

# SPRIKreports

Reports of the project *Languages in Contrast* (**Språk i kontrast**)  
<http://www.hf.uio.no/forskningsprosjekter/sprik>

No. 36, December 2005

**Explicitation, its lexicogrammatical realization, and  
its determining (independent) variables – towards an  
empirical and corpus-based methodology**

**Erich Steiner**

Lehrstuhl für Englische Sprach- und Übersetzungswissenschaft  
Institut für Angewandte Sprachwissenschaft sowie  
Übersetzen und Dolmetschen (FR 4.6)  
Postfach 15 11 50  
Universität des Saarlandes  
D- 66041 Saarbrücken

**SPRIK**  
SPRÅK I KONTRAST

# **Explicitation, its lexicogrammatical realization, and its determining (independent) variables – towards an empirical and corpus-based methodology<sup>1</sup>**

Erich Steiner

## **Contents:**

1. Specific properties of translations and “explicitation” – state of the art
2. “Explicitation” and the notion of “explicitness”
3. Explicitness, explicitation and linguistic levels
4. The metafunctional modularization of language and “explicitness”
5. Boundaries of the notion of “explicitness”
6. An operationalization of linguistic phenomena in terms of properties of lexicogrammatical constructions – micro-level operationalizations and hypotheses
7. An example: a comparison of two original texts in terms of explicitness, density and directness
8. Towards an annotation scheme
9. Independent and dependent variables

---

<sup>1</sup> This paper is part of the first phase of DFG-project STE 840/5-1 “Linguistic properties of translations – a corpus-based investigation for the language pair English-German” which started in June 2005 (<http://fr46.uni-saarland.de/croco/>). Substantial parts of this report were written during a research period with the SPRIK-project of the University of Oslo in September/ October 2005. I am grateful to SPRIK and its members for creating this possibility and for discussions of various aspects of the paper, in particular Cathrine Fabricius-Hansen, Stig Johansson, Bergljot Behrens, Hilde Hasselgard and Wiebke Ramm. I am also very much indebted to colleagues in and around the project in Saarbruecken, where Silvia Hansen-Schirra and Stella Neumann are co-authors of the project proposal, Kerstin Kunz and Mihaela Vela are constant contributors, and Alberto Gil, Hans Haller, David Horton and Andrea Kamm have provided critical comments at several points. It should also be obvious how the work of Elke Teich, now Darmstadt, has been influential in the genesis of our project in a general way. None of those mentioned, though, should be held responsible for any remaining weaknesses.

## Abstract

In the following paper, notions of “explicitness” of lexicogrammatical encoding, and of “explicitation” as a relationship between translation and original are proposed and some operationalizations in terms of properties of linguistic constructions are suggested. The categories used in these operationalizations are intended to form the basis for an annotation scheme in an empirical corpus-based project investigating properties of translated texts between English and German. After a brief overview of the state of the art in investigations of properties of translated texts, and after a short characterization of the corpora to be created, some general approaches to “explicitness” and “explicitation” in linguistics will be discussed, and the current approach will be situated relative to them. This is followed by a stratification of the notion of “explicitness” according to the linguistic levels of lexicogrammar and text. In addition to this initial specification of the notion, a modularization of “explicitness” will be proposed into not only the levels of lexicogrammar and text, but also according to linguistic metafunctions. A parallel modularization will be given for the properties of “directness” and “density”, as these interact heavily with explicitness. We shall furthermore define “explicitation” as a relationship and a process between instantiated and aligned pieces of translated or otherwise registerially closely related texts. After the general notions of “explicitness” and “explicitation” have thus been specified, an attempt will be made to outline its conceptual boundaries, discussing types of phenomena which appear to be similar at first sight, but which represent different phenomena on closer inspection and are thus outside our immediate research interest. As the concept of explicitness developed up to that point is still too abstract to be directly quantified on linguistic data in electronic corpora, a series of further micro-level operationalizations will be undertaken which are meant to bring the relevant phenomena down to an empirical level at which they can be directly identified in linguistic data. In the following sample analysis of two texts according to the variables addressed in our operationalization, the functioning of the approach will be illustrated and detailed problems and open questions of analysis discussed. We shall finish with the proposal of an annotation scheme to be used in the analysis, as well as with a summary of the independent and dependent variables characterizing the textual varieties represented in the linguistic data constituting our corpora.

In an overall perspective, the project within which the following remarks are situated aims at

- the development and application of empirical methods in humanities research, linguistics and translation studies in particular
- the narrowing of the gap between the top-down example-based hermeneutic tradition in the humanities and the bottom-up empirical inductive methodologies currently emerging in the form of electronic corpora and information extraction technologies.

An example of the “gap” referred to above is the notion of “explicitness”, sometimes also “explicitation”. As we shall see, the notion of “explicitness/ implicitness” of texts has often been used, both applied to texts/ discourses, and to sentences/ clauses. In the present context, there are then two more specific methodological questions:

- How do we operationalize the high level notion of “explicitness”, given that empirical data are very low-level (word, shallow phrase structure), at least if by “data” we mean not the analysis of a few clauses, but representative corpora?
- Is the “explicitness/ implicitness” of higher level units such as texts/ discourses simply the sum total of “explicitness/ implicitness” features of clauses, or is it rather an “emergent property” on a higher level?

The specific task of this paper is to develop the basis for an annotation scheme for the bilingual annotation on different linguistic levels. The dependent variables have to be indicators of the property of “explicitation/ explicitness” to be investigated in the project. This is to be followed by testing and refinement of the scheme and later on the larger-scale application to the corpus. This gradual approach is designed to make the scheme robust and independently applicable with sufficient inter-annotator reliability.

## 1. Specific Properties of translations and “explicitation” – state of the art

There is a small, but significant tradition of work on assumed properties of translations as text, and more recently as “text-type” or “register”. The following are to various degrees part of this tradition, without, of course, representing a complete overview:

- Levy 1963 on lexical impoverishment and explicitation
- Duff 1981 on translationese
- Berman (1984/ 2000): on rationalization, clarification, expansion, ennoblement, popularization, and other assumed properties of translations
- Blum-Kulka (1986) on explicitation in translation
- Sager 1994 (179ff): on translation text types
- Toury 1995: on growing standardization vs. interference
- Baker 1996, Laviosa -Braithwaite 1998, Olohan 2001: simplification, normalization, levelling-out, explicitation, sanitization (Kenny 1998)
- Fabricius-Hansen 1996, 1999, Ramm and Fabricius-Hansen 2005: on changing informational density in translations, and on explicitness of discourse relations
- Doherty 1991, 1996 (ed), 2004a,b: on perfect adequacy through adaptation to the stylistic principles of the target language system
- House 1977/97: on covert vs. overt translation, and “cross-cultural pragmatics” (especially informational explicitness vs. implicitness (2002:200))
- Vinay/Darbelnet 1958: 182ff
- Johansson 2004, 2005, as well as Hasselgard 2004, on properties of (English-Norwegian) translations
- Englund Dimitrova 2005: on level of expertise as one determinant of explicitation in the translation process

While “explicitness/ explicitation” has been postulated as one possible property of translated texts by most of those just mentioned, some have suggested other phenomena and relationships, as indicated in our list above. Yet even those properties will at least contribute to “explicitness/ explicitation” in the sense assumed here. The very different approaches to translation of Doherty and House have been particularly influential for our work, even though the property of “explicitation” as such plays a significant role only in the latter’s work.

House postulates “explicitness vs. implicitness” (cf. House 2002: 200) as properties of texts within her cross-cultural pragmatics, alongside “directness vs. indirectness; orientation towards self vs. orientation towards other; orientation towards content vs. persons; ad-hoc formulations vs. verbal routines”. This is where our efforts clearly meet, with the difference that we are concentrating more on the micro-structural realizations of what a possible property of “explicitness” may mean for texts as wholes.

Doherty (e.g. 1991, 1996, 2004a,b) may be seen to represent a sort of “null-hypothesis”, which would be that translations do not have (should not have) specific textual properties. Through her rich and controlled set of theoretically inter-related language-specific parameters of and constraints on information distribution she attempts to describe and explain how an optimal translation is a perfect text in its target language. The work to be undertaken in our

project here should lead to a framework for empirically testing such claims: Whereas our initial hypothesis is that translated texts may be more explicit and/ or dense and/ or direct than registerially parallel texts in their target language, it is a hypothesis and may well be disconfirmed. Ultimately, and on a general level, our assumption is that translated texts may indeed be somewhat different from their parallel texts in their target language, though in ways which do not make them inferior, but interestingly different texts, and thus potential catalysts in situations of language contact and language change. However this may be, our main goal is to create possibilities for empirically testing such claims.

After this very brief review of some main lines in the investigation of properties of translated texts, let us add a few remarks about the corpus architecture which the current project is designed to develop:

Earlier exploratory work leading up to the current attempts is published in Baumann et al 2004, Hansen 2003, Neumann 2003, Steiner 2001, 2004a,b, 2005 a,b, Steiner and Teich 2004, Teich 2003, Teich et al 2001. In this strand work, we have developed different aspects of corpus architecture, and of variables to be investigated. The corpus which we are building up at the moment consists of a background of cross-register reference corpora in English and German, together with both independent and aligned register-specific corpora of originals and their translations, all of these on the basis of representative excerpts from full texts. The corpora are represented as layered XML-data and indexed for correspondences by layer of representation and as originals and translations. This is illustrated in Figures 1 to 3 below (taken from Neumann and Hansen-Schirra 2005). The question of the operationalization of the property of explicitness (sometimes in interaction with those of density and directness) poses itself independently of any particular corpus-design, but is certainly aimed at the type of multi-level annotated electronic corpus of the type exemplified here. The corpus itself is described in more detail in Neumann 2005.

Figure 1 shows our corpus architecture, consisting of the cross-register reference corpora (ER, GR), the register-controlled corpora of originals and their translations into German and English (EO/ GTrans and GO/ ETrans). Note that the corpus itself consists of randomly sampled excerpts, rather than full texts.

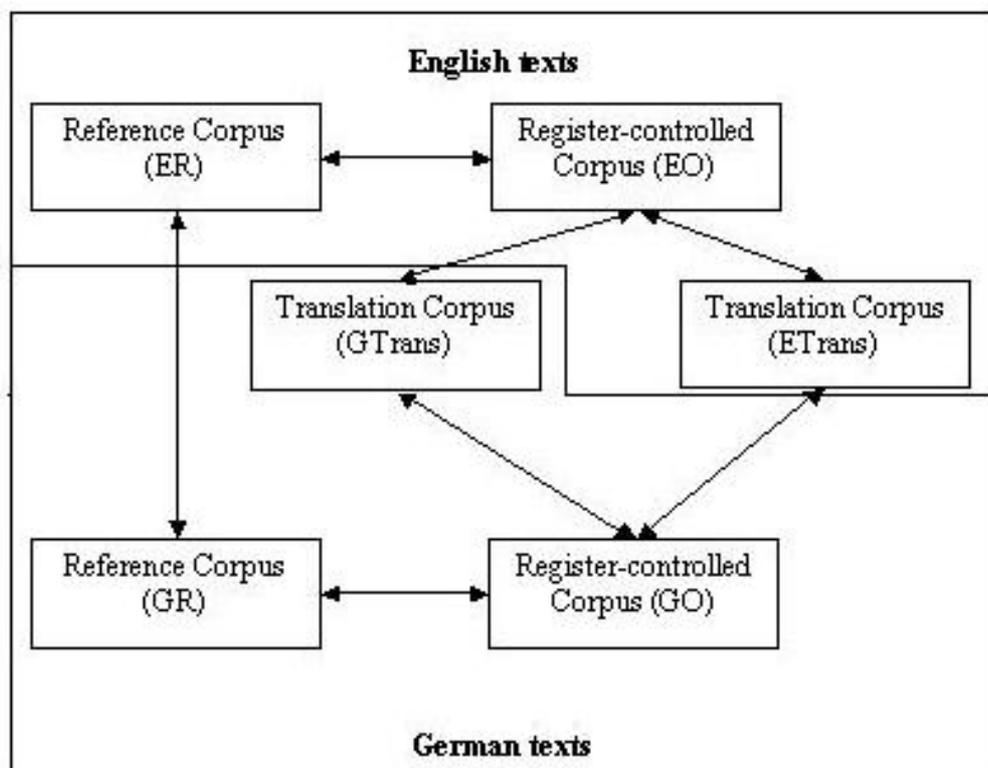


Figure 1: The CroCo Corpus design (Neumann and Hansen-Schirra 2005)

Figures 2 and 3 show multi-layered annotated corpus structures, both in aligned and non-aligned constellations. With this architecture, we are aiming at inclusion and control of our theoretically motivated annotations on various layers in our sub-corpora (for more detail cf. Neumann and Hansen-Schirra 2005). Note that in the aligned corpora, we align grammatical units. “Translation units” may emerge as a result of our work, but are at this stage not taken as something established in our annotations.

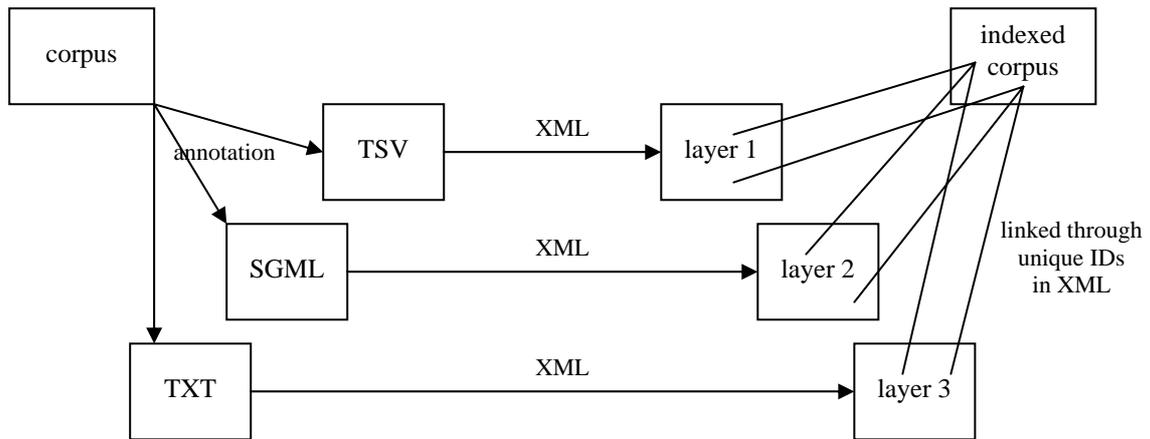


Figure 2. Multi-layer annotation in XML stand-off mark-up (Neumann and Hansen-Schirra 2005)

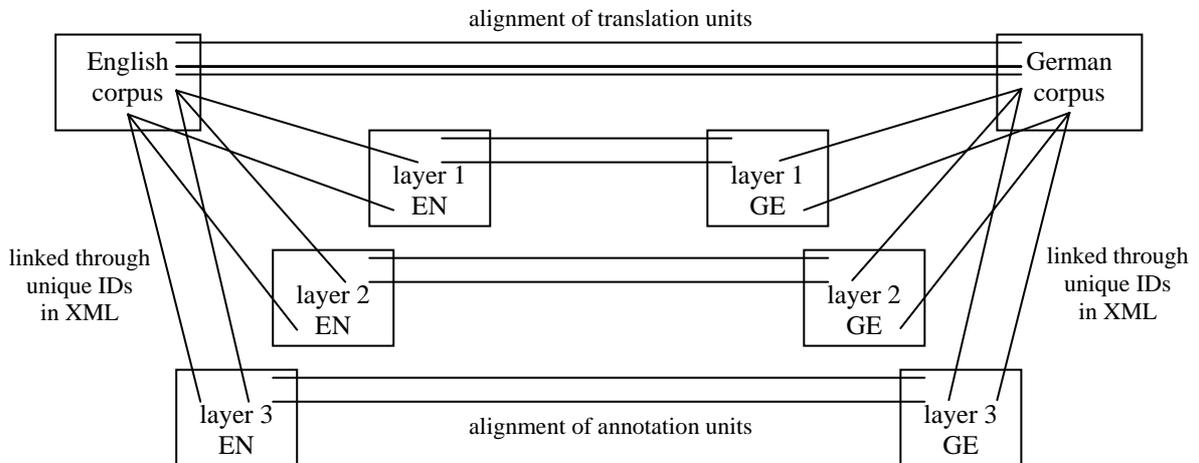


Figure 3. Alignment of source and target language annotation (Neumann and Hansen-Schirra 2005)

## 2. “Explicitation” and the notion of “explicitness”

The notion of “explicitness”, and its counterpart “implicitness”, represent a challenge in several respects. For one thing, they are very general, central to some models of language, especially for a philosophically anchored semantics (cf. e.g. Carston 2002, Burton-Roberts 2005), and highly complex in any case. Furthermore, a methodologically empirical project will have as data not high-level interpretations of utterances by human interpreters, but text corpora with relatively low-level lexicogrammatical and cohesive categories in multi-level annotations. The data thus yield information about *properties* of encoding, rather than about high level *interpretations*. Precisely the former is the focus of the current project – the attempt to empirically enquire into properties of encoding which have to do with “explicitness”, rather than adding yet another set of example-based discussions of (interpretations of) the data.

While we have developed notions leading up to this discussion in earlier papers (cf. references in section 1), we have only very sketchily compared them to other traditions in linguistics. Therefore, some remarks situating our notion of “explicitness/ explicitation” in a wider discussion may be in place here:

“Explicitation” and “explicitness” are both highly general and intuitively attractive notions for any model of text and discourse. However, if applied without further focus, they do not have the status of concepts triggering specific research hypotheses and being amenable to empirical testing. We shall therefore discuss a few relevant general definitions, gradually focussing in on more specific notions of “explicitness” which are not restricted to, but firmly anchored in, lexicogrammatical realization and thus amenable to empirical investigation.

An initial distinction is therefore suggested between “explicitation” and “explicitness”: “explicitation” is a process, or a relationship, which assumes that some meaning “is made explicit” in moving from one text or discourse to some other one. It also assumes that in some sense, whatever is “explicitated” must have been “implicit” in the other variant. As a result of such “processes”, some textual variant may be more or less “explicit” than the other, and it is to this term that we turn to next.

Linke and Nussbaumer (2000: 435ff) in “Concepts of Implicitness: Presuppositions and Implicatures” (my translation) anchor their discussion in the wide-spread metaphor, or allegory, which conceptualizes texts as “icebergs”: only a smaller part of them is visible, the larger part is hidden from perception. The visible part (of form and meaning) is called “explicit”, the invisible part “implicit”.

More specifically, they draw a distinction between meanings which are *implicit, non-literal, dependent on use* (the province of pragmatics) on the one hand, and those meanings which *are literal and independent of use* (the province of semantics). Only within the latter do they distinguish between *implicit (non-realized) and explicit*.

In a first attempt at situating our own concept of “explicitness” here, it appears as if our classification cuts across theirs, even though the two can be related:

First, in our corpus-based research design, we can only investigate meanings which are *explicit* in one of the registerial variants compared to some other, or else can be grammatically

or cohesively related *as explicit/ implicit variants* to our data. What remains outside of our methodology is simple *adding or dropping* of meanings without any grammatical or cohesive relationships between variants.

Second, the meanings which we investigate do not have to be *literal*, they may, indeed, be (grammatically or lexically) metaphorical, provided they are “explicit” in one of our variants (registers, translations).

And, finally, the meanings which we investigate are dependent on use in the sense that the data are drawn from linguistic instantiations, i.e. texts. However, our operationalizations in terms of lexicogrammatical or cohesive realization will bias our observations towards whatever is grammaticalized and lexicalized, or at least highly conventionalized (cohesive relations, rhetorical relations), and in that sense we will appear quite system- and grammar-oriented.

The reason why our perspective seems to cut across that of Linke and Nussbaumer is that by being corpus-based, and thus product-based, rather than interpretation-based and process-based, we are methodologically forced to link our investigations to lexicogrammatical realization, so that any meanings which they call “pragmatic” and which are not linked systematically to realization appear invisible to our perspective – which is not the same as saying they are unimportant. They will feature in our hermeneutic example-based interpretations, but only there. Outside our particular research design, they can, and have, become the object of investigation through psychological testing (for which cf. Hansen 2003).

Let us furthermore attempt to situate our own methodology relative to a discussion contrasting “Relevance Theory” with Gricean Pragmatics, this time taken from Burton-Roberts’ (2005: 389ff) review of Carstons’s (2002) *Thoughts and utterances: the pragmatics of explicit communication*:

Quite in parallel to what we said about the Linke/ Nussbaumer model above, we would, like Burton-Roberts (and Carston), say that our explicit vs. implicit distinction cuts across at least several Gricean dichotomies: (A) semantics vs. pragmatics, (B) what is said vs. what is implicated, (C) explicit vs. implicit, (D) linguistically en(/de)coded vs. not linguistically en(/de)coded, (E) context-free vs. context-sensitive, (F) truth conditional (entailment) vs. non-truth-conditional (non-deductive). Furthermore, and addressing Carstons’s (2002:117) and at this point also Burton-Roberts’ (2005: 391) position, we would also claim that variants (1) (a) to (d) below cannot simply be contrasted on a binary “explicit vs. implicit” dichotomy:

- 1 (a) Mary Jones put the book by Chomsky on the table in the downstairs sitting room.
- (b) Mary put the book on the table.
- (c) She put it there.
- (d) On the table.

According to Carston and Burton-Roberts, any of 1 (a-d) above “could be used, in different contexts, to communicate explicitly one and the same proposition (or thought or assumption)” (Carston 2002: 117). This appears to be true – what we are investigating with our research design, however, is *not a communication (and interpretation)* situated in a specific context, but rather *properties of the encoding* (“explicitness”, alongside “directness” and “density”). In our terms (cf. Figure 4 below, and its application to the text analysis later on), 1 (a) to (d) are partly identical as far as ideational and interpersonal explicitness are concerned. There is also no difference between them in terms of directness, but there are differences along several dimensions in density, and there are differences in explicitness on the interpersonal and textual dimensions and in terms of some sub-parameters of cohesion.

With respect to Relevance Theory, then, the characteristic of our approach is that we measure explicitness as a property of the encoding, not as a property of the communicative act as such. And although the latter is of great significance for any attempt at understanding communication, our own approach focussing on textual encoding provides a necessary prerequisite for investigations of communication.

A distinction which seems somewhat closer to our own modelling is that of Polenz (1988: 24ff, 40ff, 92ff, 202ff): He draws a basic distinction between “elliptical, compressed/compact, and implicating” modes of expression, and their respective corresponding “full, expanded and explicating” counterparts. In his classification, what we are investigating through our methodology would be the difference between

- “compressed/ compact” modes of expression and their “expanded” counterparts,
- plus the difference between “elliptical” textures which can be related through grammar or cohesion to non-elliptical “full” counterparts,
- plus the difference between “implicit” textual configurations and their “explicit” counterparts.

It has to be said, though, that von Polenz frequently uses “explicit” as an opposite term to all three of “compressed/ compact, elliptical, implicit” (1988:24ff). Where our methodology is more constrained than von Polenz’ terms would suggest is in that we would restrict our notion of “realization” to “lexicogrammatical and cohesive realization”. We would demand some sort of lexicogrammatical reflex for an assumed “elliptical, compact/ compressed, implicit” meaning, rather than allowing any meaning as potentially “implicit” which is “addable” to the piece or discourse in question without violating coherence.

Now, in comparison to Linke/ Nussbaumer, to Carston, to Burton-Roberts, and to a lesser extent to von Polenz, our methodology appears restrictive in the sense of being tied to formal realization. However, all of the realizational patterns are only considered to be signals, instructions, to the full (inter-)textual meaning, and in that sense, we are opening the door towards a fuller view, ultimately taking into account the invisible part of the “iceberg” as well. Methodologically, though, we can only do this in our example-based hermeneutic interpretations of individual examples, not in the empirical part of our investigations. Philosophically, this is, of course, only to be expected if we assume that understanding texts is ultimately not an empirical, but an interpretative hermeneutic exercise.

Methodologically, we are in some respects closer to e.g. Biber (1988: 142ff on “explicit vs. situation-dependent reference”, but also Biber 1995: 157ff, 161ff), than to the authors above. However, we do believe that it is possible to develop a linguistically richer and theoretically better motivated notion of “data” than is used by Biber, while building on his achievements in making linguistic enquiry a more empirical discipline than before. The linguistically richer conceptual tools to be outlined below and based on the notions of “grammatical metaphor” and on metafunctional diversification (Halliday and Matthiessen 1999, 2004, Matthiessen 2004) are intended to narrow the gap between the more conceptual and hermeneutic top-down and the more empirical bottom-up approaches.

### 3. Explicitness, explicitation and linguistic levels

After having clarified the present notions of “explicitation” and “explicitness” by relating them to some discussions in the literature, let us now attempt a stratification of the notion of “explicitness” in terms of the linguistic levels of lexicogrammar and text. There are relevant phenomena intuitively classified under the general notion of “explicitness” on either level, but they are not of the same degree of abstraction from the data: it is not the case that textual explicitness is simply the sum of the explicitness of clauses, and whereas both are structural phenomena, inside the clause they are grammaticalized (and in a general sense cohesive) phenomena, but outside and beyond it, they are cohesive only.

“Explicitness” on the lexicogrammatical (structural) level is conceptually related to “density” and “directness”. These three are properties of (lexico-)grammatical constructions (cf. Steiner 2004b, 2005 a,b). 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;” and similar ones in the sense of those referred to in our list in section 1. The “explicitness” of higher level units such as texts/discourses is not simply the sum total of “explicitness” features of clauses. 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. “Explicitness” – a property of lexicogrammatical or cohesive structures and configurations - is measured through operationalizations as in our Figure 4 below.

“Explicitation”, as we have already said at the beginning of the previous section, is a process or a product, not a property. The products resulting from “explicitation” are more “explicit” lexicogrammatically and cohesively than their counterparts. In terms of instantiated discourse, the product of an explicitation is, of course, a more explicit discourse. Explicitation is thus defined on instantiated, indexed and aligned pieces of discourse/ text, translations in particular.

**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.

Note that the final part of this definition is meant to exclude the indefinitely many possibilities through which meaning can simply be added to some text/ discourse, without being in any sense “implicit” in the source variant. We shall discuss these phenomena more extensively below under “boundaries of the notion of explicitness”.

It should also be said that the data in our non-aligned corpora will initially only show which texts are more or less explicit, more or less dense, or compact. Whether or not “explicitation” in our further sense obtains can only be determined through the co-indexation in our aligned translation units, in principle including translation between different variants of a text within a language. “Explicitness”, and even more so “explicitation”, are thus inherently relative terms, presupposing the comparison of two or more variants.

#### **4. The metafunctional modularization of language and “explicitness”**

In the following Figure 4, we shall intersect the properties of “explicitness, density, and directness” with the linguistic metafunctions according to Systemic Functional Grammar (SFL) (Halliday 1978, Halliday and Matthiessen 1999, 2004, cf. also Steiner 2005b). Figure 4 expresses two key assumptions. The first of these is that languages modularize their lexicogrammar along a small set of functional dimensions (ideational, interpersonal, textual, and their sub-dimensions), and the second is that on each of these levels, the mapping of semantics onto lexicogrammar will vary in terms of explicitness, directness and density. Our focus here is clearly on “explicitness”, but as this property is closely related to density and directness, we would at least give an operationalization of all of these. It should also be said that, whereas the metafunctional hypothesis is a basic assumption developed and elaborated in *Systemic Functional Grammar*, the further assumption of properties of encoding, or of the mapping between semantics and grammar, is not foregrounded in current formulations of the model. It is, however, in my view directly derivable from SFL.

A discussion of a number of specific questions to do with operationalizations, as well as an application to two sample texts will be postponed until section 7 below. At this point, we shall present the operationalizations without further discussion.

Properties Metafunctions	Explicitness	Directness	Density
Ideational: Experiential	No. of explicit functions per discourse unit: No. of implicit functions	No. of directly mapped experiential clause functions : No. of indirectly/ metaphorically mapped clause functions	No. of functions: Grammatical Unit CC: CI ranking CI ranking: CI embedded Clauses finite: Clauses non-finite Clause: Clause elements Phrases: Phrase elements Groups: Group elements
Ideational: Logical <sup>2</sup>	No. of explicit functions: No. of implicit functions	No. of directly mapped logical functions per unit: No. of indirectly/ metaphorically mapped functions	No. of functions: Discourse-Unit
Interpersonal: Mood	No. of explicit Mood-markers per discourse unit: No. of implicit Mood-markers	No. of directly/ congruently marked Mood options: No. of indirectly marked Mood-options	No. of Mood-markers: Discourse Segment
Interpersonal: Modality	No. of explicit Modality-markers: No. of implicit Modality-markers	No. of directly/congruently marked Modality options: No. of indirectly marked Modality-options	No. of Modality-markers: Discourse Segment
Textual: Theme	No. of auto-semantic Themes: No. of syn-semantic (phoric) Themes	Not applicable	No. of Themes: Clause Functions
Textual: Information	Not applicable	Not applicable	No. of NEW elements: Discourse Segment No. of marked NEW elements: Discourse Segment
Cohesion: Reference	No. of implicit referents: No. explicit referents; No. of phoric phrases: No. of autosemantic phrases	Not applicable	No. of newly introduced referents per discourse segment
Cohesion: Ellipsis	No. of cohesive ellipses: Discourse Segment	Not applicable	Not applicable
Cohesion: Substitution	No. of cohesive substitution: Discourse Segment	Not applicable	Not applicable
Cohesion: Lexical	No. of lexical words: No of grammatical words	Not applicable (lexical metaphor is not yet counted here)	Type-token ratio; Lexical words per clause
Cohesion: Conjunction	No. of explicit relations: No. of Figures	No. of cohesively realized relations: No. of non-realized RST realizations	No. of relations

**Figure 4:** Properties of lexicogrammatical encoding across metafunctions.

<sup>2</sup> Grammaticalized/ lexicalized expressions only, i.e. not counting here conjunctive cohesive ties

The operationalizations given in Figure 4 are meant to constitute a frame for classifying texts, and constructions, in terms of three properties and along all of the SFL-metafunctions as well as in terms of cohesion. Some of these can be directly queried in our tagged corpora, while others will at this stage still have to be computed by human interpreters based on the data in our corpora.

If we apply a scheme as here in 4 to aligned pairs (or tuples) of textual variants, e.g. translations, what it measures directly is explicitness, density or directness *of constructions*. Only if we then trace the information in co-indexed (instantiated) matches do we find traces of whether or not any particular quota of meaning is explicitated (or otherwise) in one of the variants. The former can be done on the basis of fully or semi-automatic information extraction, whereas the latter is, for the time being, only accessible through human interpretation of the aligned variants.

## 5. Boundaries of the notion of “explicitness”

So far, a selective overview has been given of relevant investigations of properties of translated texts, and some relevant approaches to “explicitness” and “explicitation” in linguistics and adjacent areas have been discussed. This was followed by a stratification of the notion of “explicitness” into the levels of lexicogrammar and text. In addition to this initial diversification of the notion, a modularization of “explicitness” has been proposed according to linguistic metafunctions and phenomena have been additionally divided between those which are lexicogrammatical, and those that are cohesive. A parallel modularization has been given for the properties of “directness” and “density”, as these heavily interact with explicitness (cf. Figure 4). “Experiential Explicitness”, however, still seems to present some particular problems: One of these is that much of what is called “explicitness” in the literature elsewhere, is covered by “logical” and generally interpersonal and textual explicitness in our multifunctional framework. So, explicitness is in no way only explicitness of experiential meaning. Another is that experiential directness and experiential density cover additional aspects of what is often referred to under the phenomenon of explicitness, albeit in an indirect way. However, there is a residual category of truly “experiential explicitness”, appearing in the top left-hand corner of our Figure 4, which needs further clarification at this point:

First, it is not the case that any quantum of information “added” to a given piece of discourse is a form of “explicitation”. Very often it is simply “adding information and meaning”. In order for the term of “explicitation” to make sense, we need to have a notion of “implicit information”. Most approaches to discourse representation have such a notion, for example in the work of “discourse representation theory” (Kamp and Reyle, 1993; Asher 1993; applied to problems of translation cf. Fabricius-Hansen 1996). Usually, it means something like “information which must be added to linguistically explicitly expressed information in order for (the interpretation of) a piece of discourse to become complete and coherent”. I find this notion intuitively appealing (cf. also Steiner 1991: 80ff), yet difficult to apply in empirical work, unless we tie it down to linguistic clues for implicitness. Where there are no such clues, our methodology will not identify any implicit meaning.

At the other (and very “grammatical”) end of the spectrum, there are notions of lexicogrammatically encoded types of implicitness, for example in non-finite constructions unrealized participant roles, logico-semantic relators (conjunctions, prepositions), tense, aspect, number or in lexicogrammmatical “cryptotypes” in general. Staying more on the word-level, and with English as a specific language, grammarians (e.g. Dixon 1991: 68-71) have

noted the optional dropping of complementizers, relative pronouns, copulas from complement clauses. In all of these cases, of course, it can be argued that the (highly generalized) grammatical meaning signalled by the absence of the word is there, at least in the features of the construction, if not in lexicogrammatical functions. It can be brought out by contrasting the construction with its agnates. However, this notion of “implicitness” is very grammar-oriented and thus also very language specific. It will be visible to our methodology, and will be used as a relevant indicator – although not necessarily of experiential, but often of logical, interpersonal or textual meaning.

Then there are SFL-based notions of “implicitness/ explicitness”, as in accounts of modality and modulation (Halliday and Matthiessen 2004: 620ff), or inferred/ implicit discourse relations, often triggered by genre or register (Halliday and Matthiessen 2004: 363ff). A further context for the notion of “implicitness” is cohesive ellipsis (Halliday and Hasan 1976: 142ff). And there is, of course, the important notion that grammatical metaphor, at least the type involving relocation in rank between semantics and grammar, has far-reaching influences on how much and what kinds of information are made explicit (Halliday and Matthiessen 1999: 231ff; 258; 270; Halliday and Martin 1993). These are our starting points for recognizing more and higher-level types of implicit meaning, but operationalizations at the borderlines (i.e. those to do with genre and register) are not far enough advanced to admit reliable quantification in all cases.

In order to make the boundaries of our notion of “experiential explicitness” clearer, let us give three examples of phenomena, all of which are sometimes discussed under the label of “explicitness”, or “explicitation”. Our aim here will be to illustrate which types of phenomena are captured by our operationalizations:

The following ((2) and (3)) are two short excerpts from popular scientific prose in English and German (cf. House 2002: 205). These are meant to illustrate phenomena which, however frequent they may be in real-world translations, and however much they may contribute to the cultural and contextual acceptability of a piece of text, are not subsumed under our notion of “experiential explicitness”

(2.1.) *Groundbreaking work that began more than a quarter of a century ago has led to ongoing insights about brain organization and consciousness.*

(2.2.) *Jahrzehntelange Studien an Patienten mit chirurgisch getrennten Großhirnhälften haben das Verständnis für den funktionellen Aufbau des Gehirns und das Wesen des Bewußtseins vertieft.*

(M. Gazzangia: The Split Brain Revisited, *Scientific American*, July 1998 – Rechtes und linkes Gehirn: Split-Brain und Bewußtsein, *Spektrum der Wissenschaft*, December 1998.)

My own word-for-word back-translation of the German:

(2.3.) *Decade-long studies on patients with surgically-separate brain-halves have the understanding for the functional structure of the brain and the essence of consciousness deepened.*

Cases such as these, not infrequent in translations, have referents and properties added and/ or missing in the translation relative to the source text (marked by boldface). This adding of experiential meaning is not explicitation, but simply adding (or in the opposite case dropping) of information – which may or may not be closely triggered by something in the immediate or wider context. If it is so triggered, as in the frequent cases of “compensation”, in which units of information have to be translated non-locally, they are again not cases of explicitation, but simply cases of non-local translations. The classification of “non-local” depends on our choice of relevant translation unit in each case.

Additionally, and very typically, we find cases of different degrees of “specificity” and/ or “vagueness” between original and translation, as in the translation “*more than a quarter of a century ago*” by “*jahrzehntelang*”. Increasing specificity of a sign is a case of explicitation, but interestingly one which does not necessitate any addition of structure, but may simply result in choosing the same structure with a more specific feature set.

While we are thus excluding cases such as those above from the notion of “explicitation” as used here, all the variants involved above do have properties along all the dimensions covered in Figure 4. Yet these measure the explicitness, density or directness of a lexicogrammatical or cohesive configuration, rather than textually or discursively instantiated “explicitation” in principle.

Our second excerpt, again taken from House 2002, shows the same phenomenon:

(3.1.) *Treatment may reduce the chance of contracting HIV infection after a risky encounter.*

(3.2.) *Eine **sofortige** Behandlung nach Kontakt mit einer Ansteckungsquelle verringert unter Umständen die Gefahr, dass sich das Human-Immunschwäche-Virus im Körper festsetzt. **Gewähr gibt es keine, zudem erwachsen eigene Risiken.***

(S. Buchbinder: Avoiding Infection after HIV Exposure, *Scientific American*, July 1998 – Prävention nach HIV-Kontakt, *Spektrum der Wissenschaft*, Oktober 1998.)

My own word-for-word back-translation of the German:

(3.3.) *An **immediate** treatment after contact with a source of infection reduces under circumstances (possibly) the chance, that itself the human-immune-weakness syndrome in the body implants. **Guarantee there is none, additionally grow specific risks.***

In the original and published German translation above, the entire second sentence is added, apparently without any clear trigger in the original. Again, this would not show up as “explicitation” in our data: In terms of instantiated and indexed discourse entities, there is no match for the added sentence in its original, and therefore explicitation would not apply. In terms of the lexicogrammatical and cohesive constructions, we would of course register different degrees of explicitness, density and directness as specified in Figure 4, but these

would be outside any translation relation, which would also be shown in our (missing) alignments for the added fragments. The translation of “risky” by “Ansteckungsquelle” would also fall outside of what we are counting, unless we find a generalizable way of handling this as a form of “specification”. The same might hold for the translation of “infection” by “im Körper festsetzt”.

A different type of example ((4), cf. Fabricius-Hansen (1996:522)) would seem to fall more easily into our category of “explicitation”:

(4.1.) *Frankreich trauert über den Tod eines sehr bekannten französischen Schauspielers*

‘France mourns the death of a very famous French actor’

(4.2.) *Ein französischer Schauspieler ist gestorben. Er war sehr bekannt. Frankreich trauert über seinen Tod.*

‘A French actor has died. He was very famous. France mourns his death.’

In terms of properties of the constructions (Figure 4 above), version (4.2) here is more explicit than version (4.1) along several dimensions (although least experientially), and less dense and more direct, as we shall argue below. In addition, in terms of instantiated discourse, and as shown in alignments of translation units, the two variants are close to equal in terms of explicitness, except that the tense selections in version (2) can be seen as a real explicitation of information implicit in version (1).

A third example, constructed by me from example (2.1), will show a different, though related, type of clear “explicitation” (cf. (5) below):

(5.1.) *Groundbreaking work that began more than a quarter of a century ago has led to ongoing insights about brain organization and consciousness.*

(5.2) *Because they have broken new ground by working for more than a quarter of a century, people have become able to understand increasingly how the brain is organized and what consciousness is.*

(5.3.) *Weil man neues Terrain erforscht hat indem man mehr als ein Viertel Jahrhundert gearbeitet hat, konnten die Leute zunehmend verstehen, wie das Gehirn organisiert ist und was Bewusstsein ist.*

The example in (5) illustrates a deliberately constructed and unambiguous case of both differences in explicitness and instantial explicitation on the constructional and instantiated

discourse level: (5.2.) would be more explicit, more direct and less dense than (5.1.) on the construction level, and it would be an explicitation in instantial terms, as shown by aligned and indexed versions of the two variants (5.1.) and (5.2.).

## **6. An operationalization of linguistic phenomena in terms of properties of lexicogrammatical constructions – some micro-level operationalizations and hypotheses**

The property of “explicitness”, and the process of “explicitation”, have frequently been postulated on a merely intuitive level in hermeneutically-based work, relying on the analysis of individual, and in the best cases, representative examples. In other, more empirically-oriented approaches, they have been reduced to low-level criteria, such as words per text, words per sentence, realization of individual conjunctions or pronouns, or individual cohesive devices, lexical density etc. (cf. for example Baker 1996, Laviosa-Braithwaite 1998, Olohan 2001). We are attempting a theoretically more motivated operationalization in this paper by defining explicitness and explicitation, by stratifying it in terms of different linguistic levels, by tightening its boundaries, and by modularizing it in a multifunctional perspective. Explicitness and explicitation, however, still need to be directly connected to the data in our electronic corpora. Therefore, we shall move onwards to formulating research hypotheses first in theory-neutral terms, and then in terms of the more SFL-based notions of Figure 4. Finally, in order to test such hypotheses, we need a level of annotation, at which they can be expressed and checked against data.

The following are relatively theory-neutral operationalizations in terms of which we attempt to investigate “explicitness”:

- The proportionality form : content words
- Average number of words per clause
- Biber’s informational : involved production
- The proportions between the following lexical categorical types: conjunction: preposition; verb: noun; adverb. Adjective; finite : non-finite<sup>3</sup>
- Proportionalities between ranks (levels of projection): clause complexes: clauses: groups/ phrases: words)
- Proportionalities between directly verbally governed to directly nominally governed phrases
- Degree of specificity of lexical items
- Number of grammatical units (clauses, phrases/ groups, words) per discourse segment,<sup>4</sup> where the number correlates positively with explicitness;

---

<sup>3</sup> Note that these oppositions are ultimately scalar, rather than binary

<sup>4</sup> This is not the same as the number of intermediate phrase types *per clause*, which we shall use below in our operationalizations of density and metaphoricity in (H3) and (H4).

- Explicitness of grammatical categories, such as person, number, gender, but also diathesis, (including voice), relativization, complementation, etc.;
- Explicitness of cohesive relations, especially reference and conjunction.
- To these operationalizations, we must add those used in Figure 4 above, although from then on we cease being theoretically neutral. Figure 4 is also at a different level of abstraction, constituting an interpretation already of some of the more shallow data formulated in the earlier part of our list. And finally, the operationalizations in Figure 4 are for “explicitness”, but also for “density” and “directness”, which constitute properties of their own, even if they interact with “explicitness”.

Using these operationalizations, we are comparing the relative “explicitness” of linguistic encoding in sub-corpora, or individual texts, along the independent variables language, register, translation. And in those cases where we compare aligned texts (translations), we additionally investigate “explicitation” as a relationship between instantiated pieces of discourse, because in translationally related pairs of texts, both exemplars are assumed to refer to experientially identical (or at least very similar) instances.

The operationalizations suggested here presuppose corpora tagged, annotated and query-able for variables which we shall list in section 8 of this paper.

But let us now, by way of exemplification, illustrate how we can proceed from assumptions about “explicitness” and its related properties to research hypotheses making reference to the types of information presupposed in our operationalizations above (cf. Steiner 2005a for an earlier formulation of these).

Globally, our strategy is designed to

- 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.

Let us consider a couple of examples, some of them already used above, to make our point clearer (examples under 4 are from Fabricius-Hansen 1996, those under (6) to (16) are slightly altered from original examples by Halliday and Matthiessen 1999: 231ff):

- (4.1) *Frankreich trauert über den Tod eines sehr bekannten französischen Schauspielers*  
 ‘France mourns the death of a very famous French actor’

- (4.2) *Ein französischer Schauspieler ist gestorben. Er war sehr bekannt. Frankreich trauert über seinen Tod.*  
 ‘A French actor has died. He was very famous. France mourns his death.’
- (6) Lung cancer death rates are clearly associated with increased smoking.  
 (7) (It is clear that) if more people smoke, then more people die of lung cancer.  
 (8) (It is clear that) some people smoke more, so they die faster of lung cancer.
- (9) A rising number of people smoke.  
 (10) The number of people dying from lung cancer is increasing.  
 (11) A rising number of people smoke. *As a consequence*, the number of people dying from lung cancer is increasing.  
 (12) *Because* a rising number of people smoke, the number of people dying from lung cancer is increasing.  
 (13) *Because of* an increase in the number of smokers, the number of people dying from lung cancer is increasing.  
 (14) The increase in smoking *leads to* an increase in death rates from lung cancer.  
 (15) Increasing lung cancer death rates and the *causally related* increase in smoking ...  
 (16) *The cause of* increasing lung cancer death rates in increased smoking ...

Focussing on ‘explicitness’ first, we consider the following operationalizations:

- number of grammatical units (clauses, phrases/ groups, words) per discourse segment,<sup>5</sup> where the number correlates positively with explicitness;
- explicitness of grammatical categories, such as person, number, gender, but also diathesis, (including voice), relativization, complementation, etc. per discourse segment;
- explicitness of cohesive relations, especially reference and conjunction per discourse segment.

The notion of ‘explicitness’, in the context of this discussion, is restricted to meaning ‘lexicogrammatically and/ or cohesively realized’. Furthermore, ‘realized’ may mean ‘phorically’, rather than fully lexically realized. Formulating two hypotheses now for our three notions of ‘information structure’, ‘informational density’ and ‘grammatical metaphoricity’, we would say that while information structure has to be directly annotated or parsed in a corpus of running text, for informational density and for grammatical metaphoricity we assume that

(H1) The more informationally dense and the more grammatically metaphorical a stretch of text, the less it will be explicit grammatically (and cohesively).

(H2) The more informationally dense and the more grammatically metaphorical a stretch of text, the more the explicit grammatical and cohesive marking will be of the general nominal, rather than verbal type.<sup>6</sup>

---

<sup>5</sup> Note: this is not the same as the number of intermediate phrase types *per clause*, which we shall use below in our operationalizations of density and metaphoricity in (H3) and (H4).

By way of exemplification, we would like to point to our examples (6) to (16), but also (4.1.), (4.2), all of which positively illustrate our hypotheses (H1) and (H2). In particular, the more highly metaphorical and the informationally denser variants are related to their respective opposites as

- less explicit to more explicit, in terms of absolute number of grammatical units per discourse segment,
- less explicit to more explicit, in terms of absolute number of grammatical features per discourse segment,
- more nominal to more verbal in terms of type (quality) of phrases.

Turning to ‘grammatical density’ next, as a second operationalization of informational density and of grammatical metaphoricity, let us emphasize that here ‘density’ is a grammatical notion, not a semantic or discourse notion. Consider sample sentences such as (17) and (18), which illustrate the phenomenon in the area of translation, that is to say, across languages (from Doherty 1991):

(17) The suspicion that volcanic eruptions are the primary source of aerosols in the upper atmosphere has been around for many years.

(18) Seit vielen Jahren vermutet man schon, dass die Aerosole in den höheren Schichten der Atmosphäre vor allem aus Vulkanausbrüchen stammen  
(Lit: ‘Since many years suspects one already that the aerosols in the higher layers of the atmosphere above all from volcanic eruptions stem.’)

If we align translationally corresponding phrases and then consider the proportions between types of phrase, then we find both quantitative and qualitative effects, predicted by the following hypotheses (H3) and (H4):

(H3) The more informationally dense and the more grammatically metaphorical a stretch of text, the higher the proportion of ‘intermediate phrase types’ (groups, phrases, rather than words or clauses) per clause.

The resulting figures for examples (17) and (18) are:

((number of groups + phrases) : number of clauses) : number of clause complexes

English Original: ((10:2):0) = infinite<sup>7</sup>

German Translation: ((8:2):1) = 4

---

<sup>6</sup> Cross-linguistically, grammatical explicitness will be directly dependent on the morphological type of language, so that for any comparative work across languages, figures have to be standardized for type of language and for type of register (for some cautionary remarks along these lines cf. Steiner 2001).

<sup>7</sup> The value “indefinite” results from the fact that the English original has no clause complex and thus undergoes division by zero. This particular value is not very meaningful if derived from an individual example such as the one here. In general, however, it is obvious that the numerical value of the informationally denser and grammatically more metaphorical variants will be higher because of the higher percentage of groups and phrases per clause and per clause complex than in the case of the less metaphorical and less dense variants.

According to the model of grammatical metaphor, we should not simply count the relationships of ranks to each other, giving us a quantitative effect, but rather look at the differences between types of phrases and groups, dependent on the class of their lexical heads, a qualitative effect. This will give us at least one more type of hypothesis:

(H4) The more informationally dense and the more grammatically metaphorical a stretch of text, the higher the proportion of phrases with a nominal head relative to phrases with a verbal head per clause.

The figures for our sample sentences are:

(number of units with a nominal head) : (number of units with a verbal head)

English Original: 9:5

German Translation: 7:7

Similar results are obtained for the same test on our other examples (4) to (16) in this paper. They will again be illustrated in our text analysis in section 7 below under “experiential density”

We would also stipulate that

(H5) The more informationally dense and the more grammatically metaphorical a stretch of text, the higher the number of grammatical features per unit.

Hypothesis (H5) is also borne out by our examples (4) to (18). Note that, in this case, we are not considering the number of grammatical features per discourse segment, which is a measure of explicitness, but the number of features per grammatical unit.

We thus expect that one central type of grammatical metaphoricity, more precisely the one involving relocation in rank between semantics and grammar, seems to be straightforwardly mirrored in grammatical density as we are using the term here. And to the extent that informational density is related to this type of metaphoricity, the same would apply.<sup>8</sup>

One implication of the type of ‘density’ illustrated above is that for the encoding of a given discourse segment, or part of it, density would increase as we move ‘downwards’ on the rank scale, i.e. from clause complex into clause, phrase/group, word, and morpheme. We need to be aware of a simplification, though: ‘some semantic phenomenon’, once it is re-encoded on a different rank, does not simply remain ‘the same’. Instead, it gets expanded or reduced, according to the particular systemic options valid at the given rank. A process plus modality plus tense plus phase configuration (verbal group), if it is re-expressed metaphorically as a nominal group, is no longer anchored in temporal deixis, modality and phase, but now through expressions of identifiability, personal deixis etc. – whatever is expressed in the grounding part of the nominal group (cf. Langacker 1999, Davidse 1998, Taverniers in press). It should furthermore be clear that, as we have said before, any direct comparison across languages requires the use of comparative figures for languages (and registers) against which to standardize our findings.

---

<sup>8</sup> There is another type of grammatical metaphor, though, the type involving re-arrangement of semantics-to-grammar mapping not in terms of rank, but in terms of class on one rank, which has to be treated separately (see Steiner 2001, 2004 a for detailed coverage).

Let us finally turn to directness. By this we mean the sense in which (7) and (8) above are more direct than (6), (11) is more direct than any of (12) to (16), (18) is more direct than (17), and even (4.2) is more direct than (4.1), although probably to a lesser extent (for extensive usage of the concept, see e.g. Hawkins 1986: 53ff, Doherty 1996: 604ff, and Halliday and Matthiessen 2004, chapter 10; Halliday and Matthiessen 1999: 231ff under the labels of (non-)congruent realization). Directness is a graded property of the semantics-to-grammar mapping, for example between participant (semantic) roles and grammatical functions, between expressions of modality in different lexicogrammatical categories, or between logical relationships, such as causation, and their lexicogrammatical or cohesive expression. In order to be able to use the concept of ‘directness vs. indirectness’ in the relationship between semantics and lexicogrammar, that is to say *between levels within one language*, we need to further operationalize them. A direct encoding of a given semantic meaning is a 1:1 realization into its corresponding and thus transparent and motivated lexicogrammatical category. Often – though by no means necessarily – the less metaphorical version will be the more explicit lexically, morphologically and/or grammatically. All of this would be an operationalization of ‘directness’ from within one linguistic system. Because of this, it is also gradable and relative across languages. It has to be said, though, that directness can at this stage only be counted on hand-coded corpora.

The properties of ‘explicitness, grammatical density, and directness’ are assumed to be relatively shallow evidence of (aspects of) informational density and grammatical metaphor. Note that these are properties of semantics-to-grammar mappings, whereas the realizations of these mappings are syntactic, lexical, and finally phonological and phonetic categories. These properties have to be kept conceptually distinct from each other in their influence on the overall phenomenon of information distribution. In principle, they can be shown to vary independently, even though there is co-variation in naturalistic data. A further methodological issue is their applicability across languages: whereas ‘explicitness, density, and directness’ can be reasonably well defined as properties of semantics-to-grammar mapping within a language, it requires additional clarification to apply them cross-linguistically. Finally, information structure as the third realization of information distribution does not become empirically visible through explicitness and grammatical density directly, but through word order, phoricity, definiteness, and intonation, and therefore has to be read off directly from the parsed corpus (cf. Baumann et al. 2004).

## 7. An example: a comparison of two original texts in terms of explicitness, density and directness

In the following, we shall reproduce two texts which are both examples of relatively “autonomous” discourses, in the sense that they are written by a native language author in each case, are mainly addressed to a native speaking readership, and are also both specimens of well-established registers and genres in their respective communities. Both excerpts are the opening section from the monograph in question, written at approximately the same time and with similar fields, modes and tenors of discourse.

We shall apply the operationalizations of explicitness, directness and density, dispersed

- by metafunction
- by stratum (grammar vs. cohesion)

which are introduced in Figure 4 above.

Obviously, the comparison of two single texts, and additionally across languages, does not in itself provide evidence of anything, other than of methodological properties and questions. Furthermore, we are contrasting explicitness, density and directness as properties of encodings. We are not investigating either instantial explicitation, as defined in earlier parts of this paper, nor are we in this example using any alignment – which would be impossible because of the lack of a translation- or otherwise register relationship.

We shall reproduce the German sample text first, then give a word-order-preserving “translation” into English, and then reproduce the English sample text. The comparisons will then be between the German and the English sample texts, using the operationalizations from our 4, and some additional ones from our chapter 6 above. In an overall and integrated interpretative framework, we are using

- a) the data in the columns of Figure 5 below
- b) to generalize about textual properties “explicitness, directness, density”
- c) “explicitness, directness, density” in b) as indicators of (resulting from) information structure, informational density and grammatical metaphoricality of constructions in the texts.

Additionally, we shall raise a number of methodological issues in footnotes. One of these deserves to be mentioned more globally here: All categories of analysis, starting even from word-class tags, extending through phrasal categories, and certainly categories such as “Theme” and “Rheme”, are anchored in necessarily language-specific grammars. In our work, in which we necessarily code data across languages, we hope that categories have a better chance of being interlingual the more general they are, and the closer the languages are typologically. That is to say, a part of speech coding in terms of “nominal” and “verbal” is somewhat less problematic than one in terms of “phrasal verb” vs. “prepositional verb” across languages, and a grammatical analysis in terms of “process, participant role, circumstantial role” is less problematic than one in terms of any of their finer sub-classes. And finally, English and German are easier to compare with a cross-linguistic tag-set, than are, say, Chinese and German. However, the problem of interlinguality of analytic categories is one that we share with all attempts at comparing texts in different languages.

Sample text (1) Siegfried Jäger. 1993. *Kritische Diskursanalyse. Eine Einführung*. p. 9ff

## Einleitung

Es geht mir in dieser Einführung in die kritische Diskursanalyse darum, einen neuen integrierten **sozialwissenschaftlich-linguistischen** (emphasis in original) Ansatz für eine Diskurstheorie und - darauf aufbauend - eine Methode von Diskursanalyse zu entwickeln, der die traditionellen primär strukturalistisch orientierten Ansätze der Sprachwissenschaft, die heute noch bis in die Textlinguistik hinein dominieren, ebenso überwindet wie solche Ansätze, die im Rahmen qualitativer Sozialforschung entwickelt worden sind.

Da beide Disziplinen, ebenso wie die Diskursanalyse, auch den Anspruch stellen, Grundsätzliches über den Zusammenhang von Gesellschaft und Sprache/Kommunikation auszusagen, werde ich mich im ersten Teil dieses Textes exemplarisch kritisch mit einigen Grundannahmen der (Sozio-)Linguistik und der (qualitativen) Sozialforschung auseinandersetzen. Zunächst jedoch einige Vorbemerkungen zur **Linguistik** allgemein:

Aufgefallen ist mir insbesondere, dass linguistische Konzepte aller Art dazu neigen, die mit Hilfe sprachlicher Mittel transportierten **Inhalte** auf der Mikro- und Makroebene zu vernachlässigen bzw. programmatisch aus der Linguistik auszuschließen.

Mit den Inhalten wird aber zugleich alles Gesellschaftliche aus der Linguistik vertrieben. ...

Diese Position gilt es zu überwinden, auch wenn sie im Selbstverständnis der meisten Linguisten noch zutiefst verankert ist.

So schreibt etwa Bernhard Sowinski in seiner Einführung in die Textlinguistik, dass die Sprachwissenschaft sich in den letzten Jahren zwar verstärkt der Untersuchung von Texten zugewandt habe. Die Behandlung von Inhalten sei jedoch eine Sache anderer, teils benachbarter Disziplinen, etwa der Theologie, der Juristerei, der Literaturwissenschaft, der Geschichte oder auch der Soziologie etc., nicht aber eine Aufgabe der Linguistik.

English (preserving word-order as far as possible):

### Introduction:

It concerns me in this introduction to the critical discourse analysis about, a new integrated socialscientific-linguistic approach for a discourse theory and – on that building – a method of discourse analysis to develop, which the traditional primarily structuralistically oriented approaches of the linguistics, which today still into the textlinguistics dominate, supersedes as well as such approaches, which have been developed in the framework of qualitative social science research.

Because both disciplines, in the same way as the discourse analysis, also make the claim to proclaim fundamentals about the relationship between society and language/communication, will I myself in the first part of this text exemplificatorily critically with some basic assumptions of the (socio-) linguistics and the (qualitative) social research engage. At first however some preliminary remarks on the linguistics:

Impressed has me particularly, that linguistic concepts of all kinds tend towards the with help of linguistic means transported contents on the micro- and macro-levels to neglect respectively programmatically from the linguistics to exclude.

With the contents is however at the same time everything social from the linguistics expelled. ....

This position necessitates it to overcome, even if it in the self-concept of the most linguists still deeply anchored is. Thus writes for example Bernhard Sowinski in his introduction into the textlinguistics that the linguistics may have turned in the last years more forcefully to the examination of texts. The treatment of contents, however, be a task of other, partly neighbouring disciplines, for example of the theology, the law, the literary studies, the history or the sociology etc., not however the task of the linguistics.

## Sample Text 2: Norman Fairclough. Discourse and Social Change

Today individuals working in a variety of disciplines are coming to recognize the ways in which changes in language use are linked to wider social and cultural processes, and hence are coming to appreciate the importance of using language analysis as a method for studying social change. But there does not yet exist a method of language analysis which is both theoretically adequate and practically usable. My main objective in this book, therefore, is to develop an approach to language analysis which can contribute to filling this gap – an approach which will be particularly useful for investigating change in language, and will be usable in studies of social and cultural change.

To achieve this, it is necessary to draw together methods for analyzing language developed within linguistics and language studies, and social and political thought relevant to developing an adequate social theory of language. Among the former, I include work within various branches of linguistics (vocabulary, semantics, grammar), pragmatics, and above all the “discourse analysis” that has been developed recently mainly by linguists (the various senses of “discourse” and “discourse analysis” are discussed shortly); and I include among the latter the work of Antonio Gramsci, Louis Althusser, Michel Foucault, Jürgen Habermas and Anthony Giddens (see references). Such a synthesis is long overdue, but there are various factors which have militated against it being satisfactorily achieved so far. One is the isolation of language studies from other social sciences, and the domination of linguistics by formalistic and cognitive paradigms.

Properties M-functions	Explicitness	Directness	Density
Experiential	No of explicit functions per discourse unit: No of implicit functions  G: 230:6 E: 358:13 <sup>9</sup>	No of directly mapped experiential clause functions: Number of indirectly/ metaphorically mapped clause functions  G: 66: 10 E: 82: 13 <sup>10</sup>	No of functions: Unit <sup>11</sup>  <b>English:</b> CC (8) Cl ranking: 12 Cl embedded: 17 Clauses finite: 7 Clauses non-finite: 10 Clause elements: 95 Phrases : 33 Phrase elements 66 Groups : 136 Group elements: 197 Words : 248  <b>German :</b> CC : 7 Cl ranking : 12 Cl embedded : 6 Clauses finite : 3 Clauses non-finite 3 Clause elements : 76 Phrases: 25 Phrase elements:48 Groups : 125 Group elements : 206 Words : 233
Logical <sup>12</sup>	No explicit functions: No implicit functions <sup>13</sup>  G: 6 (2 conj + 4 preps): 0 <sup>14</sup> E: 5 (3 conj + 2) :0	No of directly mapped logical functions per unit: number of indirectly/metaphorically mapped functions  G: 6:0 E: 5:0 <sup>15</sup>	No. Functions: Discourse Unit G: 6:1 E: 5:1

<sup>9</sup> Counting only implicit functions resulting from non-finites. Cohesive ellipsis is counted further below

<sup>10</sup> Operationalization of “congruent/ direct” is difficult in many cases. Clause functions only are counted at this point, and only realized Functions. Clauses under clause combining (tactic) relations are also not counted as clause elements.

<sup>11</sup> Here we are counting CC, Cl (ranking vs. embedded), all phrases and groups plus their elements (recursively).

<sup>12</sup> Grammaticalized/ lexicalized expressions only; conjunctive cohesive ties are counted under “cohesion” below.

<sup>13</sup> Non-finite clauses without explicit linkage, certain prepositions

Mood	No of explicit Mood-markers per discourse unit: No of implicit Mood-markers E: 0:10 G: 0:9	No of directly/ congruently marked Mood options: No of indirectly marked Mood-options E: 10: G: 9:0	No of Mood-markers <sup>16</sup> : Discourse Segment E: 10 G: 9 <sup>17</sup>
Modality	No of explicit Modality-markers: No of implicit Modality-markers E: 1:0 G: 0:1	No of directly/ congruently marked Modality options: No of indirectly marked Modality-options <sup>18</sup> E: 0:1 G: 1:0	No of Modality-markers: Discourse Segment E: 1:0 G: 1:0
Theme	No of auto-semantic Themes: No of syn-semantic (phoric) Themes <sup>19</sup> G: 7:0 E: 16:0	Not applicable	No Themes: Clause Functions G: 7: 76 E: 16: 95
Information	Not applicable	Not applicable	No NEW elements <sup>20</sup> : Discourse Segment Unavailable at this stage No of marked NEW elements: Discourse segment G: ca. 8 E: 4-5
Cohesion: Reference	No. implicit referents: No. of explicit referents G: 7:99 E: 16: 97 <sup>21</sup> No. phoric phrases: No. autosemantic phrases <sup>22</sup> : G: 86 E: 86 G: 13 phoric (+28 definite articles) E. 11 phoric (+ 8 definite articles)	Not applicable	No of newly introduced referents <sup>23</sup> per discourse segment E: 55 (per 29 Figures (verbs)) G: 64 (per 20 Figures (verbs))

<sup>14</sup> Projection is not counted here, but conjunctions and prepositions encoding logical meaning are (i.e. excluding prepositions with only case assigning function, or such with only locational and temporal meaning). Two instances of “und” in the German text are not counted, nor are 14 in English – although the discrepancy there is interesting in itself, yet the functions of “und/ and” are extremely varied.

<sup>15</sup> Difficult notion, because of the wide usage of spatio-temporal prepositions for logical relations

<sup>16</sup> Mood options/ markers are not counted for embedded clauses (i.e. those without choice). Note also that there are ranking (non-finite) clauses without mood selection. Finally, lexicogrammatical encoding of mood counts as “implicit”, vs. the “explicit” version by “speech-act verbs”.

<sup>17</sup> In 2 German projecting clauses, I have counted the Mood-option preselected by the verbs (*auffallen, schreiben*)

<sup>18</sup> I am not sure about the precise operationalizations here, and about the interactions with “appraisal”. We could use Halliday’s features here (Halliday and Matthiessen 2004: 613ff).

<sup>19</sup> We only count realized Themes in finite main clauses; the numerical difference is due to the frequency of multiple Themes in English.

<sup>20</sup> We haven’t counted NEW elements for lack of a spoken version. As marked NEW, we counted any constituent which was so marked through word-order or lexical means.

Cohesion: Ellipsis	Number of cohesive ellipses G: 1 E: 3	Not applicable	Not applicable
Cohesion: Substitution	Number of cohesive substitution G: hardly applicable (0) <sup>24</sup> E: 0	Not applicable	Not applicable
Cohesion: Lexical	Ratio lexical/ grammatical words G: 106 lexical/110 grammatical E: 130/ 117 <sup>25</sup>	Not applicable (lexical metaphor is not yet counted here)	Lexical words per clause: G: 106:18 E: 130:29  Type token ratio?
Cohesion: Conjunction (cohesive relations only) <sup>26</sup>	No explicit relations: No of Figures G: 13:20 E: 9:29	No cohesively realized relations: No of non-realized RST realizations <sup>27</sup>	No of relations <sup>28</sup> G: 13 E: 9

**Figure 5:** Properties of lexicogrammatical encoding across metafunctions for two register-parallel texts.

<sup>21</sup> The number of referents is much lower than the number of phrases, because by no means all phrases refer (they may predicate, for example.). The number of implicit referents here is small, assuming that dropped Subjects in cases of non-finites have already been counted under “experiential”

<sup>22</sup> Including proper names, i.e. individuals; noun-noun compounds are counted per entity introduced

<sup>23</sup> Here we are counting referents, NOT concepts or lexical items. So, a referent is only counted if newly introduced into the discourse. Nominalizations are also not counted twice, unless they introduce different referents. The discourse unit here is the “Figure” (cf. Halliday and Matthiessen. 1999).

<sup>24</sup> German has hardly any “substitution” in the strict sense of Halliday/ Hasan 1976.

<sup>25</sup> Interesting, because contrary to first expectations deriving from the morphologically analytical nature of English vs. German – although the more extensive compounding of German points in the opposite direction (cf. Steiner 2001 for remarks on these issues).

<sup>26</sup> Grammaticalized expressions are treated under “logical” above. The proportionalities between grammaticalized and cohesive ways of expression of logical relations are an interesting matter in themselves. This may also be the place to comment on the counting of German “pronominal adverbs” (*darauf, darum, etc.*): when they are part of the prepositional or otherwise phrasal verb, they are syntactically obligatory and thus are not counted, just like their English counterparts. Where they have a cohesive function, they are counted under reference (due to the demonstrative character of their first part), and additionally as conjunctive relation, where there is one involved (*deswegen, dafuer* etc. where they are not part of the verb).

<sup>27</sup> Difficult to operationalize

<sup>28</sup> Ultimately counted as a proportion per Figure. The mere numbers hide the fact that in the English version, 5 out of 9 are of the type simple additive or adversative types, whereas this is true of hardly any one of the German. The adversative types in German are not simple, but more marked, specialized.

With a reminder that our analysis undertaken here serves as an illustration only, let us briefly go through our findings, adding a few comments where appropriate:

**Experiential:** counting implicitness only in a restricted grammatical sense, we see that the German text has a proportion of 6: 230 (roughly 0,026) , while the English has one of 13: 358 (roughly 0,039). Furthermore, we see that both texts exhibit approximately the same degree of “directness” as far as their clause functions are concerned. The German text, however, comes out as more dense than the English one, expressed by the proportions of clause elements per clause (G: 4,2 vs. E: 3,29) and the proportion of group elements per group (G: 1,64 vs. E 1,44). It also comes out as more dense than the English one in terms of proportionalities between ranks (levels of projection) suggested in section 6. The ratio of

clause complexes: clauses: groups/ phrases: words

is for the English text : 8 : 29 : 169 : 248

for the German text: 7: 18: 150: 233

The German text has a substantially higher proportion of phrases/ groups per clause, and of words per clause than the English one. The percentages of words per groups/phrases, however, are not so different – and all this in spite of the fact that English, being the more morphologically analytic language, has morphological properties which should bias it generally towards a higher proportion of words per higher unit as compared to German.

On the other hand, the English has many more embedded clauses (both finite and non-finite clauses functioning as group elements), which results in the overall higher number of grammatical functions and clauses in absolute terms. In other words, the information in the English text is much more expressed in clauses than that of the German, which seems to condense the information into clause and group elements.

But let us emphasize at this point again that the figures obtained here are figures for two registerially parallel texts, not for a translation pair. We are thus measuring density (and explicitness and directness) of constructions, not the relative explicitness or explicitaion between two related variants (translations). Figures such as those obtained here will, however, figure as part of the reference profile for English and German for the given register and will then be statistically factored out of any comparisons of translation pairs in order to obtain possible effects of the process of translation and the translation relationship as such.

**Logical:** The German text is slightly more explicit in its encoding of logical functions, it is also more direct and more dense. If we take these figures together with those under cohesive conjunction below, we can see a picture of a logically much more explicit and denser German text.

**Mood:** There are hardly any differences here in any of the variables explicitness, directness and density. This is likely to be a consequence of the register of our texts here.

**Modality:** Again, the numbers are much too small for any conclusions. However, the few instances that we find differentiate the two versions along the dimensions of “explicitness” and “directness”.

**Theme:** For both texts, their Themes are all explicit, in the sense of auto-semantic, which may be partly the consequence of the fact that both passages come from the opening of the books. The English version is more explicit in having more multiple Themes, but this seems typologically motivated. “Directness” does not apply, but “density” is again much higher in English because of its obligatory “SV” properties and the associated frequency of multiple Themes.

**Information:** In this case, the notions of “explicitness” and “directness” may not be applicable, because of the nature of the textual metafunction: either some constituent is encoded as NEW or it isn’t, and if it is, there is no immediate sense in which it could be encoded “indirectly”. One might think of an operationalization of “explicitness” in terms of lexical vs. word-order and intonational encoding. There may be a valid notion of “density”, operationalized as “number of NEW elements per discourse unit”. We have not counted this variable here, simply because of time constraints involved in having native speakers produce plausible readings of the texts. What we have tentatively done, though, is count the number of (lexico-grammatically) marked NEW units, which gives a denser picture for the German text.

**Reference (Cohesion):** Here two operationalizations come to mind: One would be the proportion of implicit to explicit referents. This is for German 7:99 (approx. 0,07), for English 16: 97 (0,16). The other is the proportion between phoric and autosemantic referents (G: 13: 86; E: 11:86). Along this dimension, the two texts appear relatively similar at first sight. However, we notice a far higher percentage of NPs with definite articles in German (28:86) than in English (8:86). Of course this has to do with the systemic differences between the two languages as far as the expression of “genericity” with abstract NPs, and of countability and boundedness, are concerned. Furthermore, German often uses NP-attributes where English prefers non-finite ones.<sup>29</sup> There may be additional effects due to a higher percentage of phrases in German being treated as “recoverable” from co(n)-text, but this remains to be investigated. As far as “directness” is concerned, we do not see an immediate sense in which it could be used, other than in the sense of “directness of encoding of experiential meaning”, which is covered above. For “density”, though, we are counting the number of newly introduced referents per discourse segment, or rather, per semantic “Figure”: (E: 55:29); (G: 64:20). These proportions would classify the German text as substantially more “dense” than the English one.

**Ellipsis (Cohesion):** Under this category, a high number of cohesive ellipsis would contribute to low “explicitness”. Absolute numbers are, again, low here, but the English text is less explicit on this dimension, than the German one. “Directness” does not seem to play a role here, nor “density” – although one could count the number of “ellipsis” relative to clause or phrase functions.

**Substitution (Cohesion):** once more, the categories of “directness” and “density” do not seem to be applicable. A high number of “substitution”, especially through general nouns and verbs, would contribute negatively to “explicitness”, but is not given in our present text. One might suspect that general high frequency nouns in the English text (individuals, ways, processes, factors) play a role here, but this will be taken up again under “lexical cohesion”.

**Lexical Cohesion:** In an initial step, we have operationalized “explicitness” here as the proportionality of lexical to grammatical words. Interestingly, this relationship is counter to first expectations (G: 106:110 vs. E: 130:117), but is perhaps less surprising if we remember

---

<sup>29</sup> Further systemic differences in English and German NPs relevant to reference can be found in Doherty (2004a,b) and in Königs (2000: 479ff).

the role of compounding (16 in our short German text) in the two writing systems. “Directness” is not applicable, or rather, would require an analysis of lexical metaphor, which we haven’t done here. Finally, the number of fully lexical words per clause shows German (106: 18) denser than English (130: 29) (cf. Halliday and Martin 1993:69ff). The type-token-ratio has not been counted here, because it would be meaningless over such a short stretch of text. However, it can be taken to be a measure of overall lexical density.

**Conjunction (Cohesion):** The explicitness of the German text is much higher than that of the English (G 13:20) as against English (E 9:29). This is even strengthened by the fact that the German is also higher in its explicitness in the expression of grammaticalized logical meanings (cf. above). In terms of “directness” we do not find any non-realized rhetorical relations. The Figures for “density” look similar, but the semantic subtypes and the semantic spread of the relations are quite different.

In summary, then, our German text would be

- experientially more explicit and denser
- logically more explicit
- thematically less explicit and less dense
- informationally denser
- referentially more explicit and denser
- lexically denser
- conjunctively more explicit

than its English counterpart. Otherwise, they would be classified as similar. However, and this is of paramount importance, our summary comparison creates the wrong impression, i.e. that there are cross-linguistic properties of explicitness, density and directness which can be compared in their absolute numbers and proportionalities directly. This is not the case: if English and German texts for the register which they exemplify here were different generally – numerically and proportionally – in the way that our two texts are, then the two texts would be of equal (un-)markedness in their registers, and in their cultural communities. Any figures obtained in direct comparisons must therefore be interpreted against the statistical background of their registers and their languages, which is what our corpus-design is intended to allow.

## 8. Towards an annotation scheme

In summary, and repeating something we have already said in section 6, we have suggested the following relatively theory-neutral operationalizations, or categories of analysis, in terms of which we shall attempt to investigate “explicitness”:

- The proportionality form : content words
- Average number of words per clause
- Biber’s informational : involved production
- The proportions between the following lexical categorical types: conjunction: preposition; verb: noun; adverb. Adjective; finite : non-finite<sup>30</sup>
- Proportionalities between ranks (levels of projection): clause complexes: clauses: groups/ phrases: words)
- Proportionalities between directly verbally governed to directly nominally governed phrases
- Degree of specificity of lexical items
- Number of grammatical units (clauses, phrases/ groups, words) per discourse segment,<sup>31</sup> where the number correlates positively with explicitness;
- Explicitness of grammatical categories, such as person, number, gender, but also diathesis, (including voice), relativization, complementation, etc.;
- Explicitness of cohesive relations, especially reference and conjunction.
- To these operationalizations, we must add those used in Figure 4 above, although from then on we cease being theoretically neutral. Figure 4 is also at a different level of abstraction, already constituting an interpretation of some of the more shallow data formulated in the earlier part of our list. And finally, the operationalizations in Figure 4 are for “explicitness”, but also for “density” and “directness”, which constitute properties of their own, even if they interact with “explicitness”.

In order for the operationalizations above to work, our corpora should be tagged, or otherwise annotated, with the following categories:

- number of words
- form words
- content words
- lemmatized lexical items
- morphological items and their grammatical features
- number of constituents at some level
- realized constituents at some level
- non-realized constituents at some level (grammatical)
- non-realized constituents at some level (cohesive)

---

<sup>30</sup> Note that these oppositions are ultimately scalar, rather than binary

<sup>31</sup> Note: this is not the same as the number of intermediate phrase types *per clause*, which we shall use below in our operationalizations of density and metaphoricity in (H3) and (H4).

- Biber's textual factors and dimensions (1988:104ff), in particular dimension 1 (informational : involved production) and dimension 3 (explicit vs. situation-dependent). Features for German need to be developed where necessary.
  - Part-of-Speech-Tagging (POS)
  - A phrase-structure-analysis, preferably involving several levels of projection (ranks)
  - Tagging of heads for lexical categories
  - Topological analysis into something like *Vorfeld-Mittelfeld-Nachfeld*
  - Degree of specificity of lexical items (parsed with word-net)
- 
- Functions of grammatical units
  - Mood-features of clauses: finiteness, clause-moods
  - Modality realizations: (semi-)auxiliaries, other lexical categories
  - Realizations for marked NEW: extrapositions, clefts, focussing adverbs and particles
  - Indexing for (newly introduced) discourse referents
  - Substitution (clausal, nominal, verbal), depending on the language
  - Cohesive conjuncts

What is of paramount importance to be observed is the fact that we cannot only retrieve information which is directly covered by our tagging and further annotation, but additionally information which can be queried in our (tagged, enriched, annotated) corpus, so that the information surplus in our corpus, as compared to a raw corpus, derives from enriched data plus theoretically guided querying. A first sorting of the information in our annotations depending on the source of that information is given below:<sup>32</sup>

### Corpus processing and querying steps

Raw statistics (no annotation needed)

- number of words
  - Type-token-ratio
  - Average sentence length
- 
- Part-of-speech-tagging
  - TnT-Tagger

Perl-script "Lexical Density"

- form words and content words (not segmented into clauses) (builds on PoS-Tagging)

MPro<sup>33</sup> (incl. phrase-chunking)

- lemmatization and morphological analysis
- number of constituents at some level
- realized constituents at some level
- A phrase-structure-analysis, preferably involving several levels of projection (ranks)

---

<sup>32</sup> Thanks to Stella Neumann for suggesting this first classification, which may still be elaborated, extended and modified

<sup>33</sup> MPro has been developed by the IAI in Saarbruecken, the help of which is hereby gratefully acknowledged. For a general description cf. Maas 2005

- Mood-features of clauses: finiteness

#### Wordnet

- Degree of specificity of lexical items (parsed with word-net?)

#### Topology parser<sup>34</sup>

- Topology (*Vorfeld-Mittelfeld-Nachfeld*)

#### Manual analysis

- non-realized constituents at some level (grammatical) (builds on chunking)
- non-realized constituents at some level (cohesive) (builds on chunking)
- Tagging of heads for lexical categories (builds on chunking)
- Indexing for (newly introduced) discourse referents (builds on chunking)
- Mood-features of clauses: clause-moods
- Substitution (clausal, nominal, verbal)
- Functions of grammatical units (builds on chunking)

#### Query building on annotation

- Biber's textual factors and dimensions (1988:104ff), in particular dimension 1 (informational : involved production) and dimension 3 (explicit vs. situation-dependent). Features for German need to be developed where necessary.
- Realizations for marked NEW: extrapositions, clefts, focussing adverbs and particles
- Modality realizations: (semi-)auxiliaries, markers of other lexical categories
- Cohesive conjuncts (difficult, because open-set)

In an overall and integrated interpretative framework, we are using

- a) the data in the columns of Figure 4 (and 5), derived from corpus tags, annotations and queries
- b) to generalize about textual properties "explicitness, directness, density" (and others)
- c) "explicitness, directness, density" in b) as indicators of (resulting from) information structure, informational density and grammatical metaphoricality of constructions in the texts.

All of a) to c) are dependent variables, however on different levels of abstraction from the data. These dependent variables, or their behaviour, then need to be explained in terms of a few independent variables as outlined below.

---

<sup>34</sup> cf. Braun 1999

## 9. Independent and dependent variables

Our overall research architecture is an attempt to link the following independent and dependent variables:

Independent Variables:

- (1) **Language:** here we are comparing cross-register corpora in English with cross-register corpora in German (a comparison of the reference corpora as in Figure 1)
- (2) **Register:** here we are comparing texts of one register A with texts of one register B in English and German, respectively (language-internal comparison of registers)
- (3) **Translation:** here we are comparing translations of a constant register A of English and German texts with originals of the same register in the respective language (language-internal comparison of translations and originals)
- (4) **Translation**, in interaction with language (and possibly register): here we are comparing source texts in English with their translations into German (direct comparison of source texts and target texts)

In (1), (2) and (3) above, the investigation is one of properties of constructions, i.e. “explicitness/ implicitness” in the sense defined in this paper. In (4) only, where the investigation will be one of translations and their originals, the object will be aligned and ultimately also referentially indexed texts, investigating “explicitations” in the sense defined here, alongside with “explicitness” of constructions.

Dependent variables will be those operationalized in the previous sections, ultimately interpreted within a framework such as in Figure 4 (or some other linguistically-based model). By then comparing the differences found in the comparisons under (1) to (4) above, we hope to pair individual independent and dependent variables.

## References

- Aijmer, Karin and Hasselgard, Hilde. eds. 2004. *Translation and Corpora. Selected Papers from the Goteborg-Oslo Symposium*. Goteborg: Acta Universitatis Gothoburgensis. 163-188
- Asher, Nicholas. 1993. *Reference to abstract objects in discourse*. Dordrecht: Kluwer
- 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. , Brinckmann, C., Hansen-Schirra, S., Kruiff, G.-J., Kruiff-Korbayova, I. Neumann, S., Steiner, E., Teich, E.& Uszkoreit, Hans 2004. "The MULI Project: Annotation and Analysis of Information Structure in German and English", in: *Proceedings of LREC 2004*. 26-28 May. Lisbon, Portugal
- Berman, Antoine. 1985/ 2000. "La Traduction comme épreuve de l'étranger » , translated by Lawrence Venuti as « Translation and the trials of the foreign » in : Venuti. ed. 2000. 284-297.
- 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*. Norwood, N.J.: Ablex
- Biber, Douglas. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press
- Biber, Douglas. 1995. *Dimensions of register variation: a cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Blum-Kulka, Shoshana. 1986. "Shifts in Cohesion and Coherence in Translation". In: House, Juliane et al. eds. 1986. *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies* Tuebingen: Narr. pp. 17-35
- Braun, Christian 1999: *Flaches und robustes Parsen deutscher Satzgefüge*. Diplomarbeit, Universität des Saarlandes, Saarbrücken.
- Burton-Roberts, Noel 2005 "Review article: Robyn Carston on semantics, pragmatics and 'encoding'". In: *Journal of Linguistics* 41 (2005), 389-407
- Caffarel, Alice, Martin, James R. & Matthiessen, Christian M.I.M. (eds.). 2004. *Functional Typology*. Amsterdam: Benjamins.
- Carston, Robyn 2002 *Thoughts and utterances: the pragmatics of explicit communication*. Oxford: Blackwell
- Davidse, Kristin. 1998 "The Dative as participant role versus the Indirect Object. On the need to distinguish two layers of organization" In: Van Langendonck, W. & Van Belle, W. (eds.) *The Dative. Vol. 2: Theoretical and Contrastive Studies*. (Case and Grammatical Relations across Languages, 3.) Amsterdam: Benjamins, 143–184.
- Dixon R.M.W. 1991. *A new approach to English grammar, on semantic principles* Oxford: Clarendon Press
- Doherty, Monika. 1991. "Informationelle Holzwege. Ein Problem der Übersetzungswissenschaft. " *Zeitschrift für Literaturwissenschaft und Linguistik*, 21(84).
- Doherty, Monika. (ed.) 1996. *Information structure as a problem of translation*. Special Issue of *Linguistics*. 34.
- Doherty, Monika. ed. 1999. *Sprachspezifische Aspekte der Informationsverteilung*. Berlin: Akademie-Verlag.
- 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
- Duff, Alan 1981 *The Third Language: Recurrent Problems of Translation into English*. Oxford: Pergamon Press
- Englund Dimitrova, Brigitta. 2005. *Expertise and Explicitation in the Translation Process*. Amsterdam: John Benjamins
- Fabricius-Hansen. Cathrine. 1996. "Informational density: a problem for translation theory". Doherty (ed.) 1996: 521–565.
- Fabricius-Hansen. C. 1999. Information packaging and translation: Aspects of translational sentence splitting (German – English/ Norwegian). In Doherty (Hrsg.) 1999: 175 - 214
- 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. 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
- Hasselgard, Hilde. 2004. "Spatial linking in English and Norwegian", in: Aijmer and Hasselgard. Eds. 2004
- 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
- House, Juliane. 1977/ 1997. *A model for translation quality assessment. A Model revisited*. Tuebingen: Gunter Narr (revised model of the earlier 1977 version)
- House, Juliane. 2002. "Maintenance and convergence in translation – some methods for corpus-based investigations" in Hasselgard et al. (eds.) 2002. 199-212.
- Johansson, Stig 2004 "Why change the subject? On changes in subject selection in translation from English into Norwegian." In: *Target*16.1 29-52
- Johansson, Stig 2005 "Sentence openings in translations from English into Norwegian", in: Norsk Lingvistik Tidsskrift 23. 2005 3-35
- Kamp, Hans and Reyle, Uwe. 1993. *From Discourse to Logic*. Dordrecht: Kluwer
- Kenny, Dorothy 2001. *Lexis and Creativity in Translation. A corpus-based study*. Manchester: St. Jerome.
- Königs K., 2000 *Übersetzen Englisch Deutsch* München und Wien: Oldenbourg,
- Langacker, Ronald. 1999. *Grammar and Conceptualization*. Berlin/New York: Mouton de Gruyter
- Laviosa-Braithwaite Sarah (ed.), 1998. *Meta. Special Issue: The corpus-based approach*. (43)4.

Levy, Jirzy. 1969. *Die literarische Übersetzung. Theorie einer Kunstgattung*. Frankfurt.

Linke, Angelika and Nussbaumer, Markus. 2000. Konzepte des Impliziten: Präsuppositionen und Implikaturen. In: Brinker, Klaus/ Antos, Gerd/ Heinemann Wolfgang und Sager, Sven F. eds. 2000. *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*. Halbband 1. Berlin, New York: Walter de Gruyter. pp. 435-448

Maas, Dieter 2005. *Multilingualism in MPro*. <http://www.iai.uni-sb.de/docs/mmpro.pdf>

Matthiessen, C.M.I.M. 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

Neumann, Stella & Silvia Hansen-Schirra, 2005, The CroCo Project. Cross-linguistic corpora for the investigation of explicitation in translations. In: *Proceedings from the Corpus Linguistics Conference Series*. Vol. 1 no. 1, ISSN 1747-9398. <http://www.corpus.bham.ac.uk/PCLC>

Neumann, Stella 2005 The CroCo Corpus design. Deliverable No. 1.1. <http://fr46.uni-saarland.de/~croco>

Olohan, Maeve 2001 „Spelling out the optionals in translation: a corpus study“, in: UCREL Technical Paper No. 13. Special Issue: Proceedings of the Corpus Linguistics 2001 Conference. Lancaster. UK. 29 March – 2 April 2001. pp. 423 – 432

Polenz P., 1988. *Deutsche Satzsemantik. Grundbegriffe des Zwischen-den-Zeilen-Lesens* Berlin: Mouton de Gruyter

Ramm, Wiebke and Fabricius-Hansen, Cathrine 2005. "Coordination and discourse structural salience from a cross linguistic perspective" SPRIK Reports Nr. 30. University of Oslo <http://www.hf.uio.no/forskningsprosjekter/sprik/publikasjoner/sprikreports.html>. pp.1-11

Sager, Juan C. 1994 *Language Engineering and Translation: Consequences of automaton*. Benjamins Translation Library. Vol.1 Amsterdam: John Benjamins

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, 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

- Steiner, Erich. 2005a. Some properties of texts in terms of 'information distribution' across languages. In: *Languages in Contrast 5.1. (2004-2005)* pp.49-72
- Steiner E. 2005b. "Some properties of lexicogrammatical encoding and their implications for situations of language contact and multilinguality." In: Franceschini, Rita. Hrsg. 2005. *In einer anderen Sprache. Sondernummer der Zeitschrift für Literaturwissenschaft und Linguistik* Jahrgang 35, Heft 139. pp. 54-75 Stuttgart: Metzler Verlag
- Steiner, E. and Vandenberg, A. eds. in press. Special issue of *Language Sciences* on *Functional approaches to discourse: perspectives, interactions, and recent developments*. in press
- Taverniers, Miriam. in press. "Subjecthood and the notion of instantiation" In: Steiner, E. and Vandenberg, A. eds. in press. Special issue of *Language Sciences* on *Functional approaches to discourse: perspectives, interactions, and recent developments*
- Teich, Elke. , Hansen, Silvia & Fankhauser, Peter.. 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)
- Toury, Gideon. 1995. *Descriptive Translation Studies and beyond*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Venuti, Lawrence. ed. 2000. *The Translation Studies Reader*. London and New York: Routledge
- Vinay, Jean-Paul and Jean Darbelnet. 1958. *Stylistique comparée du français et de l'anglais*. Paris: Les Éditions Didier. Page references are to the English translation 1995 in Venuti. ed. 2000. 284-297.