Tense, Modality and Polarity: The Finite Verbal Group
in English and German Newsgroup Texts

Anke Schulz

Abstract
This paper describes work in progress on a corpus-based study, comparing seemingly similar registers in two languages: English and German newsgroup texts, collected in the Bremen Translation Corpus. Systemic Functional Grammar (SFG, Halliday 1994 [1985]) provides a theoretical framework for categorizing empirical findings. I will focus on three systems of the finite verbal group, i.e. tense, modality and polarity, to describe these registers on the textual and the interpersonal metafunction levels.

The use of tense, apart from its logical function, can also reflect the textual metafunction, since different tenses are preferred in spoken and written discourse (see Biber et al. 1999 or Duden 2005). Do newsgroup text authors favour one tense over the other? And is tense choice the same in both languages? On the interpersonal metafunction level, I analyse the form and function of modal auxiliaries, and look at how modal auxiliaries combine with process types. I then investigate the position of the syntactic negation markers in English and German clauses.

My aim is to provide a thorough description of the realization of finite verbal groups in English and German newsgroup texts as a preliminary step towards research on variation in parallel translations using the Bremen Translation Corpus.

1. Introduction and theoretical background
1.1 Introduction
My research project has several purposes. The first and most central is to describe original texts of naturally occurring English and German language so that later on I can investigate variations in parallel translations on the basis of the similarities and differences between English and German language systems. An interesting by-product of my research is the SFG description of a relatively new register, asynchronous computer-mediated communication (CMC) or, more precisely, newsgroup discourse. These newsgroup texts are collected in the Bremen Translation Corpus (BTC). The third aim is to add to the SFG description of German, to see whether and in how far English categories, e.g. modality and process types, can be transferred to an analysis of German.

Odense Working Papers in Language and Communication vol. 29
(ISSN 0906-7612, ISBN: 978-87-90923-47-1)
The theoretical background provided by SFG is explained in the next section. Section 2 will give a description of the design and purpose of my corpus, the Bremen Translation Corpus (BTC), which consists of newsgroup texts in English and in German and five parallel translations of each text into the other language. In that section I also give a brief outline of how I processed the data.

In section 3, some first exploratory results will be presented. Part 3.1 focuses on the frequency of different tense forms in the original newsgroup texts. I will argue that, apart from the logical function of tense forms, the frequency of tense forms reflects to a certain extent whether we are dealing with written or spoken texts, making tense a feature of the textual metafunction as well as of the interpersonal metafunction. As a ground for transfer comparison of the distribution of tense forms in different registers, I refer to the *Longman Grammar of Spoken and Written English* (*LGSWE*) (Biber et al. 1999). The *LGSWE* uses a 40 million word corpus, the *LSWE corpus*, to compare tense form distribution in different written and spoken registers. With an investigation of the tense form distribution in my corpus of newsgroup texts I hope to show the affinity of these texts to either written or spoken discourse.

Section 3.2 then is dealing with modality, more specifically with the function of modal auxiliaries in my English and German texts, comparing the results from the two language corpora. Do English and German authors express modality by the same means, i.e. modal auxiliaries, and to the same extent? I will also look at the kind of process types (following the Cardiff Grammar categorisation) that occur in combination with a modal auxiliary.

In section 3.3 I am concerned with the position of syntactic negation markers in the English and German clause complexes. Where in a clause are negation markers typically placed in my newsgroup text corpus?

A conclusion is provided in section 4, with an evaluation of the work presented here as well as an outlook into future work following in section 5.

### 1.2 Theoretical background

Systemic Functional Grammar (Halliday 1994 [1985]) provides the theoretical framework for this study. The value of SFG as the underlying theory for the description and comparison of different languages has been shown, e.g. in Caffarel et al. (2004) on a number of languages ranging from French and German to Chinese, Vietnamese and Telugu, or in Teich (2001),
comparing English and German. Following Caffarel et al. (2004: 15), I will take the practical heuristic approach in applying the method of transfer comparison:

However, the type of approach where no assumptions are made based on other languages and where the description of the lexicogrammatical system is built up from observations of discursive instances takes a considerable amount of time, so as a practical heuristic, it may be helpful to model the description of one language on the description of another – this is the method of transfer comparison [...]. (Caffarel et al. 2004: 15)

I will describe the realization of finite verbal groups in the corpus of English and German newsgroup texts, taking descriptions of English as the starting point. The focus will be on the textual and interpersonal metafunctions of SFG.

“Traditional” grammars, e.g. Biber et al. (1999) or Duden (2005), see tense as being used differently in spoken and written discourse. In addition to their function on the ideational and interpersonal levels, we can therefore consider finite verbal groups as a component of the textual metafunction in SFG. Among other things, the interpersonal metafunction is reflected in the use of modal operators and negation markers. With these the authors of the newsgroup texts reveal their stance towards the problems they discuss. Due to limitations of space the ideational metafunction will not be considered in this paper.

In their paper ‘Metafunctional profile of the grammar of German’, Steiner & Teich (2004) describe modality and polarity in German, though they only dedicate one page to it. Still they postulate a difference in the realization of modality between English and German:

[...] meanings of phase and modality are typically realized in prepositional phrases and adverbial groups in German, rather than within the verbal group. We shall not have the space here to go into this very interesting area, as we will focus on the clause, rather than on the verbal group. (Steiner & Teich 2004: 144)

The present research, however, will focus only on the verbal group. It can be valuable to test how far modality is realized within the verbal group in German, and how the realization of modality, polarity and tense differs from English.

The present research will also be a small step towards a more comprehensive description of German which may finally serve to evaluate the validity of SFG categories for describing this language and language in general, as suggested by Caffarel et al. (2004):
Whether or not such descriptive categories can be applied in the description of several languages (or even all languages) is an empirical question, not a theoretical one; and it is a question to be decided only after the categories have been independently motivated in comprehensive descriptions of every language. (Caffarel et al. 2004: 11)

2. Corpus design and processing

2.1 Corpus design

The research uses the Bremen Translation Corpus (BTC) to investigate finite verbal groups, focusing on the use of tense, modal auxiliaries and syntactic negation in naturally occurring language. The corpus was collected at Bremen University, Germany, in 2004-5. It consists entirely of asynchronous computer-mediated communication (CMC), i.e. newsgroup texts, where “users do not have to be online at the same time to communicate” (Beißwenger & Storrer to appear: 2). Synchronous CMC, e.g. chats or Instant Messaging, is not included in the BTC at the moment. The part consisting of original newsgroup texts from the internet has about 16,000 words. The choice of topics follows a vertical design (Zitzen 2004), i.e. two topics only are covered: eating disorders and relationship problems. Half of these texts are taken from newsgroups in English, the other half from newsgroups in German. Each one of these texts, approximately 250 words in length, has then been translated by five different non-professional translators who were native speakers of the target language, German or English, respectively. The translated part of the corpus amounts to approximately 80,000 words, meaning that the entire corpus consists of approximately 96,000 words. Figure 1 below shows the structure of the BTC.
The need for a corpus of parallel translations has been expressed before, e.g. by Mauranen (2002: 166):

One problem with translation corpora that has been pointed out by Malmkjær (1998) is that they only provide one translation solution for every SL instance, which conceals the variation in translations that would ensue if we had available versions of the same source text by different translators.
With a corpus of just one translation of one original text, as for example in the English-Norwegian Parallel Corpus (see Johansson & Hofland 1994), it is not possible to answer questions such as: how much variation occurs when different people translate the same text? Where in the text/sentence/clause does the variation occur? Can variation in parallel translations be explained by differences in the language systems, or is it due to differences in language use? The BTC was built to make possible linguistic research of all these aspects (and more). The following examples (1, 2) show two sentences from the corpus and the five corresponding parallel translations:

1 Ich denk-e mir halt,
   I think-1SG.PRS.TR me-PRON.1SG.DAT just-ADV,
  ‘I just think’,

   jed-e bereits vergeben-e Frau
  every-DET.3SG.F already spoken_for-ADJ.3SG.F.NOM woman-NN.NOM
  ‘every woman who is already spoken for’

   müsste bei mein-en Annäherungsversuch-en
  should-AUX.SBJV at my-DET.DAT.PL advances-NN.DAT.PL
  ‘should immediately put a stop’

   sofort Einhalt gebieten.
  immediately put_a_stop_to-VB
  ‘to my advances.’

   A I just think that this woman has to reject my advances
   B I just think that every unavailable woman must immediately stop my approaches
   C I had thought that every already-taken women must immediately back off at my attempts to get closer
   D I think that every woman who is already attached should react to advances immediately
   E I think every woman who’s spoken for has always left me hanging when I tried something.

2 Yeah, I'd love to give up the negative behaviours someday.

   A Ja, ich würde die negativ-en
  Yes, I would-AUX.SBJV the-DET.ACC.PL negative-ADJ.ACC.PL

   Verhaltensweise-n lieben-d gerne eines Tages ablegen.
  behaviours-NN.ACC.PL loving-ADJ like_to-ADV one day throw_off-VB.

  ‘Yes, I would really love to throw off the negative behaviours someday.’
‘Yes. I would really love to give up my bad customs someday.’

‘Of course I want to give up the negative habits sooner or later.’

‘Well, certainly I want to give up the negative habits someday’.

‘Yes, I would like to throw off the bad customs someday’.

As a starting point I focus only on the original texts to see whether the use of tense, modality and polarity is different in English and in German. The results can serve in future research as grounds for a comparison of the parallel translations.
Additional to all that could be worth investigating in the parallel translations, the original newsgroup texts are interesting in themselves in terms of *mode* (Halliday & Hasan 1989). They can be described as written-as-if-spoken; the texts show features of spoken discourse, e.g. interjections and discourse particles, however, the discourse is written and distributed in written form on the internet. Beißwenger & Storrer (to appear: 14) note that “[s]ince this dichotomy is crucial for the categorization in speech and text corpora, it is difficult to decide whether CMC discourse should form part of text or speech corpora.” The research presented here is meant to investigate the affinity of CMC to written and to spoken discourse.

### 2.2 Corpus processing

Since all original texts in the corpus are taken from the internet, they can conveniently be saved in their digital format and be computationally processed. Analysis of tense, modality and polarity eventually has to be done manually. I therefore use the Systemic Coder (O’Donnell 2005) to support computer-assisted manual annotation (CAMA). The texts are saved in UTF-8 format, then processed using the Systemic Coder which can split a text into sentences automatically using punctuation to identify sentence boundaries. Sentences are then split up into clauses containing a verbal group. Since annotation has to be done manually, only small amounts of text can be analysed for a first exploratory investigation. I focus, therefore, on the two sub-corpora of originals discussing relationship problems. The English sub-corpus contains 4,500 words, which can be split into 906 clauses with an average of 5 words per clause. The German equivalent contains 4,080 words and 667 clauses with an average of 6 words per clause.

For the purpose of finding all instances of syntactic negation, i.e. all clauses that include a negation marker like in English for example *no, not, never*, and contracted forms of primary and modal auxiliaries like *aren’t, doesn’t, couldn’t, wouldn’t* or in German *nicht, nichts, nie, kein*, *niemals*, all texts of the English and of the German sub-corpus of originals on relationship problems are processed using a concordance programme, in this case WConcord (Martinek 1998). After extracting only those clauses carrying a marker of syntactic negation, these clauses are annotated, again with the aid of the Systemic Coder. Likewise, all clauses containing a modal auxiliary are traced with WConcord and then processed by Systemic Coder for annotation with a coding scheme. For analysis of tense, all clauses have to be annotated individually, as described in the following section.
3. Some exploratory results

3.1 Tense

The primary function of the verbal group, including the finite element, is of course to connect the discourse to the speaker’s or writer’s immediate reality, or the context. The verbal group “… relates the proposition to its context in the speech event.” (Halliday 1994 [1985]: 75). In addition to this logical function on the ideational metafunction level, however, tense can also be an indicator for spoken or written discourse. The Longman Grammar of Spoken and Written English (LGSWE) (Biber et al. 1999) shows in a study based on the 40 million word LSWE corpus that in different registers, preference for the use of present, past or future tense differs. Thus, the tense system has an additional function on the textual metafunction level.

In order to be able to show the affinity of my corpus of newsgroup texts to spoken or to written discourse, a quantitative analysis of the tense forms in the corpus is necessary. I start by splitting up sentences into clauses containing a verbal group (excluding nominalisations). These clauses are annotated with the tense form they display, using the system networks shown below.

![Diagram of tense system in English](Figure 2: System network for tense in English)
As shown in table 1, English newsgroup texts have a higher amount of non-finite clauses, 17 %, compared with 5 % of non-finite clauses in German. The amount of modal auxiliaries is slightly lower in English, 9 %, as compared to 12 % in German. In the following tables, the number in brackets is the total number of clauses displaying a feature. A sum of more than 100 % results from rounding up results to full numbers.

<table>
<thead>
<tr>
<th>Clause type</th>
<th>English (906 clauses)</th>
<th>German (667 clauses)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tensed</td>
<td>(658) 73 %</td>
<td>(530) 79 %</td>
</tr>
<tr>
<td>Modal auxiliary</td>
<td>(78) 9 %</td>
<td>(78) 12 %</td>
</tr>
<tr>
<td>Non-finite</td>
<td>(153) 17 %</td>
<td>(30) 5 %</td>
</tr>
<tr>
<td>Minor clause</td>
<td>(17) 2 %</td>
<td>(29) 5 %</td>
</tr>
</tbody>
</table>

Table 1: Clause type in English and German newsgroup texts
For the investigation of tense, we will concentrate on the tensed clauses: 658 in English and 530 in German. With the method of transfer comparison, I try to apply to both my English and German texts what an English grammar, i.e. *LGSWE* (Biber et al. 1999), says about the frequency of tense forms in written and spoken discourse. Unfortunately, I do not have any data available about the frequency of tense forms in different registers in German, which would be essential for valid conclusions. I assume, however, that present, past and future tense has the same functions of logically connecting the discourse to the context in both English and German. Table 2 displays the results of the computer-assisted manual annotation with Systemic Coder using the system networks shown in Figures 1 and 2.

<table>
<thead>
<tr>
<th>Tense type</th>
<th>English (658 clauses)</th>
<th>German (530 clauses)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Present</td>
<td>(376)</td>
<td>(315)</td>
</tr>
<tr>
<td></td>
<td>57 %</td>
<td>59 %</td>
</tr>
<tr>
<td>Present Perfect</td>
<td>(39)</td>
<td>(106)</td>
</tr>
<tr>
<td></td>
<td>6 %</td>
<td>20 %</td>
</tr>
<tr>
<td>Present Progressive</td>
<td>(37)</td>
<td>--</td>
</tr>
<tr>
<td>Simple Past</td>
<td>(187)</td>
<td>(74)</td>
</tr>
<tr>
<td></td>
<td>28 %</td>
<td>14 %</td>
</tr>
<tr>
<td>Past Perfect</td>
<td>(6)</td>
<td>(3)</td>
</tr>
<tr>
<td></td>
<td>1 %</td>
<td>1 %</td>
</tr>
<tr>
<td>Past Progressive</td>
<td>(8)</td>
<td>--</td>
</tr>
<tr>
<td>Future</td>
<td>(5)</td>
<td>(23)</td>
</tr>
<tr>
<td></td>
<td>1 %</td>
<td>4 %</td>
</tr>
<tr>
<td>Subjunctive mood</td>
<td>--</td>
<td>(9)</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>2 %</td>
</tr>
</tbody>
</table>

Table 2: Tense types in English and German newsgroup texts

The *LGSWE* (Biber et al. 1999: 456) notes that “about 85 % of all finite verbal groups in the LSWE Corpus are tensed”. Looking at the results in table 1, we can conclude that of the originals in the BTC on relationship problems 69 % of all English finite verbal groups and 79 % of the German finite verbal groups are tensed which is less than in the LSWE Corpus.

Both the English and the German corpus display a strong use of the simple present tense, suggesting the proximity of newsgroup texts to conversation. Biber et al. (1999: 457) state that “[t]he preference for present tense verbs is particularly strong in conversation […], the reliance on present tense reflects speakers’ general focus on the immediate context”.

We can also compare the results in table 2 with what Biber et al. (1999: 456) state about the use of different tenses in different registers: “Conversation and academic prose are alike in showing a strong preference for present tense forms. Fiction shows the opposite pattern, with a strong preference for past tense verbs.” Our newsgroup texts, like conversation (and
academic prose) in the LSWE Corpus, also show a strong preference for present tense forms (simple present + present perfect + present progressive). The English texts have 69 % present tense forms in the finite verbal groups, in the German counterpart, present tense forms amount to 79 %. With only 30 % past tense verbs in the English texts and only 15 % in the German texts, we cannot identify a strong preference for past tense verbs in the newsgroup texts. Therefore, these newsgroup texts can be said to be more similar to conversation than to written narration (fiction) if we look at the use of tense in isolation.

While the second most frequent tense in English is the simple past, the second most frequent tense in German is the present perfect. Again, this reflects the similarities with spoken discourses in German. Quoting the standard German grammar *Duden* (2005: 519-20), in German, the simple past is the unmarked tense for narration in written discourse, while the present perfect is the tense that is typically chosen for narration in spoken discourse. In addition, the results show gaps in the language systems: German does not mark verbs for aspect, while English has no formally distinguishable subjunctive mood. German marks modal auxiliaries for present and past tense, and English does not.

Remaining fully aware of the dangers involved in comparing a 4,500 word corpus with a 40 million word corpus, if we convert our results to the frequency in one million words, they can be visualised as shown in figure 4 below.

![Figure 4: Frequency of present tense, past tense and modal verbs across registers, including English and German newsgroup texts, adapted from Biber et al. 1999: 456](image)

German newsgroup texts (G NG) seem to be more similar to English conversation than the English newsgroup texts (E NG). We also see that there are differences between the use of
tenses in German and English newsgroup texts. The results presented here, however, must be verified with more data from a much larger corpus to allow us to form any conclusions about their significance. My exploratory research suggests that it might be worth carrying out such a comparison of English and German use of tenses in a relatively new register.

3.2 Modality
The next aspect that catches the eye when exploring the BTC is the divergence between the parallel translations of modal auxiliaries. If we think back to example 1, the German modal auxiliary müsste ‘must’ (subjunctive mood) has been translated into English as has to (once), must (twice) and should (once), and in the last translation it appears without a modal auxiliary. Again I start with an exploration of the originals before investigating the parallel translations. With only 9 % (English) and 12 % (German) of all clauses in the sub-corpora on relationship problems carrying a modal auxiliary (that is, 78 clauses in both languages) the study is easily feasible. The results, however, must naturally be treated with caution due to the limited size of the corpus. Once again the clauses are manually annotated with support from the Systemic Coder according to the main functions of modality (Halliday 1994 [1985]) and using the system network shown in figure 5. The results are shown in table 3 below.

![Figure 5: System network for modality in English and in German](image)

<table>
<thead>
<tr>
<th>Modality type</th>
<th>English (78 clauses)</th>
<th>German (78 clauses)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modalization</td>
<td>70 %</td>
<td>49 %</td>
</tr>
<tr>
<td>-probability</td>
<td>87 %</td>
<td>100 %</td>
</tr>
<tr>
<td>-usuality</td>
<td>13 %</td>
<td>0 %</td>
</tr>
<tr>
<td>Modulation</td>
<td>30 %</td>
<td>51 %</td>
</tr>
<tr>
<td>-obligation</td>
<td>48 %</td>
<td>60 %</td>
</tr>
<tr>
<td>-inclination</td>
<td>52 %</td>
<td>40 %</td>
</tr>
</tbody>
</table>

Table 3: Function of modal auxiliaries in English and German newsgroup texts
We can see that in the German texts the clauses are evenly distributed between modalization (probability and usuality) and modulation (obligation and inclination). In English, however, about two thirds of the clauses display modalization, only one third of the clauses showing some kind of modulation. We might say that the German writers have a stronger sense of obligation or inclination, i.e. they think more about what they should do, or what other people should do. We might even claim that the German writers actually think that other people should be doing something (60% obligation) rather than feeling the necessity to act themselves (40% inclination).

Another difference lies in the subtypes of modalization: whereas in English texts 13% of modal auxiliaries express usuality, in the German texts there are none. Either, by coincidence, none of these specific German texts actually expressed usuality, or maybe this is a sign that in the German language system modal auxiliaries cannot express usuality. The results, however, seem to support the statement in Steiner & Teich (2004: 151):

> Generally, the expression of meanings of modality, modulation and also of phase is dissociated from the Finite element: rather than employing auxiliaries for realizing such meanings, German encodes them in Adjuncts. In addition to modal verbs and adverbs, German has a larger number of modal particles, which add other kinds of interpersonal meanings to a clause.

Differences in the function of modal auxiliaries may consequently result from the fact that in German other means are used to express what English expresses using modal auxiliaries, e.g. modal particles or adverbs. This possibility may be worth pursuing in future research.

The second aspect of the investigation deals with the combination of modal and lexical verbs. What kind of process types are modified using a modal auxiliary? For the annotation I use the system network for process types following the Cardiff Grammar (Fawcett, in preparation). Since event-relating or environmental processes are found in neither the English nor the German texts, table 4 demonstrates the results for the remaining five process types. I annotate the German texts with the re-expression tests for identifying participant roles and, thereby, the process type that Fawcett suggests for analysing process types in English. This is an interesting and thought-provoking exercise, to say the least. The evaluation of the validity of the English re-expression tests for other languages certainly deserves more time and attention than it can receive here.
Unsurprisingly, mainly action processes carry additional modality. In my English corpus, other process types are modified only to a small extent, between 5 and 16%. At the same time, in the German texts a quarter of all clauses with a modal auxiliary contain a mental process. This seems to suggest that German writers feel more uncertain about what they think than English writers. These results have to be compared with the overall frequency of process types in the corpus; this has not yet been done due to lack of time. Only then can we see whether the results reflect an even distribution across the process types or whether one or more process types are modified significantly when compared to other process types.

### 3.3 Polarity

The investigation of polarity, or negation markers, is of a slightly different nature. It is inspired by Sinclair (1991) and his statement about where the phrasal verb “set in” was most likely to appear in a clause complex: “A number of the clauses are subordinate. With the samples available, it is not possible to assign status in every case, and there are some of clear main clauses; but I think the tendency to lower status should be noted.” (Sinclair 1991: 74).

While I carried out the above research, I had the impression that markers of syntactic negation (no, not, never and contracted forms like e.g. aren’t, doesn’t, couldn’t, wouldn’t in English or nicht, nichts, nie, kein*, niemals in German) appeared most often in subordinate clauses. To verify this hypothesis, all negated clauses in English and German from the sub-corpora on relationship problems are annotated using the system network shown in figure 6 below.

<table>
<thead>
<tr>
<th>Process type with modal verbs</th>
<th>English (78 clauses)</th>
<th>German (78 clauses)</th>
</tr>
</thead>
<tbody>
<tr>
<td>action</td>
<td>56 %</td>
<td>51 %</td>
</tr>
<tr>
<td>relational</td>
<td>16 %</td>
<td>11 %</td>
</tr>
<tr>
<td>mental</td>
<td>11 %</td>
<td>26 %</td>
</tr>
<tr>
<td>mental 3role cogn.= verbal</td>
<td>12 %</td>
<td>9 %</td>
</tr>
<tr>
<td>influential</td>
<td>5 %</td>
<td>3 %</td>
</tr>
</tbody>
</table>

Table 4: Frequency of process types with modal auxiliary
Figure 6: System network for position of negation marker

The results are shown in table 7 for the 78 sentences in English and 86 sentences in German carrying negative polarity. We see that in the English newsgroup texts less than one third of all negation markers appear in a sentence consisting of a simple clause. In the German newsgroup texts, even less than a quarter of all syntactic negations are in a simple clause.

<table>
<thead>
<tr>
<th>Position</th>
<th>English (78 sentences)</th>
<th>German (86 sentences)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple clause</td>
<td>28 %</td>
<td>22 %</td>
</tr>
<tr>
<td>Main clause in clause complex</td>
<td>32 %</td>
<td>31 %</td>
</tr>
<tr>
<td>Co-ordinated clause</td>
<td>27 %</td>
<td>33 %</td>
</tr>
<tr>
<td>Sub-ordinated clause</td>
<td>41 %</td>
<td>36 %</td>
</tr>
</tbody>
</table>

Table 5: Position of negation marker in clause complex

The distribution across clauses in a sentence consisting of more than one clause, i.e. a clause complex, is fairly even. Slightly more negated finite verbal groups are positioned in a sub-ordinate clause, both in English and in German, thus verifying my hypothesis. There seems to be a tendency to “hide” negative polarity in a sub-ordinate clause, or to elaborate and maybe justify negative polarity using dependent clauses. But, as with the combination of modal auxiliaries and process types, the frequency of negation appearing in a certain type of clause has to be compared with the overall frequency of these different types of clauses in the entire corpus to allow any statement on the significance of these results.

4. Conclusion
In my exploration of the Bremen Translation Corpus, Systemic Functional Grammar has been a valuable theoretical background for an analysis of tense, modality and polarity. On the level of the textual metafunction, I have shown that differences in the use of tense can reflect
whether we are dealing with spoken or written discourse. This is an interesting feature when studying a fairly new register such as newsgroup texts from the internet. With a frequency of 69% (English) and 79% (German) of present tense verbs, newsgroup texts show their proximity to conversation, which also displays a preference for present tense verbs (Biber et al. 1999: 457). But differences can also be detected between the English and German newsgroup texts; the second most frequently used tense in the English texts is the simple past, whereas in the German texts it is the present perfect.

On the level of the interpersonal metafunction, modality and polarity provide us with categories to study how writers express their stance towards the problems they discuss in these newsgroup texts. While in the German texts the distribution between modalization and modulation is fairly even, in the English texts more than two thirds of all modal auxiliaries express modalization. Another finding that may point to a difference between the systems of modal auxiliaries in English and German is that in the German sub-corpus not one modalizing auxiliary expressed usuality. The process types modified most frequently in both languages are action processes, followed in the English sub-corpus by a rather equal distribution between relational, mental and three-role-cognition (verbal) processes. In the German sub-corpus, however, mental processes are clearly the second most frequently modified processes.

The research on polarity is of a different nature. I study the position of negation markers in the clause (complex). There seems to be a tendency for syntactic negation to appear in clause complexes rather than simple clauses, and syntactic negation seems most often to be placed within a subordinate clause.

5. Evaluation and Outlook

After all that could be shown with my exploratory research, further interesting questions emerged. What are the criteria for annotating tenses correctly in English and in German? How can, for example, the will-future be distinguished from will as a modal auxiliary? How can we distinguish simple from historical present, or simple present future?

Is SFG’s categorisation of modality types (Halliday 1994 [1985]: 91) equally valid for other languages, e.g. German? What other means, apart from modal auxiliaries and adverbs, does German have to express modality? How does German express usuality? How can the different process types be analysed in German? Can the criteria for English process type analysis, for example the Cardiff Grammar’s re-expression-tests, be transferred to German? If
not, how can we analyse process types in German, if we want to carry out contrastive studies based on SFG?

The major problem has been the small size of the two sub-corpora of original texts on relationship problems in the BTC. With such small numbers, no statements can be made about the relevance of the results, or even about their truthfulness. The research presented here has constituted what Sinclair (1991: 137) has called “only the first dipping of an inquisitive toe into the vast pool of language texts”. The limited size of the data has, however, made exploratory research feasible on a number of different aspects, thus revealing interesting features in the newsgroup texts and in the comparison of English and German.

As a first step in the future, the corpus of original newsgroup texts will be extended to 10,000 words in English and also 10,000 words in German. According to Biber (1995), “[...] it is possible to represent the distributions of many core linguistic features, both within and across registers, based on relatively short text samples (as short as 1,000 words) and relatively few texts from each register (as few as ten texts)” (Biber 1995: 131).

Since a comparison of registers is not the aim of this research project, I feel confident that ten times 1,000 words in each language will suffice to provide reliable results in an investigation of the lexicogrammatical systems of German and English, while still remaining a feasible size for computer-assisted manual annotation. Furthermore, the results from the BTC must be compared to results from a reference corpus, e.g. the British National Corpus (BNC) for English and the Cosmas II for German. This procedure will enable us to see more clearly how newsgroup texts from the internet differ from other registers.

Work then has to be undertaken to develop guidelines for the annotation of tenses, process types and modality in both languages to make results comparable. When detailed descriptions of the finite verbal groups in the two languages represented in the BTC have been completed, we can start to analyze how these phrases are translated. My ultimate aim is to study what kind of variation exists between the five parallel translations, and how this variation can be explained.

Anke Schulz, Research Assistant
Institut für Sprach- und Literaturwissenschaft
Technische Universität Darmstadt, Germany
schulz@linglit.tu-darmstadt.de

Systemic Functional Linguistics in Use, OWPLC 29, 2008
References


Systemic Functional Linguistics in Use, OWPLC 29, 2008 715
