

Erich Steiner

Saarbrücken

**Empirical studies of translations as a mode of language contact - “explicitness”
of lexicogrammatical encoding as a relevant dimension**

Contents

1. Clarification of terms
 - 1.1. Language Contact
 - 1.2. Multilinguality
 - 1.3. A Multifunctional and feature-based perspective
2. A methodology for the empirical investigation of translations
 - 2.1. Corpus architecture, types of contrast, hypothesis formation, querying techniques
 - 2.2. Some “low-level” results
3. Explicitness of encoding, directionality of change, frequency of use

Ohnehin wäre zu überlegen, ob die Mittelmäßigkeit von Sprachkenntnissen nicht etwas Erhaltenswertes ist. Würde einen ein großer Wortschatz nicht um das Vergnügen jener harmlosen, naiven und zarten kurzen Gespräche bringen, die man auf Reisen mit anderen Touristen führt? Eine Dame aus Bangladesh, die mich bei Starbucks an den Niagara-Fällen ersucht hatte, einen Blick auf ihre Tasche zu werfen, solange sie auf der Toilette weilte, und die der Meinung war, es läge der Geruch von Schnee in der Luft, fragte mich, ob es in Deutschland auch schneie. Ich antwortete:

„Yes, there is some snow, but in the northern part where I come from there isn't much, and when it snows, it quickly turns into brown mud, which I don't like very much, but in the mountain areas, there is sometimes quite a bit of snow, and I think that's quite pretty.”

Max Goldt, *Vom Zauber des seitlich dran Vorbeigehens*, Reinbek bei Hamburg, Rowohlt, 2005, S. 96

1.1. Language Contact

Language contact is the situation in which languages, or rather, instantiations of language systems in texts through their speakers, influence each other synchronically in shared socio-semiotic contexts. This is complementary to the historical axis, along which genetically related languages are in contact through time (cf. Weinreich 1953, Oesterreicher 2001, Thomason and Kaufmann 1988). Language contact applies to varieties within languages, as it does to different standard languages.

Frequent **topics of research** are:

- the interplay between synchronic contact and genetic inheritance
- linguistic vs. socio-cultural constraints on interference
- analytic frameworks for contact-induced language change (linguistic levels of change; borrowing vs. interference through shift; predictive power of the frameworks, external vs. internal explanations)
- language maintenance
- normal vs. exceptional transmission (creolization, pidgins).

In an attempt to generalize on the strength of language contact, Thomason and Kaufmann (1988, pp. 65ff) postulate a ***borrowing scale***, ranging

- from only lexical borrowing
- through slight structural borrowing
- through moderate structural borrowing
- to heavy structural borrowing.

Many studies of language contact have focussed on lexical items and/ or grammatical structures, rather than on features or properties of the linguistic systems and instances (discourses, texts) involved, although both perspectives have often been acknowledged as relevant.

1.2. Multilinguality

- if we say that **speakers** are multilingual, then it means they are different neurophysiologically, and/or cognitively, and/ or developmentally from monolingual ones. If none of these implications, or others of their kind, held for the speakers, then the notion of *multilingual speakers* could be reduced to “speakers who entertain multilingual discourse”.
- if we say that **communities** are multilingual, then it means that they are different in their institutions, hierarchies, identities etc.. If not, then again the notion should be reduced to one of “communities which entertain multilingual discourses”.
- if we say that **discourses** are multilingual, then it means that they show special *discourse properties* of directness vs. indirectness, orientation towards self vs. other, orientation towards content vs. interaction, explicitness vs. implicitness, routine-orientedness vs. ad-hoc formulation, ultimately to be realized as lexicogrammatical properties such as interference, borrowing, code-/language switching, special metafunctional orientations in terms of ideational, interpersonal, or textual biases, directness, density, explicitness, and others. These discourses thus instantiate specific contact varieties, or registers

1.3. A Multifunctional and feature-based perspective

In terms of methodology, a perspective of the kind advocated here will give due consideration

- to systems alongside structures,
- to the instance alongside the system,
- to more abstract (and at the same time, more empirical) types of contrast than have often been in the centre of theorizing,
- to the metafunctional modularization of language

1.3.1. Consideration of systems alongside structures

In analysing a particular translation, say, from English into German, a **structural** account might for example explore, among other features of the German translation, the relaxing of constraints between semantics-to-grammar mapping. A **systemic account would additionally** specify how:

- the range of realizations of transitivity types is extended
- the polarity between transitive and ergative types/ features in the system is weakened
- modal responsibility (a term in the interpersonal metafunction) can now be assigned to discourse phenomena that wouldn't have received that assignation before
- new mechanisms of making an entity thematic are developing in the textual metafunction
- the directness of semantics-to-grammar mapping is generally weakened within and across metafunctions
- markedness is changing, in the sense that under the influence of English, the notion of markedness in German will become weaker for *personification*-type grammatical metaphors, and stronger for non-Subject thematicity.

1.3.2. Consideration of the instance alongside the system

Teich (2003, pp. 207ff) generalizes as follows:

- “Where the TL has fewer options than the SL (e.g. less delicacy in a particular system), the TL needs to compensate; where the same compensatory means is used sufficiently frequently, we get TL normalization.
- Where the TL has more options than the SL (e.g., more delicacy in a particular system), the TL can afford to let the SL shine through.
- Where the SL and the TL are sufficiently different system-wise, there is no danger of interference, so there are no shining through effects (there may, however, be other effects).”

She thus attempts a systematic relating of observed properties of texts in her sub-corpora to possible explanations, emphasising strongly the effects of the language systems involved, expressed through/ accompanied by register differences, where this is applicable. She also identifies those cases in which an explanation in terms of the translation process appears possible, or even compelling.

Hansen (2003) in another corpus-based study exploring specific properties of translated texts, attempts to isolate the different independent variables and thus explanations with the help of different methodologies. She also works on different corpora, using different sets of features, but also adds psycholinguistic experimentation as a methodology.

- both the translation process and the language pair involved are major causes for observed properties of translated texts such as Baker's features (normalization etc.), different degrees of readability, different degrees of explicitness etc.
- translated texts are both more normalized and more neutralized if compared to their register-parallel English texts, depending on whether the focus is on typical or untypical features in the sense of Biber. This would explain the often claimed 'typicalness' (in terms of register) and at the same time 'unmarkedness' (in terms of the general language) of translated texts.
- the properties observed in the first part are partly explainable by the interference from the source language and texts, and thus also vary with the specific language pairs involved.
- psycholinguistic experiments show the influence of the **translation process** in terms of grammatical metaphorization, specifically the variables of degree of expertise and time constraints. But it also shows that the particular **language pair** involved still has a strong influence, and that there are influential processes of 'unpacking' in the **process of understanding**, followed by incomplete re-packaging, depending on experience and time constraints. **Explicitation and simplification** can be shown to be reflexes of these facts.

Neumann 2003 establishes controlled corpora of travel guides and cross-register texts, investigating them for the variables *transitivity*, *circumstances*, *modality*, *voice*, *theme*, and then interprets the differences found in terms of *register*.

- Development of a clear and transparent architecture for linking lexicogrammatical (micro-)variables and types of register (macro-) variables in a theoretically motivated way.
- Operationalizations of the notions of size and representativity of samples of a register (2003:70ff).
- Hypotheses about the influences of the respective registers and languages on each other.
- Application of Biber's (1995) notions of positive and negative indicators of a register to the range of higher-level features characteristic to the work of our group (2003: 211ff).

1.3.3. More abstract (and at the same time, more empirical) types of contrast than have often been in the centre of theorizing.

We suggest to (Steiner 1991, 1994, 2001, 2005a,b,c)

- a) partly reduce an intuitive notion of 'information distribution' in texts and sentences to more technical and better understood notions of *information structure*, *informational density* and *grammatical metaphoricity*, and
- b) operationalize these latter notions in such a way as to make them empirically testable on electronic corpora, using the 'shallow' concepts of *explicitness*, *density*, and *directness* as properties of semantics-to-grammar mapping in sentences.

1.3.4. The metafunctional modularization of language

Several linguistic frameworks have postulated a modularization of linguistic structures along different dimensions, usually adopting some diversification into, roughly,

- Referential/ ideational/ propositional
- Interactional/ interpersonal
- Textual/ organizational

The latter is an 'enabling' function yielding structure in terms of

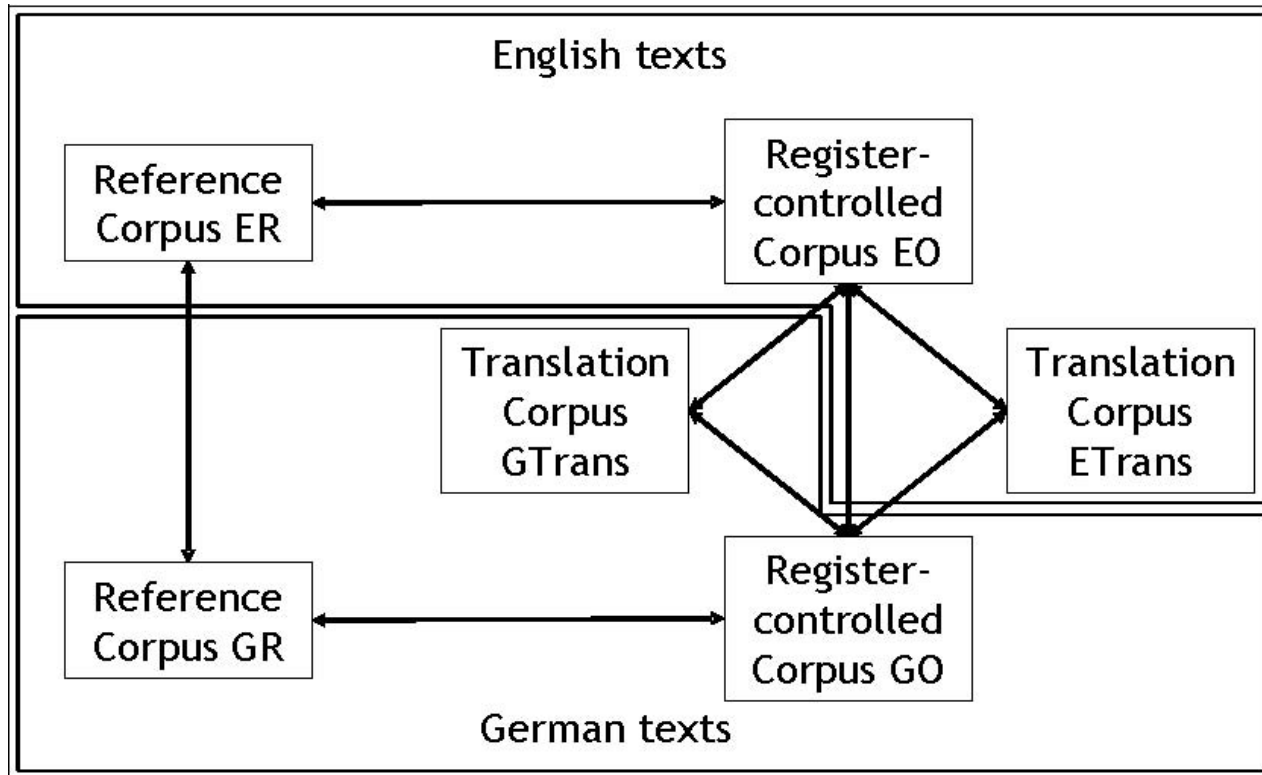
- Theme vs. Rheme,
- Topic vs. Comment
- Given vs. New information.

A model giving prime architectural place to these distinctions is Systemic Functional Grammar (cf. Halliday 1978; Halliday and Matthiessen 2004; Caffarel et al. 2004).

A significant step in the more system-based and multifunctional direction has, outside of SFL, been taken in work by Traugott (1982; Traugott and Dasher 2005: 94), who focus on semantic-pragmatic change and grammaticalization. They hypothesize semantic change in grammaticalization to proceed along the following cline:

- *Propositional towards textual towards expressive.*

2.1. Corpus architecture



As corpora to be contrasted, we have:

Contrast C1 (reference corpora):

English Reference Corpus (ER) 34 000 words (17 registers X 2000 words)

German Reference Corpus (GR) 34 000 words (17 registers X 2000 words)

Contrast C2 (register controlled)

C2.1. (within one register, between languages, differentiated into 8 sub-contrasts by register): 31,250 words per register, each sample comprises at least 10 texts resulting in a computed length of 3,125 words per text, 250 000 words altogether

English register controlled corpus (EO)

popular-scientific texts
tourism leaflets
prepared speeches
political essays on economics
fictional texts
corporate communication
instruction manuals
websites

German register controlled corpus (GO)

popular-scientific texts
tourism leaflets
prepared speeches
political essays on economics
fictional texts
corporate communication
instruction manuals
websites

C2.2.: (between registers, within each of the languages English and German, yielding 8 contrasts within each of the languages)

Contrast C3:

Translations vs. originals within each of the two languages:

EO vs. ETrans

GO vs. GTrans

Contrast C4:

Originals and their translations

English Translations of GO (ETrans)

German Translations of EO (GTrans)

C4.1. Sub-corpora and texts as wholes (without alignment)

Ok for sub-corpora, possibly in the future for texts as wholes, cf. above.

C.4.2. On aligned corpora, i.e. explicitation by translation units

Hypothesis formation

Hypotheses 1. to 7. below are examples of hypotheses about cohesion to be tested on the data: For either a given pair of non-aligned text segments globally, or else for a given aligned source – target fragment of two texts in a translation relationship, we expect differences along the following parameters:

1. the proportion of explicit to implicit referents;
2. the proportion of phoric to fully lexical (auto-semantic) phrases;
3. the number of newly introduced discourse referents per discourse segment;
4. the amount of cohesive ellipsis and substitution;
5. the strength of lexical chains as measured by various ratios between content and function words, and as measured by type-token relationships;
6. the strength (internal connectivity) of lexical chains as measured by average number of items per lexical chains;
7. the ratio between explicit and implicit encoding of conjunctive relations.

Observe that in comparing any text fragments which are not in a unit-of-translation-relationship, as in our registerially parallel sub-corpora of originals, we are testing for the global property of (relative) explicitness. However, whenever we are comparing a specific aligned and instantiated source-target (translation) unit, we are testing “explicitation” (or its opposite, implicitation).

(cf. Hansen-Schirra, Neumann and Steiner 2006).

querying techniques

English original: ... a palmist, inferring the future out of his own lined flesh
German translation: ... ein Handleser, der seine Zukunft aus den eigenen Linien ableitete
(a palmist who his future out-of the own lines inferred)

token index file

```
<token id="t64" strg="ein"/>
<token id="t65" strg="Handleser"/>
<token id="t66" strg=","/>
<token id="t67,, strg="der"/>
<token id="t68" strg="seine"/>
<token id="t69" strg="Zukunft"/>
```

part-of-speech annotation

```
<token pos="art" xlink:href="#t64"/>
<token pos="nn" xlink:href="#t65"/>
<token pos="yc" xlink:href="#t66"/>
<token pos="prels" xlink:href="#t67"/>
<token pos="pposat" xlink:href="#t68"/>
<token pos="nn" xlink:href="#t69"/>
```

word alignment

```
<token> <align xlink:href="#t55"/>
<align xlink:href="#t66"/> </token>
<token> <align xlink:href="#t56"/>
<align xlink:href="#undefined"/> </token>
<token> <align xlink:href="#tundefined"/>
<align xlink:href="#t67"/> </token>
```

Figure 1: XML corpus annotation and alignment on word level including empty links

For the investigation of explicit pronominal referents in German relative clauses vs. implicitly encoded English referents, all German tokens with the part-of-speech tag *prels* (for relative pronoun) have to be extracted which are not aligned on the word level (since the pronominal reference is encoded in the English participle). The respective XQuery is shown in Figure 2.

```
for $k in $doc//tokens/token
let $fileName := $doc//translations/translation[@n='1']/@trans.loc
let $fileNameNew := replace($fileName,"tok","tag" )
where ($k/align[1][@xlink:href != "#undefined"] and $k/align[2]
[@xlink:href = "#undefined"] and doc($fileNameNew)//token
[@xlink:href eq $k/align[1]/@xlink:href][@pos eq "prels"])
```

Figure 2: XQuery for relative pronouns with empty links

The output of this query are sentences like the ones displayed in Figure 1. This example (taken from the fiction sub-corpus) is interpreted as explicitation since participant role (and thus the reactivation of the referent), tense and mood are explicitly realised in the finite relative clause of the German translation, whereas they are implicit in the English original.

```
<result no="13"><ori_en>Baker Hughes Business Support Services has assumed accounting, payroll, benefits and IT support duties for many of the company's U.S. operations, eliminating duplicate efforts by division personnel. </ori_en>
<trans_ge>Baker Hughes Business Support Services hat die Buchführung, Gehalts- und Sozialleistungen sowie IT-Aufgaben für viele Niederlassungen des Unternehmens in den Vereinigten Staaten übernommen, wodurch doppelte Arbeit durch das Personal in den Tochterunternehmen vermieden werden konnte. </trans_ge></result>
<result no="14"><ori_en>In this environment, Baker Hughes revenue declined 22% to $4.5 billion for 1999, compared to $5.8 billion in 1998. </ori_en>
<trans_ge>Vor diesem Hintergrund sanken die Umsatzerlöse von Baker Hughes im Jahre 1999 um 22% auf 4,5 Mrd. Dollar, während sie 1998 noch 5,8 Mrd. Dollar betragen hatten. </trans_ge></result>
```

Figure 3: Results for conjunctions with empty links

(Cf. Hansen-Schirra, Silvia, Stella Neumann, and Mihaela Vela 2006)

(Cf. Hansen-Schirra, Silvia, Stella Neumann and Erich Steiner 2006)

2.2. Some “low-level” results

CroCo Statistics

Type-Token-Ratio

	English		German		Referenzkorpus	
	Ori	Trans	Trans	Ori	Englisch	Deutsch
Essay	9,37	9,51	13,09	16,44		
Fiction	13,00	12,43	16,31	16,37		
Instruction	7,09	6,87	10,11	12,59		
Popsci	11,15	10,53	15,77	15,34		
Share	9,56	8,36	14,37	13,84		
Speech	10,04	8,68	13,11	13,73		
Tourism	12,90	13,43	18,31	21,58		
Web	11,21	11,63	16,31	19,03		
TOTAL	∅ 10,54	∅ 10,18	∅ 14,67	∅ 16,12		
TOTAL_L ANG	∅ 10,36		∅ 15,40			
TOTAL	4,29	4,12	7,56	8,37	14,04	19,86
TOTAL_L ANG	3,06		6,40			

Lexical Density (as proportion of lexical words!)

	English		German		Referenzkorpus	
	Ori	Trans	Trans	Ori	Englisch	Deutsch
Essay	62,84	62,87	59,96	61,99		
Fiction	55,83	55,73	54,80	55,76		
Instruction	63,97	66,60	58,80	63,16		
Popsci	62,49	58,83	61,31	59,05		
Share	62,99	61,84	61,01	61,14		
Speech	60,82	59,89	58,73	59,14		
Tourism	64,15	63,93	63,10	63,99		
Web	63,51	62,24	61,79	62,15		
TOTAL_REG	62,07	61,56	59,90	60,81	61,28	61,82
TOTAL_LAN G	61,80		60,36			

PoS-Verteilung - Englisch (in Prozent)^[1]

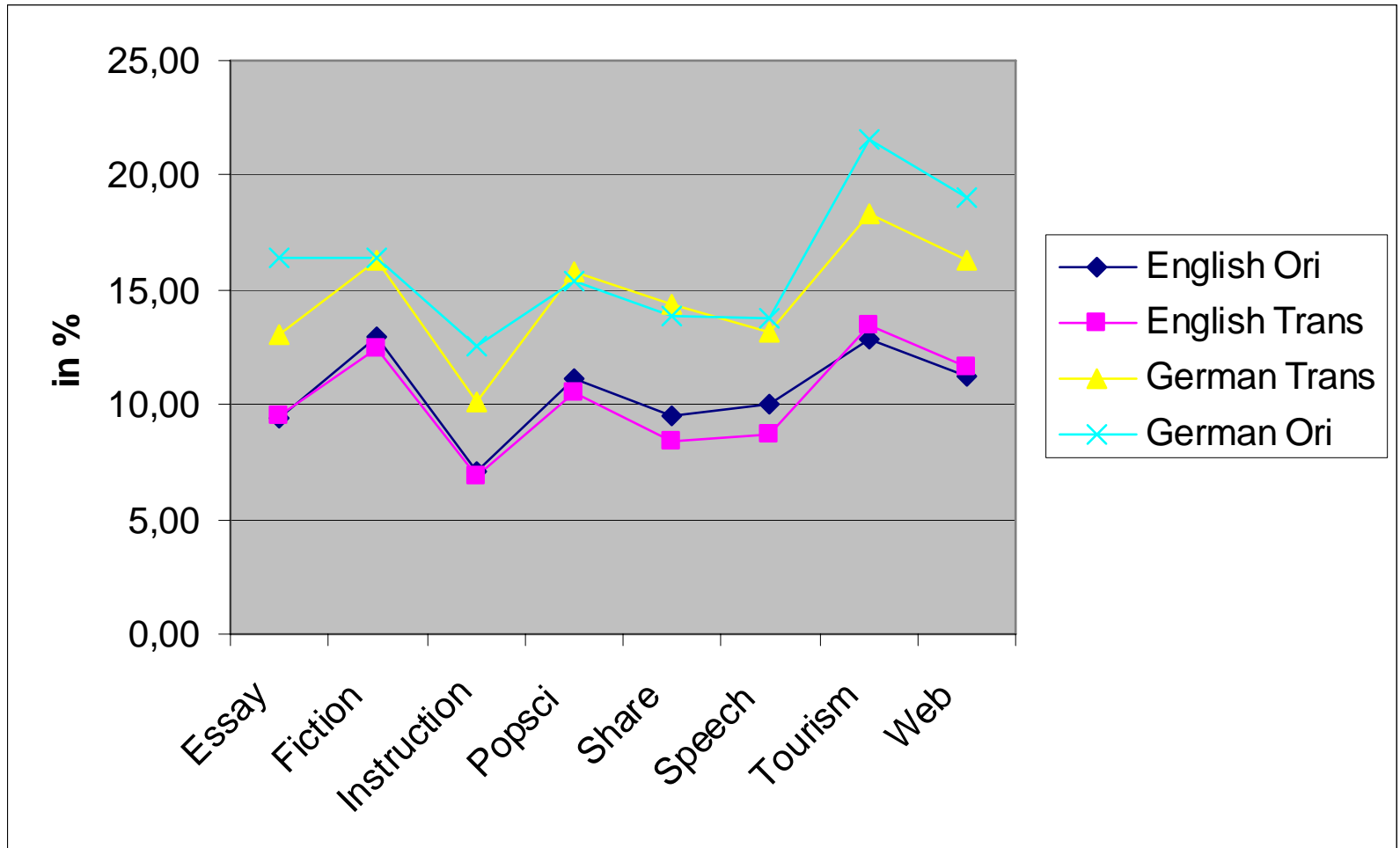
		Nomen	Pron.	Verb	Adj.	Adverb	Adpos.	Artikel	Konj.	Partikel	Zahlen	Interj.	Komp	Fremd	NichtW.	
E N G L I S C H	O R I S T I N A L	Essay	30,80	3,99	17,01	9,18	4,06	12,13	11,66	6,15	3,23	1,74	0,01	0,01	0,00	0,03
		Fiction	22,21	11,69	19,37	6,17	6,76	11,22	11,80	6,66	2,79	1,16	0,09	0,00	0,02	0,07
		Instruc.	33,79	4,11	18,09	5,92	3,62	10,42	12,92	5,45	3,14	2,29	0,06	0,02	0,00	0,17
		Popsci	28,45	3,49	17,99	8,64	5,54	12,58	12,83	6,15	2,45	1,65	0,00	0,01	0,00	0,20
		Share	32,65	5,98	15,49	7,81	3,53	12,33	10,10	6,15	2,44	3,31	0,00	0,03	0,01	0,17
		Speech	28,30	5,91	17,62	9,51	3,93	12,10	11,31	6,50	3,35	1,37	0,03	0,00	0,01	0,06
		Tourism	36,02	2,79	12,38	9,53	4,13	13,09	12,30	5,83	1,84	2,04	0,01	0,00	0,01	0,03
		Web	33,01	4,92	16,40	7,65	3,92	11,88	10,85	6,18	2,66	2,44	0,04	0,01	0,01	0,04
		TOTAL	30,65	5,37	16,80	8,05	4,43	11,97	11,72	6,13	2,74	2,00	0,03	0,01	0,01	0,10
	I N T E R N A T I O N A L	Essay	29,73	2,86	15,40	9,70	5,65	13,28	13,44	5,31	2,23	2,31	0,00	0,00	0,01	0,06
		Fiction	23,21	11,58	17,98	5,50	7,58	11,66	12,23	5,62	3,18	1,29	0,08	0,01	0,03	0,06
		Instruc.	33,73	2,32	18,71	6,13	4,57	10,77	13,49	4,80	2,02	3,12	0,07	0,01	0,00	0,26
		Popsci	24,96	5,34	17,20	8,24	6,25	12,95	14,12	6,39	2,36	1,72	0,00	0,00	0,03	0,43
		Share	30,34	5,77	15,47	8,70	4,53	13,57	11,54	4,79	2,49	2,70	0,00	0,00	0,04	0,05
		Speech	25,64	5,93	17,49	9,27	6,11	12,36	13,08	5,71	3,03	1,36	0,01	0,00	0,01	0,01
		Tourism	36,09	2,49	12,05	8,65	4,15	13,57	13,31	5,09	1,61	2,80	0,00	0,01	0,14	0,06
Web		31,05	4,71	14,57	9,71	4,50	13,28	12,39	5,41	1,97	2,23	0,03	0,01	0,09	0,06	
TOTAL	29,46	5,06	16,04	8,27	5,39	12,70	12,95	5,38	2,35	2,20	0,02	0,00	0,04	0,12		
TOTAL_EN		30,02	5,20	16,40	8,17	4,94	12,36	12,37	5,73	2,53	2,11	0,03	0,01	0,03	0,11	
REF_EN		27,95	6,20	17,86	7,09	5,26	11,99	11,72	5,81	2,99	2,74	0,07	0,01	0,21	0,10	

[1] Reliability for English and German 96,7 % as reported in Brants 2000:

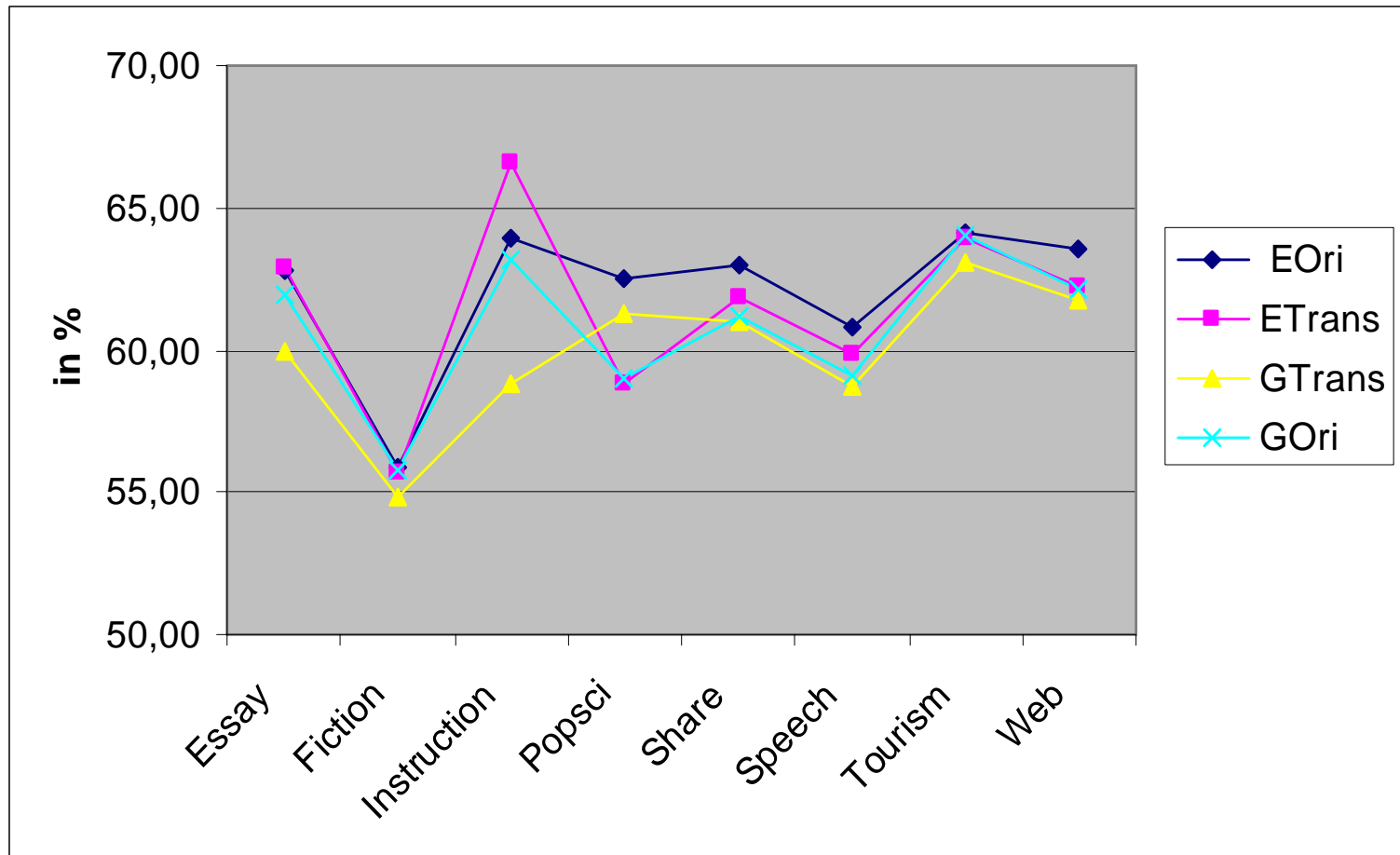
PoS-Verteilung - Deutsch (in Prozent)

		Nomen	Pron.	Verb	Adj.	Adverb	Adpos.	Artikel	Konj.	Partikel	Zahlen	Interj.	Komp.	Fremd	NichtW.	
D E U T S C H	T	Essay	28,69	9,15	13,66	12,61	3,49	9,43	13,32	6,08	2,05	0,97	0,00	0,24	0,23	0,06
		Fiction	20,33	17,34	17,92	8,90	6,72	8,19	9,46	7,32	2,89	0,56	0,03	0,08	0,12	0,13
		Instruc.	29,66	10,13	15,02	8,75	3,19	9,40	12,84	6,20	2,63	0,97	0,00	0,34	0,16	0,70
	R	Popsci	26,10	10,06	14,60	12,16	6,86	9,77	11,45	5,37	2,05	1,22	0,01	0,18	0,06	0,13
		Share	28,15	10,27	12,80	10,63	4,85	10,54	10,49	5,93	1,75	3,47	0,00	0,40	0,67	0,04
	N	Speech	27,99	10,87	14,46	10,87	4,26	9,23	12,58	6,34	2,25	0,75	0,00	0,17	0,20	0,03
		Tourism	33,56	7,81	10,27	11,84	3,91	10,32	11,92	5,77	1,09	1,47	0,00	0,11	1,40	0,53
	S	Web	30,40	10,13	14,23	10,50	3,35	9,66	9,89	6,31	2,22	1,70	0,00	0,27	1,16	0,19
		TOTAL	28,06	10,76	14,14	10,77	4,58	9,56	11,48	6,18	2,12	1,40	0,01	0,23	0,50	0,22
I N G L I S C H	O	Essay	27,83	8,05	13,36	12,43	5,85	9,58	12,83	5,53	2,02	1,82	0,00	0,37	0,29	0,04
		Fiction	21,04	16,26	17,28	8,82	7,48	8,66	10,47	5,69	3,16	0,63	0,07	0,08	0,18	0,19
		Instruc.	30,79	6,98	15,75	9,03	4,03	9,66	12,92	5,28	2,00	2,44	0,00	0,42	0,13	0,57
	R	Popsci	23,55	11,64	14,74	11,40	7,55	8,84	12,52	5,74	2,21	1,15	0,00	0,17	0,18	0,30
		Share	28,25	10,35	13,00	11,94	4,80	10,35	11,50	4,96	1,70	2,45	0,00	0,46	0,23	0,02
	I	Speech	24,91	12,32	15,29	11,25	6,28	8,92	11,53	5,84	2,25	0,86	0,00	0,21	0,26	0,09
		Tourism	32,85	5,65	9,77	12,48	4,85	11,58	11,73	5,83	1,23	2,55	0,01	0,48	0,82	0,18
	S	Web	30,70	8,94	12,51	11,93	4,42	9,81	11,49	5,95	1,66	1,65	0,00	0,31	0,49	0,14
		TOTAL	27,51	10,00	13,95	11,17	5,65	9,68	11,88	5,60	2,02	1,70	0,01	0,31	0,32	0,19
TOTAL_DE		27,79	10,38	14,04	10,97	5,12	9,62	11,68	5,89	2,07	1,55	0,01	0,27	0,41	0,21	
REF_DE		26,69	9,84	15,19	10,72	5,85	9,41	11,51	5,21	2,21	2,64	0,02	0,16	0,27	0,29	

Type Token Ratio

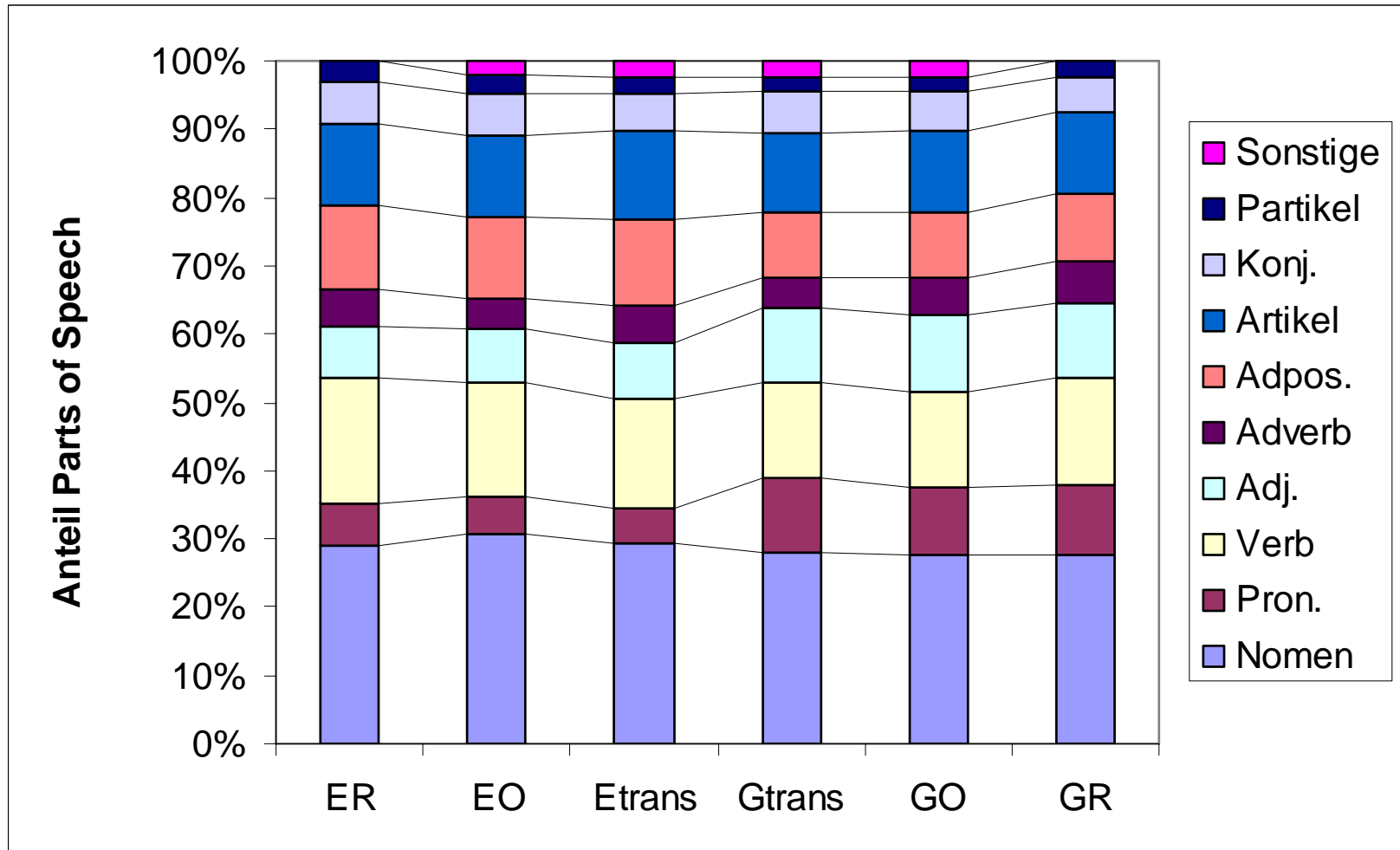


Lexikalische Dichte

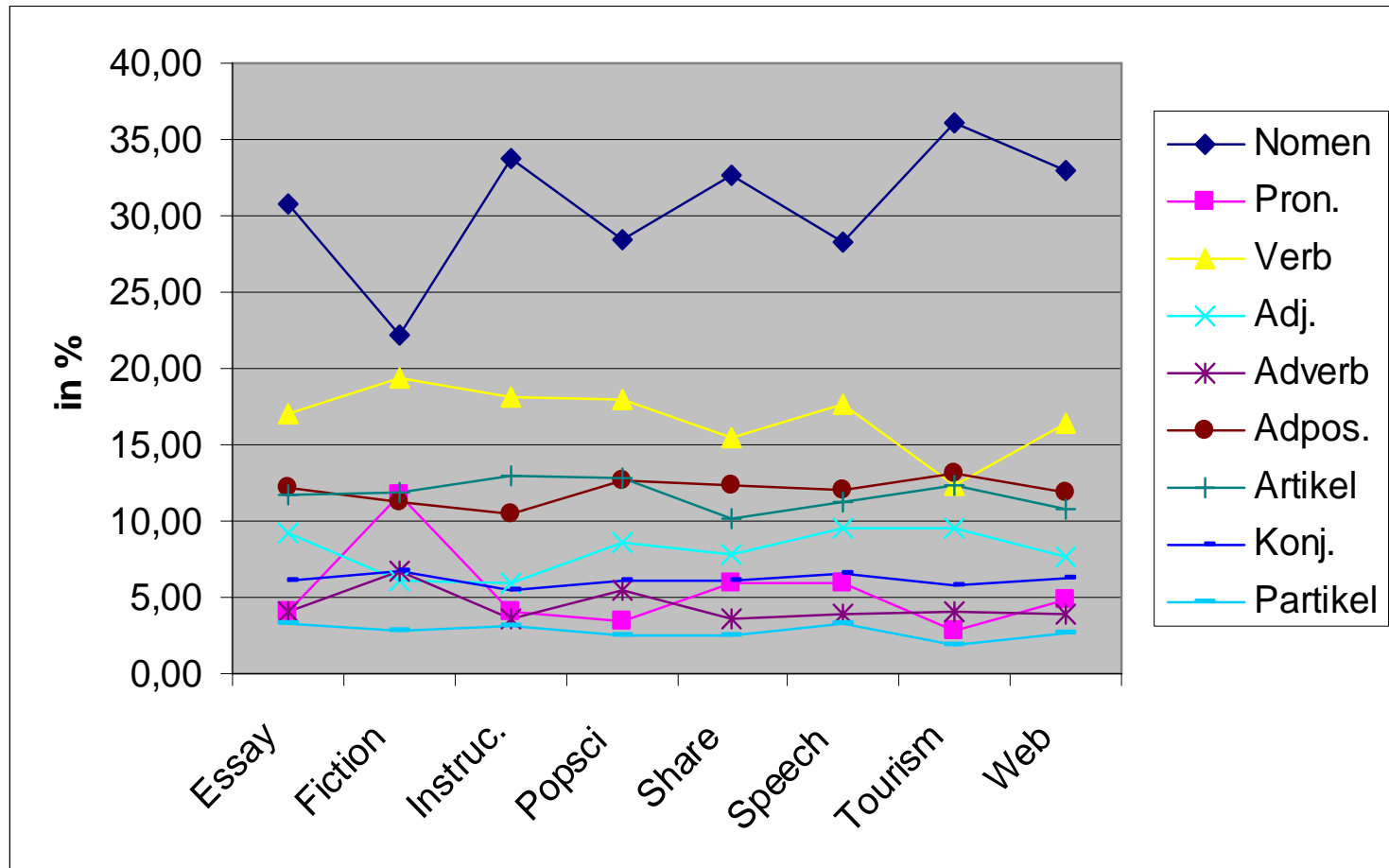


Achtung! Y-Achse beginnt erst bei 50 %!

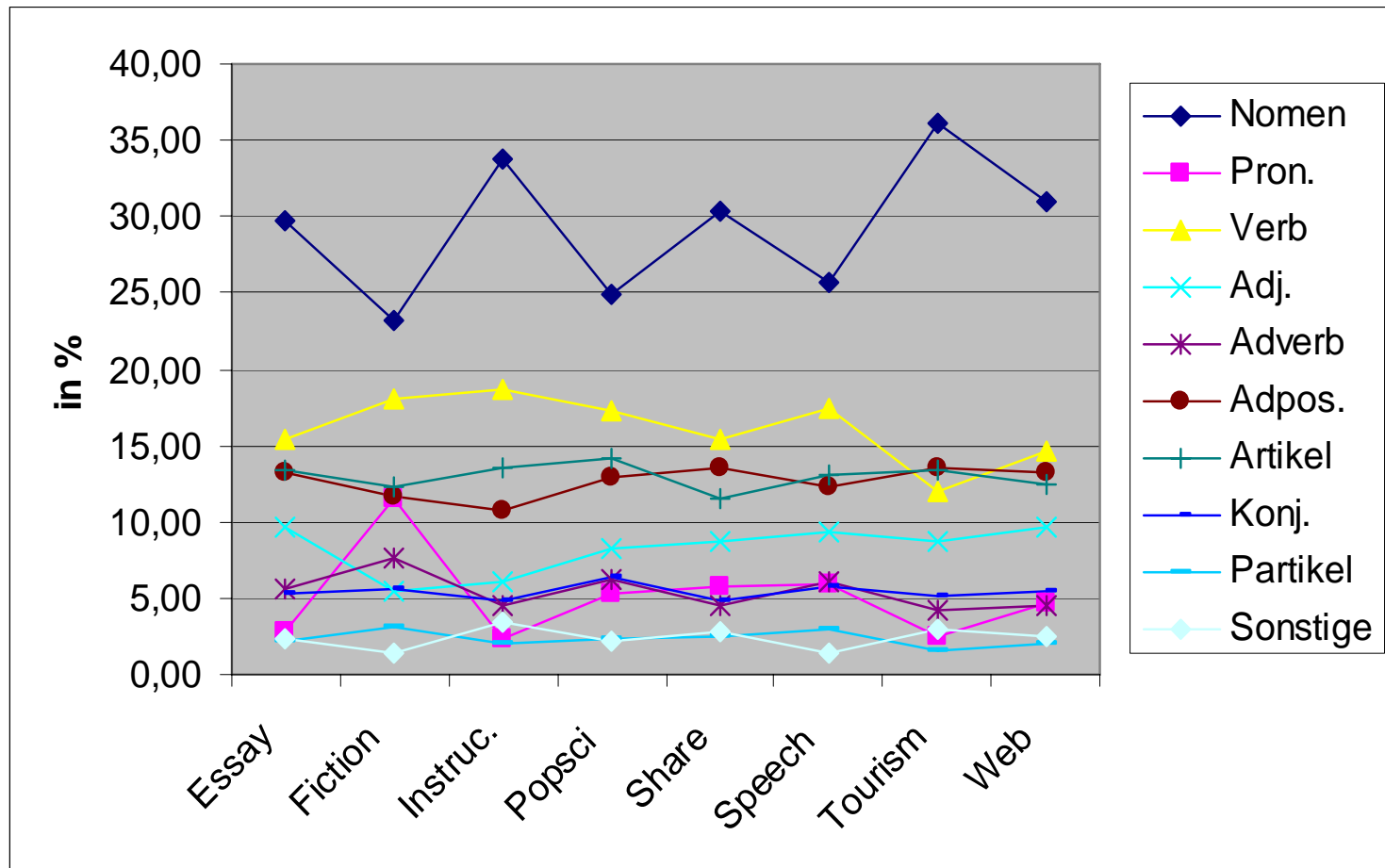
PoS Vergleich Subkorpora



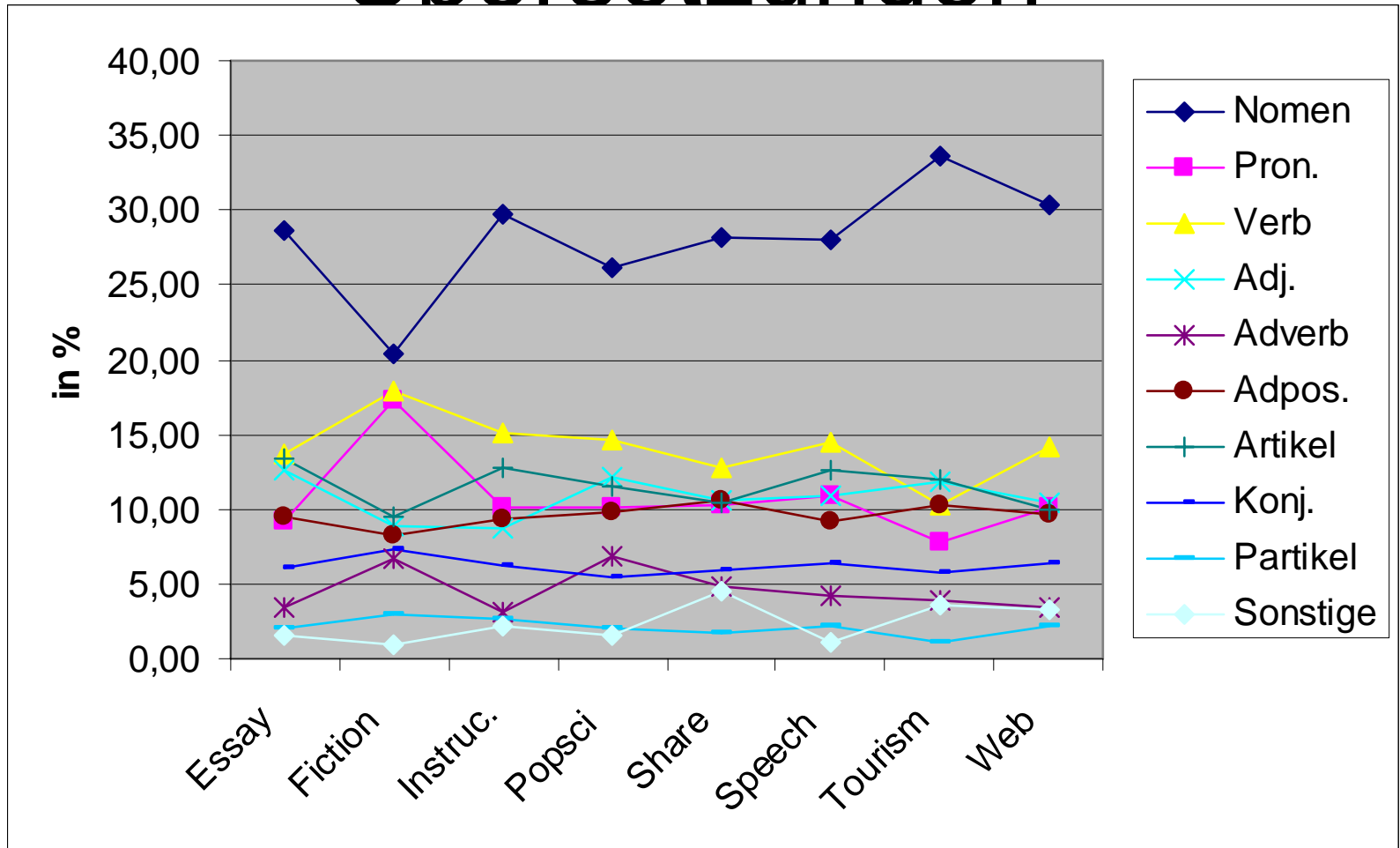
PoS Vergleich Englische Originale



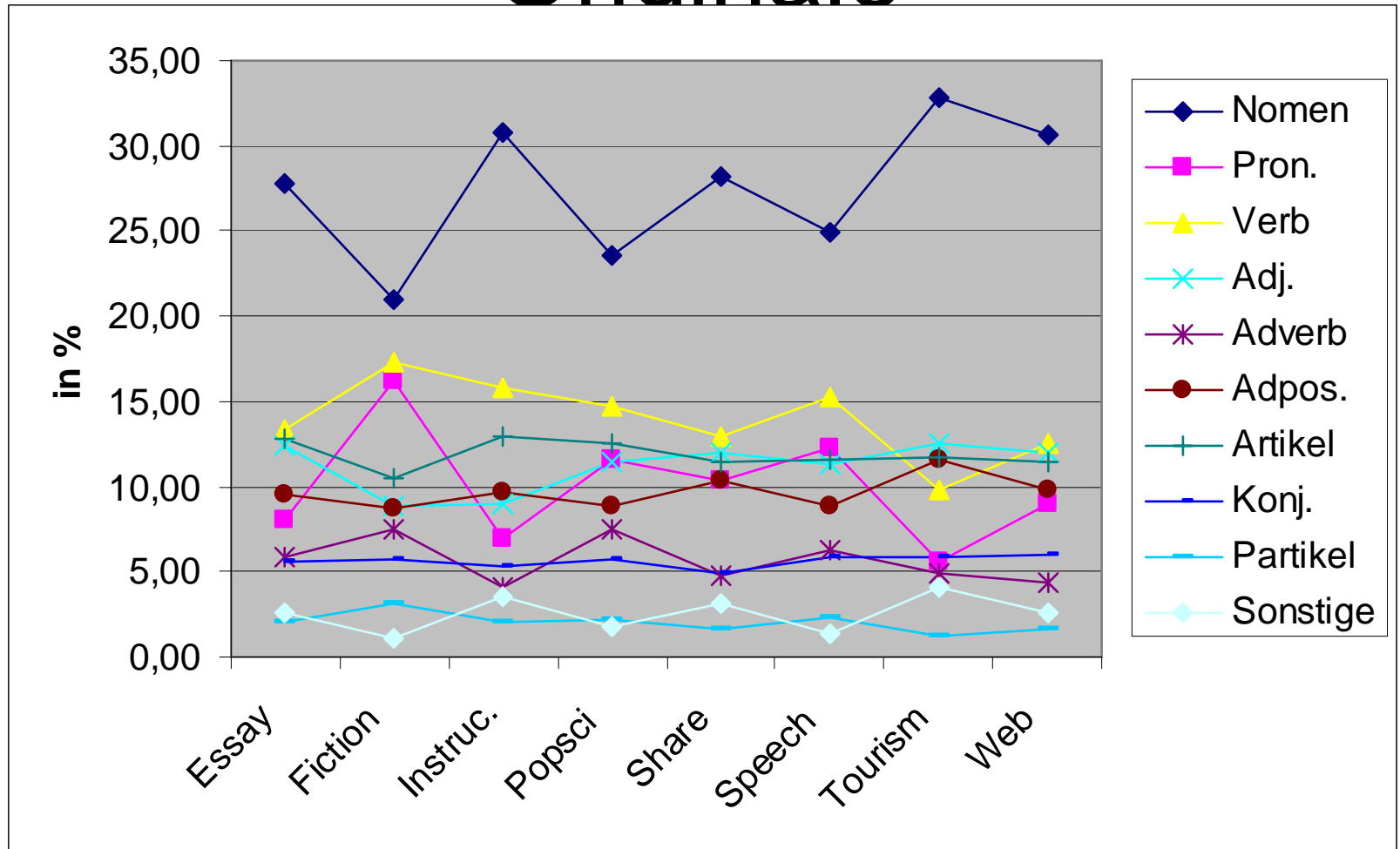
PoS Vergleich Englische Übersetzungen



PoS Vergleich Deutsche Übersetzungen



PoS Vergleich Deutsche Originale



“Directionality of change”

- “*Directionality of change*” and different types of structure (morphologically synthetic and analytic language types, types of basic word order, types of marking of relations such as head-marking, dependent marking, mixed-marking etc.)
- In more recent times, directionality has sometimes been linked to multi-functional, or multi-dimensional models of language, as for example in the work of Traugott and Dasher (2005), on *subjectivity*, *intersubjectivity* and *objectivity* in language (use) (2005: 19ff), or in the work of Halliday and others (Halliday and Hasan 1976; Halliday and Matthiessen 1999: 227ff, 507ff changes between *ideational*, *interpersonal* and *textual* encodings).
- Traugott (originally in 1982), had postulated a unidirectional development of semantic change in the form of *propositional* > (*textual*) > *expressive*
- *subjectivity*, *intersubjectivity* and *objectivity* as properties of grammatical constructions on one level of abstraction with *explicitness*, *density* and *directness*
- Our role relative to this line of research is in
 - a comparison of the kinds of abstraction we are making,
 - in their relationship to the multifunctional hypothesis,
 - empirical research methodologies based on electronic corpora.
 - investigations of situations of multilinguality *and translation* and their impact on language contact and change
 - investigations of particular registers as sites of language contact and change

Explicitness

A *stratification* of the notion of “explicitness” in terms of the linguistic levels of lexicogrammar and text.

- “*Explicitness*” on the *lexicogrammatical (structural) level* is conceptually related to “density” and “directness”. These three are properties of (lexico-)grammatical constructions. The opposite of “explicit” here is “lexicogrammatically not realized”, but still part of the construction (unrealized participant roles, unrealized features in non-finite constructions, grammatical ellipsis, projection onto different grammatical categories, grammatical metaphor, transcategorization, etc.).
- “*Explicitness*” on the *textual level* is conceptually related to properties such as “lexically impoverished, rationalized, clarified, expanded, ennobled, popularized, standardized, simplified, normalized, levelled-out, sanitized , direct vs. indirect; oriented towards self vs. oriented towards other; oriented towards content vs. persons.” (For the latter of these, cf. House 2002). It is an “emergent property” on a higher level in the sense that the properties on text level are perceived as a result of the interaction of clause level features, such as “explicitness, directness, density”, with textual features such as cohesion, markers of genre, register. All of the latter will, in turn, be realized as lexical and/ or grammatical patterns, but their function is not accounted for by lexicogrammar. “Explicitness” on this level can furthermore be a result of global textual patterns (such as type-token ratio, lexical density, etc.), which are “epiphenomena” of lexicogrammatical patterns, but not lexicogrammatical themselves.

Explicitation

“Explicitation” is a process or a product, not a property. The products resulting from “explicitation” are more “explicit” lexicogrammatically and cohesively than their counterparts. Explicitation is defined on instantiated, indexed and aligned pieces of discourse/ text, translations in particular, which share all or some of their meaning.

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.

Frequency

Frequency of use can be found in several of the studies in Halliday 2005. In Bybee and Hopper (2001:2f), frequency is accorded a key role in the *emergence* of structure in discourse (Cf. in particular McWhinney 2001: 449ff).

Review of argumentation

Objects of investigation

- Brief review of (structure-based) notions of **language contact**
- “**Multilinguality**” as a property of discourses

A *multifunctional* and *feature-based* **perspective** on language contact and multilinguality would add much-needed depth to considerations of

- systems alongside structures, and thus of the linguistic potential
- the instance alongside the system, and thus to empirical research methodologies
- more abstract types of contrast than have often been in the centre of theorizing, allowing us to reflect on properties, rather than structures
- the metafunctional modularization of language, thus creating an interface to “functional/ communicatively oriented” models of language and culture.

Description of a **methodology** for the empirical investigation of translations

- corpus architecture
- types of contrast
- hypothesis formation
- querying techniques
- some first results on low-level properties of our texts (lexical density, type-token statistics, part-of-speech profiles), which can be interpreted as to their contribution to explicitness and explicitation.

Outlook:

- “directionality of change” (Traugott and Dasher 2005)
- “frequency of usage” (Bybee and Hopper 2001)

in empirical studies of language contact (and change)

Multilingual texts, translations in particular, have specific properties such as

- Normalization and interference (Toury)
- Normalization and shining-through (Teich 2003)
- Normalization, simplification, explicitation, levelling-out, sanitization (Baker et al)
- Interpersonal, ideational orientations, explicitness, vs. implicitness, density vs. linearity of information packaging (House 2002)
- Characteristic degrees of grammatical metaphoricity and / or other characteristic register properties (Hansen 2003, Neumann 2003)
- Higher explicitness, lower density, different kinds of directness (Steiner 2001, 2005a,b,c; within a different approach also Fabricius-Hansen 1996; several contributions in Hasselgard, Johansson, Behrens and Fabricius-Hansen. eds. 2002)
- Characteristic types of subjectivity, intersubjectivity, objectivity (Traugott 1982, Traugott and Dasher 2005; Halliday and Matthiessen 1999)

References

- Baker, Mona. 1996. "Corpus-based translation studies: The challenges that lie ahead." In: Somers, H. (ed.) 1996. *Terminology, LSP and translation: Studies in Language Engineering in Honour of Juan C. Sager*. Amsterdam: John benjamins
- Baumann, S. C. Brinckmann, s. Hansen-Schirra, G.-J. Kruiff, I.Kruiff-korbayova, S. Neumann, E.Steiner, E.Teich & Hans Uszkoreit. 2004. "The MULI Project: Annotation and Analysis of Information Structure in German and English", in: *Proceedings of LREC 2004*. 26-28 May. Lisbon, Portugal
- Berry, Margaret, Butler, Christopher, Fawcett, Robin and Huang, Guowen (eds.) 1996 *Meaning and Form: Systemic Functional Interpretations. Meaning and Choice in Language: Studies for Michael Halliday*
- Biber, Douglas. 1995. *Dimensions of register variation: a cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Bybee, Joan and Hopper, Paul. Eds. 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins
- Caffarel, Alice, Jim.R. Martin & Christian M.I.M. Matthiessen (eds.). 2004. *Functional Typology*. Amsterdam: Benjamins.
- Doherty, Monika. (ed.) 1996. *Information structure as a problem of translation*. Special Issue of *Linguistics*. 34.

Doherty, Monika. 2004a. Reorganizing Dependencies SPRIKreports Reports of the project *Languages in Contrast* (**Språk i kontrast**) <http://www.hf.uio.no/german/sprik> No. 23, October 2004. pp 18

Doherty, Monika. 2004b. SAP - a strategy of attributive parsimony and its constraints characterizing translations between English and German SPRIKreports Reports of the project *Languages in Contrast* (**Språk i kontrast**) <http://www.hf.uio.no/german/sprik> No. 24, October 2004. pp 22

Ellis, Jeffrey 1966 *Towards a General Comparative Linguistics* The Hague: Mouton

Fabricius-Hansen. Cathrine. 1996. "Informational density: a problem for translation theory". Doherty (ed.) 1996: 521–565.

Franceschini, Rita. in press „Weshalb braucht die Linguistik mehrsprachige Sprecher?“ In: *Revue française de linguistique appliquée" dédié au plurilinguisme*; éd.: Gaetano Berruto

Halliday, M.A.K. 1978. *Language as social semiotic. The social interpretation of language and meaning*. London: Edward Arnold

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

Halliday, M.A.K. and Martin, J. 1993. *Writing science: Literacy and Discursive Power*. London: The Falmer Press

Halliday, M.A.K. 2005. *Computational and Quantitative Studies*. Vol. 6 of Collected Works of M.A.K. Halliday. London: Continuum

Halliday, M.A.K. and Christian.M.I.M. Matthiessen. 1999. *Construing Experience Through Meaning. A Language-Based Approach to Cognition*. London: Cassell.

Halliday, M.A.K. and Christian.M.I.M. Matthiessen. 2004. *An Introduction to Functional Grammar* London: Arnold (earlier versions by Halliday in 1985/ 1994).

Hansen, Silvia. 2003. *The Nature of Translated Text. An interdisciplinary methodology for the investigation of the specific properties of translations*. Saarbrücken: Saarbrücken Dissertations in Computational Linguistics and Language Technology. Vol. 13

Hansen-Schirra, Silvia, Stella Neumann, and Mihaela Vela: 2006 *Multi-dimensional Annotation and Alignment in an English-German Translation Corpus*. To be presented at the EACL Workshop on Multi-dimensional Markup in Natural Language Processing (NLPXML-2006), Trento, Italy, April 2006. Proceedings of EACL

Hansen-Schirra, Silvia, Stella Neumann and Erich Steiner 2006. Cohesion and explicitation in an English-German translation corpus. SPRIK-Report

Haspelmath, Martin.; König Ekkehard., Oesterreicher, Wulf. and Raible, Wolfgang. (eds.) 2001. *Language Typology and Language Universals*. [Handbücher zur Sprach- und Kommunikationswissenschaft 20.1. und 20.2.] Berlin, New York: Mouton De Gruyter.

Hasselgard, Hilde., Johansson, Stig, Behrens, Bergljot. and Fabricius-Hansen, Cathrine. (eds.) 2002. *Information Structure in a Cross-Linguistic Perspective* Amsterdam, New York: Rodopi.

Hawkins, John 1986. *A comparative typology of English and German*. London: Croom Helm

Hopper, Paul J. and Thompson, Sandra A. (eds.) 1982. *Studies in transitivity* (Syntax and semantics 15.). New York: Academic Press

House, Juliane. 1997. *A model for translation quality assessment. A Model revisited*. Tuebingen: Gunter Narr

House, Juliane. 2002. "Maintenance and convergence in translation – some methods for corpus-based investigations" in Hasselgard et al. (eds.) 2002. 199-212.

House, Juliane and Rehbein, Jochen (eds.) 2004. *Multilingual Communication*. Amsterdam: John Benjamins

Langacker, Ronald. 1999. *Grammar and Conceptualization*. Berlin/New York: Mouton de Gruyter

Lehmann, Christian. 1995. *Thoughts on Grammaticalization*. München – Newcastle: Lincom Europa (originally 1982)

Lindquist, Hans and Mair, Christian (eds.) 2004. *Corpus Approaches to Grammaticalization in English*. Amsterdam: John Benjamins

Macwhinney, Brian. 2001. "Emergentist approaches to language". In: Bybee and Hoper. eds. 2001: 449ff.

Mahlberg, M. 2004. *English General Nouns: a Corpus Theoretical Approach* PhD Thesis Philosophische Fakultät II, Universität des Saarlandes

Matthiessen 2004 "Descriptive Motifs and Generalizations" in: Caffarel, Alice, J.R. Martin & Christian M.I.M. Matthiessen (eds.). 2004. 537-674

Neumann, Stella Maria 2003 *Die Beschreibung von Textsorten und ihre Nutzung beim Übersetzen. Eine systemisch-funktionale Korpusanalyse englischer und deutscher Reiseführer*. Frankfurt/M. etc.: Peter Lang

Oesterreicher, W. 2001. „Historizität – Sprachvariation, Sprachverschiedenheit, Sprachwandel“. in: Haspelmath et al. (eds.) 2001. Vol. 2. 1554 – 1596

Steiner, Erich 1991 *A Functional Perspective on Language, Action and Interpretation* Berlin: Mouton De Gruyter

Steiner, Erich. 1994. "A fragment of a multilingual transfer component and its relation to discourse knowledge." In: W. Ramm. (ed.) 1994. *Text and Context in Machine Translation: aspects of Discourse representation in discourse processing. Studies in Machine Translation and natural Language processing*. Brussels and Luxemburg; commission of the European Community. Also published in Berry, Butler, Fawcett and Huang (eds.) 1996: 601-636

Steiner, Erich. 2001. "Translations English - German: investigating the relative importance of systemic contrasts and of the text-type "translation"". Papers from the 2000 Symposium on Information Structures across languages, in: SPRIK-Reports No.7. Reports from the Project "Languages in Contrast" University of Oslo. <http://www.hf.uio.no/german/sprik/english/reports.shtml> pp. 1-48; updated version in Steiner 2004a

Steiner, Erich and Yallop, Colin. (eds.) 2001. *Exploring Translation and Multilingual Text Production: Beyond Content*. Berlin: Mouton De Gruyter

Steiner, Erich. 2004a. *Translated Texts: Properties, Variants, Evaluations*. Frankfurt/M.: Peter Lang Verlag

Steiner, Erich. 2004b. "Ideational grammatical metaphor: exploring some implications for the overall model." In: *Languages in Contrast*. Vol. 4:1 (2004), 139-166

Steiner, E. 2005a. Some properties of texts in terms of 'information distribution across languages'. *Languages in Contrast* 5:1 (2004-2005): 49-72.

Steiner, E. 2005b. Some properties of lexicogrammatical encoding and their implications for situations of language contact and multilinguality. In: Franceschini, R. (ed.), *Zeitschrift für Literaturwissenschaft und Linguistik* Jahrgang 35, Heft 139. Stuttgart: Metzler Verlag. 54-75.

Steiner, E. 2005c. Explicitation, its lexicogrammatical realization, and its determining (independent) variables – towards an empirical and corpus-based methodology. *SPRIK-reports* no. 36. Oslo.

Steiner, Erich. and Teich, Elke. 2004. “Metafunctional profile for the grammar of German” in: Caffarel, Alice, J.R. Martin & Christian M.I.M. Matthiessen (eds.). 2004. 139-184

Taverniers, Miriam. 2006. “Subjecthood and the notion of instantiation” In: Steiner, E. and Vandenberg, A. eds. 2006. Special issue of *Language Sciences on Functional approaches to discourse: perspectives, interactions, and recent developments*

Teich, Elke. , Silvia.Hansen & Peter. Fankhauser. 2001. “Representing and querying multi-layer annotated corpora.” In: *Proceedings of the IRCS Workshop on Linguistic Databases*. Philadelphia. 228-237

Teich, Elke. 2003. *Cross-linguistic Variation in System and Text – A Methodology for the Investigation of Translations and Comparable Texts*. Berlin, New York: Mouton De Gruyter (published version of Habilitationsschrift of 2001)

Thomason, Sara G. and Kaufmann, Terrence. 1988. *Language Contact, Creolization, and Genetic Linguistics* Berkeley: University of California Press

Traugott, Elisabeth G. & Dasher, Richard. 2005. *Regularity in Semantic Change*, Cambridge: CUP

Weinreich, Uriel. 1953 *Languages in Contact* The Hague: Mouton