Conceptual Annotations Preserve Structure Across Translations:
A French-English Case Study

Elior Sulem
Institute of Computer Science
Hebrew University of Jerusalem
eliors@cs.huji.ac.il

Omri Abend
School of Informatics
University of Edinburgh
oabend@inf.ed.ac.uk

Ari Rappoport
Institute of Computer Science
Hebrew University of Jerusalem
arir@cs.huji.ac.il

Abstract

Divergence of syntactic structures between languages constitutes a major challenge in using linguistic structure in Machine Translation (MT) systems. Here, we examine the potential of *semantic* structures. While semantic annotation is appealing as a source of cross-linguistically stable structures, little has been accomplished in demonstrating this stability through a detailed corpus study. In this paper, we experiment with the UCCA conceptual-cognitive annotation scheme in an English-French case study. First, we show that UCCA can be used to annotate French, through a systematic type-level analysis of the major French grammatical phenomena. Second, we annotate a parallel English-French corpus with UCCA, and quantify the similarity of the structures on both sides. Results show a high degree of stability across translations, supporting the usage of semantic annotations over syntactic ones in structure-aware MT systems.

1 Introduction

Structural information, be it syntactic or semantic, has the potential to address long-standing problems in Statistical Machine Translation (SMT), such as phrase-level (rather than word-level) reordering and discontiguous phrases. Structure-aware models[1] (Chiang, 2005; Liu et al., 2006; Mi et al., 2008) aim to address these and other problems by taking into account the hierarchical structure of language. However, while structure-aware

---
[1] We use the term "structure-aware" rather than "syntax-based" so to include any type of hierarchical structure.

models are effective at improving reordering at the phrase level, they are limited in their ability to map between arbitrarily divergent structures. Cross-linguistic divergences therefore pose a difficult problem for the integration of structural knowledge into statistical models (Dorr, 1994; Ding and Palmer, 2004; Zhang et al., 2008).

Consequently, an annotation scheme that assigns similar structures to translations has direct applicative value for structure-aware MT systems. Such structures can be used either as features in phrase-based systems, yielding more robust decoding, or as a structural scheme which directs the translation, replacing the PCFG trees often used today. Using more stable schemes is likely to result in simpler MT systems, avoiding structure modifications like pseudo-nodes (Marcu et al., 2006) or tree sequences (Zhang et al., 2008) used in syntax-based systems to handle cross-linguistic divergences.

Semantic annotation is an appealing avenue for constructing cross-linguistically stable structures, since a major goal of translation is to preserve the meaning of a sentence. Cross-linguistically stable schemes have further benefits for applications such as knowledge projection across languages (Kozhevnikov and Titov, 2013), the induction of cross-lingual semantic relations (Lewis and Steedman, 2013), or in translation studies (Lembersky et al., 2013) (see Section 7.3). A recent example of a semantic scheme aiming to be cross-linguistically stable is AMR (Abstract Meaning Representation) (Banarescu et al., 2013) which uses elaborate hierarchical structures in order to abstractly represent semantic information and presents promising preliminary results for SMT improvement (Jones et al., 2012). Nevertheless, the stability of semantic annotation across

11

translations is seldom addressed and has yet to be adequately supported (see Section 2), a gap we address in this paper using a detailed analysis of a semantically annotated parallel corpus.

Universal Cognitive Conceptual Annotation (UCCA) is a coarse-grained semantic annotation scheme which builds on typological and cognitive linguistic theory (Abend and Rappoport, 2013a; Abend and Rappoport, 2013b). The scheme aims to be applicable cross-linguistically, to abstract away from specific syntactic forms and to directly represent semantic distinctions. These properties make UCCA an appealing source of structural annotation which is cross-linguistically stable. We give an overview of UCCA in Section 3.

This paper focuses on the case study of English-French, a well studied language pair in MT. We demonstrate through this language pair both UCCA's *portability*, namely its ability to be applied to different languages, and its *stability*, namely its ability to preserve structure across translations. We conduct both type-level and token-level experiments to support our claim.

To verify UCCA's portability to French, we first conduct a type-level analysis by systematically examining UCCA's applicability over all major grammatical phenomena in French. We find that UCCA is fully applicable to French as exemplified in the case of French-specific phenomena like pronominal verbs (Section 4.1). Further in the type-level, we apply UCCA to a published inventory of structural divergences, and find that UCCA abstracts away from almost all of them (Section 4.2).

For a token-level analysis, we manually UCCA-annotated a parallel French-English corpus of over 25K tokens, which we make publically available, and compare the similarity between the UCCA structures in the two languages to the corresponding similarity between syntactic annotations. We find that UCCA is considerably less divergent than syntactic annotation (Section 6). We expect the relative stability of UCCA compared to syntactic schemes to be even greater in language pairs that are more syntactically different than the relatively similar English-French.

Finally, we analyze the semantic correspondence between the annotations on both sides of the parallel corpus (Section 7). We find remarkably high semantic correspondence between the two languages. For instance, over 92% of the Scenes (a similar notion to a "frame"; see Section 3) in both languages have a correspondent in the other. We analyze the non-corresponding units in the two languages according to various parameters, and show that many of them are due to ambiguity or semantic changes. These results offer a better understanding of UCCA's stability and suggest paths for further improvements.

2 Related Work

We begin by discussing previous work that studied the portability and stability of semantic schemes. We then briefly survey the means in which semantic information is integrated into MT systems.

Portability of semantic annotation. Several works addressed the portability of semantic annotation schemes, namely whether the same scheme, often originally developed for English, can be applied to other languages.

Burchardt et al. (2009) addressed the application of the English FrameNet (Baker et al., 1998) to German. They found that about a third of the verb senses identified in the German corpus were not covered by FrameNet. Their analysis further revealed that the English category set is not always sufficient, resulting in the introduction of a new category for German. Van der Plas et al. (2010) addressed the application of English Prop-Bank (Palmer et al., 2005) to French, and found that while the scheme can be applied to French, the annotation requires proficiency in both languages. Samardzic et al. (2010a; 2010b) also studied the portability of the English PropBank to French, and found that the overwhelming majority of the French verbal predicates in the corpus correspond to a verb sense in the PropBank lexicon. The portability of PropBank was also examined in the case of English-Chinese through the construction of annotated parallel corpora used in the OntoNotes project (Weischedel et al., 2012).

Portability has also been studied in the context of more elaborate hierarchical structures (Dorr et al., 2010; Banarescu et al., 2013), often with the intention of producing an *inter-language* – a representation independent of any specific language, which exhaustively accounts for the meaning of the sentence. Dorr et al. (2010) studied portability through the construction of a set of annotated parallel corpora in six languages, as part of the IAMTC project. Portability has also been investigated through the construction of annotated parallel treebanks such as the Prague Czech-English

Dependency Treebank[2], enabling a subsequent valency stability study (Urešová et al., 2015).

Stability of semantic annotation. Another line of work focused on the stability of specific schemes, i.e., their ability to preserve structure across translations. Fung et al. (2006; 2007) studied the stability of semantic role annotation between arguments in English and Chinese. They found that 83% of the alignable verbal arguments in English have a role-compatible argument in Chinese, but did not address arguments that have no correspondent in the other language. This motivated the use of semantic roles in MT, but also highlighted the existence of divergences between the structures in the two languages.

Semantic role schemes used in MT are generally restricted to verbal predicates, excluding several highly frequent constructions, such as copula clauses and nominalizations, which can result in a loss of stability. Furthermore, the fine-grained information such schemes provide as to the role of the arguments can be difficult to port across languages. For further discussion, see (Abend and Rappoport, 2013b) and (Birch et al., 2013).

Abstract Meaning Representation (AMR) (Banarescu et al., 2013) is a hierarchical semantic representation scheme whose aim is to provide simple, readable semantic annotation that can be applied cross-linguistically and assist MT systems. While UCCA is encoded over the text, AMR provides a structure for each sentence that is not trivially alignable with the text (Flanigan et al., 2014). Xue et al. (2014) studied the scheme's portability and stability when applied to English-Chinese and English-Czech parallel corpora. They annotated 100 Chinese and Czech sentences translated from English, and examined the similarities and differences of the AMRs across translations. In the English-Czech comparison, 53% of the sentences are reported to be structurally different in a non-local way. They conclude that at this point AMR is not stable enough to be used as an inter-language, but should be used only either on the target or on the source side.

Focusing on closer languages, namely English-French, we employ both type-level and token-level approaches for UCCA, including a comparison to syntax and a qualitative analysis of divergences, which are likely to generalize to some extent to other semantic annotations. We report a prelimi-

---

[2]`https://ufal.mff.cuni.cz/pcedt2.0/`

nary study of the stability of AMR in our corpus.

Integrating semantics into MT systems. Widely used in early MT (Uchida, 1987; Nirenburg, 1989), the integration of semantics into SMT systems is receiving much renewed interest in recent years. The first line of research is the integration of semantic features (often semantic roles) in SMT systems. In the phrase-based SMT models, they were mainly utilized for influencing reordering (Wu and Fung, 2009; Xiong et al., 2012; Feng et al., 2012). In syntax-based SMT models, semantic roles were involved in assisting reordering models (Li et al., 2013) and in translation rules (Zhai et al., 2012; Liu and Gildea, 2010; Bazrafshan and Gildea, 2013).

The second line of research concerns the use of an inter-language as an intermediary representation in SMT. Edelman and Solan (2009), relying on the cognitive model Revised Hierarchical Model (RHM), tried to represent the network of constructions that mediates between concepts and the channels of linguistic input and output. Jones et al. (2012) conducted preliminary experiments on a geographical querying domain using AMR.

3 UCCA Annotation

UCCA is a a semantic annotation scheme, strongly influenced by typological, notably Basic Linguistic Theory (Dixon, 2010a; Dixon, 2010b; Dixon, 2012), and cognitive linguistic theories (Langacker, 2008). The scheme aims to provide a coarse-grained, cross-linguistically applicable representation by directly reflecting the major semantic phenomena represented in the text and abstracting away from specific syntactic forms. We briefly introduce the UCCA formalism and main categories. For a more elaborate presentation, as well as evidence for the accessibility of UCCA to annotators with no linguistic background, see (Abend and Rappoport, 2013a; Abend and Rappoport, 2013b).

UCCA structures are directed acyclic graphs, where the words in the text correspond to (a subset of) their leaves. The nodes of the graphs, called *units*, are either terminals or several elements jointly viewed as a single entity according to some semantic or cognitive consideration. The edges bear one or more categories, indicating the role of the sub-unit in the relation that the parent represents.

UCCA is built as a multi-layered scheme, where each layer represents a different set of distinc-

tions. In this work we use the foundational layer of UCCA, which mostly addresses predicate-argument structures and linkage relations between them.

UCCA views the text as a collection of *Scenes* and relations between them. A Scene, the most basic notion of this layer, describes a movement, an action or a state which is persistent in time. Every Scene contains one main relation, or anchor (similar to frame-evoking element in FrametNet), and is labeled as a *State* (S) or a *Process* (P).

A Scene may contain one or more *Participants* (A), which are interpreted in a broad sense, and include locations, destinations and complement clauses. Secondary relations in the Scene, such as manner or temporal descriptions, are labeled as *Adverbials* (D). For example, the sentence "He slowly ran into the park" is annotated as follows: "[He]$_A$ [slowly]$_D$ [ran]$_P$ [into the park]$_A$".

The definitions of the UCCA categories are not dependent on POS distinctions. For instance, a Scene's main relation can be an adjective ("[He]$_A$ ['s thin]$_S$") or a noun ("[John 's]$_A$ [decision]$_P$").

4 Type-Level Analysis

In this section we focus on type-level analysis and show both the portability of UCCA, examining the annotation of the French grammatical phenomena with UCCA, and its stability, assessing UCCA's influence on commonly studied structural divergences.

4.1 Portability

We examine UCCA's applicability to French by systematically examining the major grammatical phenomena in French, and verifying that UCCA categories can be applied to them. To this purpose, we use the same annotation guidelines and category set previously applied to English, and apply it to the phenomena and examples described in a French grammar book (Hawkins and Towell, 2001). Tense and agreement are not covered in the UCCA foundational layer which we use, and are therefore disregarded in this work.

We find that even for French-specific phenomena, current UCCA categories permit their annotation in the foundational layer without requiring changes in the definitions or additional categories. Due to space limitations, we only present here one case of interest. The full analysis according to the grammar book can be found in Sulem (2014) (Ap-

pendix 2) [3].

As an example, we consider reflexive pronouns, representing the applicability of UCCA to French phenomena that have no direct parallel in English. In French, in addition to the counterparts of "himself" and "themselves" ("lui-même" and "eux-mêmes"), reflexivity is also expressed through the pronouns "se", "me", "te", "nous" and "vous", which precede some verbs (termed "pronominal verbs"). For instance, "lavé" is "washed", while "s'est lavé" is "washed himself". We show that the UCCA's category definitions can be applied naturally to this phenomenon.

A key guideline in UCCA is that the annotation of a unit does not depend on its part of speech but rather on its meaning and role in the context it is situated in. We therefore distinguish between three cases based on their semantics.

First, cases where the reflexive pronoun refers to the same Participant as the subject. Here the pronoun is annotated as an A: "[Jean]$_A$ [s']$_A$ [est lavé]$_P$" ("Jean washed himself").

Second, cases where the pronoun changes the meaning of the verb in an unpredictable way, or alternatively, where the verb may only appear in a pronominal form. In these cases the formal means of reflexivity is used, but is not associated with the semantic phenomenon of reflexivity. Semantically then, the reflexive pronoun and the verb form one unanalyzable unit, as in the following example: "Il [s' est aperçu]$_P$ qu'il était tard" ("He realized that it was late").

Third, cases where the pronoun changes the meaning and the number of arguments of the verb without creating semantic reflexivity. In these cases the verb is the Center (C) of the Process, while the reflexive pronoun serves as an Elaborator (E). For example: "Je [m'$_E$ appelle$_C$]$_P$ John" ("my name is John" where "appelle" means "call").

4.2 Stability

Overcoming cross-linguistic divergences (or translation divergences) is one of the main challenges in machine translation. We briefly review the main examples of translation divergences presented in (Dorr, 1994; Dorr et al., 2002; Dorr et al., 2004), adapting the original English-Spanish examples to English-French analogues. Then, for each example, we present its annotation according

---

[3] www.cs.huji.ac.il/~eliors/papers/
elior_sulem_thesis.pdf

to UCCA. The resulting annotations show that UCCA abstracts away from almost all of these divergences and exposes the semantic similarity, demonstrating the stability of the scheme at the type-level.

Categorical divergence: Translation of words in one language into words that have different POS tags in another language. For example, "to be cold" – "avoir froid" ("to have cold"). In UCCA the expression in both languages is annotated as a State where the Center (similar to the notion of a semantic head) is "cold" / "froid".

Conflational divergence: Translation of two or more words in one language into one word in another language. For example: "to kick" – "donner un coup de pied" ("give a kick"). In UCCA, the expression describes a Process in the two languages, and the French light verb "donner" ("give") is a Function (a unit which does not introduce a relation or participant) inside the Process.

Structural[4] **divergence**: Realization of verb arguments in different syntactic configurations in different languages. For example, "to enter the house" – "entrer dans la maison" ("enter in the house"). In UCCA there is a Participant in both languages.

Thematic divergence: Realization of verb arguments in syntactic configurations that reflect different thematic to syntactic mapping orders. For example, "I like this house" – "Cette maison me plaît" ("this house pleases to me"). In UCCA there are two Participants in English as well as two Participants in French ("cette maison" / "this house" and "me" / "me").

Promotional/Demotional divergence: Promotion is the case where a modifier in the source language is promoted to a main verb in the target language (Dorr, 1990; Gola, 2012). Demotion is its mirror image, where a main verb in the source language becomes a modifier in the target language.

An example where an English adverb is promoted to a main verb is the French: "John usually goes home" – "John a l'habitude de rentrer à la maison" ("John has the habit to go home"). In UCCA, both "usually" and "a l'habitude" ("has the habit") are annotated as Adverbials.

An example where an English verb is demoted to an adverb is the French "to run in" – "entrer en courant" ("enter running"). In UCCA, the En-

glish example contains a Process ("to run") and a Participant ("in"). The annotation in French is somewhat different, where "entrer" ("enter") is a Process, while "en courant" ("running") is an Adverbial.

To summarize, aside from the case of demotional divergence, the UCCA annotation (in its foundational layer) abstracts away from canonical examples for cross-linguistic divergences. With demotional divergence, where UCCA annotation is different across languages, we note that the divergence does correspond to a semantic difference of emphasis, that is, whether the entering action or the running action is the main relation. We leave it open whether this divergence should be considered a result of a true semantic difference between the languages or a shortcoming of UCCA that fails to capture the similarity between them.

5 Parallel French-English UCCA Corpus

The parallel corpus. The French-English corpus used here is an extract from the book *Twenty Thousand Leagues Under the Sea* (*Vingt Mille Lieues Sous les Mers*), a classic science fiction novel written in French by Jules Verne (1828–1905) and first published in 1870. We use an online version of the book and the English translation by J.P. Walter (Verne, 1870; Verne, 1991). Each of the two monolingual parts of the corpus contain 583 sentences which correspond to 12.5K tokens in English and 13.1K tokens in French. The annotated corpus is publically available[5].

Initial alignment. We segment the parallel corpus into 154 bilingual pairs of aligned passages. Each passage in French corresponds to a single passage in English. The passages correspond to the paragraphs in the original texts except in a few cases of long dialogues, where we split the paragraphs into several passages. A sentence-level alignment is not necessary in our analysis since in UCCA, the text is viewed as a collection of Scenes, where sentence boundaries play no significant role.

Manual annotation. The annotation was carried out using UCCA's web application. Both French and English texts were annotated by the same annotator (one of the authors of the paper), according to UCCA annotation guidelines[6]. Re-

---

[4]Here the term "structural" refers specifically to syntax, in contrast to the broader sense used elsewhere in the paper.

[5]`www.cs.huji.ac.il/~eliors`

[6]Both the web-application and the guidelines are available in `homepages.inf.ed.ac.uk/oabend/ucca/`.

cent updates to the guidelines concerning the annotation of secondary verbs as Adverbials, are not applied here. We expect these changes to further improve the quality of the results (Section 7.3). The annotation in English and French was carried out separately in each of the languages, rather than in parallel, thus permitting cases where the same linguistic form in English and French is subject to different interpretations, leading to different annotations. This effect on the differences in UCCA annotation in English and French is discussed in Section 7.

6 Token-level Analysis

In order to demonstrate UCCA's stability at a token-level, we examine the number of UCCA units of various types in both English and French for each parallel passage in our annotated parallel corpus. We compare these numbers to those obtained through syntactic annotation. In light of our type-level analysis (Section 4), we expect these UCCA categories to be more stable cross-linguistically than syntactic ones. The number of Scenes is compared to the number of non-auxiliary verbs, and the number of Participants and Adverbials is compared to the number of noun phrases (NPs), prepositional phrases (PPs) and adverb phrases (ADVPs).

We note that English-French is a particularly challenging candidate for this type of analysis since the language pair is relatively structurally similar (e.g., measured by word reordering (Birch, 2011)). Syntactic annotation is therefore a strong baseline. We expect UCCA's relative stability to be even greater in more syntactically divergent language pairs.

We are mainly interested not in the absolute number of units/constituents of a certain type, but more in the extent to which this number diverges between languages. Minimal divergence in the number of units/constituents of a certain type between the two languages is an indication of the scheme's stability.

We compute the similarity in the number of units/constituents of each type in the two languages in the following manner. For each language $l \in \{Fre, Eng\}$ and for each unit/constituent type t, we compute the number of instances of that type $n_i^{(t,l)}$ in each passage $i = 1, .., N$. We thereby obtain for each (t, l) a vector $n^{(t,l)} = \{n_i^{(t,l)}\}_i$. For each type t, the sim-

ilarity between $n^{(t,Fre)}$ and $n^{(t,Eng)}$, which is an indication of the stability of the scheme, is computed using l_1 and l_2 norms of the difference between them.

We further compute an F-score as follows: precision and recall of the French vector against the English one are defined respectively by $P = s/f$ and $R = s/e$ when $s = \sum_i \min(n_i^{(t,Fre)}, n_i^{(t,Eng)})$, $f = \sum_i n_i^{(t,Fre)}$ and $e = \sum_i n_i^{(t,Eng)}$. The F-score F is the harmonic mean of P and R. This measure provides an upper bound of the number of aligned units in the two languages, looking at the category of the units and their appearance in aligned passages. We note that the measures described are more applicable in this context than statistical correlation measures (e.g., the Pearson correlation coefficient). This is because a stable scheme is determined by the similarity of the count vectors in absolute terms, rather than their statistical correlation.

Experimental setup. For tagging, we use the Stanford POS tagger package (Toutanova et al., 2003). We compute the number of verbs in the parallel corpus and compare them to the number of Scenes. We exclude auxiliaries since such verbs tend to differ considerably between languages. We manually correct the tagging (by a single annotator, highly proficient in both languages), and therefore expect these numbers to be comparable in quality to a gold standard[7].

The syntactic constituents we study are noun phrases (*NP*), prepositional phrases (*PP*) and adverb phrases (*ADVP* in English and *AdP* in French). We used the Stanford parser's pre-trained models for English (*englishPCFG*, (Klein and Manning, 2003)) and French (the *frenchFactored* (Green et al., 2011)), with the same manual tokenization taken from the UCCA annotation. Six passages which contain very long sentences in French and for which the parser was unable to produce a parse were omitted from this evaluation. We note that we include in our analysis Scenes marked as unanalyzable (For example: "Hello!"), but exclude Scenes appearing as remote Participants, so to avoid double counting.

In order to correct for possible biases of the parsers towards overprediction or underprediction of certain syntactic constituents, we conduct the following experiment. We manually count the

---

[7]The French tagger overestimated the number of verbs by 0.6%, while the English tagger overestimated it by 8.7%.

	l_1	l_2	F	Fr. Avg.	En. Avg.
Scenes	124	14.97	0.96	9.25	9.49
Verbs	157	18.79	0.94	9.30	9.10
Participants (As)	273	31.13	0.95	17.68	18.27
NPs and PPs	952	102.74	0.89	26.64	32.33
NPs	847	88.89	0.87	18.78	24.20
PPs	299	32.05	0.87	7.86	8.13
Adverbials (Ds)	133	17.18	0.86	3.3	3.07
Adverb Phrases	342	40.0	0.15	0.24	2.49
As + Ds	334	37.18	0.95	20.99	21.34
NPs+PPs+ADVPs	1226	127.40	0.87	26.88	34.82

Table 1: Comparison of UCCA's Scene, Participant and Adverbial stability across the two languages with the stability of verbs, NPs, PPs and ADVPs. l_1 and l_2 represent respectively the l1 and l2 norms of the difference between the French and English count vectors. The F-score F, resulting from an upper bound on the number of aligned units in the two languages, evaluates the similarity between these vectors. The Scenes and the verbs are computed over the whole corpus (154 passages), while the other categories are computed on 148 passages (see text).

number of NPs, PPs and ADVPs in the first 10 passages in English and French, according to the original guidelines of the English and French Treebanks (Bies et al., 1995; Abeillé et al., 2004). All borderline cases are counted pessimistically, i.e., in the direction that maximizes the difference between the manual and automatic counts.

Results. Our results are given in Table 1. In all cases the UCCA annotation is more stable across annotations than the syntactic counterpart. The relative similarity between the number of PPs in the two languages, as reflected in the relatively low vector distances of $n^{(PP,Eng)}$ and $n^{(PP,Fre)}$, can be explained by the fact that the presence of a preposition in French usually requires a preposition in its English translation. PPs are also less affected than NPs by nominalizations which often result in cross-linguistic syntactic divergences[8]. Table 1 also presents the average number of units/constituents of each type per passage, on the two right columns. The latter numbers cannot be seen as a measure of stability, as an excessive number of units in one passage (relative to the translation) may cancel out a deficient number of units in another.

Concerning the correction term for the parsers' biases, we find that in the first 10 passages, the English parser overpredicted NPs by 12.2% and underpredicted ADVPs by 3.8%. The same num-

---

[8]The low number of French adverb phrases is partially due to the presence of some adverbial expressions that were tagged as multi-word adverbs (MWADV). If we consider MWADV as adverb phrases as well, the l_1 value is 292 and the l_2 value is 33.05, which is still much higher than the distances for UCCA's Adverbials (133 for l_1 and 17.18 for l_2).

ber of English PPs was obtained through manual and automatic counting. In these passages the French parser overpredicted NPs by 0.9% and PPs by 11.4%. The average difference between the results of the manual and automatic counting of French adverb phrases was 0.5. The biases are in an order of magnitude less than the relative differences in the l_1 and l_2 norms. Therefore, the stability of UCCA relative to syntactic schemes is not a result of the parsers' biases.

7 Divergence Analysis and Discussion

The analysis in Section 6 provides a comparison in terms of the number of units of specific types, as opposed to corresponding numbers of syntactic constituents. In this section we define a more refined methodology (Section 7.1) for examining not only the correspondence in the number of units between the languages, but also the semantic correspondence between units (Section 7.2 and 7.3).

7.1 Defining Divergences using UCCA

We define a *correspondence* between two UCCA annotations to be a one-to-one mapping which preserves UCCA's categories and meaning. Concretely, given a parallel corpus, a unit in one language corresponds to a unit in the other language if they have the same category and if the units have the same meaning. More formally, we define a *sufficient subset* of a unit u to be a subset of e that contains its heads (the main relation in the case of a Scene, or the Centers in the case of a non-Scene). For example, "He ran" is a sufficient subset of the Scene "He slowly ran" since it contains the main relation "ran". A unit e in English and a unit f in French correspond to each other if they have the same category and any of the three following conditions hold: (1) e is a translation of f, (2) a sufficient subset of e is a translation of f, or (3) a sufficient subset of f is a translation of e. For example, the English Scene "He slowly ran" corresponds to the French Scene "Il a couru" ("He ran") since condition (2) holds.

Given a UCCA category, some of the units of that category are left unaligned between the two sides of the parallel corpus, creating a *UCCA divergence*. We classify UCCA divergences according to their category, defining Scene, Participant and Adverbial divergences. We distinguish between divergences in the English and French sides.

An example of a UCCA divergence from our French-English corpus is: "of the ship victimized by this new ramming" – "du navire victime de ce

	Property	Scene Div.		Participant Div.		Adverbial Div.	
		Eng.	Fre.	Eng.	Fre.	Eng.	Fre.
	Translation Study						
1	Similar Translation Possible	65.18	58.33	50	35.29	70.83	50.0
2	Similar Source Possible	73.21	63.89	54.55	47.06	75.0	46.15
-	None	18.75	31.94	38.64	47.06	16.67	42.31
	Annotation Study						
3	Conforming Analysis	41.96	54.16	72.72	73.53	25.0	53.85
4	Different Interpretation	10.71	1.39	25	23.53	8.33	7.69
–	None	55.36	44.44	25	20.59	70.83	46.15
	Semantic Effect of the Unaligned Unit						
5	Additional Information	38.39	18.06	25.0	20.59	37.50	0.0
6	Tense Information	8.04	5.56	–	–	–	–
7	Emphasis	19.64	8.33	31.82	26.47	41.67	3.85
–	None	50.89	80.56	61.36	64.71	58.33	96.15

Table 2: Percentage of UCCA divergences according to their types (columns) that have certain properties (rows). All numbers are percentages computed over all UCCA divergences of the given type. Note that the properties are not mutually exclusive (see text). Participant and Adverbial divergences are only evaluated on passages with no Scene divergences.

nouvel abordage". The French noun "victime" describes a result, while the corresponding English "victimized" is an action. The unaligned Scene is in English. It is therefore an *English Scene divergence*. In the example "He slowly ran"/"Il a couru" we saw above, there is no Scene divergences but the English Adverbial "slowly" is unaligned, creating an *English Adverbial divergence*.

7.2 Number of UCCA Divergences

The analysis of Scene divergences is performed manually over the entire set of passages. The analysis of Participant and Adverbial divergences is restricted to passages with no Scene divergences, i.e., with a perfect Scene to Scene correspondence (57 passages of the total 154). This permits the capture of lower level divergences which are not just consequences of the divergences at the Scene level.

We found a total of 112 English Scene divergences and 72 French ones. This amounted to 92.3% of the English Scenes having a French correspondent and 94.9% of the French Scenes having an English correspondent. Only 25% of the sentences (148 out of 583) contains any Scene divergences.

Concerning Participant divergences, we found that 694 out of 738 English Participants (94.0%) have a correspondent in French. 694 of the 728 French Participants (95.3%) have a correspondent in English. 100 out of the 124 English Adverbials (80.6%) have a correspondent in French, and 100 out of the 126 Adverbials (79.4%) have a correspondent in English. Thus, our results show low rates of UCCA French-English divergences.

We also conduct a preliminary study into the applicability of another semantic scheme, namely AMR, to our domain. We annotate 10 sentence pairs with AMR. Our analysis shows that AMR conserves the main structures in most sentences (7 out of 10), and suggests that other semantic annotations may also be structurally stable. However, semantic roles, used in PropBank and AMR, are often a source of divergences across the languages.

7.3 Properties of UCCA Divergences

In order to examine the causes and semantic types of the different divergences, we manually classified each of them according to three groups of properties, which are not mutually exclusive. The results of the divergence analysis are presented in Table 2.

Translation study: The properties in this group investigate whether a given UCCA divergence can be avoided using a different formulation closer to the one used in the other language. This approach evaluates the translator's choices and creativity. Properties #1 and #2 check whether different formulations can be used in the source and target side respectively, that would avoid the UCCA divergence. Results show that many of the divergences can be indeed ascribed to the specific translation selected. For example, only less than a third of the Scene divergences in each language could not have been avoided through a different translation. We thus speculate that in a more technical and less literary corpus, the number of UCCA divergences will be lower.

Annotation study: These properties study the influence of the annotator's preferences. Property #3 (conforming analysis) covers cases where UCCA allows another analysis which would have avoided the divergence. While both annotations are permitted, one of them is sometimes preferred, to capture a nuance of meaning conveyed by one language but not the other. Property #4 refers to

Replaced by	Scene Div.		Participant Div.		Adverbial Div.	
	Eng.	Fre.	Eng.	Fre.	Eng.	Fre.
Linker	6.25	1.39	–	–	8.33	7.69
Ground	1.79	1.39	–	–	4.17	3.85
Elaborator of Participant	–	–	0	2.94	4.17	19.23
Main Relation	–	–	20.45*	20.59*	25.0*	26.92*
Parallel Scene	–	–	13.64	2.94	–	–
Participant	–	–	–	–	4.17	11.54
Adverbial	–	–	6.82	2.94	–	–
2 Participants	–	–	11.36	2.94	–	–
2 Adverbials	–	–	–	–	4.17	0.0
None	91.96	98.21	47.73	67.65	50.0	30.77

Table 3: Analysis of divergences in terms of replacements by other UCCA categories. Columns correspond to divergence types, while rows correspond to the category, as defined in Abend and Rappoport (2013b), of the replacing unit. All numbers are given in percents. Percentage is taken over all UCCA divergences of the same type. *: In these cases, a Participant or an Adverbial in one of the languages is included in the meaning of the main relation (Process or State) in the other language.

divergences resulting from different readings (ambiguity) allowed by the text, where one meaning was selected in one language and another in the other. The results for this group (properties #3 and #4) reveal that most of the Scene and Adverbial divergences could have been avoided had a different annotation been selected. This suggests that more restrictive annotation guidelines or some post-annotation normalization can substantially reduce the number of divergences.

Effect of the unaligned unit: Divergences are often a result of a semantic or pragmatic difference between the source text and its translation. Property #5 addresses cases where additional information is conveyed by the unaligned unit. Property #6 is a sub-case of #5 that specifically addresses tense information. Property #7 addresses cases where the unaligned unit emphasizes some aspect of meaning. The results show that many divergences can be ascribed to a true semantic difference between the source and the translation.

Finally, in some cases, the UCCA divergences simply replace one UCCA category with another (Table 3). In these cases there are unaligned units in the English and the French sides that roughly correspond to one another semantically, but have different UCCA categories. Cases of replacement are common with Participant and Adverbial divergences, but fairly rare in the case of Scene divergences. In case of Adverbial divergences, many of them result from including the meaning of an Adverbial in one language in the meaning of the main relation (Process or State) in the other language. This can be seen as a generalization of demotional/promotional divergences (Dorr, 1994) discussed in Section 4.2. Annotating secondary verbs (e.g., "begin" or "try") as Adverbials instead of being part of the main relation, as was done in the latest version of UCCA's guidelines, may con-

siderably reduce this kind of divergence.

To summarize, our study sheds light on the circumstances in which UCCA divergences arise and suggests how many divergences can be avoided. This study also contributes to the understanding of the differences between original and translated texts, which can improve MT (Lembersky et al., 2013).

8 Conclusion

We showed that basic semantic structures can be stably preserved across English-French translations. This means that semantic structures may be more suitable to SMT systems than syntactic ones, which exhibit well known divergence phenomena. We used the UCCA scheme, but we expect these advantages to generalize to other structured semantic schemes. Future work will address the integration of UCCA into structure-based SMT either by adding UCCA as features to phrase-based and syntax-based systems, or by replacing existing syntactic structures with UCCA structures. We also plan to investigate related tasks that would benefit from UCCA's stability like bilingual alignment and MT evaluation.

References

Anne Abeillé, François Toussenel, and Martine Chéradame, 2004. *Corpus le Monde Annotation en constituents Guide pour les cor-*

recteurs. `http://www.llf.cnrs.fr/Gens/Abeille/guide-annot.pdf`.

Omri Abend and Ari Rappoport. 2013a. UCCA: A semantic-based grammatical annotation scheme. In *Proc. of IWCS-13*, pages 1–12.

Omri Abend and Ari Rappoport. 2013b. Universal Conceptual Cognitive Annotation (UCCA). In *Proc. of ACL-13*, pages 228–238.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley Framenet project. In *Proc. of ACL-COLING-98*, pages 86–90.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for sembanking. In *Proc. of Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186.

Mazrieh Bazrafshan and Daniel Gildea. 2013. Semantic roles for string to tree machine translation. In *Proc. of ACL-13 (Short Paper)*, pages 419–423.

Ann Bies, Mark Ferguson, Karen Katz, and Robert MacIntyre, 1995. *Bracketting Guidelines for Treebank II Style Penn Treebank Project*. Linguistic Data Consortium.

Alexandra Birch, Barry Haddow, Ulrich Germann, Maria Nadejde, Chrsitian Buck, and Philipp Koehn. 2013. The feasibility of HMEANT as a human MT evaluation metric. In *Proc. of the 8th Workshop on SMT, ACL-13*, pages 52–61.

Alexandra Birch. 2011. *Reordering metrics for statistical machine translation*. Ph.D. thesis, University of Edinburgh.

Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Padó, and Manfred Pinkal. 2009. Using FrameNet for the semantic analysis of German: Annotation, representation and automation. In Hans C. Boas (Herausgeber), editor, *Multilingual FrameNets in Computational Lexicography - Methods and Applications*, pages 209–244, New York/Berlin. Mouton de Gruyter.

David Chiang. 2005. A hierarchical phrase-based model for statistical machine translation. In *Proc. of ACL-05*, pages 263–270.

Yuan Ding and Martha Palmer. 2004. Synchronous dependency insertion grammars: a grammar formalism for syntax based statistical MT. In *Workshop on Recent Advances in Dependency Grammars, COLING-04*.

Robert M.W. Dixon. 2010a. *Basic Linguistic Theory: Grammatical Topics*, volume 2. Oxford University Press.

Robert M.W. Dixon. 2010b. *Basic Linguistic Theory: Methodology*, volume 1. Oxford University Press.

Robert M.W. Dixon. 2012. *Basic Linguistic Theory: Further Grammatical Topics*, volume 3. Oxford University Press.

Bonnie J. Dorr, Lisa Pearl, Rebecca Hwa, and Nizard Habash. 2002. DUSTer: a method for unraveling cross-language divergences for statistical word-level alignment. In *Proc.of AMTA-02*, pages 31–43.

Bonnie J. Dorr, Eduard H. Hovy, and Lori S. Levin. 2004. Machine translation: interlingual methods. In *Encyclopedia of language and linguistics, 2nd edition*. ms.939, Brown, Keith.

Bonnie Dorr, Rebecca Passonneau, David Farwell, Rebecca Green, Nizar Habash, Stephen Helmreich, Edward Hovy, Lori Levin, Keith Miller, Teruko Mitamura, Owen Rambow, and Advaith Siddharthan. 2010. Interlingual annotation of parallel text corpora: A new framework for annotation and evaluation. *Natural Language Engineering*, pages 197–243.

Bonnie Dorr. 1990. Solving thematic divergences in machine translation. In *Proc. of ACL-90*, pages 127–134.

Bonnie J. Dorr. 1994. Machine translation divergences: a formal description and proposed solution. *Computational linguistics*, 20(4):597–635.

Shimon Edelman and Zach Solan. 2009. Machine translation using automatically inferred construction-based correspondence and language models. In *Proc. of PACLIC-09*, pages 654–661.

Minwei Feng, Wiwei Sun, and Hermann Ney. 2012. Semantic cohesion model for phrase-based SMT. *Proc. of COLING-12*, pages 867–878.

Jeffrey Flanigan, Sam Thomson, Jaime Carbonell, Chris Dyer, and Noah A. Smith. 2014. A discriminative graph-based parser for the Abstract Meaning Representation. In *Proc. of ACL-14*, pages 1426–1436.

Pascale Fung, Zhojun Wu, Yongsheng Yang, and Dekai Wu. 2006. Learning of Chinese/English semantic structure mapping. In *Workshop on Spoken Language Technology, IEEE/ACL-06*, pages 230–233.

Pascale Fung, Zhaojun Wu, Yongsheng Yang, and Dekai Wu. 2007. Learning bilingual semantic frames: shallow semantic parsing vs. semantic role projection. In *Proc. of the 11th Conference on Theoretical and Methodological Issues in Machine Translation (TMI 2007)*, pages 75–84.

Francesca Gola. 2012. An analysis of translation divergence patterns using PanLex translation pairs. Master's thesis, University of Washington.

Spence Green, Marie-Catherine de Marneffe, John Bauer, and Christopher D. Manning. 2011. Multiword expression identification with Tree Substitution Grammars: a parsing tour de force with French. In *Proc. of EMNLP-11*, pages 725–735.

Roger Hawkins and Richard Towell. 2001. *French Grammar and Usage*. McGraw-Hill, 2nd edition.

Bevan Jones, Jacob Andreas, Daniel Bauer, Karl Moritz Hermann, and Kevin Knight. 2012. Semantics-based machine translation with hyperedge replacement grammars. In *Proc. of COLING-12*, pages 1359–1376.

Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proc. of ACL-03*, pages 423–430.

Mikhail Kozhevnikov and Ivan Titov. 2013. Cross-lingual transfer of semantic role labeling models. In *Proc. of ACL-13*, pages 1190–1200.

Ronald W. Langacker. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford University Press, USA.

Gennadi Lembersky, Noam Ordan, and Shuly Wintner. 2013. Improving statistical machine translation by adapting translation models to translationese. *Computational Linguistics*, 39(4):999–1023.

Mike Lewis and Mark Steedman. 2013. Unsupervised induction of cross-lingual semantic relations. In *Proc. of EMNLP-13*, pages 681–692.

Junhui Li, Philip Resnik, and Hal Daumé III. 2013. Modeling syntactic and semantic structures in hierarchical phrase-based translation. In *Proc. of NAACL-HLT-13*, pages 540–549.

Ding Liu and Daniel Gildea. 2010. Semantic role features for machine translation. In *Proc. of COLING-10*, pages 716–724.

Yang Liu, Qun Liu, and Shouxun Lin. 2006. Tree-to-string alignment template for statistical machine translation. In *Proc. of COLING-ACL-06*, pages 609–616.

Daniel Marcu, Wei Wang, Abdessamad Echihabi, and Kevin Knight. 2006. SPMT: Statistical machine translation with syntactified target language phrases. In *Proc. of EMNLP-06*, pages 44–52.

Haitao Mi, Liang Huang, and Qun Liu. 2008. Forest-based translation. In *Proc. of ACL-08 HLT*, pages 192–199.

Sergei Nirenburg. 1989. New developments in knowledge-based machine translation. In Alatis J.E., editor, *Georgetown University Round Table on Languages and Linguistics 1989: Language teaching, testing, and technology: lessons from the past with a view toward the future*, pages 344–357. Georgetown University Press.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: A annotated corpus of semantic roles. *Computational Linguistics*, 31(1):149–159.

Tanja Samardžić, Lonneke van der Plas, Goljihan Kashaeva, and Paola Merlo. 2010a. The scope and the sources of variation in verbal predicates in English and French. In *Proc. of the 9th International Workshop on Treebanks and Linguistic Theories*, pages 199–211.

Tanja Samardžić, Lonneke van der Plas, Goljihan Kashaeva, and Paola Merlo. 2010b. Variation in verbal predicates in English and French. *Generative Grammar in Geneva*, 6:109–135.

Elior Sulem. 2014. Integration of a cognitive annotation into machine translation: Theoretical foundations and bilingual corpus analysis. Master's thesis, Hebrew University of Jerusalem.

Kristina Toutanova, Dan Klein, Christopher Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proc. of HLT-NAACL-03*, pages 252–259.

Hiroshi Uchida. 1987. ATLAS: Fujitsu machine translation system. In *Machine Translation Summit*, Japan.

Zdeňka Urešová, Ondřej Dušek, Eva Fučíková, Jan Hajič, and Jana Šindlerová. 2015. Bilingual English-Czech valency lexicon linked to a parallel corpus. In *Proc. of the 9th Linguistic Annotation Workshop (The LAW IX)*, pages 124–128.

Lonneke van der Plas, Tanja Samardžić, and Paola Merlo. 2010. Cross-lingual validity of PropBank in the manual annotation of French. In *Proc. of the 4th Linguistic Annotation Workshop (The LAW IV)*, pages 113–117.

Jules Verne. 1870. *Vingt Mille Lieues Sous les Mers*. J. Hetzel. `http://fr.wikisource.org/wiki/Vingt_mille_lieues_sous_les_mers`.

Jules Verne. 1991. Twenty Thousands Leagues Under the Sea. Translated from the original French by J.P. Walter. `http://jv.gilead.org.il/fpwalter`.

Ralph Weischedel, Sameer Pradhan, Lance Ramshaw, Jeff Kaufman, Michelle Franchini, Mohammed El-Bachouti, Nianwen Xue, Martha Palmer, Jena D. Hwang, Claire Bonial, Jinho Choi, Aous Mansouri, Maha Foster, Abdel aati Hawwary, Mitchell Marcus, Ann Taylor, Craig Greenberg, Eduard Hovy, Robert Belvin, and Ann Houston. 2012. OntoNotes Release 5.0. Technical report, Linguistic Data Consortium 2013T19.

Dekai Wu and Pascale Fung. 2009. Semantic roles for SMT: a hybrid two-pass model. In *Proc. of NAACL-HLT-09 (Short Paper)*, pages 13–19.

Deyi Xiong, Min Zhang, and Haizhou Li. 2012. Modeling the translation of predicate-argument structure for SMT. In *Proc. of ACL-12*, pages 902–911.

Nianwen Xue, Ondřej Bojar, Jan Hajič, Martha Palmer, Zdeňka Urešová, and Xiuhong Zhang. 2014. Not an interlingua, but close: comparison of English AMRs to Chinese and Czech. In *Proc. of LREC-14*, pages 1765–1772.

Feifei Zhai, Jiajun Zhang, Yu Zhou, and Chengquing Zong. 2012. Machine translation by modeling predicate-argument structure transformation. In *Proc. of COLING-12*, pages 3019–3036.

Min Zhang, Hongfei Jiang, Aiti Aw, Haizhou Li, Chew Lim Tan, and Seng Li. 2008. A tree sequence alignment-based tree-to-tree translation model. In *Proc. of ACL-08*, pages 559–567.