

# Multi-level error annotation in learner corpora

Anke Lüdeling, Maik Walter, Emil Kroymann, Peter Adolphs  
Institut für deutsche Sprache und Linguistik  
Humboldt-Universität zu Berlin  
anke.luedeling@rz.hu-berlin.de

## 1. Introduction

Learner corpora – principled collections of learner language – provide interesting insights into the mechanisms by which a foreign language is acquired. For overviews over the current state of learner corpus research see Granger (2002, to appear), Nesselhauf (2004), and Pravec (2002).

Learner corpora are used to test hypotheses in the theory of acquisition in two main ways. First, learner corpora can be used for the so-called contrastive interlanguage analysis (CIA), i.e. the quantitative comparison of learner language and native language to find patterns of overuse or underuse. For CIA, a corpus does not have to be tagged.

In this article we are concerned with the second main area of learner corpus research: error tagging. While error-tagging is problematic in many theoretical respects, it is probably not controversial anymore that error-tagged learner corpora can be useful for a number of research questions if the tagging follows certain guidelines. In this paper we do not argue for the need for error annotation (see Granger, to appear, for a motivation) or discuss the theoretical problems involved but are concerned only with issues of error tagging and corpus architecture. We argue for a multi-level standoff architecture (rather than a flat token-tag architecture) for error-tagged learner corpora. By using the German learner corpus Falko as an example, we show how multi-level approaches to learner corpora can help solve some of the problems that occur in error tagging if flat annotation models are used.

### 1.1. The German learner corpora situation

While there are many learner corpora for English and some for other languages, for example French and Norwegian (see Granger, to appear, for an overview) there are only very few learner corpora for German as a Foreign Language (GFL). Most German learner corpora are small and/or not publicly available. Ursula Weinberger from Lancaster University collected a corpus (95 texts, 27635 words) of German learners with English as their L1 (Weinberger 2002). The corpus is the only error-tagged GFL corpus we are aware of but it is not publicly available. Julie Belz and her colleagues at Pennsylvania State University are building a corpus of telecollaborative data (Belz 2004) which is also not publicly available. In addition there is a well-known collection of learner errors which is available on CD (Heringer 1995). This collection is not a corpus in that it does not contain full texts.

### 1.2. Falko

Before we describe the architecture of our example learner corpus Falko in the following sections we want to say a few words about its design and content. Falko, which stands for *Fehlerannotiertes Lernerkorpus* ‘error-annotated learner corpus’ is in its building-up phase. The corpus is currently small, but growing. From summer 2005 on, Falko will be available online at <http://www.linguistik.hu-berlin.de/korpuslinguistik/projekte/falko/index.php>.

Already a number of learner language studies have used Falko data in their research (Hirschmann 2005; Lippert, in preparation; Schmidt & Walter, to appear; Walter, Schmidt & Dittmar, submitted).

Falko contains several distinct sets of data. Each text is annotated with detailed header information so that the data sets can be combined to sub-corpora according to the needs of the researcher.

Core corpus: The core corpus is a highly controlled set of summaries of academic texts written by advanced GFL learners (henceforth L2) and native German speakers (henceforth L1) under comparable conditions. All texts are produced by students of German after approximately two years of study. All of the L2 students have passed the DSH exam (*Deutsche Sprachprüfung für den Hochschulzugang ausländischer Studienbewerber*, roughly comparable to the TOEFL test for English). As of June 2005, the core corpus consists of 59 L2 texts and 41 L1 texts (together 35 949 tokens). Further data sets are collected every term. Because students often copy whole sequences from the original texts, the originals are provided in the same format for reference. The learner texts were written manually and later digitized. Up to now, error tagging has only been done on the core corpus.

In addition we have several extension text sets. Most important among these is the longitudinal data collected at Georgetown University in Washington D.C. This data consists of so-called prototypical performance tasks collected from students at the end of four consecutive curricular levels and coded for clause types (Byrnes 2002). Other extension sets are composed of summaries of the same original texts as the core corpus texts which are collected in foreign countries and essays written in advanced linguistic classes.

All texts in Falko are tokenized and tagged for part-of-speech using the TreeTagger (Schmid 1994). We evaluated the tagging error rate and found that, although it is slightly higher than the error rate for newspaper texts, it is still low enough for the data to be usable.

## 2. Error tagging and learner corpus architecture

Learner corpora, as well as other text corpora, differ with respect to how much linguistic information is added to the raw text. Whereas most available learner corpora provide a header for their texts that specifies information such as the L1 of the learner, the task, the learner history, etc., most learner corpora do not have any further linguistic annotation (Granger, to appear).

In this article we are concerned with error-annotated learner corpora. Most error-tagged learner corpora that are currently available use some kind of flat file format. We want to illustrate why this format is problematic for learner data and motivate a multi-layer standoff model as a more appropriate annotation model for learner data. Simply stated, in multi-layer standoff annotation (originally developed for speech corpora, Carletta et al. 2003) the original text is coded in a reference line and each annotation is coded in an independent level with pointers to the reference line. Before we explain the format in Section 3, we discuss three properties of learner language and error annotation that are problematic for flat annotation models.

### 2.1. Target hypothesis

All error annotation implies an interpretation on the part of an annotator. This fact has often been discussed and is, in fact, one of the main arguments against error annotation. Consider the following learner utterance with an error tag from Weinberger (2002, 25).

- (1)            die Erklärung für <MoArInGn>diese Phänomen ist einfach  
                 the explanation for these phenomenon is simply  
                 (Mo – morphology, Ar – article, In – Inflection, Gn – gender)

In this utterance *diese* ‘this’ and *Phänomen* ‘phenomenon’ do not agree. Weinberger interprets this as a gender error, most likely on the basis of the surrounding context which is not given in the description (it should be *dieses Phänomen*; *Phänomen* is neuter and the determiner should agree in gender). Without further information, the error could, however, also be seen as a number error (*diese Phänomene*, plural). In this case, the error would be marked on the noun and classified differently.

In error tagging it is impossible **not** to interpret, which leads to two problems in flat annotation models. First, the ‘target hypothesis’ or “reconstruction of those utterances in the target language” (Ellis 1994: 54) is usually implicit, as in (1). Second, encoding several different hypotheses is not possible. This is due to the fact that the error tags are coded in a flat file together with the original data.

In Falko we have chosen therefore to give an explicit target hypothesis TARGET1 for each deviant section (sequence of tokens) of the text. The errors are then coded with regard to this target hypothesis (see Section 4). Different target hypotheses TARGET2, TARGET3, etc. can be stated and errors can be coded with regard to each target. The above example would appear as follows:

utterance	<i>Die</i>	<i>Erklärung</i>	<i>für</i>	<i>diese</i>	<i>Phänomen</i>	...
TARGET1				dieses		
Error Tag				Gender		
TARGET2					Phänomene	
Error Tag					Number	

Table 1: Illustration of several target hypotheses for one learner error

## 2.2. Error exponent

Oftentimes an error is not confined to one word. Consider the following example from Falko (orthographic errors are interpreted in the gloss):

- (2) *Eine nicht informations übermittelnde Kommunikation mit nicht ernsthaften Menschen kann nur dann statt finden, wenn sie entweder sich über das Thema der Diskussion nicht geeignet haben, oder sie wollen kein Gewinn erzielen (sondern reden nur so dahin).*  
A not information transmitting communication with not serious people can only then place take, if they either REFL about the topic the-POSS discussion not agreed have, or they want no profit realize (but talk only so there)

[Translation (with interpretation): A non-information-transmitting communication with unserious people can only take place if either they have not agreed on the topic of the discussion, or they do not want to make a profit (but only chat). (Falko 1.1.L2)]

Most orthographic errors can be tied to a single token (the correction is given after the slash): *Diskusion/Diskussion, nich/nicht, erziehlen/erzielen*. Other orthographic errors should ideally be tied to two tokens: *informations übermittelnde/informationsübermittelnde, stadt finden/stattfinden*. In the second case, we have two orthographic errors: *stadt/statt* and the spatium. (The whole error could also be classified as a word formation error, see below). Other errors, like word order errors, can be marked on a sequence of several tokens. An example is [...] *weil sie entweder sich ... geeignet haben [...]* which should be [...] *weil sie sich entweder ... geeinigt haben [...]*. In this case, the reflexive is in the wrong place and the verb is incorrect. But, since the reflexive is an (obligatory) argument of the verb *einigen* ‘to agree’ (which should have been used), the whole clause should be the exponent of the error. A similar argument can be given for [...] *oder sie wollen kein Gewinn erzielen [...]* which has main clause word order but should have dependent clause word order (*oder sie keinen Gewinn erzielen wollen*) because of the conjunction *wenn*.

Since it is sometimes necessary to mark an error on a sequence of tokens, the encoding format should provide for this.

### 2.3. Independent annotation levels

All encoding schemes for error analysis use some kind of level system. Some are based on linguistic levels like morphology, syntax, orthography etc., others are based on formal properties of the errors like omission, insertion etc. or on the part-of-speech of the error exponent (for an overview see Granger 2002). In learner language it is often the case that several errors need to be coded with respect to the same token or sequence of tokens. In (2), for example, we find orthographic errors – sometimes several in one word – overlapping with word formation errors. An error encoding system must therefore allow for the possibility of an in principle arbitrary number of annotation levels.

In addition these levels must be independent of each other. This has a thematic reason as well as a formal reason. Consider the missing space in the *stadt finden/stattfinden* error in (1). This could be an orthographic error or a word formation error (the learner did not realize that *stattfinden* is a particle verb). As this cannot be decided, both possibilities should be marked. The formal reason for independent annotation layers pertains to the fact that in tree-structures like XML conflicting hierarchies cannot be coded within the same tree. This is further described in the following section.

To summarize: An annotation model for error annotation of learner corpora should ideally have the following properties:

- (a) it should be possible to encode several explicit target hypotheses and the respective errors on several independent levels
- (b) it should be possible to mark sequences of tokens as error exponents

In the following section we describe different annotation models that have been suggested for learner corpora and discuss how they can be evaluated according to these criteria. Then we introduce the model used in Falko and discuss some advantages and problems.

## 3. Annotation Models

### 3.1. Flat annotation models

#### 3.1.1 Tabular annotation models

Traditionally, text corpora are annotated on the token level, i.e. a tag is assigned to each token on each annotation layer. The example from the Verbmobil corpus (<http://verbmobil.dfki.de/>) shows, that each word is followed by a part-of-speech tag (following the STTS tagset, cf. Schiller et al. 1995) and its lemma:

(3) I/PP/I have/VBP/have got/VBN/get ././, Monday/NP/Monday or/CC/or Tuesday/NP/Tuesday off/RB/off

We call this annotation model a tabular annotation model because it could also be represented as a table where each row corresponds to one token, and each column corresponds to an annotation layer. Such tables can be easily indexed or stored in a relational database, and can be efficiently searched (for example using the IMS Corpus Workbench, Christ 1994).

Tabular annotation models have certain shortcomings with regard to the requirements stated above:

(a') The number and category of annotation layers must be decided in the corpus design phase. Thereafter it is not easily possible to add layers. Furthermore each token must be given a value for every annotation layer for the whole corpus. For error annotation this is not appropriate because one cannot determine beforehand how many errors have to be tagged with any given token.

(b') Joining adjacent cells is not possible which renders this model unsuitable for annotating sequences of words.

Another kind of token-based annotation model is the C-LEG corpus described in Weinberger (2002). Weinberger developed an error tagset for classifying errors made by learners of German. The tags are inserted in angle brackets into the raw text preceding the erroneous word. The problem is that her tagset covers both errors which can be attributed to single words and errors which can be attributed to sequences of words. Example (4) illustrates the problem: the end of the erroneous phrase is not structurally marked (note that the bold face is added by Weinberger but not encoded in the corpus):

- (4) <LxPhCh> **Es gibt eine veränderte Gesellschaft** und ...  
 there is a changed society and ...  
 Lx- lexical, Ph – phrase, Ch – incorrect choice

### 3.1.2 Tree-based annotation models

Another common annotation model is structural annotation via ordered tree structures. These are commonly represented using mark-up languages like XML or SGML. Consider an example from the French Interlanguage Database corpus FRIDA (Granger 2003).

- (5) L'héritage du passé est très <G><GEN><ADJ> #fort\$ forte </ADJ></GEN></G> et le sexisme est toujours présent.  
 Translation: The heritage of the past is very strong and the sexism is always present.  
 G – error domain gender, GEN – sub-domain, ADJ – word class

It shows a misused adjective inflection. The learner incorrectly used the form *forte*, which is marked for feminine gender, instead of the form *fort*, which is marked for masculine gender. The target hypothesis is coded in front of the erroneous word.

Tree-based annotation models overcome some of the disadvantages of token level annotation models in that they allow annotating sequences of tokens. However, it is not possible to annotate overlapping ranges on different annotation layers since these cannot be mapped on a single ordered tree.

word	aus	denen	sich	insgesamt	die	Bedeutung	und	den	Sinn	des	ganzen	Textes	erschließen	läßt
target					die Bedeutung und der Sinn des ganzen Textes erschließen lassen									
finiteness														x
agreement					x									
binding								x						

Table 2: Illustration of overlapping annotation spans.

In the example from Falko, one can see an error which must be described on multiple annotation levels. However, the errors ranges on the different layers overlap. This kind of analysis cannot be encoded in a flat annotation model.

### 3.2. Multi-layer standoff annotation models

Multi-layer standoff annotation refers to a subset of the very general annotation models which have been developed for speech corpora and multi-modal corpora. Speech data, as well as written texts, can be conceptualized as a continuous sequence of linguistic or extra-linguistic, signals emitted by a speaker/writer. The signals can be aligned along a time line.

This concept fits naturally to speech data given as a recorded audio signal. The time line is a continuous interval, which covers the recorded signal. For written texts, the concept of a time

line is, of course, metaphorical. The time line consists of a linear ordered set of discrete time points which correspond to the word boundaries in the text.

In this model, an annotation is a value which is associated to an interval on the time line. The term *standoff* refers to the fact that the annotation is not inserted into the text, as in the case of the flat annotations models discussed above. Instead, the text and the annotation are coded separately. This has a number of advantages (for instance, concerning the authenticity of the text). Here we concentrate on one example that is crucial for error annotation: Annotations can be defined for overlapping intervals of the time line. Multiple annotations that do not overlap can be encoded in one annotation layer. Overlapping annotations must be represented on multiple layers.

The multi-layer standoff model has been formalized a number of times. See for instance the annotation graph model (Bird and Liberman 2001), the ODAG Data Model (for ordered directed acyclic graphs, Carletta et al. 2003) and the model described in Dipper et al. (2004a). The annotation graph model is the model described above. The latter two models are extensions in which one can express relations between different annotations. In error-annotated learner corpora, this might be used to formally relate every error analysis to its target hypothesis. As there are currently no suitable annotation tools for these models available, this possibility will have to be investigated at a later stage.

The important point is that the annotation graph model complies with the requirements stated above. In other words, the erroneous range of a text can be explicitly coded and there can be alternative annotations for the same range or even overlapping ranges of the text and is therefore selected as the model for error annotation in Falko.

A number of ready-to-use annotation tools based on the annotation graph model have been developed in the speech community. For Falko we have chosen EXMARaLDA (Schmidt 2004, 2005), a partitur editor. EXMARaLDA visualizes the standoff annotation model in a straightforward way. The time line and multiple annotation layers are always visible.

EXMARaLDA implements three different variants of the multi-layer standoff annotation model, to represent the annotated text, basic transcription, list transcription and segmented transcription. The basic transcription model is an implementation of the AG model. It contains a time line and arbitrarily many annotation layers. The elements of this model can be manipulated directly through the EXMARaLDA user interface. The list transcription and the segmented transcription are more expressive and can represent certain relations between annotations. These relations are, however, very specific to EXMARaLDA's primary use case, the analysis of spoken discourse. List transcription and segmented transcription therefore do not seem useful for our purposes. If we are to code relations between different annotations, the problem of encoding these relations in EXMARaLDA's basic transcription model remains. It is possible to have meta-information in the form of key-value pairs for each layer. This meta-information could be exploited to express relations between layers.

#### **4. Error annotation on three levels**

This section shows how the described annotation model is used in the error annotation in Falko. At the moment not all texts in Falko are error annotated, and we have not yet completed the design of all tagsets.

In addition to the automatically tagged part-of-speech category and one or more target hypotheses, Falko's tagging scheme currently uses the following linguistic levels: orthography, word formation, agreement, government, tense, mood, word order, and expression, which codes errors of appropriacy. Examples are given in Table 3. Each linguistic level consists of at least

three sub-levels, structured according to a pattern widely used in second language acquisition research: identification, description, and explanation.

level	example of an error in Falko	target form	gloss of the target
orthography	<i>Gebeuden</i>	<i>Gebäuden</i>	<i>buildings</i>
word formation	<i>Analysenverfahren</i>	<i>Analyseverfahren</i>	<i>manner of analysis</i>
agreement	<i>dass die <u>Bedeutung</u> der "kleinen" Worte eine wichtige Bedeutung für den gesamten Text haben können.</i>	<i>dass die Bedeutungen der "kleinen" Worte eine wichtige Bedeutung für den gesamten Text haben können.</i>	<i>that the meanings of the "small" words could have an important meaning for the whole text</i>
government	<i>Bei Gespräche</i>	<i>Bei Gesprächen</i>	<i>In conversations</i>
tense	<i>Die Realisten <u>wollen</u> die Wirklichkeit nicht kopieren, was mit der Erfindung der Photographie der Fall war.</i>	<i>Die Realisten wollten die Wirklichkeit nicht kopieren, was mit der Erfindung der Photographie der Fall war.</i>	<i>The Realists did not want to copy reality, which was the case in the invention of photography.</i>
mood	<i>Die Hermeneutik <u>würde</u> in der Neuzeit erfunden und ist keine Kompensation ...</i>	<i>Die Hermeneutik wurde in der Neuzeit erfunden und ist keine Kompensation ...</i>	<i>Hermeneutics was discovered in modern times and is not a compensation ...</i>
word order	<i>dass nur er ( Josef K. ) <u>konnte</u> durch dieses Tor eingelassen werden</i>	<i>dass nur er ( Josef K. ) durch dieses Tor eingelassen werden konnte</i>	<i>that only he (Josef K.) could be admitted through this gate</i>
expression	<i>Die von F. Schlegel und F. Schleiermacher <u>genannte</u> Hermeneutik ist...</i>	<i>Die von F. Schlegel und F. Schleiermacher begründete Hermeneutik ist...</i>	<i>Hermeneutics, founded by F. Schlegel and F. Schleiermacher ...</i>

Table 3: Linguistic levels in Falko.

Error analysis has four steps according to Corder (1974) and Ellis (1994). First, the learner data is collected. Error analysis in the strictest sense starts with the identification of errors. After this, the identified deviations from the target hypothesis are described. Finally, an explanation for the deviation is provided.

As described above, we mark one or several target hypotheses. This presupposes that we can identify an error. Such an approach raises many problems, which are extensively discussed in second language acquisition research (cf. Ellis 1994: 50-54). This applies in particular to Corder's (1967) distinction between competence errors (errors) and performance errors (mistakes). Ideally, mistakes should not be marked. Since it is not always possible to decide whether a deviation is an error or a mistake, we include all deviations and mark them on the identification level. This differs significantly from classical error analysis found in second language acquisition (Ellis 1994) as well as in classroom research (Brown 1994). The distinction between errors and mistakes becomes relevant on the level of explanation.

In (6) we can, for example, identify a word order error as shown in Table 4.

(6)	<i>Am</i>	<i>Ende</i>	<i>seines</i>	<i>Lebens</i>	<i>erfährt</i>	<i>er</i>	<i>von</i>	<i>dem</i>	<i>Türhüter,</i>
	At	end	POSS	life	find out	he	from	the	doorman
	<i>dass</i>	<i>nur</i>	<i>er</i>	<i>( Josef K. )</i>	<i>konnte</i>	<i>durch</i>	<i>dieses</i>	<i>Tor</i>	<i>eingelassen</i>
	that	only	him	NAME	could	through	this	gate	admitted
	<i>werden,</i>	<i>weil</i>	<i>der</i>	<i>Eingang</i>	<i>nur</i>	<i>für</i>	<i>ihn</i>	<i>bestimmt</i>	<i>war.</i>
	AUX- PASSIVE	because	the	entrance	only	for	him	determined	COPULA- PAST

[Translation (with interpretation): At the end of his life, he finds out from the doorman that only he (Joseph K.) could be admitted through this gate, because the entrance was determined for him only. (Falko 1.1. L2)]

word	dass	nur	er	...	konnte	durch	dieses	Tor	eingelassen	werden	
target					durch dieses Tor eingelassen werden konnte						
word order identification					X						

Table 4: Identification level.

The description level includes a comparison between the learner utterances and the target hypothesis. In second language acquisition research, different systems are used for this comparison. We have adopted an L2-surface oriented approach following Dulay, Burt and Krashen (1982), who used categories such as “omission”, “addition”, “misinformation” und “misordering” (for alternatives cf. Granger, to appear). Level-specific information can be added, which means that a separate error tagset must be developed for each linguistic level. The tags sets used in Falko are descriptive and as ‘theory-neutral’ as possible.

In the subordinate clause in (6), the finite verb is not placed in the final position. We have developed a descriptive tagset based on the topological field model which is a descriptive model for constituent order in German and other Germanic languages (Höhle 1986). In this model the German sentence consists of three basic topological fields: the initial field (“Vorfeld”=VF), the middle field (“Mittelfeld”=MF) and the final field (“Nachfeld”=NF).

The middle field in the target hypothesis in (6) is limited by the complementizer *dass* and the finite verb *konnte*. These limits are referred to as left sentence bracket („linke Satzklammer“=LSK) and right sentence bracket (“rechte Satzklammer”=RSK). By comparing the learner utterance with the target hypothesis in the topological-field model, we can classify the error as MF\_RSK: The element *konnte* occurs in the middle field MF, although the correct position is in the right sentence bracket RSK.

Word	dass	nur	er	...	konnte	durch	dieses	Tor	eingelassen	werden	
Target					durch dieses Tor eingelassen werden konnte						
word order identification					X						
word order description					MF_RSK						

Table 5: Description level.

The final step, explanation, “is concerned with establishing the source of the error“ (Ellis 1994: 57). This sub-level is strongly influenced by the annotator’s theory of second language acquisition and therefore more subjective than the other levels. Three main factors are often discussed: the influence of the L1 and other foreign languages in the learner’s history, the L2 itself and the learner’s developmental behaviour (Richards 1971). Explanations in Falko consider the learner’s language biography, which is stored in the meta-data in the header for each text. In



our example, the explanation of the error could be due to a transfer problem from the learner’s L1: Verb-final structures are ungrammatical in Polish (Damerau 1992: 135), which is the L1 of the learner who produced the utterance. This explanation is annotated as “transfer”.

word	dass	nur	er	...	konnte	durch	dieses	Tor	eingelassen	werden	
target					durch dieses Tor eingelassen werden konnte						
word order identification					x						
word order description					MF_RSK						
word order explanation					transfer						

Table 6: Explanation level.

This example illustrates the way the different sub-levels of error annotation can be coded in a multi-layer architecture. The levels *identification*, *description*, *explanation* have been coded in other error annotation schemes as well. In flat systems they are often combined into a single tag (Weinberger 2002 is one example). This is problematic, however, because, as discussed above, the different levels show different degrees of theory dependence. The identification sub-level is probably the most stable or ‘objective’ level (but even here, different target hypotheses and different concepts of ‘error’ could lead to discrepancies). It is well possible that different tagsets for the descriptive sub-level might be developed for the same error sequences. For the explanation sub-level, competing analyses are highly probable. In a system like Falko it is possible to add new levels at any point in time.

## 5. Summary

In this paper we described and motivated a multi-layer standoff architecture for error analysis in learner corpora and showed how this is implemented in the German learner corpus Falko. We first discussed properties of learner language and error annotation that led to the formulation of the following criteria for an annotation model for learner data: it must be possible to code errors on arbitrarily many independent levels and it must be possible to mark sequences of tokens as error exponents. We showed that – in contrast to flat annotation models – multi-layer standoff models fulfil these requirements. In the final section we illustrated the model by showing how a three-level error annotation can be coded in Falko.

## 6. Acknowledgements

Our warmest thanks goes to Karin Schmidt for providing ideas, encouragement and a lot of coding and analysis work in the making of Falko and to Castle Sincrope for help with the English.

We also want to thank the members of the Hauptseminar ‘Phänomene’ at the Humboldt-University, Berlin, in the winter term 2004/2005 who gave much valuable input and especially Seanna Doolittle, Stefan Büch, Matthias Süß and Kerstin Zimmermann for developing a tagset for word-order errors.

## 7. References

Belz, J. A. (2004) Learner Corpus Analysis and the Development of Foreign Language Proficiency, *System: An International Journal of Educational Technology and Applied Linguistics*. 32.4: 577-591.

- Bird, S. and Liberman, M. (2001) A Formal Framework for Linguistic Annotation, *Speech Communication* 33 (1,2), 23-60.
- Brown, H. D. (1994) Contrastive Analysis, Interlanguage, and Error Analysis, in H. D. Brown (ed.) *Principles of Language Learning and Teaching*. Third edition, (Englewood Cliffs: Prentice Hall Regents), 192-225.
- Byrnes, H. (2002) The Role of Task and Task-based Assessment in a Content-oriented Collegiate FL Curriculum, *Language Testing* 19.4: 425-443.
- Carletta, J., Kilgour, J., O'Donnell, T., Evert, S. and Voormann, H.(2003) The NITE object model library for handling structured linguistic annotation on multimodal data sets, *Proceedings of the EACL Workshop on Language Technology and the Semantic Web (3<sup>rd</sup> Workshop on NLP and XML, NLPXML-2003)*, Toulouse.
- Christ, O. (1994) A modular and flexible architecture for an integrated corpus query system. *COMPLEX'94*, Budapest, 1994. Available online from <http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/#Papers> (accessed June 17<sup>th</sup>, 2005).
- Corder, S. P. (1967) The significance of learner errors. *International Review of Applied Linguistics* 5: 161-170.
- Corder, S. P. (1974) Error analysis, in J. P. Allen and S. P. Corder (eds.) *The Edinburgh Course in Applied Linguistics*, Vol. 3 (London: Oxford University Press).
- Damerau, N.(1992) *Polnische Grammatik* (Berlin & New York: Walter de Gruyter).
- Dipper, S., Faulstich, L., Leser, U and Lüdeling, A.(2004) [Challenges in Modelling a Richly Annotated Diachronic Corpus of German](http://www.deutschdiachrondigital.de/publikationen/dipper_etal_XBRAC04.pdf), *Workshop on XML-based richly annotated corpora*, Lisbon. Available online from [http://www.deutschdiachrondigital.de/publikationen/dipper\\_etal\\_XBRAC04.pdf](http://www.deutschdiachrondigital.de/publikationen/dipper_etal_XBRAC04.pdf) (accessed June 17<sup>th</sup>, 2005).
- Dulay, H. C., Burt, M. K. and Krashen, S. (1982) *Language Two* (New York: Oxford University Press).
- Dürscheidt, C. (2000) *Syntax. Grundlagen und Theorie* (Wiesbaden: Westdeutscher Verlag).
- Ellis, R. (1994) *The Study of Second Language Acquisition* (Oxford: Oxford University Press).
- Granger, S (2002) A bird's-eye view of learner corpus research, in S. Granger, J. Hung, S. Petch-Tyson (eds) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (Amsterdam: John Benjamins), 3-33.
- Granger, S. (2003) Error-tagged Learner Corpora and CALL: A Promising Synergy. *CALICO Journal*, 20 (3), 465-480.
- Granger, S (to appear) Learner corpora, in A. Lüdeling and M. Kytö (eds) *Corpus Linguistics. An International Handbook* (Berlin: Mouton de Gruyter).
- Heringer, H. J. (1995): *Aus Fehlern lernen*, Universität Augsburg, CD-ROM for Win9x/NT.

- Hirschmann, H. (2005) *Platzhalterphrasen bei fortgeschrittenen Lernern des Deutschen als Fremdsprache*, Zulassungsarbeit, Humboldt-Universität zu Berlin.
- Höhle, T. (1986) Der Begriff Mittelfeld. Anmerkungen über die Theorie der topologischen Felder, *Akten des Siebten Internationalen Germanistenkongresses*, Göttingen, 329-340.
- Lippert, E. (in preparation) *Definitheit in Nominalphrasen bei Lernern des Deutschen als Fremdsprache*, Magisterarbeit, Humboldt-Universität zu Berlin.
- Milton, J. (1998) Exploiting L1 and Interlanguage Corpora in the Design of an Electronic Language Learning and Production Environment, in S. Granger (ed.) *Learner English on Computer* (London & New York: Addison Wesley Longman), 186-198.
- Nesselhauf, N. (2004) Learner corpora and their potential for language teaching, in J. Sinclair (ed) *How to Use Corpora in Language Teaching* (Amsterdam: John Benjamins), 125-152.
- Pravec, N. A. (2002) Survey of learner corpora. *ICAME Journal* 26, 81-114. Available online from <http://nora.hd.uib.no/icame/ij26/> (accessed June 17<sup>th</sup>, 2005).
- Schiller, A, Teufel, S., Thielen, C. (1995) *Guidelines für das Tagging deutscher Textcorpora mit STTS Technical report*, IMS Stuttgart / Seminar für Sprachwissenschaften, Tübingen.
- Schmid, H. (1994) Probabilistic Part-of-Speech Tagging Using Decision Trees, *Proceedings of International Conference on New Methods in Language Processing*, Manchester.
- Schmidt, K. and Walter, M. (to appear) Hier wird die Wahl schwer, aber entscheidend. Konnektorenkontraste im Deutschen. Eine Pilotstudie, in H. J. Krumm. and P. [Portmann-Tselikas](#) (eds.) *Innovationen. Theorie und Praxis. Österreichische Beiträge zu Deutsch als Fremdsprache. Jahrbuch 9/2005* (Innsbruck: Studienverlag).
- Schmidt, T. (2005) EXMARaLDA und Datenbank 'Mehrsprachigkeit' - Konzepte und praktische Erfahrungen, in S. Dipper and M. Stede (eds.): *Heterogenity in Focus: Creating and Using Linguistic Databases*. ISIS (Interdisciplinary Studies on Information Structure), Working Papers of the SFB 632, Universität Potsdam.
- Schmidt, T. (2004) EXMARaLDA - ein System zur computergestützten Diskurstranskription, in: A. Mehler and H. Lobin (eds.): *Automatische Textanalyse. Systeme und Methoden zur Annotation und Analyse natürlichsprachlicher Texte* (Wiesbaden: Verlag für Sozialwissenschaften), 203-218.
- Taylor, G. (1986) Errors and explanations. *Applied Linguistics*: 144-166
- Richards, J. (1971) A non-contrastive approach to error analysis. *English Language Teaching Journal* 24: 205-219.
- Walter, M., Schmidt, K. and Dittmar, N. (submitted) Connectives in advanced learner varieties: A corpus-linguistic analysis, in H. Byrnes, H. Weger-Guntharp and K. A. Sprang (eds.) *GURT 2005: Educating for Advanced Foreign Language Capacities: Constructs, Curriculum, Instruction, Assessment* (Washington: Georgetown University Press).
- Weinberger, U. (2002) *Error Analysis with Computer Learner Corpora. A corpus-based study of errors in the written German of British University Students*, MA thesis, Lancaster University.