Hans C. Boas

A roadmap towards determining the universal status of semantic frames

Abstract: The Berkeley FrameNet project (http://framenet.icsi.berkeley.edu), founded in 1997, organizes the lexicon of English by semantic frames (Fillmore 1982), with valence information derived from attested, manually annotated corpus examples (Fillmore and Baker 2010). The resulting FrameNet database contains more than one thousand frames, together with more than twelve thousand lexical unites and close to 200,000 annotated example sentences. FrameNet data have been used to answer a variety of empirical research questions on the mapping from semantics to syntax and they have been employed in a number of NLP tasks such as role labeling and text summarization. Since the early 2000s, several projects have re-used the semantic frames based on English for constructing FrameNets for other languages, most notably Spanish (Subirats 2009), Japanese (Ohara 2009), German (Burchhardt et al. 2009), and Swedish (Borin et al. 2010, among others. While the tools, corpora, and databases differ from each other, the main organizing principle, the semantic frame, used for structuring the lexicon remains similar across all the FrameNets for different languages. The motivation for re-using semantic frames from English for other languages is the idea that frames are universal, similar to Fillmore's (1968) original case roles (Boas 2005a). However, there has not yet been any empirical investigation into what constitutes "universal" frames or how one can possibly determine the universal status of semantic frames. This paper proposes a systematic method for identifying semantic frames that could be labeled "universal" (based only on data from languages under investigation). We specifically address the question of how semantic frames can be used for contrastive analysis.

AU: As per guidlines the reference citation not allowed inside abstract. Please check whether this is fine.

Keywords: Frame Semantics, Contrastive Linguistics, FrameNet, Semantic Frames

1 Introduction

This paper investigates the nature of semantic frames as developed by Charles Fillmore during the 1970s and 1980s in order to determine their usability for contrastive linguistics. More specifically, this paper discusses how semantic

Hans C. Boas, The University of Texas at Austin, Department of Germanic Studies, Austin, U.S.A

https://doi.org/10.1515/9783110682588-002

frames can be used to establish cross-linguistic relationships in the context of what Granger (2003) calls the corpus approach for contrastive linguistics and translation studies (see also Gast 2015; Hasegawa et al. 2016; Hansen-Schirra et al. 2017). One of the central questions to be investigated is whether semantic frames can be used as a *tertium comparationis* (see Connor and Moreno 2005; Boas 2010a; Boas 2010b) and to what degree they might be considered universal or language-specific. In doing so, this paper also addresses the question of how semantic frames can be employed to establish comparability between languages, specifically in the context of different types of data. Because of space limitations, this paper focuses primarily on determining how semantic frames based on English can be applied to another language, specifically German. While the insights based on this comparison are potentially limited, they nevertheless provide insights into the question of whether semantic frames could potentially be considered as providing a (limited) "universal" inventory of meaning structures useful for research in Contrastive Linguistics.

The paper is structured as follows. Section 2 provides (1) an introduction to the notion of semantic frame in the Berkeley FrameNet project, and (2) an in-depth look at how FrameNet frames are used to structure and analyze the lexicon of English. Section 3 discusses how semantic frames of English have been re-used for the analysis of lexicons of other languages, most notably Spanish, Japanese, German, and Swedish. Based on ideas proposed by Heid (1996), Fontenelle (1997), and Boas (2002), Section 4 then develops systematic criteria that can be used to identify universal frames such as `Motion`, `Communication`, and `Ingestion`. I propose three sets of criteria: (1) translation equivalence; (2) valence equivalence; and (3) cultural equivalence. Section 5 shows how these criteria can be applied not only to frames that re-occur across languages, but also how they can be used to identify culture-specific frames that do not have equivalents in other languages, such as `Personal_relationship`. Finally, Section 6 summarizes the paper and provides suggestions for further research.

2 Semantic frames and the Berkeley FrameNet project

This section sets the stage for our discussion of the potentially universal status of semantic frames in the remainder of this paper. More specifically, it discusses the notion of semantic frame and how it has been implemented in the Berkeley FrameNet Project for English. The next section discusses how semantic frames derived on the basis of English have been applied to the analysis of words in other languages.

We begin with the concept of semantic frame, which can be traced back to Fillmore's (1968) seminal paper *The Case for Case*. In this paper, Fillmore proposed a limited set of semantic roles (also known as deep cases) such as Agentive, Instrumental, Dative, Locative, and Objective that were thought to be organized in a hierarchy for realizing grammatical functions. Fillmore's proposals were different from previous approaches, because they explicitly called for the identification of a restricted set of (universal) semantic roles that would be applicable to any argument of any verb. In addition, semantic roles were defined independently of verb meaning, they were regarded as unanalyzable, and each semantic role was supposed to be realized by only one argument. At the same time, each syntactic argument should bear only one semantic role and semantic roles were thought to be universal and applicable across languages. As such they were thought to be capable of capturing the lexical semantics of verbs at a level at which they could be compared across languages, while also providing language-specific hierarchies and linking rules. Fillmore's (1968) concept of semantic roles seemed initially attractive to many researchers, but during the 1970s multiple problems concerning the granularity of semantic roles and their systematic mapping properties led the research community to abandon the original concept of Fillmore's semantic roles (see Fillmore 1977; Levin and Rappaport Hovav 2005; and Boas and Dux 2017 for an overview).

In a series of publications throughout the 1970s, Fillmore revised and extended his original theory of case, eventually leading him to propose a theory called Frame Semantics. His new approach to meaning was driven by the insight that cultural and world knowledge motivate much of what we regard as "meaning" and that such knowledge is embedded in linguistic expressions. The theory of Frame Semantics (Fillmore 1982, 1985), originally developed on the basis of data from English, emphasized that solely truth-conditional semantic approaches cannot account for these aspects of word meaning, necessitating a "semantics of understanding" (see also Fillmore 1975). The core ideas underlying research in Frame Semantics are summarized in the following quote:

> A word's meaning can be understood only with reference to a structured background of experience, beliefs, or practices, constituting a kind of conceptual prerequisite for understanding the meaning. Speakers can be said to know the meaning of the word only by first understanding the background frames that motivate the concept that the word encodes.
> (Fillmore and Atkins 1992: 76–77)

While the main concepts of Frame Semantics were originally developed by Fillmore on the basis of English during the 1970s and 1980s, several other studies during the 1980s explored the application of semantic frames to languages other than English, including German (Lambrecht 1984) and Hebrew (Petruck 1986).

Fillmore and Atkins' (1992) detailed study of *to risk* focused on lexicographic and grammatical issues regarding the syntactic realization of (semantic) participants (a.k.a. roles). Their seminal research offered a detailed investigation of how the concept of 'risk' is realized linguistically by (1) identifying all participants in the risk scenario, (2) documenting how participants are formally realized in concrete linguistic expressions, and finally (3) summarizing the various ways in which the concept can be realized syntactically. They show, for instance, that 'risk' can be construed in (at least) two ways and therefore evokes two different frames, with expressions such as *take a risk* perspectivizing the risky activity carried out by the risk-taker and *put at risk* perspectivizing the entity endangered by the risky activity (see also Ohara 2009 and Boas and Dux 2017). Fillmore and Atkins' (1992) laid the groundwork for the development of FrameNet (http://framenet.icsi.berkeley.edu), which started at the International Computer Science Institute in Berkeley, California, in 1997.

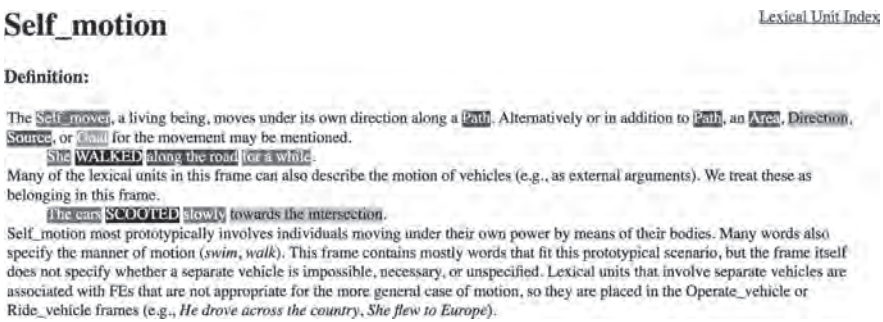
FrameNet (FN) organizes the lexicon of English by semantic frames, with valence information derived from attested, manually annotated corpus examples (Fillmore and Baker 2010, Ruppenhofer et al. 2013, 2017). FrameNet's workflow involves a number of stages starting with the selection of a target word (including multi-word expressions) and determining the frame it evokes by "characterizing schematically the kind of entity or situation represented by the frame" (Fillmore et al. 2003b: 297). To achieve this goal, FN researchers use a combination of corpus data and intuition to determine what features are necessary for the understanding of the word and assign mnemonic labels to each of the Frame Elements (FEs) defining the frame. Next, a thorough corpus search is conducted for expressions deemed semantically similar to the target word in order to determine whether they have the same frame semantics and Frame Elements, thereby arriving at a full list of lexical units for the frame (a lexical unit (LU) is a word in one of its senses).[1] For each of these lexical units, a number of representative corpus sentences are extracted and manually annotated for both syntactic and (frame-)semantic information. Specifically, the grammatical function and phrase type for each Frame Element occurring in the sentence is documented, resulting in layered annotations. The resulting FN database contains more than 1,300 frames, together with more than 13,000 lexical units and more than 200,000 annotated example sentences (see Baker, Fillmore and Cronin 2003 for an overview of the FN database).

AU: Please clarify whether this is "Fillmore, Charles J., Baker, Collin, & Beau Cronin 2003" or "Fillmore, Charles J., Johnson, Chris, & Miriam Petruck 2003."

1 FrameNet takes a splitting approach to determining the multiple senses of a word. Whenever there is enough corpus evidence available that supports the characterization of a particular use (sense) of a word as evoking a separate semantic frame, then FrameNet creates an extra entry for another LU (a word in one of its senses).

To illustrate the end-result of this workflow, consider the information about the LU *to crawl* evoking the `Self_motion` frame in the FN database. Typing *to crawl* into the search form on the FrameNet website yields three different links relevant to *crawl* in the `Self_motion` frame.[2] Clicking on the first link provides the user with detailed information about the `Self_motion` frame, as in Figures 1–3 below. The top of the frame definition in Figure 1 provides a prose description of the `Self_motion` frame, in which the target-evoking LUs are marked in black, while the Frame Elements (FEs) are marked in color.[3] The definition of the frame includes example sentences taken from the British National Corpus to illustrate how the prototypical meaning of LUs evoking the frame is realized in context. Each colored FE in the definition is a situation-specific semantic role that is defined more precisely in the remainder of the frame description, as shown in Figures 2 and 3 below.

AU: We have highlighted the cross references for author/ editor's check which will be removed in the next stage. Please check and confirm.



Self_motion

Lexical Unit Index

Definition:

The Self_mover, a living being, moves under its own direction along a Path. Alternatively or in addition to Path, an Area, Direction, Source, or Goal for the movement may be mentioned.

She WALKED along the road for a while.

Many of the lexical units in this frame can also describe the motion of vehicles (e.g., as external arguments). We treat these as belonging in this frame.

The cars SCOOTED slowly towards the intersection.

Self_motion most prototypically involves individuals moving under their own power by means of their bodies. Many words also specify the manner of motion (*swim*, *walk*). This frame contains mostly words that fit this prototypical scenario, but the frame itself does not specify whether a separate vehicle is impossible, necessary, or unspecified. Lexical units that involve separate vehicles are associated with FEs that are not appropriate for the more general case of motion, so they are placed in the Operate_vehicle or Ride_vehicle frames (e.g., *He drove across the country*, *She flew to Europe*).

Figure 1: Definition of `self_motion` frame.

Following the definition of the `Self_motion` frame, there are three more parts that make up the description of the frame. First, the definitions of so-called core FEs in Figure 2. Core FEs are those FEs which are central to a semantic frame (see Ruppenhofer et al. 2017).

Non-core Frame Elements, as in Figure 3, are FEs that are grammatically less prominent than core FEs. They can also be thought of as less semantically central to a semantic frame. For example, the non-core FE Duration (The amount of time for

2 Other results include the noun *crawl* (which also evokes the `Self_motion` frame) and a different LU *to crawl*, which evokes the `Abounding_with` frame (e.g. *The table is crawling with ants*.)

3 In this paper, names of Frame Elements and annotated sentences including Frame Elements appear in different shades of grey. The color representation refers to the online lexical entries of FrameNet.

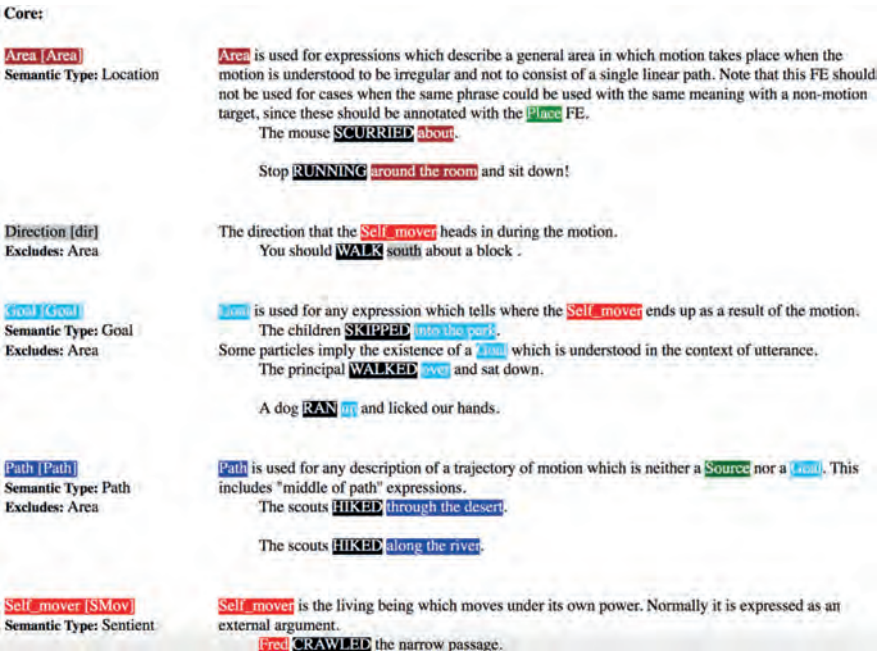


Figure 2: Core frame elements of the `self_motion` frame.

which a state holds or a process is ongoing) in Figure 4 is a non-core FE of the `Self_motion` frame. In contrast, the FE Self_mover (The living being that moves under its own power. Normally, it is expressed as an external argument) in Figure 3 is a core FE of the `Self_motion` frame. Other information provided for each frame description includes a list of frame-to-frame relations and a table of LUs that evoke the frame (see Petruck et al. 2004; Ruppenhofer et al. 2013; Boas 2017a; Boas and Dux 2017).

Recall that frame descriptions, including the definition of the frame and its FEs, are the end-result of a workflow involving several lexicographers relying on their intuitions and coming to an agreement about frame definitions that are supported by corpus evidence. As we will see in Section 3 below, these frame descriptions derived on the basis of English can be reused for the description and analysis of LUs in other languages, too. This means that frame descriptions can be thought of as a type of cross-linguistic (and possibly universal) metalanguage for lexical analysis.

Returning to our discussion of *to crawl* evoking the `Self_motion` frame, users can access two different types of reports about each LU. The Annotation Report provides the corpus sentences together with their frame-semantic annotations, the result of the manual annotation by FN annotators. These sentences form the

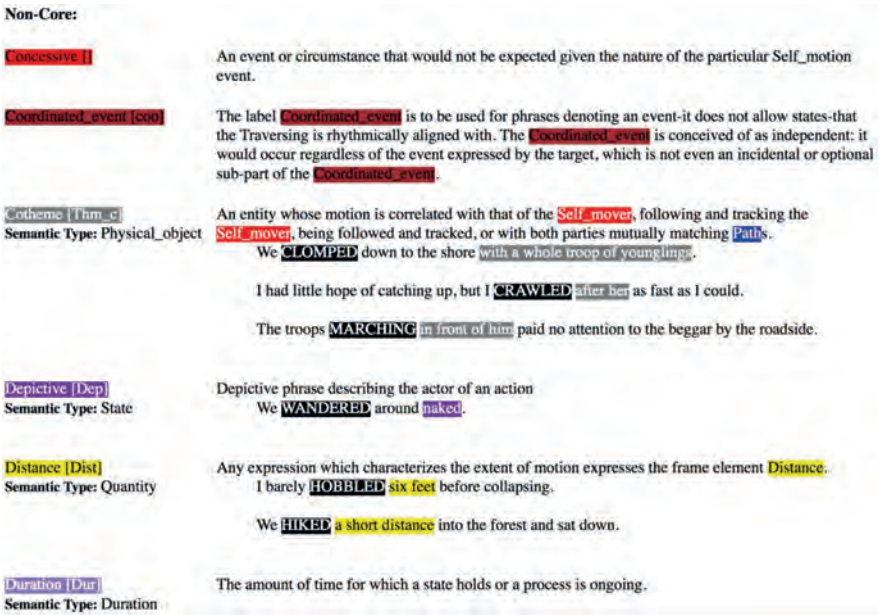


Figure 3: Non-core frame elements of the `self_motion` frame.

Purpose	(8)	VPto.Dep (7) Sfin.Dep (1)
Result	(2)	VPing.Dep (2)
Self_mover	(144)	NP.Ext (143) CNI.-- (1)
Source	(25)	PP[off].Dep (6) PP[out].Dep (11) PP[from].Dep (6) AVP.Dep (6)
Speed	(2)	AVP.Dep (1) PP[at].Dep (1)

Figure 4: First part of the lexical entry report of *to crawl* in the `self_motion` frame: Summary of FEs and their syntactic realizations (excerpt).

basis for the Lexical Entry Report, which consists of two parts. It first offers a list of how individual FEs are realized syntactically in the sentences annotated by the FN team. Figure 4 shows the different ways in which some of the FEs of the `Self_motion` frame are realized syntactically with *to crawl* (not all FEs are shown because of space limitations).

Figure 4 shows how FEs differ in their syntactic realizations in terms of phrase type (e.g. NP or PP) and grammatical function (e.g. Dep(endent)). While the FE Result has only one syntactic realization with *to crawl*, namely as a dependent VPing, other FEs exhibit a greater range of syntactic realizations: The FEs Purpose, Self_mover, and Speed each exhibit two different types of syntactic realizations while the FE Source shows four different types of syntactic realizations.

Note that FN captures not only overt syntactic realizations of FEs, but also cases in which FEs are not explicitly realized. Such cases are known as null instantiation, of which there are three different types. In Figure 4, the FE Self_mover may be null instantiated in terms of a Constructional Null Instantiation (CNI) such as the passive construction.[4] The two other types of null instantiation happen through the idiosyncratic licensing of a LU and cannot be captured in terms of higher-level generalizations such as grammatical constructions. In the partial valence pattern table for *to crawl* in the `Self_motion` frame in Figure 5 below we find one case of Indefinite Null Instantiation (INI), where the FE Path is null instantiated. INIs are instances in which FEs are merely existentially bound. In contrast, Definite Null Instantiation (DNI) are instances in which FEs are unrealized but which have to be recoverable from context (there is no example of DNI in the FN entry of *to crawl*). For more information on the different types of null instantiation, see Fillmore (1986), Lyngfelt (2012), Boas (2017b), and Ruppenhofer (2018).

Let us now turn to the second part of a LU's Lexical Entry Report, the valence pattern report, which is based on corpus examples that have been annotated by hand by FN annotators. It provides a summary of the many different ways in which combinations of FEs in sentences (so-called Frame Element Configurations (FECs)) are realized syntactically. For example, at the top of Figure 5 we find the FEC [Goal, Manner, Self_mover, Time], which is realized syntactically as [PP[to].Dep, PP[on].Dep, NP.Ext, PP[at].Dep]. While some FECs have only one particular syntactic realization, others may have multiple syntactic realizations as the third FEC [Goal, Path, Self_mover] from the top in Figure 5 shows. It has two syntactic realizations. Because of space limitations, only 6 FECs of the valence table for *to crawl* are shown in Figure 5. Overall, the valence table for *to crawl* has a total of 60 FECs with a total of 112 different syntactic realizations.

This brief overview illustrating the level of detail in FN lexical entries is important for our discussion of potential frame universality, because it shows three

4 For an overview of the so-called Constructicon, an online database of corpus-based construction entries, parallel to the FrameNet lexical database, see Fillmore (2008), Boas (2017a), and Ziem and Boas (2017).

1 TOTAL	Goal	Manner	Self_mover	Time
(1)	PP[to] Dep	PP[on] Dep	NP Ext	PP[at] Dep
1 TOTAL	Goal	Path	Purpose	Self_mover
(1)	PP[to] Dep	PP[along] Dep	Sfin Dep	NP Ext
2 TOTAL	Goal	Path	Self_mover	
(1)	PP[into] Dep	PP[up] Dep	NP Ext	
(1)	PP[to] Dep	PP[around] Dep	NP Ext	
1 TOTAL	Goal	Path	Self_mover	Time
(1)	PP[into] Dep	AVP Dep	NP Ext	PPing[before] Dep
1 TOTAL	Goal	Purpose	Self_mover	
(1)	AVP Dep	VPto Dep	NP Ext	
1 TOTAL	Goal	Purpose	Self_mover	Source
(1)	AVP Dep	VPto Dep	NP Ext	AVP Dep

Figure 5: Second part of the lexical entry report of *to crawl* in the `self_motion` frame: Summary of valence patterns showing how FEs are realized syntactically (excerpt).

things. First, each FN entry captures, among other things, the different semantic configurations of FEs and their various syntactic realizations. This idiosyncratic information differs from LU to LU evoking the same semantic frame in English. While there is some overlap in how LUs evoking the same semantic frame realize their FEs syntactically, the majority of cases of how LUs realize their FEs syntactically is idiosyncratic and cannot be captured at a more general or abstract level (see Boas 2010c; Dux 2016). Second, at the lexical level there appears to be very little predictability as to how the semantics of a frame is realized syntactically. Unlike research claiming that verbs closely related in meaning also exhibit the same patterns of syntactic distribution (Levin 1993), research on English verbs in Frame Semantics by Baker and Ruppenhofer (2002), Boas (2003b), Boas (2011b), and Dux (2018) shows that most aspects of a verb's syntactic distribution appear to be idiosyncratic (when compared to other verbs closely related in meaning). Third, even though the LUs differ so drastically in how they realize the FEs of the same frame differently, their meanings can still be captured at a somewhat general level that goes beyond the individual LU, namely the semantic frame. As I will show in Sections 3 and 4 below, this level of description and generalization does not only hold for English, but also for other languages, which means that the

concept of semantic frame should be considered as a basis for contrastive (and potentially cross-linguistic) analyses.[5]

FrameNet data have been used to answer a variety of empirical research questions on the mapping from semantics to syntax, they have been employed in a number of NLP tasks such as role labeling and text summarization, and they have been used for supporting foreign language teaching (for an overview, see Boas and Dux 2017). The next section discusses how semantic frames derived on the basis of English have been used to explore the lexicons of other languages, thereby establishing FrameNet databases for these languages. In this context it is important to keep in mind that the primary nature of these efforts is lexicographic in nature.[6] We begin by looking at some preliminary case studies that laid the theoretical groundwork for the architecture of multilingual FrameNets.

3 Semantic frames for multilingual lexicography

3.1 Exploring contrastive lexicon fragments

Exploratory studies such as Heid (1996) and Fontenelle (1997) show how English semantic frames could be applied to the analysis of the lexicons of other languages, such as French and German. The motivation for re-using semantic frames from English for other languages was the idea that frames could be universal (similar to Fillmore's 1968 original case roles) and that they could be used to create parallel lexicon fragments. Subsequent research demonstrates in greater detail how English-based semantic frames derived on the basis of English data could be employed for the analysis of polysemy structures of English verbs and their translation equivalents in other languages. One of the main goals of this research is to determine whether semantic frames could be used as a *tertium comparationis*, or what Connor and Moreno (2005: 157) call "a platform for comparison".

5 FrameNet differs from other lexical databases such as WordNet (Fellbaum 1998) in that it does not primarily rely on lexical relations such as synonymy, meronomy, etc. to structure the lexicon. Instead, it makes use of independent organizational units that are larger than words, i.e. semantic frames (see Boas 2005b). As such, FrameNet facilitates a comparison of the comprehensive lexical descriptions and their manually annotated corpus-based example sentences with those of other LUs (also of other parts of speech) (see Boas 2009b).

6 For different approaches of how semantic frames can be employed for translation studies, see Boas (2013) and Czulo (2013).

For example, following Fillmore and Atkins (2000), Boas (2001) employs semantic frames to investigate the polysemy structures of English and German motion verbs to find out whether a contrastive analysis of their polysemy structures allows for systematic predictions about translation equivalence or not. Based on examples such as those in (1) and (2), Boas (2001: 64) notes that some usages of the verbs *to run* and *to walk* evoke the same semantic frame, while other usages evoke different semantic frames: the semantics of *run* in (1a) is similar to the semantics of *walk* in (2a) in that both LUs evoke the `Self_motion` frame, in which a Self_mover moves on its own volition form a Source along a Path to a Goal.

(1) a. *Julie ran to the store*.[7]
b. *Julie ran Pat off the street*.

(2) a. *Rod walked to the door.*
b. *Rod walked Melissa to the door.* (Boas 2001: 64)

In contrast, *walk* differs from *run* in at least two respects, according to Boas. First, the manner of motion of *walk* is different from *run* in that the speed is slower, but this difference appears to have no direct influence on the type(s) of frame(s) evoked by the two verbs. Second, there is a difference in the types of semantic frames evoked by the two verbs. While both evoke the `Self_motion` frame, *run* also evokes the `Cause_motion` frame (i.e. the usage of *run* in (1b) constitutes a separate LU from the usage of *run* in (1a)), involving contact with force.[8] Note that the usage of *walk* in (2b) does not evoke the `Cause_motion` frame (there is

7 Note that the examples in (1) and (2) are representative of only one syntactic realization of a Frame Element Configuration (see above) of the verbs *to run* and *to walk* evoking the `Self_motion` and `Cause_motion` frames. The valence tables of the LUs exhibit significant differences. For example, the valence table of *to walk* in the `Self_motion` frame lists many more FECs and syntactic realizations than the valence table of the LU *to run* in the `Self_motion` frame, which lists different FECs and syntactic realizations.

8 FN definition of the `Cause_motion` frame: An Agent causes a Theme to move from a Source, along a Path, to a Goal. Different members of the frame emphasize the trajectory to different degrees, and a given instance of the frame will usually leave some of the Source, Path and/or Goal implicit. The completion of motion is not required (unlike the Placing frame, see below), although individual sentences annotated with this frame may emphasize the Goal. This frame is very broad and contains several different kinds of words that refer to causing motion. Some words in this frame do not emphasize the Manner/Means of causing the motion (transfer.v, move.v). For many of the others (cast.v, throw.v, chuck.v, etc.), the Agent has control of the Theme only at the Source of motion, and does not experience overall motion. For others (e.g. drag.v, push.v, shove.v, etc.) the Agent has control of the Theme throughout the motion; for these words, the Theme is resistant to motion due to some friction with the surface along which they move.

no contact with force), but rather the Cotheme_motion frame, in which a Theme moves together with the Cotheme in a Direction (along a Source, Path, and Goal).[9] In other words, the four LUs in (1)-(2) above evoke a total of three different semantic frames: Self_motion, Cause_motion, and Cotheme_motion. Note that this is not only relevant to these two verbs, but applies to a broader variety of verbs, too, as Table 1 shows.

AU: Please check and confirm that setting of Table and its captions are fine for all the chapters.

Table 1: Different verbs / LUs evoking different semantic frames.

	Self_motion	Cause_motion	Cotheme_motion
run	X	X	
walk	X		X
crawl	X		
hike	X		
scoot	X	X	
trail			X
slam		X	

The distribution of LUs and semantic frames evoked by them is only a small snapshot from the FN lists of LUs evoking the the three frames. But they illustrate an important point, namely that there is no direct way of predicting which LUs will evoke which frames. In other words, just because a particular LU evokes the Self_motion frame does not automatically mean that it also evokes the Cause_motion or the Cotheme_motion frames. The data thus suggest that the types of meanings cannot be systematically predicted based on frame membership alone, but that they need to be catalogued manually.[10]

9 FN definition of the Cotheme frame: This frame contains words that necessarily indicate the motion of two distinct objects. The Theme is typically animate and is expressed the same way a Self-mover is expressed in the Self-motion frame--i.e. as the subject of a target verb. The Cotheme may or may not be animate and is typically expressed as a direct object or an oblique. Source, Path, Goal, and the other frame elements common to motion words also regularly occur with the words in this frame. For more details, please see [https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Cotheme].

10 Fillmore and Atkins (2000: 103) provide a much more detailed corpus study of *to crawl*, employing corpus data to show that the different senses of motion verbs can be represented in terms of a semantic network diagram. In such a systematic representation of a verb's various meanings (in terms of frames), there is one central sense and sense extensions are represented by lines connecting the central sense and more extended senses. Fillmore and Atkins' comparison of English

Looking at the distribution of English LUs and the frames they evoke one might ask: How is this distinction relevant to a paper on contrastive linguistics? It is relevant because semantic frames are relevant not only to determining and modeling sense distinctions and polysemy networks in one language, but also across languages, thereby serving as a helpful structuring device for identifying, linking, and investigating word senses across languages. Part of this research asks the question of whether the semantic frames derived on the basis of English are also applicable for the description and analysis of other languages and whether semantic frames could be regarded as potentially universal linguistic concepts applicable across the languages of the world.

But before examining the question of how "universal" semantic frames are, let us first take a more straightforward bottom-up approach by determining how semantic frames derived on the basis of English can be applied to just one other languages. Consider, for example, the German counterparts of (1) and (2) above. Boas (2001) shows that while the basic types of situations described by *run* and *walk* in (1a) and (2a) are typically expressed by *rennen* 'run' and *gehen* 'go' (both evoking the `Self_motion` frame), thereby showing considerable syntactic and semantic overlap, there is no such overlap between *run* in (1b) above and *rennen* in (3b).

(3) a. *Tina rannte zum Geschäft.*
Tina ran to-the store
'Tina ran to the store.'
b. **Tina rannte Enno von der Strasse ab.*
Tina ran. Enno from the street off
c. *Tina drängte Enno (beim Rennen) von der Strasse ab.* (Boas 2001: 65)
Tina pushed. Enno while running from the street off
'Tina ran Enno off the street.'

The data in (3b) show that there is no LU of *rennen* that evokes the `Cause_motion` frame in parallel to run in (1b). This is a case of diverging polysemy (Altenberg and Granger 2002; Viberg 2002), in which items in two languages have different types and networks of meaning extensions. In the case of German *rennen* and English *to run*, this means that the translation equivalent of the `Cause_motion` sense evoked by *run* in (1b) is expressed by a completely different type of verb,

crawl with its French counterpart *ramper* demonstrates that even though the basic senses of the two verbs can be regarded as translation equivalents of each other, the semantic network of *ramper* with its sense extensions is very different from the semantic network of *crawl*.

namely *abdrängen* 'push aside' in (3c). Note that *abdrängen* itself does still not provide an adequate translation equivalent of the `Cause_motion` sense of *to run*, because it does not encode the manner in which the Theme (i.e. *Enno* in (3c)) has been caused to move to its end location. Information about the manner in which the caused motion took place has to be provided by a separate phrase *beim Rennen* ('by means of running'), because German *abdrängen* conforms to a different type of lexicalization pattern (cf. Talmy 1985) than English *to run*. Without this information it is not clear how the caused motion took place.

AU: The reference "Talmy 1985" is cited in the text but is not listed in the references list. Please either delete in-text citation or provide full reference details.

Similar observations can be made about the German translation equivalents of *walk* in (2) above: Boas (2001: 65) shows that the German translation equivalent of *walk* in (2a) evoking the `Self_motion` frame, the verb *gehen* in (4a), cannot be used as a translation equivalent for the Cotheme sense of *walk* in (2b). Instead, the different lexicalization in German requires that a different verb be used to express the `Cotheme` semantics, in this case *begleiten* ('to accompany').

(4) a. *Bernd ging zur Tür.*
b. **Bernd ging Anna zur Tür.*
c. *Bernd begleitete Anna zur Tür.* (Boas 2001: 65)

Observations such as these lead Boas (2001) to the conclusion that semantic frames are a useful tool for conducting a contrastive analysis of English motion verbs and their German translation equivalents. In a series of other case studies, Boas presents further case studies employing semantic frames as contrastive structuring devices to create and link parallel lexicon fragments for communication verbs in English and German (Boas 2002), English and German verbs describing operating a vehicle, affecting a person's mental state, and transportation (Boas 2003), and communication verbs in English, German and Spanish (Boas 2005a).

AU: Please clarify whether this is "Boas 2003a or Boas 2003b.

What unifies these case studies are three important insights. First, semantic frames derived on the basis of English can also be employed for the description and analysis of verbs in other languages, laying the foundation for creating parallel lexicon fragments. Second, semantic frames serve as a useful tool for linking parallel lexicon fragments between English, German, and other language pairs. Third, as the examples with *run* and *rennen* above have demonstrated, it is very difficult to predict the exact types of sense extensions (and the frames evoked by them) of a translation equivalent of an English verb. Just because the basic sense of an English motion verb evokes the `Self_motion` frame does not automatically mean that another sense of the same verb evokes the `Cause_motion` or the `Cotheme` frame (see Table 1 above). This means that in most cases parallel lexicon fragments cannot be predicted on the basic sense of a verb (i.e. one LU)

evoking a particular frame, but the sense extensions (other LUs) and the frames evoked by them (the other LUs) need to be identified and catalogued by hand and linked to their parallel lexicon fragment. We now turn to a discussion of how the idea of using English semantic frames for the description and analysis of other languages has been implemented in FrameNet-type projects for other languages.

3.2 Multilingual FrameNets: How universal are semantic frames?

Over the past 15 years, the Berkeley FrameNet database for English has served as inspiration for FrameNets for other languages. Building on insights by Heid (1996), Fontenelle (1997), Fillmore and Atkins (2000), and Boas (2001, 2002, 2005a), these other FrameNet projects differ in the types of corpora, tools, databases, workflows, and methodologies they employ (for a discussion, see Boas 2009b), but they are all similar in that they seek to create lexical entries employing semantic frames from the Berkeley FrameNet database for English.[11] In 2002, Spanish FrameNet started as the first large-scale FrameNet for a language other than English (Subirats and Petruck 2003; Subirats 2009). Since then, FrameNets for other languages, including Japanese (Ohara et al. 2004; Ohara 2009), German (Burchardt et al. 2009), Swedish (Borin et al. 2010), Brazilian Portuguese (Salomão et al. 2013), and French (Candito et al. 2014) have been applying semantic frames derived on the basis of English to the description and analysis of the lexicons of their languages.[12]

Due in large part to funding constraints, none of these FrameNets offer the same amount of coverage or continuity as the Berkeley FrameNet for English,

11 Parallel efforts are under way to create parallel repositories of construction entries, so-called constructicons for a variety of languages. See Fillmore (2008), Boas (2017), Ziem and Boas (2017), and the contributions in Lyngfelt et al. (2018) for details.

12 This does not necessarily imply that all FrameNets for other languages started out only with the semantic frames for English. Instead, each FrameNet has been using their own linguistic data in order to create their frames for their languages, while at the same time keeping an eye to how those frames compare with frames created for English by the Berkeley FrameNet project. Note that besides general-domain multilingual FrameNets, there are also domain-specific FrameNet-type projects and databases dealing with specific aspects of the lexicon, such as the Kicktionary for soccer terminology in English, French, and German (Schmidt 2009), BioFrameNet covering biomedical terminology (Dolbey et al. 2016), Bertoldi and Chishman (2012) for legal terminology, and the German Frame-based Online Dictionary, a learner's dictionary for English speakers learning German (Boas and Dux 2013; Boas et al. 2016).

which has been in operation since 1997. For example, the SALSA project for German (2002–2010) explored methods for large-scale manual frame-semantic annotation of the German TIGER Treebank (Brants et al. 2002), and multilingual approaches to inducing and verifying frame semantic annotations. The SALSA team also used the English FN frames where possible, but instead of starting with English frames and LUs and identifying equivalent German ones, they conducted full text annotation. When they ran into words for which there was no corresponding LU in the English FN database, they created so-called proto-frames, i.e. provisional frames for a single lexeme, without grouping them into larger frames. The eight years of SALSA funding resulted in roughly 20,000 annotations of verbs and 17,000 for nouns. In contrast, Spanish FrameNet (from 2002–2015) put together their own 940-million-word Spanish corpus and created their own tagging system in order to directly use the frames and frame elements from the English FN database for the vast majority of their LUs, resulting in a total of 10,334 manually annotated lexicographic examples as the basis for 1,124 LUs in 325 frames.

Since the 1980s, one of the questions asked by research in Frame Semantics has been whether frames should be regarded as "universals" of human language or whether they are language specific. Over the past 15 years, the process of employing semantic frames developed on the basis of English to develop FrameNets for other languages strongly suggest that many frames can be regarded as applicable across different languages, especially those relating to basic human experience such as eating, drinking, sleeping, and walking (see the contributions in Boas 2009a). Even some cultural practices appear to be comparable across many languages, such as commercial transaction: in many cultures, we find a specific type of exchange that can be characterized as a type of commercial transaction involving the FEs Buyer, Seller, Money, and Goods. However, to date there has been no empirical study determining the universal applicability of semantic frames across languages. This is due to at least three problems.

The first problem concerns coverage of the lexicon. English FrameNet, which so far has the largest inventory of more than 1,200 frames together with entries for more than 13,600 lexical units, does not yet provide a large coverage of the English lexicon. Recall that unlike traditional dictionaries, which are organized alphabetically, FrameNet describes and analyzes the English lexicon frame by frame. While with traditional dictionaries we can estimate their coverage by looking at how many words they have covered under each letter of the alphabet, this is somewhat more difficult with FrameNet because we do not have a clear understanding of how extensive FrameNet's coverage of the English lexicon really is. Researchers estimate that average speakers of English have an average active knowledge of about 20,000 words (and a passive knowledge of about 40,000 words) (cf. Na &

Nation 1985).[13] But even this estimate is somewhat unreliable because the notion of "word" itself is problematic. More specifically, lexicographers do not always agree on how many senses a word has, depending on whether they are lumpers or splitters (see Kilgarriff 1997). Our short discussion nevertheless suggests that currently English FrameNet does not adequately cover the average active vocabulary of a speaker of English. This, in turn, means that if we employ the current inventory of semantic frames based on English to explore their universal applicability to the description and analysis of other languages, we have to keep in mind that there are large gaps.[14] As such, getting closer to answering the question of how universal semantic frames really are will first require a more elaborate coverage of English FrameNet or a FrameNet for another language.[15]

The second problem concerns the methodology: Which frames should we select for our investigation? To date, we have no solid empirical criteria to measure how universal a frame is, let alone how to go about identifying which frames we should investigate. More specifically: Should we employ the Swadesh list of 207 basic concepts that are intended to cover those areas central to human life, and if we find corresponding frames based on empirical evidence in all the world's languages, should we then speak of a set of "universal" frames? What role should the concept of frequency play? These are all open questions (besides others) that will need to be

13 Note that there is no single agreed-upon list of criteria of what constitutes the basic vocabulary of a language. This point alone makes it difficult to evaluate the potential "universal" applicability of semantic frames. Some researchers argue that the core of the lexicon consists of those words that are most useful for the speaker and hearer depending on whether (1) they are most frequent in texts of different genres, (2) they designate concepts that are central to human life, or (3) they suffice to paraphrase and explain all the other words in the lexicon. See Goddard (2001) and Lehmann (2018) for a discussion of the so-called Swadesh list of 207 basic concepts and Ogden's (1930) Basic English list containing 850 items.

14 Note that other approaches, such as the Natural Semantic Metalanguage (Wierzbicka 2005), claiming to have found a universal inventory of semantic descriptors, face similar issues related to coverage.

15 Despite the perceived lack of coverage, FrameNet has come a long way during its more than 20 years of analyzing the English lexicon. The FrameNet database is so far unmatched in terms of level of detail of how the meanings of semantically related words are expressed syntactically (see Boas 2005b, 2017a). Its corpus-based methodology, resulting in more than 200,000 manually annotated example sentences, is time and labor-intensive, but the proof of concept and its underlying methodology are well-established. The current lack in coverage by FrameNet is due in large part to funding constraints. This situation can be compared with a traditional dictionary losing its funding after having completed entries covering only the relevant words starting with the letters A-N. With the remaining words starting with O-Z not being covered, such a traditional dictionary would also be regarded as lacking coverage. This means that if there were enough funding available for FrameNet, it would be relatively easy to solve the perceived lack of coverage.

addressed. Another related issue in trying to establish whether semantic frames are universal or not, is that we need to keep in mind that there might be many different types of frames and while some might be considered "universal", others might not, while others might only be considered partially universal (as is the case when typological restrictions in a language preclude the explicit morpho-syntactic coding of certain semantic categories). Consider, for example, the importance of culture-specific words, frames, and ways of thinking that are deeply embedded in the beliefs, customs, and practices of particular cultures, but not other cultures. As the Modern Language Association ad Hoc Committee on Foreign Languages (2007: 2) points out:

> Expressions such as the 'pursuit of happiness', 'liberté, egalité, fraternité' and 'la Raza' connote cultural dimensions that extend well beyond their immediate translation. [...] deep cultural knowledge and linguistic competence are equally necessary if one wishes to understand people and their communities.
>
> (MLA ad Hoc Committee on Foreign Languages 2007: 2)

AU: The reference "MLA ad Hoc Committee on Foreign Languages 2007" is cited in the text but is not listed in the references list. Please either delete in-text citation or provide full reference details.

The third problem concerns the idea of universality itself. Much linguistic research, in particular in the generative and typological paradigms of the second half of the 20th century, focused on establishing universal categories, patterns, or generalizations to arrive at a "universal" theory about language (for a critique see Croft 2001). But the empirical basis for many of the claims about universality are difficult to prove or to falsify because of a lack of data. Most claims about a particular universal aspect of language rests on a limited set of data from a limited set of languages, whether they are 5, 10, or 100 languages.[16] But even with larger numbers of languages being covered, the claim about a supposedly "universal" aspect of language cannot be upheld until we have solid descriptions of all of the world's 6,000 or so languages (many of which are endangered to various degrees, see Crystal 2000). Thus, when using the term "universal" throughout the remainder of this paper it should be interpreted as "potentially" universal.

More specifically, in what follows I will focus on investigating different ways of establishing criteria for re-using semantic frames derived on the basis of English that can be employed to describe and analyze words and concepts in other languages. For the most part, this will involve contrastive comparisons only, thereby laying the foundation for further comparisons down the road. This bottom-up approach is only a very tiny first step towards establishing a methodology for investigating the potential "universal" status of semantic frames.

16 See the Leipzig Valency Classes Project, which aims at arriving at cross-linguistic generalizations about how valency is expressed in the world's languages: https://www.eva.mpg.de/lingua/valency/files/project.php

4 Towards a methodology for identifying "universal" frames

To illustrate how frames can be identified as potential candidates for "universal" frame-hood, let us consider the Berkeley FrameNet `Questioning` frame. The words in this frame have to do with a Speaker asking an Addressee a question which calls for a reply (as opposed to making a request which calls for an action on the part of the Addressee). LUs evoking the `Questioning` frame include *to ask*, *to inquire*, *to question*, *inquiry*, *question*, etc. In order to determine whether the same frame can be applied to the description and analysis of other languages it does not suffice to just take the frame description from English and apply it to other languages.

Instead, we need to determine whether we find in the other language, in this case German, translation equivalents or near-translation equivalents corresponding to the English LUs of the "original" `Questioning` frame. Because of a lack of space, I will not be able to discuss possible translation equivalents of all LUs evoking the `Questioning` frame, let alone translation equivalents for all FECs found in the valence table of a single English LU. Instead, I will focus here for illustrative purposes only on discussing a single straightforward case of a German translation equivalent of one English LU, namely the verb *to ask* evoking the `Questioning` frame. After discussing this one example, I will address the issues surrounding finding translation equivalents for other frame element configurations, and other LUs evoking the same frame.[17]

The valence information in the FrameNet entry of the LU *to ask* in the `Questioning` frame contains a total of 15 FECs with a total of 69 syntactic realizations (the various syntactic realizations of frame elements, also known as mini-constructions; Boas 2003). Because of a lack of space, Figure 6 only illustrates a subset, namely 8 FE configurations with 16 valence realizations. Of these 16 syntactic realizations, let us take a look at only one syntactic realization, namely the one in which *to ask* appears with an external NP, an object NP, and a PP headed by *about* (marked by an arrow in Figure 6 below) as in the sentence *The immigration authorities asked her about her profession* (based on Boas 2011b).

Using the information in this one syntactic realization, we are able to map the form information [NP.Ext, *ask*.V, NP.Obj, PP_*about*.Comp] to the frame-semantic meaning of the `Questioning` frame, so that the external NP is identified as the Speaker FE, the object NP as the Addressee FE, and the PP headed by *about* as

17 Because of space limitations, we cannot discuss other translation equivalents for other languages. As such, the current paper is intended to serve only as a case study.

2 TOTAL	Addressee	Speaker	Time	
(1)	DNI --	NP Ext	AVP Dep	
(1)	NP Obj	NP Ext	AVP Dep	
1 TOTAL	Addressee	Speaker	Time	Topic
(1)	NP Obj	NP Ext	AVP Dep	PP[about] Dep
16 TOTAL	Addressee	Speaker	Topic	
(4)	DNI --	NP Ext	NP Obj	
(1)	DNI --	NP Ext	PP[about] Dep	
(1)	DNI --	NP Ext	PPing[about] Dep	
(1)	DNI --	NP Ext	Sfin Dep	
(1)	DNI --	NP Obj	PP[about] Dep	
(1)	NP Ext	CNI --	PP[about] Dep	
(2)	NP Obj	CNI --	PP[about] Dep	
(1)	NP Obj	CNI --	PPing[about] Dep	
(1)	NP Obj	NP Ext	NP Dep	
(1)	NP Obj	NP Ext	PP[about] Dep	
(2)	NP Obj	NP Ext	PPing[about] Dep	

Figure 6: Valence table of *to ask* in the `questioning` frame (excerpt).

the Topic FE. With this mapping of the English form of the sentence *The immigration authorities asked her about her profession* we now have an approximate frame-semantic representation of its meaning based on our knowledge of the `Questioning` frame including its definition. Recall that the `Questioning` frame under discussion is derived on the basis of English corpus data. We are now interested in determining how this very same frame can be reused for analyzing other languages in order to determine to what degree semantic frames are useful for contrastive analysis, and, more broadly, to cross-linguistic analysis.

To show how this can be achieved, we focus here on just one syntactic realization of the FEC [Speaker, Addressee, Topic] to just one corresponding syntactic realization of one corresponding FEC in another language, in this case German. Figure 7 illustrates how this parallel mapping of syntactic realizations of FECs

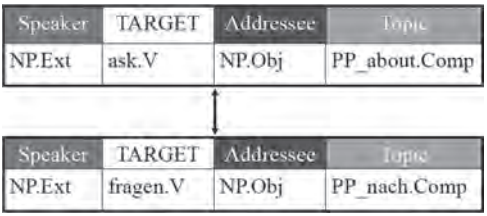


Figure 7: Parallel lexicon fragment: Cross-linguistic identification of *to ask* and *fragen* based on valence realizations of the same semantic frame (see Boas 2011b).

can be implemented using semantic frames derived on the basis of English. The relevant German counterpart of English *to ask* is German *fragen* 'to ask'.

The result is a correspondence between the FEs Speaker, Addressee, and Topic between English and German, representing the English sentence *The immigration authorities asked her about her profession* and its German counterpart *Die Einwanderungsbehörde fragte sie nach ihrem Beruf*. This example shows that employing semantic frames for mapping between an English syntactic realization of an FEC contained in a verb's valence table and its German counterpart is feasible (see also Boas 2002, 2011b). It also demonstrates that the `Questioning` frame derived on the basis of English is applicable to German in a straightforward way.

In this paper, I call cases such as illustrated in Figure 7, in which there is one clear match between the syntactic realization of one LU in one language and a corresponding LU in another language evoking the same frame, "surface translation equivalence". The term "surface" is not to be understood as in the generative transformational paradigm, but rather as denoting a situation in which there is a clear match between two syntactic realizations of the two valence tables, such that it appears at the "surface" as if they are translation equivalents. As such, "surface translation equivalence" is a much weaker version of what Viberg (2002) calls translation equivalence.[18]

18 Because of space limitations I leave aside here a discussion of differences in lexicalization patterns, which are relevant when dealing with typologically different languages and how they realize the semantics of a frame. For example, Subirats (2009) discusses lexicalization differences in Spanish and English for emotion predicates, and Subirats and Sato (2004) report on constructional differences between English and Spanish motion verbs. These differences need to be addressed within a broader context of typological differences such as expressional differences in motion events between Germanic and Romance languages (Slobin 1996). However, it is important to remember that such typological differences do not put in question the usability of semantic frames (see also Ohara 2009 on differences between English *to risk* and Japanese *kakeru*). Schmidt (2009) provides an insightful discussion of some typological differences in football language in English, German, and French.

Our example also suggests that semantic frames can be useful tools for establishing (1) translation equivalence (the English and German sentences convey the same meanings) and (2) valence equivalence (there is a one-to-one mapping between FEs, phrase types, and grammatical functions), but only at a low level. A systematic comparison of cultural equivalence would require a larger in-depth study involving parallel corpus data to determine to what degree the English LU *to ask* and the German LU *fragen* evoking the `Questioning` frame can be considered to be equivalent at a cultural level (more on that issue below). Our discussion so far suggests that semantic frames derived on the basis of English are useful for comparing and contrasting the lexicons of other languages. As such, semantic frames could be considered as possible candidates for translation universals.

However, a number of disclaimers are in place. Note, first, that our example of a corresponding parallel English-German lexicon fragment in Figure 7 represents only one syntactic realization of one of 15 FECs with a total of 69 syntactic realizations of English *to ask* evoking the `Questioning` frame.[19] We have not addressed the 5 other syntactic realizations of the same FEC, let alone the 63 other syntactic realizations of the 14 other FECs. This is why I tentatively labeled the situation depicted in Figure 7 "surface translation equivalence." In order to establish the degree to which *to ask* and *fragen* really evoke the same semantic frame and the degree to which there is more of a correspondence between the syntactic realizations in the valence tables of the two LUs we need to repeat the same procedure for each syntactic realization.

Cases in which there is a high degree of correspondence between the syntactic realizations of FEs of the semantic frame in one language with the syntactic realizations of FEs in another language are called "valence equivalence." Whenever we find more cases in which the syntactic realizations found in the valence tables of two LUs thought to evoke the same frame are equivalents of each other, the higher the degree of valence equivalence. Put differently, "surface translation equivalence" like the one depicted in Figure 7 above is the lowest degree of "valence equivalence", because we have so far only one clear case of two matching syntactic realizations across languages. Note that most likely there are no cases in which there is complete "valence equivalence" between the valence tables of two LUs from different languages and that as such the notion

19 Dux (2016, 2018) points out the importance of paying close attention to verb valence patterns across languages to formulate frame-semantic classes. He shows that the types of verb classes resulting from a frame-semantic classification do not always present an exact overlap between different languages. This insight leads him to propose a more fine-grained approach that uses verbal valency for the formulation of verb-valency classes together with semantic frames.

of valence equivalence is one of degree. In other words, the greater the number of corresponding syntactic realizations in the valence tables of two LUs from different languages that evoke the same frame is, the larger the degree of "valence equivalence."

Unfortunately, we cannot establish whether there is a high degree of valence correspondence between *to ask* and *fragen* evoking the `Questioning` frame because of limited space in this paper. This would require us to do an in-depth investigation to determine possible valence equivalents of each of the remaining 68 syntactic realizations of *to ask*. Note, however, that there are a few important points we can briefly review here, which may serve as the starting point for a future paper investigating the degree of cross-linguistic correspondence between two LUs from different languages evoking the same semantic frame.

5 Culture-specific semantic frames

So far, we have only reviewed the concepts of translation equivalence and valence equivalence to determine the degree to which there is overlap between two LUs in different languages evoking the same frame. The third concept relevant in this context is what I call "cultural equivalence," and this concept may be the most difficult to define, identify, and measure. Cultural equivalence concerns cases in which two LUs from different languages evoking the same semantic frame can be used in the same contexts with the same cultural connotations. One way of going about determining cultural equivalence (or perhaps, to a lesser degree, cultural correspondence) is to adopt insights from Wierzbicka's (2005) theory of cultural scripts, which seeks to systematically account for cultural values in the semantics of words in a given language.

VanNoy (2017) presents a first account of how Wierzbicka's (2005) cultural scripts can be combined with Fillmore's semantic frames in order to highlight and investigate cultural similarities and differences of words in two languages thought to evoke the same frame. For example, VanNoy (2017) provides an analysis of the English noun *friend* and its German counterpart *der Freund/ die Freundin* (male/female). Noting that both nouns have the same Germanic root and that both nouns can be used in many of the same contexts in contemporary English and German denoting friendship as they evoke the `Personal_relatioinship` frame, VanNoy also points out that there are a number of significant differences (see also Atzler 2011). Following the ideas underlying Fillmore and Atkins' (2000) semantic network analysis, VanNoy uses a combination of data from monolingual dictionaries, bilingual dictionaries, and corpus data to show that the two nouns

differ in a number of important aspects. For example, the English and German nouns differ in the types of personal relationships they denote, specifically the intensity of the relationship, the duration of the relationship, and whether the relationship is romantic and/or intimate or not.

Based on collocational information for the English noun *friend* and its German counterpart *der Freund/die Freundin*, VanNoy points out that such important cultural differences are not included in frame-semantic descriptions and that existing semantic frames derived on the basis of English should be augmented by more fine-grained cultural information. These observations lead her to propose two related frames for German that are more specific than the general `Personal_relationship` frame derived on the basis of English. More specifically, she proposes for German a `Platonic_Personal_Relationship` sub-frame (VanNoy 2017: 185)[20] and a `Non_Marital_Personal_Relationship` sub-frame (VanNoy 2017: 189), augmenting each with German-specific cultural scripts emphasizing the different levels of intensity, duration, and exclusivity of relationships. VanNoy shows that augmenting existing English frames with cultural scripts makes it possible to capture the entirety of cultural connotations using Frame Semantics at different levels of granularity within and across languages.

What have we learned from our short discussion of "cultural equivalence"? First, recall that our motivation for re-using semantic frames from English for other languages is the idea that frames could possibly be universal, similar to what has been claimed about Fillmore's (1968) original case roles. To establish degrees of equivalence or correspondence between two LUs from different languages thought to evoke the same semantic frame, I proposed three different levels of equivalence, namely translation equivalence, valence equivalence, and cultural equivalence. Second, as shown above, it is possible to determine translation equivalence and valence equivalence relatively straightforwardly by comparing and contrasting the valence tables of two LUs from different languages thought to evoke the same frame. Third, there are cases in which semantic frames derived on the basis of English are not fine-grained enough to be re-usable for the analysis of corresponding LUs in other languages. To identify and measure such cases of cultural equivalence I argued for adopting VanNoy's (2017) proposals to combine insights from Fillmore's Frame Semantics with that of Wier-

20 An example of VanNoy's (2017: 185) use of semantic scripts, following Wierzbicka (2005), is additional information augmenting the `Platonic_Personal_Relationship` sub-frame: X is someone like this: X is someone I know; X is someone I have met before. When I think of X I think: I know this person, but I do not know much about this person and I do not feel close to this person. Many people think like this: I have many of X because there are many people I have met I know. Evoking LU: *Bekannter* ('acquaintance').

zbicka's theory of cultural scripts. Note that we discussed only one example from the Personal_relationship frame, but the literature on Wierzbicka's cultural scripts is full of similar cases (e.g. particular words expressing politeness, personal distance, worldview, customer service, etc.) that merit a further in-depth investigation in the context of determining the possibility universality of frames (see Goddard 2001; Wierzbicka 2005).

Returning to our discussion of *to ask* and its German counterpart *fragen* above, it is important to note that there does not seem to be a need for the inclusion of cultural scripts to augment particular sub-frames to the general Communication_questioning frame derived on the basis of English. But as our example of the Personal_relationship frame has shown, there are differences between frames when it comes to cultural equivalence. How many semantic frames derived on the basis of English will require the explicit formulation of sub-frames with corresponding cultural scripts is open to further empirical investigation. This will depend on the number and types of semantic frames as well as the number and types of LUs and languages under investigation. Combining the insights of Fillmore's Frame Semantics with Wierzbicka's theory of cultural scripts may bring us one step closer to determining what types of frames are truly applicable cross-linguistically to the degree that we might eventually call them universal and what types of frames require fine-tuning depending on individual languages and cultures.

One final point worth mentioning concerns the important roles of paraphrase relations, polysemy, and translational equivalence. When determining translation, valence, and cultural equivalencies we have so far only focused on finding correspondences within single sentences and not across broader contexts. Thus, in finding an adequate equivalent for *to ask* in *The immigration authorities asked her about her profession* above we used the default German translation equivalent of *fragen* for *to ask*. While this may work in most default contexts, it is important to note that English *to ask* has multiple German translation equivalents, each of which evoke the Questioning frame, depending on context.[21] The difference in translations is typically triggered by specific background information provided by the sentence or the broader context in which the sentence containing *to ask* occurs: *ausfragen* 'to quiz somebody about something' highlights a particular aspect of the Questioning frame by focusing on a detailed, intense, or curious manner of questioning; *befragen* 'to interrogate someone' highlights

21 To determine the extent to which German translation equivalents of *to ask* differ from each other it will be necessary to conduct a detailed corpus-based investigation into how each syntactic realization in the valence table of *to ask* is realized by potentially different German LUs.

a particularly intense or authoritative manner of questioning; *bezweifeln* 'to challenge something' highlights the attitude of the Speaker of the Questioning frame, who does not believe the Addressee; *prüfen* 'to investigate' focuses on an investigative aspect of the Questioning frame, which the Speaker puts to the test whether something that the Addressee is claiming is true or not.

6 Conclusions and outlook

The procedures and proposals for identifying potentially universal semantic frames presented in this paper are only of a very preliminary nature and need to be significantly refined by future research. In this paper I have outlined how cross-linguistic correspondences between two LUs from different languages evoking the same semantic frame can be established. But note that this step alone, which itself appears to be quite labor intensive, addresses only two corresponding LUs evoking the same semantic frame. Above we discussed this procedure in the context of *to ask* and *fragen* evoking the Questioning frame. To determine the degree to which the Questioning frame could be considered a possible candidate for a universal frame would first require repeating the same procedure for the remaining 11 English LUs evoking the Questioning frame. Once we know which English LUs have German counterparts we can propose a common Questioning frame evoked and shared by both English and German LUs. Note, however, that this frame would only cover English and German, not any other languages. In other words, we would need to repeat the same procedure for all other known languages, i.e. identifying corresponding LUs with equivalent meanings, in order to see whether the same frame is evoked by LUs across the board.

If we were indeed to find corresponding LUs with roughly equivalent meanings evoking the same Questioning frame across all languages (with expected minor typological differences), then it should be possible to claim that the Questioning frame can be considered a type of "universal" frame in the sense that it is evoked by LUs from all languages. Other likely candidates that could be considered universal include the Motion, Communication, Ingestion, and Bodyparts frames. Of course, the question of universality might turn out to be a gradual notion in the case of semantic frames, because more likely than not we will find that a strict one-to-one correspondence between pairs of LUs from two language or across languages will be the exception. This means that future research needs to address in more detail the complex relationship between translation equivalence, valence equivalence, and cultural equivalence.

Note that the procedures outlined above do not rely on translation studies as a means of establishing cross-linguistic relationships, as is described in Granger (2003) and Johansson (2007). Instead, I proposed to begin with established semantic frames from the Berkeley FrameNet for English and to then use a combination of corpus-driven techniques and linguistic intuition to find and identify possible translation equivalents in monolingual corpora in other languages. While this procedure avoids some of the problems that arise from working with parallel corpora (e.g. interference between the language of the source-text and the translated text, see McEnery and Xiao 2008), it, too, is not free of problems. For example, the use of English frames based on English data to identify LUs in languages other than English raises the question of whether such frames might be too Anglo centric.

In this paper, I proposed a first step towards identifying possible universal semantic frames based on existing frames in English FrameNet. To achieve this goal I argued for a particular systematic procedure that begins with picking a specific English FrameNet frame such as `Questioning` and to pick a language other than English. The second step involves taking all LUs evoking the semantic frame in English and find translation equivalents in the other language. This step requires that we use a mix of corpus data and linguistic intuition to find for each LU corresponding valence equivalents and culture equivalents. Based on annotated corresponding corpus examples we are then in a position to create parallel lexicon fragments for English and the other language. Establishing these correspondences is a first step towards establishing a potentially universal frame. The same procedure should then be repeated for all other known languages.

Our discussion of culture and language specific words has shown that not all semantic frames derived on the basis of English are good candidates for universal frame-hood. The case of German *Freund/Freundin* has shown that there are cases in which it is necessary to define more fine-grained semantic sub-frames and augment these with more specific cultural information using Wierzbicka's theory of cultural scripts. Culture-specific words evoking particular semantic frames are likely the most difficult cases to investigate as they require a great deal of linguistic intuition and corpus data illustrating collocational restrictions and contextual requirements on the proper use of such words. While this paper has only outlined a roadmap for determining potential candidates for universal frame-hood, it has shown how this can be done in a systematic way using semantic frames based on English. Clearly, much research remains to be done in order to further explore the roadmap laid out in this paper.

References

Altenberg, Bengt & Sylviane Granger. 2002. *Introduction*. In B. Altenberg & S. Granger (eds.), *Lexis in Contrast*, 3–50. Amsterdam/Philadelphia: John Benjamins.

Atzler, Judith. 2011. *Twist in the list: Frame Semantics as Vocabulary Teaching and Learning Tool*. Unpublished Ph.D. dissertation, University of Texas at Austin.

Baker, Collin & Josef Ruppenhofer. 2002. FrameNet's Frames vs. Levin's Verb Classes. In J. Larson and M. Paster (eds.), *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society*, 27–38. UC Berkeley: Berkeley Linguistics Department.

Bertoldi, Anderson & Rove Chishman. 2012. Frame Semantics and legal corpora annotation: Theoretical and applied challenges. *Linguistic Issues in Language Technology*, *7*(1).

Boas, Hans C. 2001. Frame Semantics as a framework for describing polysemy and syntactic structures of English and German motion verbs in contrastive computational lexicography. In P. Rayson, A. Wilson, T. McEnery, A. Hardie & S. Khoja (eds.), *Proceedings of Corpus Linguistics 2001*. 64–73.

Boas, Hans C. 2002. Bilingual FrameNet dictionaries for machine translation. In M. González Rodríguez & C. Paz Suárez Araujo (eds.), *Proceedings of the Third International Conference on Language Resources and Evaluation*, Vol. IV, 1364–1371. Las Palmas, Spain.

Boas, Hans C. 2003a. *A constructional approach to resultatives*. Stanford: CSLI Publications.

Boas, Hans C. 2003b. A lexical-constructional account of the locative alternation. In L. Carmichael, C.-H. Huang, & V. Samiian (eds.), *Proceedings of the 2001 Western Conference in Linguistics. Vol. 13*, 27–42.

Boas, Hans C. 2005a. From theory to practice: Frame Semantics and the design of FrameNet. In S. Langer & D. Schnorbusch (eds.), *Semantik im Lexikon*, 129–160. Tübingen: Narr.

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

Boas, Hans C. (ed.). 2009a. *Multilingual FrameNets in Computational Lexicography*. Berlin/New York: Mouton de Gruyter.

Boas, Hans C. 2009b. Recent trends in multilingual lexicography. In Boas, Hans C. (ed.), *Multilingual FrameNets in Computational Lexicography*, 1–26. Berlin/New York: Mouton de Gruyter.

Boas, Hans C. (ed.). 2010a. *Contrastive Studies in Construction Grammar*. Amsterdam/Philadelphia: Benjamins.

Boas, Hans C. 2010b. Comparing constructions across languages. In H.C. Boas (ed.), *Contrastive Studies in Construction Grammar*, 1–20. Amsterdam/Philadelphia: Benjamins.

Boas, Hans C. 2010c. Linguistically relevant meaning elements of English communication verbs. *Belgian Journal of Linguistics 24*. 54–82.

Boas, Hans C. 2011a. A frame-semantic approach to syntactic alternations with *build*-verbs. In P. Guerrero Medina (ed.), *Morphosyntactic alternations in English*, 207–234. London: Equinox.

Boas, Hans C. 2011b. Constructing parallel lexicon fragments based on English FrameNet entries: Semantic and syntactic issues. In H. Hedeland, T. Schmidt, & K. Woerner (eds.), *Multilingual Resources and Multilingual Applications. Proceedings of the German Society for Computational Linguistics and Language Technology (GSCL) 2011, Hamburg*. University of Hamburg: Center for Language Corpora. 9–18.

Boas, Hans C. 2013. Frame Semantics and translation. In A. Rojo & I. Ibarretxte-Antunano (eds.), *Cognitive Linguistics and Translation*, 125–158. Berlin/New York: Mouton de Gruyter.

Boas, Hans C. 2017a. Computational Resources: FrameNet and Constructicon. In B. Dancygier (ed.), *The Cambridge Handbook of Cognitive Linguistics*, 549–573. Cambridge: Cambridge University Press.

Boas, Hans C. 2017b. What you see is not what you get: Capturing the meaning of missing words with Frame Semantics. In *Proceedings of the Chicago Linguistics Society* 52. 53–70.

Boas, Hans C. & Ryan Dux. 2013. Semantic frames for foreign language education: Towards a German frame-based dictionary. *Veridas On-line. Special Issue on Frame Semantics and its Technological Applications*. 82–100. http://www.ufjf.br/revistaveredas/edicoes- 2013

Boas, Hans C. & Ryan Dux. 2017. From the past into the present: From case frames to semantic frames. In *Linguistics Vanguard* 2017. 1–14. DOI: 10.1515/lingvan-2016-0003.

Boas, Hans C., Ryan Dux, & Alexander Ziem. 2016. Frames and constructions in an online learner's dictionary of German. In S. De Knop & G. Gilquin (eds.), *Applied Construction Grammar*, 303–326. Berlin: de Gruyter.

Borin, Lars, Dana Dannells, Markus Forsberg, Maria Toporowska Gronostaj & Dimitros Kokkinakis. 2010. The past meets the present in Swedish FrameNet++. *Proceedings of EURALEX 2010*. 269–281.

Brants, S., Dipper, S., Hansen, S., Lezius, W., & Smith, G. 2002. The TIGER treebank. In *Proceedings of the workshop on treebanks and linguistic theories* (Vol. 168).

Burchhardt, Aljoscha, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Padó, & Manfred Pinkal. 2009. Using FrameNet for the semantic analysis of German: annotation, representation, and automation. In H.C. Boas (ed.), *Multilingual FrameNets: Methods and Applications*, 209–244. Berlin/New York: Mouton de Gruyter.

Candito, M., Amsili, P., Barque, L., Benamara, F., De Chalendar, G., Djemaa, M., Haas, P., Huyghe, R., Mathieu, Y.Y., Muller, P. & Sagot, B., 2014, May. Developing a french framenet: Methodology and first results. In *LREC-The 9th edition of the Language Resources and Evaluation Conference*.

Connor, U. M., & Moreno, A. I. 2005. *Tertium comparationis: A vital component in contrastive rhetoric research*.

Croft, William. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press.

Crystal, David. 2000. *Language Death*. Cambridge: Cambridge University Press.

Czulo, Oliver. 2013. Constructions-and-frames analysis of translations: The interplay of syntax and semantics in translations between English and German. *Constructions and Frames* 5(2). 143–167.

Dolbey, Andrew., Ellsworth, Michael, & Jan Scheffczyk. 2006. BioFrameNet: A Domain-Specific FrameNet Extension with Links to Biomedical Ontologies. In *KR-MED* (Vol. 222).

Dux, Ryan. 2016. *A usage-based account of verb classes in English and German*. Unpublished Ph.D. dissertation, The University of Texas at Austin.

Dux, Ryan. 2018. Frames, verbs, and constructions: German constructions with verbs of stealing. In H.C. Boas & A. Ziem (eds.), *Constructional Approaches to Syntactic Structures in German*, 367–406. Berlin/Boston: de Gruyter Mouton.

Fellbaum, Christiane (ed.). 1998. *WordNet*. Boston, MA: MIT Press.

Fillmore, Charles J. 1968. The case for case. In E. Bach & R. Harms (eds.), *Universals in Linguistic Theory*, 1–90. New York: Rinehart & Winston.

Fillmore, Charles J. 1975. An alternative to checklist theories of meaning. *Proceedings of the First Annual Meeting of the Berkeley Linguistic Society (BLS)*. 123–131.

Fillmore, Charles J. 1977. Topics in Lexical Semantics. In P. Cole (ed.), *Current Issues in Linguistic Theory*, 76–136. Bloomington: Indiana University Press.

Fillmore, Charles J. 1982. Frame Semantics. In: *Linguistics in the Morning Calm*, ed. Linguistic Society of Korea, 111–38. Seoul: Hanshin.
Fillmore, Charles J. 1985. Frames and the Semantics of Understanding. *Quaderni di Semantica* 6. 222–254.
Fillmore, Charles J. 1986. Pragmatically controlled zero anaphora. *Proceedings of the Berkeley Linguistics Society*. 95–107.
Fillmore, Charles J. 2008. Border Conflicts: FrameNet meets Construction Grammar. *Proceedings of the XIII EURALEX International Congress* (Barcelona, 15–19 July 2008). 49–68.
Fillmore, Charles J. & Beryl T.S. Atkins. 1992. Toward a Frame-based Lexicon: The Semantics of RISK and its Neighbors. In A. Lehrer & E. Kittay (eds.), *Frames, Fields and Contrasts: New Essays in Semantic and Lexical Organization*, 75–102. Hillsdale: Erlbaum.
Fillmore, Charles J. & Berly T.S. Atkins. 2000. Describing polysemy: The case of crawl. In Y. Ravin & C. Leacock (eds.), *Polysemy: Linguistic and computational approaches*, 91–110. Oxford: Oxford University Press.
Fillmore, Charles J. & Colin Baker. 2010. A frames approach to semantic analysis. In B. Heine & H. Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*, 313–340. Oxford: Oxford University Press.
Fillmore, Charles J., Baker, Collin, & Beau Cronin 2003. The structure of the FrameNet database. *International Journal of Lexicography* 16.3. 281–296.
Fillmore, Charles J., Johnson, Chris, & Miriam Petruck. 2003. Background to FrameNet. *International Journal of Lexicography* 16. 235–251.
Fontenelle, Thierry. 1997. Using a bilingual dictionary to create semantic networks. *International Journal of Lexicography* 10.4. 275–303.
Gast, Volker. 2015. On the use of translation corpora in contrastive linguistics: A case study of impersonalization in english and german. *Languages in Contrast*, *15*(1). 4–33.
Goddard, Cliff 2001. Lexico-semantic universals: A critical overview. *Linguistic Typology* 5. 1–65.
Goddard, Cliff. 2005. *Semantic Analysis*. Oxford: Oxford University Press.
Granger, Sylviane. 2003. The corpus approach: a common way forward for Contrastive Linguistics and Translation Studies. *Corpus-based approaches to contrastive linguistics and translation studies*, *20*. 17.
Hansen-Schirra, Silvia, Nitzke, Jean, & Katharina Oster. 2017. Predicting cognate translation. In S. Hansen-Schirra, O. Czulo, & S. Hoffmann (ed.), *Empirical modelling of translation and interpreting*, 3–23. Berlin: Language Science Press.
Hasegawa, Yoko, Lee-Goldman, Russell, & Charles J. Fillmore. 2016. On the universality of semantic frames. In M. Hilpert & J.-O. Oestman (eds.), *Constructions across grammars*, 34–66. Amsterdam/Philadelphia: John Benjamins.
Heid, Ulrich. 1996. Creating a multilingual data collection for bilingual lexicography from parallel monolingual lexicons. In *Proceedings of the VIIth EURALEX International Congress*, Gothenburg 1996. 573–559.
Johansson, Stig. 2007. *Seeing through multilingual corpora*. Amsterdam/Philadelphia: John Benjamins.
Kilgarriff, Adam. 1997. I don't believe in Word Senses. *Computers and the Humanities* 31.2. 91–113.
Lambrecht, Knud. 1984. Formulaicity, frame semantics, and pragmatics in German binomial expressions. *Language*. 753–796.
Lehmann, Christian. 2018. Basic Vocabulary. [https://www.christianlehmann.eu/ling/ling_meth/ling_description/lexicography/basic_vocabulary.html]. Last accessed April 26, 2018.

Levin, Beth. 1993. *English verb class and alternations*. Chicago: Chicago University Press.
Levin, Beth & Malka Rappaport Hovav. 2005. *Argument Realization*. Cambridge: Cambridge University Press.
Lyngfelt, Benjamin. 2012. Re-thinking FNI. On null instantiation and control in Construction Grammar. *Constructions and Frames* 4.1. 1–23.
Lyngfelt, Benjamin, Bori, Lars, Ohara, Kyoko & Torrent, Tiago. (eds.). 2018. *Constructiography: Constructicon development across languages*. Amsterdam/Philadelphia: John Benjamins.
McEnery, Tony & Xiao, Richard. 2008. Parallel and comparable corpora: what is happening? In G. Anderman & M. Rogers (eds.), *Incorporating corpora: The linguist and the translator*, 18–31. Clevedon: Multilingual Matters Ltd.
Na, Liu & I.S.P. Nation. 1985. Factors affecting guessing vocabulary in context. *RELC journal*, *16*(1). 33–42.
Ohara, Kyoko. 2009. Frame-based contrastive lexical semantics in Japanese FrameNet: The case of risk and kakeru. In H.C. Boas (ed.), *Multilingual FrameNets: Methods and Applications*, 163–182. Berlin/New York: Mouton de Gruyter.
Ohara, Kyoko, Fujii, Seiko, Ohori, Toshio, Suzuki, Ryoko, Saito, Hiroaki, & Shun Ishizaki. 2004. The Japanese FrameNet Project: An introduction. In: *Fourth International Conference on Language Resources and Evaluation (LREC 2004). Proceedings of the Satellite Workshop "Building Lexical Resources from Semantically Annotated Corpora"*, 9–11.
Ogden, Charles K. 1930, *Basic English. A general introduction with rules and grammar*. London: Paul Treber (2nd impr. 1940). Cf.: http://ogden.basic-english.org/basiceng.html
Petruck, Miriam. 1986. Body part terminology in Hebrew: A study in lexical semantics. Ph.D. dissertation, University of California, Berkeley.
Petruck, Miriam R. L., Charles J. Fillmore, Collin Baker, Michael Ellsworth, & Josef Ruppenhofer. 2004. Reframing FrameNet data. In *Proceedings of the 11th EURALEX International Congress*, Lorient, France. 405–416.
Ruppenhofer, Josef. 2018. Argument omission in multiple German corpora. In H.C. Boas & A. Ziem (eds.), *Constructional Approaches to Syntactic Structures in German*, 204–244. Berlin/Boston: de Gruyter Mouton.
Ruppenhofer, Josef, Boas, Hans C., & Collin Baker. 2013. The FrameNet approach to relating syntax and semantics. In R.H. Gouws, U. Heid, W. Schweickhard & H.E. Wiegand (eds.), *Dictionaries. An International Encyclopedia of Lexicography*, 1320–1329. Berlin/New York: Mouton.
Ruppenhofer, Josef, Hans C. Boas, and Collin F. Baker. 2017. FrameNet. In P. A. Fuertes-Olivera (ed.), *The Routledge Handbook of Lexicography*, 383–398. New York: Routledge.
Salomão, Maria Margarida Martins, Tiago Timponi Torrent & Thais Fernandes Sampaio. 2013. A Linguística de Corpus Encontra a Linguística Computacional: Notícias do Projeto FrameNet Brasil. *Cadernos de Estudos Linguísticos*, 55(1),7–34.
Schmidt, Thomas. 2009. The Kicktionary – A multilingual lexical resource of football language. In H.C. Boas (ed.), *Multilingual FrameNets: Methods and Applications*, 101–134. Berlin/New York: Mouton de Gruyter.
Slobin, Dan. 1996. Two ways to travel: Verbs of motion in English and Spanish. In M. Shibatani and S. Thompson (eds.), *Grammatical constructions: Their form and meaning*, 195–217. Oxford: Oxford University Press.
Subirats, Carlos. 2009. Spanish FrameNet: A frame-semantic analysis of the Spanish lexicon. In: H.C. Boas (ed.), *Multilingual FrameNets: Methods and Applications*, 135–162. Berlin/New York: Mouton de Gruyter.

Subirats, Carlos & Miriam Petruck. 2003. Surprise: Spanish FrameNet! *Proceedings of CIL 17*. CD-ROM. Prague: Matfyzpress.

Subirats, Carlos & Hiroaki Sato. 2004. Spanish FrameNet and FrameSQL. In *Proceedings of {LREC} 2004. Workshop on Building Lexical Resources from Semantically Annotated Corpora*, Lisbon, Portugal.

VanNoy, Annika. 2017. *Culture specific aspects of semantic frames in multilingual frame descriptions*. Unpublished Ph.D. dissertation, The University of Texas at Austin.

Viberg, Åke. 2002. Polysemy and disambiguation cues across languages. In B. Altenberg & S. Granger (eds.), Lexis in Contrast, 119–150. Amsterdam/Philadelphia: John Benjamins.

Wierzbicka, Anna. 2005. *English. Meaning and Culture*. Oxford: Oxford University Press.

Ziem, Alexander & Hans C. Boas 2017. Towards a Constructicon for German. In *Proceedings of the AAAI 2017 Spring Symposium on Computational Construction Grammar and Natural Language Understanding*. Technical Report SS-17-02, Stanford University. 274–277.