose just four.”
  - Verbal ellipsis: ellipsis within the verbal group (modal/auxiliary/operator or lexical verb, often accompanied by the omission of related elements such as objects) - e.g. “A little town that is often missed by travellers but shouldn’t be [ ].”
  - Clausal ellipsis: omission of a part of a clause (broadest subcategory) - e.g. “Has he brought you presents? What kind of presents [ ]?”
  - Co-occurrence of nominal+verbal+clausal - e.g. “How many slices do you want?” “[ ]” “[ ]” “[ ]”
- cohesive ellipses refer endophorically to textual antecedents (ideally not in the same clause so that a textual link between different clauses or sentences is created)

- ii) Annotation of endophoric, text-forming ellipses (remnants) and their antecedents with MMX2 based on annotation guidelines

- ellipsis remnants and their antecedents are annotated with open-source tool MMX2: nominal ellipsis, verbal ellipsis, clausal ellipsis, nominal+verbal/clausal (mixed)

- within these categories, it can be distinguished between:
  - non-cohesive (e.g. exophoric, situational)
  - cohesive (cross-clausal reference to antecedent)
  - clause-internal

- problematic cases that might require further discussion, e.g. borderline cases or ambiguous structures can be treated separately -> annotated as “problematic”

- iii) Data extraction and interpretation

- some additional annotation categories cover other types of fragments, non-clausal units and omission phenomena that might look similar to ellipsis (e.g. anacoluthon, headlines...) but actually need different analysis
- manual annotation currently more accurate than automatic methods (numerous theoretical omission possibilities in different syntactic environments and POS tagging sometimes work wrong in ellipsis environments due to deficient/non-standard syntax)
- annotations can be used to identify typical syntactic patterns of ellipsis contexts to improve (semi-)automatic annotation methods for certain types (e.g. nominal e. after adjectives, clausal e. in question-answer pairs)

- some frequency distributions:

- different between certain registers and between written and spoken language greater than those between English and German in general
- ellipsis frequency lower in translations compared to that in originals of the same language

REFERENCES

Karolin Menzel, k.menzel@mx.uni-saarland.de

Institute of Applied Linguistics, Translation and Interpreting, Saarbrücken, Germany