Translation Shifts and Frame-Semantic Mismatches: A Corpus Analysis

Sebastian Padó
Universität Stuttgart
Stuttgart, Germany

Katrin Erk
University of Texas
Austin, Texas

Translation shifts form an important area of research in translation studies, and knowledge about shifts can be beneficial for both practical translation and multilingual language processing. However, manual quantitative assessment of shifts is quite expensive. At the same time, the semantic analysis of text in terms of word sense and semantic roles is an active research topic in computational linguistics. Since semantic roles reflect predicate-argument structure, it is an attractive hypothesis that cross-lingual mismatches of sense and semantic role annotation in a parallel corpus can be interpreted as (semantic) translation shifts.

This paper reports on a study testing this hypothesis on a 1000-sentence parallel corpus (English–German) with FrameNet frames and semantic roles. The results are mixed: A significant portion, but overall a minority, of cross-lingual semantic mismatches indicate semantic translation shifts. We identify two major sources of spurious mismatches. The first one is annotator error, which may be eliminated in future studies. The second one is more fundamental, namely the difficulty to formulate criteria that consistently distinguish between semantic categories.

1 Introduction

One of the most fundamental observations in the area of translation studies is that a translation can differ from its original in almost all linguistic respects – lexical, syntactic, discursive – and still be a "good" translation. These changes are termed *translation shifts*, and there is a large body of work that offers different accounts of the relationship between the process of translation and translation shifts (for an overview, see Bakker *et al.*, 1998). Our article assumes the linguistics-oriented perspective of Catford (1965). Catford defines a concept of *formal correspondence* which holds between two linguistic categories in the source and target languages if the target language category occupies "the same place in the economy of the target language as the given source language category in the source language". Shifts are then simply cases of translations where formal

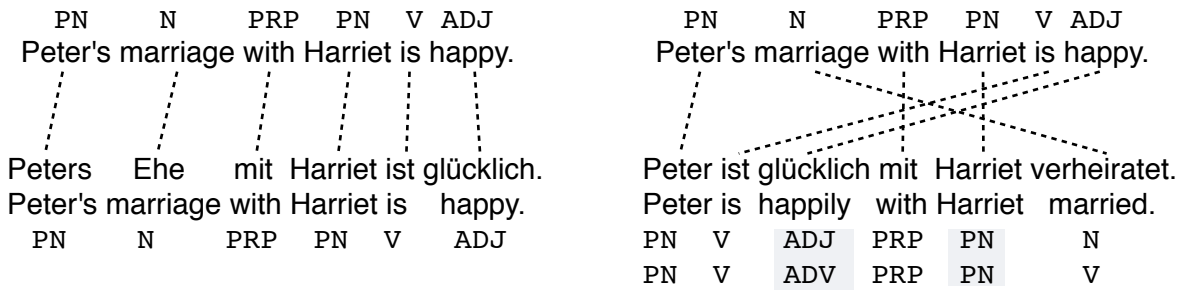

Figure 1: Annotation projection for parts of speech (left: formal correspondence; right: translation shift)

correspondence does not hold. A number of classifications for semantic shifts have been proposed (Catford, 1965; Vinay and Darbelnet, 1958; van Leuven-Zwart, 1989). In this article, we will only make a very broad distinction (corresponding to Cyrus's (2006) top-level classes) between *grammatical shifts* (such as changes in syntactic category, number change, pronominalization etc.) and *semantic shifts* such as the deletion/insertion of words or changes in word meaning.

Knowledge about translation shifts is clearly important for translators to help them choose appropriate translation strategies (Baker, 1996) and avoid translation mistakes as well as undue influence of the source language on the target language (Teich, 2003). Parallel and comparable corpora play an important role in this research program in that they can not only serve for the "grounding" of intuitions in actual examples, but also allow the extraction of quantitative tendencies about shifts and the targeted investigation of genre- and domain-specific shifts (Čulo *et al.*, 2008).

Knowledge about translation shifts is also highly valuable for *computational* linguistics studies concerned with multilingual data. For example, Collins *et al.* (2005) have shown that English-German Machine Translation can profit from making the word order of German sentences more like their English counterparts. Another important application is *annotation projection* (Yarowsky and Ngai, 2001), a technique which is used to create language technology resources for new languages by transferring linguistic annotations from source language sentences onto target language sentences using information about the alignment of linguistic units in parallel sentences, such as words, word sequences, or constituents. Annotation projection has been applied to, among others, WordNet word senses (Bentivogli and Pianta, 2005), parts of speech (Yarowsky and Ngai, 2001), and semantic role annotation (Padó and Lapata, 2009). Figure 1 shows an example (alignments at the level of individual words shown as dashed lines) that demonstrates erroneous projections due to translation shifts. In the left-hand case, projection proceeds without problems. The right-hand case, however, demonstrates how a grammatical shift (the replacement of a noun with an adverbial modifier by an adjective with an adverbial modifier) leads to wrong annotations: Due to its alignment to *marriage*, the target language adjective *verheiratet (married)* is assigned the part of speech noun (N).

Naturally, quantitative knowledge about translation shifts requires the availability

of large bilingual corpora that are annotated in a way that allows the compilation of statistics on the phenomena of interest. In this respect, the study of translation shifts is dependent on the ability of linguistics to provide reliable and informative corpus annotations. In practice, this becomes more difficult with "deeper" annotation. While there are automatic part of speech taggers that can process large corpora with high accuracy, the study of grammatical contrasts by Čulo *et al.* (2008) has relied on manual annotation of syntactic chunks and top-level grammatical functions.

The situation appears to be even more dire for semantic shifts. Cyrus (2006) has proposed to build a German-English corpus of translation shifts at the predicate-argument level. She found it necessary to use bilingual human annotators to mark different types of shifts. This reliance on manual analysis means that the creation of a translation corpus of reasonable size requires annotators with a rather specific qualification, takes a long time, and is expensive.

In this paper, we investigate an alternative idea: instead of annotating semantic shifts explicitly, we test to what extent the strategy of using generic parallel corpora with (monolingual) semantic annotation carries over from the domain of grammatical shifts onto semantic shifts. This idea, should it work, would have significant benefits. In particular, it would allow the study of semantic shifts to draw on the pool of parallel corpora constructed in computational linguistics over the last years. It would also allow easier creation of new corpora (e.g., for new language pairs) because each new language could be annotated individually.

Structure of the paper. Section 2 gives a short introduction to Frame Semantics and states our hypotheses about the relationship between frame semantics and translation shifts. Section 3 describes the corpus with frame-semantic annotation that we use for our analysis. Section 4 provides a detailed analysis, classification, and discussion of the results. We conclude with a general discussion in Section 5.

2 Semantic Role Annotation

2.1 Frame Semantics and FrameNet

Frame Semantics (Fillmore, 1985) is a usage-based theory of meaning which focuses on the role of conceptual background knowledge in comprehension. For example, Frame Semantics argues that a hearer cannot understand the predicate *promise* without being familiar with the typical "commitment" situation in which one party makes a pledge to a second party to follow some future course of action. Such situations are represented in terms of so-called *frames*. On the linguistic level, a frame is a semantic class containing all predicates which are capable of expressing the situation in question. Occurrences of these predicates in text are said to *evoke* the frame; thus, the predicates are called *frame-evoking elements (FEEs)*. Each frame also specifies the "participants and props" of the situation it describes, the so-called *frame elements (FEs)*. Frame elements are the Frame Semantics instantiation of semantic roles. Frame Semantics assumes that

Frame: COMMITMENT		
Def.	A Speaker makes a commitment to an Addressee to carry out some future action. This may be an action desirable (as with promise) or not desirable (as with threaten) to the Addressee.	
Frame Elements	SPEAKER	The SPEAKER is the person who commits him/herself to do something.
	ADDRESSEE	The SPEAKER's commitment can be made to an ADDRESSEE.
	MESSAGE	An expression of the commitment made by the SPEAKER.
	TOPIC	The topic about which the SPEAKER makes a promise.
	MEDIUM	The MEDIUM is the physical entity or channel used to transmit the MESSAGE.
FEEs	consent.v, covenant.n, covenant.v, oath.n, vow.n, pledge.n, pledge.v, promise.n, promise.v, swear.v, threat.n, threaten.v, undertake.v	
Ex.	[Democratic audiences]$_{\text{Speaker}}$ had to **consent** [to this approach]$_{\text{Message}}$.	
	[The politicians]$_{\text{Speaker}}$ made vague **promises** [about independence]$_{\text{Topic}}$.	
	["I'll be back , "]$_{\text{Message}}$ [he]$_{\text{Speaker}}$ **threatened**.	

Table 1: Frame COMMITMENT: Definition and annotated example sentences.

under normal circumstances, there is a direct parallelism between the conceptual and the linguistic level: the participants of the situations are realisable as arguments of the frame-evoking elements.

The Berkeley FrameNet project (Baker *et al.*, 1998; Fillmore *et al.*, 2003) has been compiling a Frame Semantics-based semantic lexicon for English. It it still under development; its current release (1.3) contains about 800 frames and 10,000 lexical items. The lexicon is available from `http://framenet.icsi.berkeley.edu/~framenet`. Each frame is provided with a definition in natural language; these often list properties of the described situation and presuppositions of the situation and the frame elements. Membership of a predicate in a frame is determined mainly on two grounds: (a), the predicate must describe the frame and share its semantic properties and (b), it must be able to realise all of the frame's roles; see Fillmore (2002) for details. Table 1 shows the FrameNet entry for the COMMITMENT frame referred to above.

The conceptual definition of FrameNet frames results in a considerable degree of *cross-lingual interpretability* of the frames. Being defined mainly on the conceptual level, they are often appropriate to describe not only English predicate-argument structures, but also those of other languages. In fact, frames have been proposed as largely language-independent meaning representations (Boas, 2005). Several projects investigate the use of English FrameNet frames for the semantic analysis of other languages, usually coupled with the annotation of corpora for the new languages (e.g., Subirats and Petruck, 2003; Burchardt *et al.*, 2006).

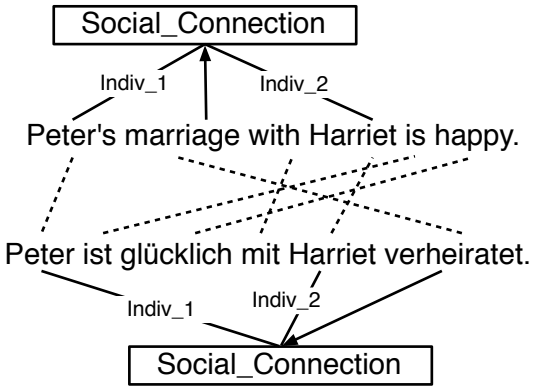

Figure 2: Annotation projection for frame-semantic annotation

2.2 Frame-semantic annotation and translation shifts

Figure 2 shows the bilingual sentence pair from Figure 1 with frame-semantic annotation.[1] In the English sentence, the predicate *marriage* introduces the frame SOCIAL_CONNECTION. This frame has two roles, INDIVIDUAL_1 and INDIVIDUAL_2, which are realised in the sentence by *Peter* and *Harriet*, respectively.

Note that the annotation for the German sentence is completely parallel. This illustrates two important points. The first one is the cross-lingual interpretability of FrameNet frames and their frame elements. The second one is that the shift in part of speech between *marriage* and *verheiratet* does not lead to change in annotation. In fact, frame-semantic analysis is, through the way in which frames are assigned to predicates (cf. Section 2.1), invariant to the phenomena subsumed under grammatical shifts. In particular, passivized verbs receive the same frame-semantic analysis as verbs in the active voice; pronouns are annotated with the same semantic roles that full noun phrases are; number change does not have an impact either.

Semantic shifts clearly have an impact on frame-semantic analysis. This seems plausible from the fact that the frames themselves can be seen as *word senses* for predicates; thus, semantic modifications of predicates in translation are likely to lead to frame changes. The roles describe the *valency* of the predicate, and are in turn affected by semantic shifts such as deletion/insertion of arguments. Finally, straight deletion/insertion of predicates leads to frames that are present in one language but not on the other.

These considerations lead us to the main hypothesis that we investigate in this article, namely, that each semantic translation shift is accompanied by a cross-lingual *mismatch* in frame-semantic annotation, and conversely, that a cross-lingual mismatch in frame-semantic annotation indicates a semantic shift. This hypothesis would make frame-semantic annotation an excellent basis for the corpus-based study of semantic shifts. The rest of this article will test this hypothesis on corpus data.

3 A Parallel Corpus With Frame-Semantic Annotation

For our study, we use a parallel corpus that was constructed for a study on the cross-lingual projection of semantic roles (Padó and Lapata, 2009). The annotations are freely available for research purposes and can be downloaded from `http://www.nlpado.de/~sebastian/srl_data.html`. This section reviews the construction of the corpus.

The corpus consists of 1,000 English–German sentence pairs drawn from the current release of the EUROPARL corpus, Release 2. EUROPARL contains around seven years' worth (1996-2003) of professionally translated debates of the European Parliament (Koehn, 2005) in the 11 official languages of the European Union at that time. An automatic word alignment for the English-German bitext was computed using the publicly available GIZA++ software (Och and Ney, 2003). From this bitext, 1,000 sentence pairs were drawn. Since FrameNet is as yet incomplete, sentences that are sampled completely at random may contain predicates or readings of predicates that are not covered in the current FrameNet frame inventory. To avoid this issue, sentences were chosen subject to the following constraint: The English side of each sentence pair had to contain a predicate covered by the English FrameNet lexicon (release 1.2) that was word-aligned with a German predicate covered by SALSA, a FrameNet-based corpus and lexicon resource for German (Burchardt *et al.*, 2006). In this way, all sentence pairs in the sample could be annotated reliably. Note that due to the objectives of the original study, no attention was paid to the source language: sentences could be translations from English, translations from German, or translations from any other language.

Annotation procedure. The bitext was extended with frame-semantic annotation. In each sentence pair, we annotated the pair of predicates that led to the sentence being included in the sample (cf. above). Annotation was performed manually by two undergraduate students, each of which had native-level proficiency in both German and English. For every predicate, the annotation task involved two steps: (a) selecting a single appropriate frame (which amounts to word sense annotation) and (b) assigning the semantic roles the frame instantiates to the constituents of the sentence. Annotators were provided with detailed guidelines, available in the appendix of Padó (2007). The annotation proceeded in three phases: a training phase (40 sentence pairs), a calibration phase (100 sentence pairs), and the annotation of the main dataset (1,000 sentence pairs). During training, annotators were acquainted with the annotation style and received feedback on their annotation decisions. In the calibration phase, each sentence pair was doubly annotated to assess the inter-annotator agreement. Finally, in the main annotation phase, each of the 1,000 sentence pairs in the main dataset was split and each half randomly assigned to one of the coders for single annotation. This is a crucial safeguard, ensuring that no annotator saw both parts of any sentence pair in order to guarantee independent annotation of the two halves of each sentence pair. Annotation proceeded predicate-wise to encourage consistency; each judge annotated approximately the same amount of data in English and German.

Measure	English	German
Frame Match	0.90	0.87
Role Match	0.95	0.95
Span Match	0.85	0.83

Table 2: Monolingual inter-annotator agreement on the calibration set

Inter-annotator agreement. Recall that our hypothesis from Section 2.2 assumes a direct correspondence between frame mismatches and semantic shifts. Clearly, this hypothesis is only viable if frame-semantic annotations can be assigned *within* the individual languages with high reliability. To test this property, we estimated inter-annotator agreement on the calibration set, shown in Table 2. We computed the ratio of frames common between two sentences (Frame Match), the ratio of common roles (Role Match), and the ratio of roles with identical spans (Span Match). As can be seen from the table, annotators generally agree well, although not perfectly, for the frame-semantic annotation task. Agreement is highest for the decision of what roles to assign (0.95). Agreeing on exact role spans is a harder problem, since it involves additional syntactic decisions. Disagreements in frame assignment are mainly due to difficult distinctions between related frames; we will come back to this issue in Section 4.4.

4 Analysis of the Parallel Corpus

As far as we know, our corpus is the only available parallel corpus with semantic role annotation to date.[2] Our corpus it not big enough to obtain quantitative tendencies in translation such as those obtained from the CroCo corpus by Čulo et al. (this issue). In contrast, it provides *deeper information* than that available from, e.g., CroCo. It is also large enough for performing an empirically grounded analysis of frame-semantic mismatch cases in the corpus, and to obtain first frequency estimates for the different classes. We see our study as a pilot for annotation-driven investigation of translation shifts at the semantic level, a complementary goal to that pursued, e.g., by Čulo et al..

Our analysis proceeded in two steps. First, we developed a hierarchical classification schema for sentence pairs. In the second step, the two authors revisited each sentence pair from the corpus and assigned it to one of the categories, to obtain frequency estimates for all classes. We also identified instructive examples which demonstrate common phenomena and difficulties for annotation.

4.1 A hierarchical classification scheme for sentence pairs

Figure 3 shows our hierarchical classification scheme. The design of the scheme was driven by the requirements of corpus-based investigations of translation shifts. For each sentence pair with its two frames and two sets of annotated roles, we ask a series of questions that are designed to determine the status of this sentence with respect to two crucial issues:

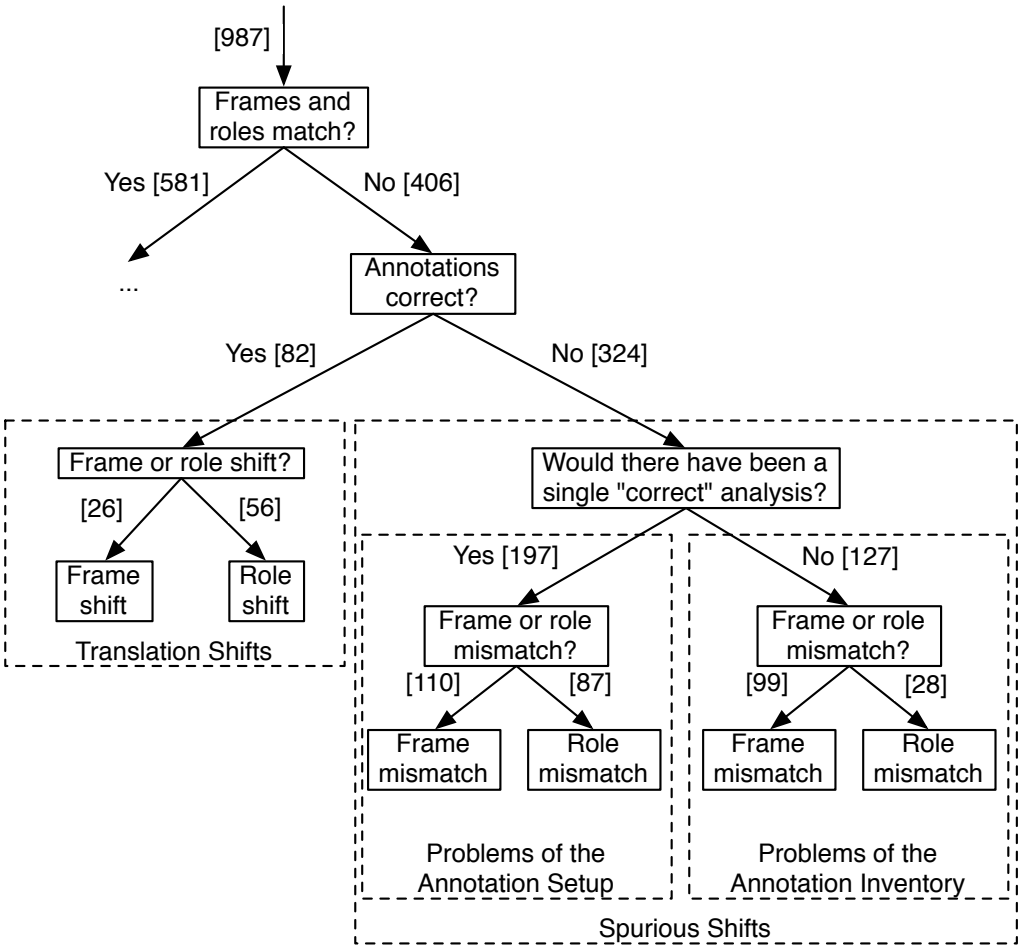

Figure 3: Classification schema for bilingual frame-semantic annotation in a parallel corpus (number of sentences in each class shown in square brackets on edge labels)

- The first one is our "main hypothesis" from Section 2.2: Can cross-lingual annotation mismatches be interpreted as translation shifts?
- The second one deals with the situation when the hypothesis fails: If the mismatch is not due to a translation shift, what is the source of the "spurious" mismatch? Can the annotation be amended?

The obvious first question we ask is whether the frame/role annotation match between English and German, as in Figure 2, or not. Since this article is concerned primarily with mismatches, we ignore the matches here; see Padó and Lapata (2009) for a discussion. Note we count two frame-semantic annotations as matching if they are *semantically* identical, that is, if they involve the same frame and the same set of overtly realized roles. We do not take into account whether these roles are realized syntactically in the same way. For example, in Figure 2, INDIV_1 is realized as a genitive specifier of a frame-evoking noun in English, while it is the subject of an auxiliary governing the frame-evoking predicate in German. Nevertheless, following our line of reasoning in

Section 2.2, we see this case as a match.

We next ask whether the two annotations (for the two languages) are correct – that is, whether they are appropriate for the respective monolingual sentences. If this is the case, the sentence pair exhibits a legitimate semantic shift. Examples of such shifts are shown below in Section 4.3.

If the annotations are however not correct, or their correctness is unclear, we are faced with sentence pairs whose annotation mismatches cannot be interpreted, a priori, as semantic shifts – in other words, as *spurious shifts*, or false positives in the sense of our main hypothesis. This class forms the big box on the right-hand side of Figure 3. This class is discussed in Section 4.4.[3]

For spurious shifts, we further ask whether there is a single, clearly correct annotation for both sentences in terms of the FrameNet inventory. This question allows us to distinguish two classes of differing severity. Cases which have a correct analysis, which has however not been annotated, are indicative of problems of our specific *annotation setup*, such as shortcomings in the annotation guidelines or annotator oversight. In these cases, a better controlled annotation setup would have a good chance of assigning correct annotations. The second class consists of cases where we felt that it was already problematic to decide on a single correct annotation for each monolingual sentence. These cases involve more fundamental issues of *characterizing* word sense and semantic roles which we will discuss below.

Finally, we subdivide each of the three main classes – translation shifts, problems due to the annotation setup and problems due to the annotation inventory – into two subclasses: one for shifts at the frame level, and one for shifts at the role level. These two subclasses differ substantially in their linguistic as well as their translational nature. Recall from Section 2.2 that frames can be seen as word senses or abstract concepts. Thus, in the case of true shifts, frame mismatches indicate *conceptualisations* during translation, or "modifications/mutations" according to Cyrus' (2006) taxonomy of semantic shifts. In the case of annotation errors, frame mismatches indicate at least that word sense disambiguation of the two predicates is difficult. In contrast, semantic roles describe the props and participants of the predicates. Role mismatches that represent true semantic shifts therefore correspond to *valency changes*, where arguments are added, dropped, or change their semantic nature ("addition" and "deletion" in Cyrus' (2006) classification). Role mismatches that are annotation errors indicate problems in recognizing predicates' argument, or in distinguishing between different types of arguments.

4.2 Frequencies of Classes

The number of sentences in each class is also shown in Figure 3, attached in square brackets to each label. We first note that out of a total of 987 annotated sentence pairs, 581 (59%) have matching frame-semantic annotation, while 406 (41%) do not. The high number of matching annotations across languages provides practical evidence for our assumption from Section 2.2 that FrameNet annotations generalize over a large number of translation shifts, at least for related languages like English and German (specifically, grammatical shifts).

Out of the 406 mismatches, we found 82, or roughly 20%, to be actual semantic shifts. The whole rest – 324 or about 80% – consists of spurious shifts. This appears to be a strong negative result with regard to our initial hypothesis that frame-semantic mismatches are interpretable as semantic shifts. However, the news is not all bad. We classified more than 60% of the spurious shifts (197 of them) as stemming from problems related to the annotation setup. We have argued above that these instances may be annotated properly in future by improving annotation guidelines; see the next section for concrete examples which support this claim.

It is noticeable that the ratio between frame-related mismatches and role-related mismatches varies considerably between the three major classes. Among true semantic shifts, we see only half as many frame shifts than role shifts, indicating that at least for the present language pairs, few translations involve proper conceptualisation's, while a large number exhibits changes at the level of argument realization. In the second class, annotation setup-related problems, frame mismatches outnumber role mismatches, but only by 20%. In the final class, semantic characterization-related problems, we find over three times more frame mismatches than role mismatches. In sum, 33% of all role mismatches correspond to true translation shifts, but only 11% of the frame mismatches do. These numbers, which correspond to the patterns we found for the inter-annotator agreement in Section 3, give a clear indication that the main problem in semantic annotation is the frame assignment. However, even for frame assignment, being able to resolve the "practical" annotation setup-related problems would improve the ratio of true shifts substantially.

4.3 True Translation shifts

In the current and the following section, we show a number of examples from our corpus and discuss their frame-semantic analysis. Since we cannot provide definitions for all FrameNet frames, we refer the reader to the FrameNet lexicon which is available online at `http://framenet.icsi.berkeley.edu`.

Examples are presented in three lines, with the German corpus sentence (marked with "G:") first, followed by an English gloss, and finally the English corpus sentence ("E:"). Sentences are annotated with their indices in the downloadable version of the parallel corpus (cf. Section 3).

Frame-level shifts. Frame-level shifts are instances where two predicates which form a translation pair nevertheless introduce two different frames. Among these instances, we did not see any cases where the shift was "mandatory" in the sense that it would not have been possible to find a translation that would have preserved the frame. Most of the frame-level shifts were cross-lingual changes in perspective used to describe some event. Consider Ex. (1):

(1) G: Man muss auch lernen, [...] zu **schauen**, was sich dahinter verbirgt.
(PERCEPTION_ACTIVE) [s231]

One must also learn [...] to **watch** what itself behind hides.
E: One must also learn [...] to **look for** the hidden meaning (SEEKING)

German uses the verb *schauen* (*to watch*) which refers to general cognition and is analysed with the frame PERCEPTION_ACTIVE. The English translation *look for* specifically emphasizes the active search component and is analysed with the SEEKING frame.

Changes of perspective can also arise through the influence of arguments on the meaning of their predicates, as in the following example:

(2) G: [... wir] dürfen unsere in der Vergangenheit gewonnenen Erfahrungen, die wir so teuer **bezahlt** haben, nicht ignorieren. (COMMERCE_PAY) [s404]
[... we] may our in the past gained experiences, which we so dearly **paid** have, not ignore.
E: [...] we should not disregard the experience that we have **gained** from the past at a very heavy **price**. (COMMERCE)

Here, English uses a rather generic predicate, *to gain*, which could in principle even be analysed with the GETTING frame. It is the *at*-PP which indicates an 'exchange'-oriented reading of *gain*, and leads to its analysis with the COMMERCE frame. In the German version, this specification has been taken another step further: The argument *price* has been incorporated into the predicate, so that *to gain [something] at a price* is conceptualised as *paying* for it, and is labeled with the frame COMMERCE_PAY.

Role-level shifts. Role-level shifts are translation shifts where a two predicates which form a translation pair evoke the same frame but realise different sets of semantic roles. In contrast to the group of frame-level shifts, here we found a considerable number of "mandatory" shifts that could not be avoided in translation. These cases arise from constructions in one language (mostly German) that are not available in the other language. An example is the German Dative grammatical case, as in the following role deletion/addition example:

(3) G: Ich bin [... ihm] eine Erklärung schuldig, die ich nun [dem ganzen Parlament]$_{\text{Recipient}}$ **geben** will. (GIVING) [s10]
I am [...] him an explanation owing, which I now [the whole parliament-DAT]$_{\text{Recipient}}$ **give** will.
E: I owe [... him] an explanation, and I am going to **give** this explanation [before the whole House]$_{\text{NONE}}$. (GIVING)

Both English *give* and German *geben* (*give*) are assigned the frame GIVING. The difference arises for the realization of the RECIPIENT frame element. In German, the availability of morphologically marked accusative and dative case makes it possible to realize both the THEME (the explanation) and the RECIPIENT as NP arguments of the predicate, with the Recipient receiving Dative case. The use of the corresponding English double object construction is much more constrained, and in particular long NPs (like the one in this example) are strongly dispreferred as indirect objects (Bresnan *et al.*,

2007). Instead, the English sentence uses a *before*-PP which functions as an adjunct, and therefore does not fill any role.

A second group of "mandatory" shifts arises from subordinating conjunctions in German that do not have a direct correspondence in English, such as in this example:

(4) G: Der letzte Vorfall ereignete sich am Montag, 19. Mai, an dem sieben spanische Lkw [...] verbrannt wurden, [wobei]$_{\text{Cause}}$ außerdem ein spanischer Lkw-Fahrer beträchtlich **verletzt** wurde. (CAUSE_HARM) [s535]

The last incident happened itself on Monday 19 May, on which seven Spanish lorries burnt were, [in-the-course-of-which]$_{\text{Cause}}$ also a Spanish lorry-driver seriously **injured** was.

E: In the most recent incident on Monday, 19th May they [...] set fire to seven Spanish lorries [...], and a Spanish lorry driver was seriously **injured**. (CAUSE_HARM)

Subordinating conjunctions like *wobei (in the course of which)* or *wodurch (by/through which)* are widely used devices in German texts to subordinate one sentence to another while establishing a quite specific discourse relation between them. Within the subordinate sentence, they are available as role-bearing elements that refer to the superordinate sentences (here, *wobei* fills the CAUSE role of the CAUSE_HARM frame). In English, these specific subordinating conjunctions have to be expressed phrasally, or remain unexpressed as in our example, which leads to a role deletion.

Of course, not all shifts arise from constructions that are not directly translatable. In the following example, both languages express a causal connection through a construction that would have been available in the respective other language as well. Nevertheless, they make different choices in doing so:

(5) G: [Schon wenn sie ihr Gesicht oder ihr Haar nicht ordnungsgemäß verhüllen,]$_{\text{NONE}}$ werden sie auf der Straße **geschlagen**. (CAUSE_HARM) [s898]

[Already when they their face or their hair not correctly cover,]$_{\text{NONE}}$ are they on the street **beaten**.

E: They can be **beaten** in the street [simply for not covering their faces and hair correctly.]$_{\text{Cause}}$

The English sentence realizes the causal connection in a *for*-PP, which fills the CAUSE role of CAUSE_HARM. The German version uses the discourse connective *wenn* (*when*) in a subordinate clause. Since *when*-Phrases are primarily temporal specifications and not causal ones, the clause is not assigned a semantic role.

Finally, the following instance shows of a role modification that arises through constructional choice.

(6) G: Ich hatte auch eine spezifische **Frage** gestellt, [nämlich welche Qualifikation die Angestellten von Fluggesellschaften und auf Flugplätzen haben]$_{\text{Message}}$. (QUESTIONING) [s134]

I had also a specific **question** asked, [namely which qualification the employees of airlines and on airports have]$_{\text{Message}}$.

E: I also asked a specific **question** [concerning the powers of airline and airport employees]$_{\text{Topic}}$. (QUESTIONING)

Both sentences evoke the QUESTIONING frames, which provides two frame elements describing the content of the question: MESSAGE (for indirect speech), and TOPIC (for higher-level descriptions of the content). The English choice of a verb phrase – *concerning the powers...* – is best analysed as a TOPIC, while the German version introduces a full subordinate clause, which fills the MESSAGE role.

4.4 Spurious translation shifts

4.4.1 Cases with a clear correct annotation

This category comprises 197 cases (110 frame mismatches and 87 role mismatches). The majority of role mismatches was due to roles missed by annotators, which frequently involved long-distance dependencies (e.g., control constructions or extrapositions), such as in the following example, where the German DONOR was not identified:

(7) G: Jetzt einen Vertrauensvorschuss zu **geben**, wird [vielen]$_{\text{MISSED}}$ schwerfallen. (GIVING) [s784]
Now an advance-of-confidence to **give**, will [many-DAT]$_{\text{MISSED}}$ find-difficult.
E: [Many people]$_{\text{Donor}}$ will find it difficult to **give** a vote of confidence at this juncture. (GIVING)

There was also a number of erroneously annotated roles, mostly adjuncts that were mistaken as "proper" semantic arguments in situations similar to Ex. (5) above. Since almost all of these distinctions are fairly uncontroversial and can be made on the basis of syntactic information, we attribute these errors to shortcomings of the training received by the annotators. The guidelines that were made available to the annotators only contained abstract characterizations of semantic roles, and were short on concrete examples. In particular, they did not contain contrastive examples for "good" and "bad" role annotations. The 100-sentence training phase was apparently too short for annotators to develop a robust intuition on these points as well.

On the frame level, we identified two major problems. The first one was metaphor annotation:

(8) G: Es geht jedoch um mehr. Dazu werde ich gleich **kommen**. (ARRIVING) [s60]
It goes however about more. To-this will I soon **come**.
E: But there is more than that, which I will **come** to in a moment. (STATEMENT)

In frame semantics, metaphors can be analysed on two layers. The annotator can either use a frame that describes the literal meaning of the expression in question – for *kommen (come)*, that would be the frame ARRIVING (coming to a place). Alternatively, a frame describing the actually conveyed meaning can be annotated (STATEMENT for **kommen**). Burchardt *et al.* (2006) call these two annotation choices the *source* and *target frame*,

and note that while the source frame annotation is always available, it is not always possible to find a frame that would describe the metaphorical meaning. In cross-lingual annotation, mismatches can arise when the metaphor is the *same* in both languages, but one annotator chooses to annotate the source, the other the target frame, as was the case in Ex. (8).

An analogous problem arises with multi-word expressions, as exemplified in Ex. (9).

(9) G: Ich wiederhole nochmals, wir können verstehen, [...] dass unsere ungarischen [...] Freunde von uns Signale erwarten, ohne die wir unsere Glaubwürdigkeit in **Frage** stellen. (QUESTIONING) [s100]
I repeat again, we can understand [...] that our Hungarian [...] friends from us signals expect, without which we our credibility into **question** put.
E: Once again, we can understand [...] our Hungarian [...] friends asking us for signs. Without these, we shall be calling our credibility into **question**. (CERTAINTY)

The multi-word expression is virtually the same in both languages: *call something into question/put something into question* in the sense of expressing uncertainty about something. However, the German annotator chose to annotate *question* with the frame QUESTIONING, which is appropriate for that lemma in isolation. In contrast, the English annotator characterized the meaning of the complete multi-word expression through the frame CERTAINTY.

The annotation guidelines specified that for both metaphors and multi-word expressions, the understood/"target" reading should be annotated whenever there was an appropriate FrameNet frame. We draw two conclusions from the fact that annotators still tended to choose frames for the literal reading even where a frame for the understood reading would have been available. First, the annotation of non-literal meaning is a subtle task that annotators should ideally receive specific training for, for example by re-annotating a corpus of such cases. Second, annotators need to be familiar with several hundred frames from FrameNet to decide whether there is an appropriate frame for a particular target reading of a metaphor or multi-word expression. This also means that annotators have to be trained extensively on a range of lexical items before being able to annotate data in "production mode" quality.

An alternative strategy that would simplify the treatment of metaphors and multi-word expressions is the consistent annotation of the literal/"source" reading. However, from the point of view of translation studies, this annotation strategy is also problematic. It results in false positives (mismatches that do not correspond to semantic shifts) for metaphors that are translated into non-metaphoric expressions.

4.4.2 Cases with an unclear correct annotation

The final category is made up of sentences where a frame or role mismatch incorrectly signals a translation shift, but where we were unable to decide on a single, indisputably correct annotation. In the case of frames, the two frames assigned in the two languages often corresponded to *complementary meaning facets*, such as in the following example:

(10) G: Für mich ist es unerheblich, ob es ein Serbe, ein Roma, ein Bosnier oder ein Albaner ist, der im Kosovo **bedroht** oder getötet wird.(ENDANGERING) [s6]
For me is it irrelevant, whether it a Serb, a Roma, a Bosnian or an Albanian is, who in Kosovo **threatened** or killed is.
E: It is irrelevant, as far as I am concerned, as to whether it is Serbs, Roma, Bosnians or Albanians who are **threatened** or killed in Kosovo. (COMMITMENT)

The English *threatened* and German *bedroht* are literal translations. The German annotator chose the frame ENDANGERING which describes a situation where "an Agent or Cause is responsible for placing a Valued_entity at risk". The annotator of the English sentence decided for the frame COMMITMENT, according to which "a Speaker makes a commitment to an Addressee to carry out some future action". Both of these annotations are appropriate: each frame describes a meaning facet that is clearly entailed by the sentences, but that is not entailed by the respective other frame.

In the following case, the discourse has apparently influenced annotation:

(11) G: Wir **erwarten** konkrete Vorschläge vom Kommissar. (REQUEST) [s254]
We **await** concrete proposals from-the Commissioner.
E: We **await** concrete proposals from the Commissioner. (EXPECTATION)

Again, the English *await* is a fairly literal translation of German *erwarten*, but the annotators have chosen different frames. English *await* was assigned the frame EXPECTATION, which is about a cognizer's beliefs about future events, while the German annotator chose REQUEST, a frame in which a speaker asks an addressee for something. While EXPECTATION corresponds more closely to the literal meaning of *await*, the German annotation takes the role of this sentence in the discourse into account: the speech act of explicitly stating an expectation practically corresponds to a REQUEST situation. The use of REQUEST can also be justified on the role level: the *from/von*-PPs realize an addressee who is supposed to play an agentive role in bringing about the future events. Such an addressee/agent is not included in the EXPECTATION frame, but it is in the REQUEST frame.

At first glance, it might seem that these problem arise from insufficient context to distinguish between frames, or at least from the FrameNet sense inventory. If that were true, the situation could be improved by adopting a different sense inventory. However, this does not seem to be true. Similar problems have been reported for WordNet (see Palmer *et al.* (2007) for an overview), probably the most widely used inventory for word senses, even though it uses sense distinction criteria that are considerably different from FrameNet. Furthermore, a number of studies agrees that classifying word usage into a set of exhaustive and mutually exclusive categories is generally hard, or even impossible. Hanks (2000) suggests that word senses may better be viewed as prototypes with graded membership. Cruse (2000) states that "in principle word meaning may be regarded as infinitely variable" before setting out to study "regions of higher semantic 'density"'. Kintsch (2007) calls word meaning "fluid and flexible" and suggests that there may not be a clear best number of dictionary senses for a word. Erk *et al.* (2009) have conducted

a study where annotators assigned graded ratings to multiple WordNet senses instead of choosing a single sense for each instance in context. They indeed found that annotators made use of the entire graded scale and frequently assigned more than one sense.

The 28 instances of unclear annotations also involved vague or metonymically related role pairs:

(12) G: Aber wir müssen auch sehen, dass [neue Technologie]$_{\text{Cause}}$ die Lebensmittelsicherheit **gefährden** können. (ENDANGERING) [s447]
But we must also see that [new technologies]$_{\text{Cause}}$ the food-safety **endanger** can.
E: It must be clear to us as well, though, that [new technologies]$_{\text{Agent}}$ can endanger food safety. (ENDANGERING)

Both annotators chose the ENDANGERING frame, but the subject phrase *new technologies* was analysed differently. The German annotator chose a CAUSE annotation, in line with ontological criteria that define CAUSES as inanimate, while the English annotator assigned the AGENT label. This can be justified either by interpreting *technologies* as metonymically standing in for inventors or technology companies, or by attributing come level of volitionality to the technologies themselves. At any rate, there appears to be a broad "grey area" between clear CAUSES (like *gravity*) and clear AGENTS (like *milkman*) that includes, for examples, animals (*dog*), companies (*Microsoft*), aggregates of agents (*the crowd*). The same true is for a number of other role pairs, like the SPEAKER of a communication event and its MEDIUM, i.e. the channel through which communication occurs: Here, any written message that is realized as the subject of a communication event (*the* ***report/study/minutes*** *state that...*) is difficult to assign to one or the other, since it carries features of both.

Similar to the frame level, the problems in assigning FrameNet roles are only manifestations of a more general issue, this time the problem of defining consistent criteria for distinguishing (any) set of semantic roles. Given the lack of success in this field, Dowty (1991) proposed viewing semantic roles as prototypes, in a similar manner to Hanks' suggestion of viewing word meanings as prototypes. Dowty defines a Proto-Agent and a Proto-Patient role, each as bundles of entailments, which makes it possible that an occurrence of an argument can be assigned both labels, to different degrees. Petukhova and Bunt (2008) extend this approach to an entailment-based definition of further roles, and Gotsoulia (2008) proposes to annotate arguments directly with fine-grained entailments instead of semantic roles. To our knowledge, none of these schemes have however been used in practice to produce "graded" role annotations.

5 Conclusions

In this article, we have considered the investigation of translational semantic shifts in parallel corpora. Our proposal was to identify semantic shifts as cross-lingual mismatches between the frame-semantic annotations for each side in a bilingual sentence pair. We have tested this proposal on a 1000-sentence English–German parallel corpus sampled

from EUROPARL and annotated with word senses (frames) and frame-semantic roles. To analyse the patterns in the annotation, we have developed a hierarchical classification schema, shown in Figure 3. Application of the schema to the corpus has shown a mixed result with regard to our original proposal. While we were able to determine that roughly 3% of the sentences involved semantic shifts in the form of reconceptualisations, and 6% shifts with respect to arguments, we also found a very high number of mismatches in sentences with annotation errors, due to which the mismatches cannot be interpreted confidently as translation shifts.[4]

We have argued that annotation mismatches that do not represent semantic shifts can be further divided into two subclasses. The first class consists of annotation errors due to the annotation setup, in particular problems with the consistent annotation of non-literal language and multi-word expressions, as well as lack of familiarity with the large inventory of frames offered by FrameNet. We believe that the size of this class can be reduced greatly by refining the annotation guidelines and training the annotators more extensively.

The second class is made up of instances that present a more fundamental issue, namely the impossibility of assigning a single semantic annotation to each instance. In instances of this type, frames or roles are either systematically related (by a common vague area, or through metonymy), or meaning facets of more than one category are activated by the sentence when considered in context. We have argued that this is neither a problem specific to FrameNet, nor a problem specific to bilingual annotation, but a more fundamental issue that arises in exhaustive meaning annotation with some fixed set of semantic categories that are supposed to be exhaustive and mutually exclusive.

This might seem a rather negative conclusion for our study. However, we think that a clear view of the limitations of categorical annotation can actually be an interesting starting point to explore novel ways of characterizing meaning in a cross-lingual setting. We conclude this paper with outlines of two avenues of research in this direction.

The first direction is the characterization of translation shifts in terms of *graded* semantic annotation. As outlined in Section 4.4.2, the semantic literature has proposed deconstructing categories (both word senses-type categories like frames, and argument class-type categories like semantic roles) into prototypes. In prototype annotation, more than one category can be annotated to an instance, and each categories can be assigned a degree to which it applies. This style of annotation is an attractive proposition since it sidesteps the issues of categorization and results in a richer annotation (Erk *et al.*, 2009). It also seems appropriate for the semantic translation shifts, where Cyrus' (2006) study in explicit shift annotation found very similar problems to us, noting that "It is rather difficult to find objective criteria [...]". However, the formulation of a graded version of the hypothesis from Section 2.2, to relate the *degree* of semantic divergence to the presence of a semantic shift, raises a number of questions: Should the presence of shifts still be treated as binary, or should there be a graded notion of "mild" and "strong" shifts? Can such a notion be operationalised? Do all mismatches contribute equally towards shifts, or are some mismatches more serious than others? How can graded annotation be done efficiently and reliably?

The second direction aims at exploiting mismatches in categorical semantic anno-

tation to gain a better understanding of gradedness and disjointness in word senses.[5] Interestingly, this type of study can use both mismatches from true translation shifts and the "ill-behaved" mismatch instances arising from annotation difficulties, since both reflect the fact that the word uses in question cannot be described by just one semantic category. In Ex. (1), the translation shift in the verb (*schauen/look at* versus *look for*) points to different meaning facets in the situation described. In Ex. (10) and Ex. (11), it was exactly the mismatching frame annotations that reveal the presence of multiple meaning facets in *endanger/expect* in the given context. The role of translation in this type of inquiry is to make meaning components explicit that might not have been overt in the original language, and thus provide a richer meaning characterization.

In another study (Padó and Erk, 2005), we have exploited this idea for a systematic study of meaning facets in a parallel corpus through translational shifts. We collected and analysed occurrences of the English transitive verb *to increase* aligned with the German adjective *höher* (*higher*). Since the transitive use of *increase* involves an actor, which cannot be realized by the adjectival *höher*, some shift must take place in order to accommodate the agent in some other fashion. We found that *X increase Y* was almost invariably translated as *X verb [ein höheres Y]$_{NP}$* (*X verb [a higher Y]$_{NP}$*), with a range of verbs that showed a gradient of causation from direct causation to a vague "causal connection" expressed by constructions like *verbunden sein mit* (*be related/connected to*). Our results suggest that such translation shifts could be used for the automatic identification of paraphrases.

Notes

[1]This example is simplified. A complete analysis of the sentence would also assign frames to *happy* and *glücklich*.

[2]Similar corpora are under development by Cyrus (2006) and Volk and Samuelsson (2007).

[3]Note that the class of spurious shifts can still contain sentence pairs that in fact contain translation shifts, which are obscured by invalid annotation. However, we call them spurious in the sense that we cannot rely on their frame-semantic analysis, which makes it impossible to distinguish them from annotation errors by anything short of reannotation.

[4]Our method is currently limited to parallel corpora, which is regrettable since comparable corpora often contain more natural language than parallel corpora. However, we think that our analysis can be generalized to comparable corpora by comparing quantitative tendencies of frame and role realizations across larger "close" comparable corpora – e.g., collections of newswire articles which report on the same events.

[5]In this context, note also the two cross-lingual tasks at SemEval 2010, `http://semeval2.fbk.eu`: cross-lingual lexical substitution, and cross-lingual word sense disambiguation.

References

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The Berkeley FrameNet project. In *Proceedings of the joint Annual Meeting of the Association for Computational Linguistics and International Conference on Computational Linguistics*, pages 86–90, Montreal, QC.

Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers, editor, *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*, pages 175–186. Benjamins.

Bakker, M., Koster, C., and van Leuwen-Zwart, K. (1998). Shifts of translation. In M. Baker, editor, *Encyclopedia of Translation Studies*, pages 226–231. Routledge.

Bentivogli, L. and Pianta, E. (2005). Exploiting parallel texts in the creation of multilingual semantically annotated resources: the MultiSemCor Corpus. *Journal of Natural Language Engineering*, **11**(3), 247–261.

Boas, H. C. (2005). Semantic frames as interlingual representations for multilingual lexical databases. *International Journal of Lexicography*, **18**(4), 445–478.

Bresnan, J., Cueni, A., Nikitina, T., and Baayen, H. (2007). Predicting the dative alternation. In G. Bouma, I. Krämer, and J. Zwarts, editors, *Cognitive Foundations of Interpretation*, pages 69–94. Royal Netherlands Academy of Science, Amsterdam.

Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S., and Pinkal, M. (2006). The SALSA corpus: a German corpus resource for lexical semantics. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, Genoa, Italy.

Catford, J. C. (1965). *A Linguistic Theory of Translation: An Essay in Applied Linguistics*. Oxford University Press.

Collins, M., Koehn, P., and Kučerová, I. (2005). Clause restructuring for statistical machine translation. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 531–540, Ann Arbor, MI.

Cruse, D. A. (2000). Aspects of the microstructure of word meanings. In Y. Ravin and C. Leacock, editors, *Polysemy: Theoretical and Computational Approaches*, pages 30–51. Oxford University Press, Oxford, UK.

Čulo, O., Hansen-Schirra, S., Neumann, S., and Vela, M. (2008). Empirical studies on language contrast using the English-German comparable and parallel CroCo corpus. In *Proceedings of the Workshop Building and Using Comparable Corpora held at LREC*, Marrakech, Morocco.

Cyrus, L. (2006). Building a resource for studying translation shifts. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, Genoa, Italy.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, **67**, 547–619.

Erk, K., McCarthy, D., and Gaylord, N. (2009). Investigations on word senses and word usages. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 10–18, Singapore.

Fillmore, C. J. (1985). Frames and the semantics of understanding. *Quaderni di Semantica*, **IV**(2), 222–254.

Fillmore, C. J. (2002). FrameNet and the linking between semantic and syntactic relations. In *Proceedings of the 19th International Conference on Computational Linguistics*, pages 28–36, Taipei, Taiwan.

Fillmore, C. J., Johnson, C. R., and Petruck, M. R. (2003). Background to FrameNet. *International Journal of Lexicography*, **16**, 235–250.

Gotsoulia, V. (2008). An entailment-based approach to semantic role annotation. In *Proceedings of the Second Linguistic Annotation Workshop held in conjunction with LREC 2008*, Marrakech, Morocco.

Hanks, P. (2000). Do word meanings exist? *Computers and the Humanities*, **34**(1–2).

Kintsch, W. (2007). Meaning in context. In T. Landauer, D. McNamara, S. Dennis, and W. Kintsch, editors, *Handbook of Latent Semantic Analysis*, pages 89–105. Erlbaum, Mahwah, NJ.

Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. In *Proceedings of the MT Summit X*, Phuket, Thailand.

Och, F. J. and Ney, H. (2003). A systematic comparison of various statistical alignment models. *Computational Linguistics*, **29**(1), 19–52.

Padó, S. (2007). *Cross-lingual annotation projection models for role-semantic information*. Ph.D. thesis, Saarland University, Saarbrücken, Germany.

Padó, S. and Erk, K. (2005). To cause or not to cause: Cross-lingual semantic matching for paraphrase modelling. In *Proceedings of the EUROLAN Workshop on Cross-Linguistic Knowledge Induction*, Cluj-Napoca, Romania.

Padó, S. and Lapata, M. (2009). Cross-lingual annotation projection of semantic roles. *Journal of Artificial Intelligence Research*, **36**, 307–340.

Palmer, M., Dang, H. T., and Fellbaum, C. (2007). Making fine-grained and coarse-grained sense distinctions, both manually and automatically. *Natural Language Engineering*, **13**(2), 137–163.

Petukhova, V. and Bunt, H. (2008). LIRICS semantic role annotation: Design and evaluation of a set of data categories. In *Proceedings of the Sixth International Language Resources and Evaluation*, Marrakech, Morocco.

Subirats, C. and Petruck, M. R. L. (2003). Surprise! Spanish FrameNet! In *Proceedings of the Workshop on Frame Semantics, XVII. International Congress of Linguists*, Prague, Czech Republic.

Teich, E. (2003). *Cross-linguistic variation in system and text. A methodology for the investigation of translations and comparable texts*. Walter de Gruyter.

van Leuven-Zwart, K. M. (1989). Translation and original: Similarities and dissimilarities. *Target*, **1**(2), 151–181.

Vinay, J. and Darbelnet, J. (1958). *Stylistique comparée du français et de l'anglais*. Éditions Didier, Paris.

Volk, M. and Samuelsson, Y. (2007). Frame-semantic annotation on a parallel treebank. In *Proceedings of the NODALIDA Workshop on Building Frame Semantics Resources for Scandinavian and Baltic Languages*, Tartu, Estonia.

Yarowsky, D. and Ngai, G. (2001). Inducing multilingual POS taggers and NP bracketers via robust projection across aligned corpora. In *Proceedings of the 2nd Annual Meeting of the North American Chapter of the Association for Computational Linguistics*, pages 200–207, Pittsburgh, PA.