DUSTer: A Method for Unraveling Cross-Language Divergences for Statistical Word-Level Alignment

Bonnie J. Dorr, Lisa Pearl, Rebecca Hwa, Nizar Habash

Institute for Advanced Computer Studies
University of Maryland, College Park, MD 20740
{bonnie,llsp,hwa,habash}@umiacs.umd.edu
http://umiacs.umd.edu/labs/CLIP

Abstract. The frequent occurrence of *divergences*—structural differences between languages—presents a great challenge for statistical word-level alignment. In this paper, we introduce DUSTer, a method for systematically identifying common divergence types and transforming an English sentence structure to bear a closer resemblance to that of another language. Our ultimate goal is to enable more accurate alignment and projection of dependency trees in another language without requiring any training on dependency-tree data in that language. We present an empirical analysis comparing the complexities of performing word-level alignments with and without divergence handling. Our results suggest that our approach facilitates word-level alignment, particularly for sentence pairs containing divergences.

1 Introduction

Word-level alignments of bilingual text (bitexts) are not only an integral part of statistical machine translation models, but also useful for lexical acquisition, treebank construction, and part-of-speech tagging [26]. The frequent occurrence of *divergences*—structural differences between languages—presents a great challenge to the alignment task.[1] In this paper, we introduce DUSTer (Divergence Unraveling for Statistical Translation), a method for systematically identifying common divergence types and transforming the structure of an English sentence to bear a closer resemblance to that of another language.[2] Our ultimate goal is to enable more accurate projection of dependency trees for non-English languages without requiring any training on dependency-tree data in those languages. (For ease of readability, we will henceforth refer to *non-English* as *foreign*.) The bitext is parsed on the English side only. Thus, the projected trees in the foreign language may serve as input for training parsers in a new language.

A divergence occurs when the underlying concepts or gist of a sentence is distributed over different words for different languages. For example, the notion

[1] The term *divergence* refers only to differences that are relevant to predicate-argument structure, i.e., we exclude constituent re-orderings such as noun-adjective swapping which occurs between Spanish and English. See [25] for an approach that involves syntactic reorderings of this type.

[2] See http://www.umiacs.umd.edu/labs/CLIP/DUSTer.html for more details.

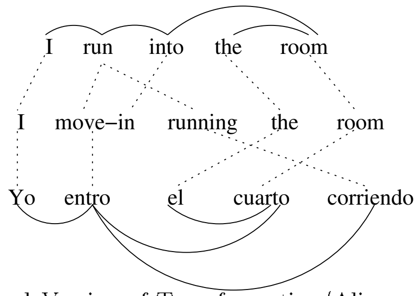

Fig. 1. Idealized Version of Transformation/Alignment/Projection

of running into the room is expressed as *run into the room* in English and *move-in the room running* (*entrar el cuarto corriendo*) in Spanish. While seemingly transparent for human readers, this throws statistical aligners for a serious loop. Far from being a rare occurrence, our preliminary investigations revealed that divergences occurred in approximately 1 out of every 3 sentences.[3] Thus, finding a way to deal effectively with these divergences and repair them would be a massive advance for bilingual alignment.

The following three ideas motivate the development of automatic "divergence correction" techniques:

1. Every language pair has translation divergences that are easy to recognize.
2. Knowing what they are and how to accommodate them provides the basis for refined word-level alignment.
3. Refined word-level alignment results in improved projection of structural information from English to another language.

This paper elaborates primarily on points 1 and 2. Our ultimate goal is to set these in the context of 3, i.e., for training foreign-language parsers to be used in statistical machine translation.

DUSTer transforms English into a pseudo-English form (which we call E') that more closely matches the physical form of the foreign language, e.g., "run into the room" is transformed to a form that roughly corresponds to "move-in the room running" if the foreign language is Spanish. This rewriting of the English sentence increases the likelihood of one-to-one correspondences which, in turn, facilitates our statistical alignment process. In theory, our rewriting approach applies to all divergence types. Thus, given a corpus, divergences are identified, rewritten, and then run through the statistical aligner of choice.

The idealized version of our transformation/alignment/projection approach is illustrated for an English-Spanish pair in Figure 1. Dependencies between English words (**E**) are represented by the curves above the words—these are produced by the Minipar system [15, 16]. Alignments are indicated by dotted lines. The dependency trees are transformed into new trees associated with E', e.g., *run* and *into* in **E** are reconfigured in E' so that the sentence in E' has a one-to-one correspondence with the sentence of the foreign language **F**. The final step—outside the scope of this paper—is to induce foreign-language dependency

[3] This analysis was done using automatic detection techniques—followed by human confirmation—on a sample size of 19K sentences from the TREC El Norte Newspaper (Spanish) Corpus, LDC catalog no LDC2000T51, ISBN 1-58563-177-9, 2000.

trees automatically using statistical alignment of the E' words with those of the foreign-language sentence (e.g., using Giza++ [1, 20]).

The next section sets this work in the context of related work on alignment and projection of structural information between languages. Section 3 describes the range of divergence types covered in this work—and analyzes the frequency of their occurrence in corpora (with examples in Spanish and Arabic). Section 4 describes an experiment that reveals the benefits of injecting linguistic knowledge into the alignment process. We present an empirical analysis comparing the complexities of performing word-level alignments with and without divergence handling. We conclude that annotators agree with each other more consistently when performing word-level alignments on bitext with divergence handling.

2 Related Work

Recently, researchers have extended traditional statistical machine translation (MT) models [4, 5] to include the syntactic structures of the languages [2, 3, 23]. These statistical transfer systems appear to be similar in nature to what we are proposing—projecting from English to a foreign-language tree—but both the method of generation and the goal behind these approaches are different from ours. In these alternative approaches, parses are generated simultaneously for both sides whereas, in our approach, we assume we only have access to the English parses and then we automatically produce dependency trees in another language *without training*.[4] From these noisy foreign-language dependency trees, we then induce a parser for translation between the foreign language and English. The foreign-language parse is a necessary input to a generation-heavy decoder [8], which produces English translations from foreign-language dependency trees.

It has been shown that MT models are significantly improved when trained on syntactically *annotated* data [25]. However, the cost of human labor in producing annotated treebanks is often prohibitive, thus rendering manual construction of such data for new languages infeasible. Some researchers have developed techniques for fast acquisition of hand-annotated Treebanks [7]. Others have developed machine learning techniques for inducing parsers [9, 10], but these require extensive collections of complex translation pairs for broadscale MT.

Because divergences generally require a combination of lexical and structural manipulations, they are handled traditionally through the use of transfer rules [12, 13]. Unfortunately, automatic extraction of such rules relies crucially the availability of scarce resources such as large, aligned, and parsed, bilingual corpora [14, 18, 19, 22].

Our approach requires parsing on only the English side of the aligned bilingual corpora—the foreign language need not be parsed. We detect and handle divergences using linguistically motivated techniques to transform the English

[4] It is important to note that rewriting the English structure as a structure in the foreign language is not intended to be an MT transfer process unto itself, but rather it is a first step in constructing a (noisy) foreign-language treebank for training a new parser for MT.

Type	English	E'	Foreign Equivalent
Light Verb	fear	have fear	tiene miedo
	try	put to trying	poner a prueba
	make any cuttings	wound	تجرحوا
	our hand is high	our hand heightened	يدنا قد عظمت
	he is not here	he be-not here	إنه ليس هنا
Manner	teaches	walks teaching	anda enseñando
	is spent	self goes spending	se va gastando
	he sent his brothers away	he dismissed his brothers	صرف إخوته
	spake good of	speak-good about	يثني على
	he turned again	he returned	رجع
Structural	after six years	after of six years	después de seis años
	and because of that in other parts	and for that in other parts	y por ello en otras partes
	I forsake thee	I-forsake about-you	اتخلى عنك
	we found water	we-found on-water	عثرنا على
Categorial	I am jealous	I have jealousy	tengo celos
	I require of you	I require-of you	te pido
	(he) shall estimate	according-to (his)-estimate	حسب تقدير
	(how long shall) the land mourn	(stays) the-land mourning	الارض نائحة
	he went to	his-return to	عودته الى
Head-Swapping[5]	walked out	move-out walking	salió caminando
Thematic	I am pained	me pain they	me duelen
	He loves it	to him be-loved it	le gusta
	it was on him	he-wears-it	كان يرتديه

Table 1. Examples of True English, E', and Foreign Equivalent

lexical and syntactic representation to match the physical form of the foreign language more closely—thus improving alignment. The ultimate goal is to bring about more accurate dependency-tree projection from English into the foreign language, thus producing a significantly noise-reduced dependency treebank for training foreign-language parsers.

3 Frequency of Divergences in Large Corpora

We investigated divergences in Arabic and Spanish corpora to determine how often such cases arise.[6] Our investigation revealed that there are six divergence types of interest. Table 1 shows examples of each type from our corpora, along with examples of sentences that were aligned with the foreign language sentences in our experiment (including both English and E').

Space limitations preclude a detailed analysis of each divergence type, but see [6] for more information. In a nutshell, Light Verb divergence involves the translation of a single verb to a combination of a "light" verb (carrying little or no specific meaning in its own right) and some other meaning unit (perhaps a noun) to convey the appropriate meaning. Manner divergence involves

[5] Although cases of Head Swapping arise in Arabic, we did not find any such cases in the small sample of sentences that we human checked in the Arabic Bible.

[6] For Spanish, we used TREC Spanish Data; for Arabic, we used an electronic, version of the Bible written in Modern Standard Arabic.

Spanish	**Arabic**
hacer (do)	ليس (be-not)
dar (give)	عبر (go-across)
tomar (take)	أحسن (do-good)
tener (have)	اكتفى (make-do)
poner (put)	أخرج (take-out)
ir + X-progressive (go X-ing)	رجع (come-again)
andar + X-progressive (walk X-ing)	غرّب (go-west)
salir + X-progressive (leave X-ing)	أسرع (do-quickly)
pasar + X-progressive (pass X-ing)	جرح (make-cuttings)
entrar + X-progressive (enter X-ing)	عظم (become-high/great)
bajar + X-progressive (go-down X-ing)	صرف (send-away)
irse + X-progressive (leave X-ing)	ارتعب (be-afraid)
soler (usually)	أثنى + على (speak-good + about = laud)
gustar (like)	بحث + عن (search + for = seek)
bastar (be enough)	أوصى + ب (command + with = command)
disgustar (dislike)	تخلى + عن (abandon + of = forsake)
quedar (be left over)	عثر + على (find + on = find)
doler (hurt)	عودة$_{noun}$ + الى (return$_{verb}$ to)
encantar (be enchanted by)	
importar (be important)	
interesar (interest)	
faltar (be lacking)	
molestar (be bothered by)	
fascinar (be fascinated by)	

Table 2. Common Search Terms for Divergence Detection

translating a single manner verb (e.g., *run*) as a light verb of motion and a manner-indicating content word. Structural divergence involves the realization of incorporated arguments such as subject and object as obliques (i.e. headed by a preposition in a PP). Categorial divergence involves a translation that uses different parts of speech. Head swapping involves the demotion of the head verb and the promotion of one of its modifiers to head position. Finally, a thematic divergence occurs when the verb's arguments switch thematic roles from one language to another.

In order to conduct this investigation, we developed a set of hand-crafted regular expressions for detecting divergent sentences in Arabic and Spanish corpora (see Table 2).[7] The Arabic regular expressions were derived by examining a small set of sentences (50), a process which took approximately 20 person-hours. The Spanish expressions were derived by a different process—involving a more general analysis of the behavior of the language—taking approximately 2 person-months. We want to emphasize that these regular expressions are *not* a sophisticated divergence detection technique. However, they do establish, at the very least, a conservative lower bound for how often divergences occur since the regular expressions pull out select cases of the different divergence types.

[7] The regular expressions are overgenerative in the current version, i.e., the system detects more seemingly divergent sentences than actually exist. Thus, we require human post-checking to eliminate erroneous cases. However, a more constrained automated version is currently under development—to be released in 2002—that requires no human checking. We aim to verify the accuracy of the automated version using some of the test data developed for this earlier version.

Language	Detected Divergences	Human Confirmed	Sample Size (sentences)	Corpus Size (sentences)
Spanish	11.1%	10.5%	19K	150K
Arabic	31.9%	12.4%	1K	28K

Table 3. Divergence Statistics

In our investigation, we applied the Spanish and Arabic regular expressions to a sample size of 19K Spanish sentences from TREC and 1K Arabic sentences from the Arabic Bible. Each automatically detected divergence was subsequently human verified and categorized into a particular divergence category. Table 3 indicates the percentage of cases we detected automatically and also the percentage of cases that were confirmed (by humans) to be actual cases of divergence.

It is important to note that these numbers reflect the techniques used to calculate them. The Arabic regular expressions were constructed more compactly than the Spanish ones in order to increase the number of verb forms that could be caught with a single expression. For example, a regular expression for a transitive verb includes the perfect and imperfect forms of the verb with various prefixes for conjugation, aspect, and tense and suffixes for pronominal direct objects. Because the Spanish regular expressions were derived through a more general language analysis, the precision is higher in Spanish than it is in Arabic. Human inspection confirmed approximately 1995 Spanish sentences out of the 2109 that were automatically detected (95% accuracy), whereas whereas 124 sentences were confirmed in the 319 detected Arabic divergences (39% accuracy).

On the other hand, the more constrained Spanish expressions appear to give rise to a lower recall. In fact, an independent study with more relaxed regular expressions on the same 19K Spanish sentences resulted in the automatic detection of divergences in 18K sentences (95% of the corpus), 6.8K of which were confirmed by humans to be correct (35% of the corpus). Future work will involve repeated constraint adjustments on the regular expressions to determine the best balance between precision and recall for divergence detection; we believe the Arabic expressions fall somewhere in between the two sets of Spanish expressions (which are conjectured to be at the two extremes of constraint relaxation—very tight in the case above and very loose in our independent study).

4 Experiment: Impact of Divergence Correction on Alignment

To evaluate our hypothesis that transformations of divergent cases can facilitate the word-level alignment process, we have conducted human alignment studies for two different pairs of languages: English-Spanish and English-Arabic. We have chosen these two pairings to test the generality of the divergence transformation principle.

Our experiment involves four steps:

i. Identify canonical transformations for each of the six divergence categories.
ii. Categorize English sentences into one of the 6 divergence categories (or "none") based on the foreign language.

iii. Apply the appropriate transformations to each divergence-categorized English sentence, renaming it E'.
iv. For each language:
   - Have two humans align the true English sentence and the foreign-language sentence.
   - Have two different humans align the rewritten E' sentence and the foreign-language sentence.
   - Compare inter-annotator agreement between the first and second sets.

We accommodate divergence categories by rewriting dependency trees produced by the Minipar system so that they are parallel to what would be the equivalent foreign-language dependency tree. Simultaneously, we automatically rewrite the English sentence as E'. For example, in the English-Spanish case of *John kicked Mary*, our system rewrites the English dependency tree as a new dependency tree corresponding to the sentence *John gave kicks to Mary*. The resulting E' (which would be seen by the human aligner in our experiment) is: 'John LightVB kick Prep Mary'.

The canonical transformation rules that map an English sentence and dependency tree to E' (and its associated dependency tree) are shown in Table 4. These rules fall into two categories, those that facilitate the task of alignment *and* enable more accurate projection of dependency trees (light verb, manner, and structural)—and those that *only* enable more accurate projection of dependency trees with minimal or no change to alignment accuracy (categorial, head-swapping, and thematic). This paper focuses on the first of these two categories.[8] In this category, there are of two types of rules: "expansion rules," applicable when the foreign language sentence is verbose relative to the English one, and "contraction rules," applicable when the foreign language sentence is terse relative to English.[9]

I. Rules Impacting Alignment and Projection
(1) **Light Verb**
Expansion: [Arg1 [V]] → [Arg1 [LightVB] Arg2(V)]
Ex: "I fear" → "I have fear"
Contraction: [Arg1 [LightVB] Arg2] → [Arg1 [V(Arg2)]]
Ex: "our hand is high" → "our hand heightened"
(2) **Manner**
Expansion: [Arg1 [V]] → [Arg1 [MotionV] Modifier(V)]
Ex: "I teach" → "I walk teaching"
Contraction:
[Arg1 [MotionV] Modifier] → [Arg1 [V-Modifier]]
Ex: "he turns again" → "He returns"
(3) **Structural**
Expansion: [Arg1 [V] Arg2] → [Arg1 [V] Oblique Arg2]
Ex: "I forsake thee" → I forsake of thee"
Contraction: [Arg1 [V] Oblique Arg2] → [Arg1 [V] Arg2]
Ex: "I search for him" → "I search him"

II. Rules Impacting Projection Only
(4) **Categorial:**
[Arg1 [V] Adj(Arg2)] → [Arg1 [V] N(Arg2)]
Ex: "I am jealous" → "I have jealousy"
(5) **Head-Swapping:**
[Arg1 [MotionV] Modifier(Direction)]
→ [Arg1 [V-Direction] Modifier(Motion)]
Ex: "I run in" → "I enter running"
(6) **Thematic:**
[Arg1 [V] Arg2] → [Arg2 [V] Arg1]
Ex: "He wears it" → "It is-on him"

Table 4. Transformation Rules between E and E'

[8] The impact of our approach dependency-tree projection will be reported elsewhere and is related to ongoing work by [11].

[9] Our empirical results show that the expansion rules apply more frequently to Spanish than to Arabic, whereas the reverse is true of the contraction rules. This is not

For each language pair, four fluently bilingual human subjects were asked to perform word-level alignments on the same set of sentences selected from the Bible. They were all provided the same instructions and software, similar to the methodology and system described by [17]. Two of the four subjects were given the original English and foreign language sentences; they served as the control for the experiment. The sentence given to the other two consisted of the original foreign language sentences paired with altered English (denoted as E') resulting from divergence transformations described above. We compare the inter-annotator agreement rates and other relevant statistics between the two sets of human subjects. If the divergence transformations had successfully modified English structures to match those of the foreign language, we would expect the inter-annotator agreement rate between the subjects aligning the E' set to be higher than the control set. We would also expect that the E' set would have fewer unaligned and multiply-aligned words.

In the case of English-Spanish, the subjects were presented with 150 sentence pairs from the English and Spanish Bibles. The sentence selection procedure is similar to the divergence detection process described in the previous section. These sentences were first selected as potential divergences, using the hand-crafted regular expressions referred to in Section 3; they were subsequently verified by the experimenter as belonging to a particular divergence type. Out of the 150 sentence pairs, 97 were verified to have contained divergences; moreover, 75 of these 97 contain expansion/contraction divergences (i.e., divergence transformations that result in altered surface words). The average length of the English sentences was 25.6 words; the average length of the Spanish sentences was 24.7 words. Of the four human subjects, two were native Spanish speakers, and two were native English speakers majoring in Spanish literature. The backgrounds of the four human subjects are summarized in Table 5.

In the case of English-Arabic, the subjects were presented with 50 sentence pairs from the English and Arabic Bibles. While the total number of sentences was smaller than the previous experiment, every sentence pair was verified to contain at least one divergence. Of these 50 divergent sentence pairs, 36 of them contained expansion/contraction divergences. The average English sentence length was 30.5 words, and the average Arabic sentence length was 17.4 words. The backgrounds of the four human subjects are summarized in Table 6.

Inter-annotator agreement rate is quantified for each pair of subjects who viewed the same set of data. We hold one subject's alignments as the "ideal" and compute the precision and recall figures for the other subject based on how many alignment links were made by both people. The averaged precision and recall

surprising because, in general, Spanish is verbose relative to English, where as Arabic tends to be more terse. Such differences in verbosity are well documented in the literature. For example, according to [21], human translators often make changes to produce Spanish sentences that are longer than the original English sentence—or they generate sentences of the same length but *reduce* the amount of information conveyed in the original English.

[10] In computing the average number of alignments per word, we do not include unaligned words.

	data set	native-tongue	linguistic knowledge?	ease with computers
Subject 1	control	Spanish	yes	high
Subject 2	control	Spanish	no	low
Subject 3	divergence	English	no	high
Subject 4	divergence	English	no	low

Table 5. Summary of the backgrounds of the English-Spanish subjects

	data set	native-tongue	linguistic knowledge?	ease with computers
Subject 1	control	Arabic	yes	high
Subject 2	control	Arabic	no	high
Subject 3	divergence	Arabic	no	high
Subject 4	divergence	Arabic	no	high

Table 6. Summary of Backgrounds of English-Arabic Subjects

	# of sentences	F-score	% of unaligned words	Avg. alignments per word
E-S	150	80.2	17.2	1.35
E'-S	150	82.9	14.0	1.16
E-A	50	69.7	38.5	1.48
E'-A	50	75.1	11.9	1.72

Table 7. Results of Two Experiments on All Sentence Pairs[10]

	# of sentences	F-score	% of unaligned words	Avg. alignments per word
E-S	97	81.0	17.3	1.35
E'-S	97	83.8	13.8	1.16
E-A	50	69.7	38.5	1.48
E'-A	50	75.1	11.9	1.72

Table 8. Results for Subset Containing only Divergent Sentences

	# of sentences	F-score	% of unaligned words	Avg. alignments per word
E-S	75	82.2	17.3	1.34
E'-S	75	84.6	13.9	1.14
E-A	36	69.1	38.3	1.48
E'-A	36	75.7	11.5	1.67

Table 9. Results for Subset Containing only Sentence Pairs with Expansion/Contraction Divergences

figures (F-scores)[11] for the the two experiments and other relevant statistics are summarized in Table 7. In both experiments, the inter-annotator agreement is higher for the bitext in which the divergent portions of the English sentences have been transformed. For the English-Spanish experiment, the agreement rate increased from 80.2% to 82.9% (error reduction of 13.6%). Using the pair-wise t-test, we find that the higher agreement rate is statistically significant with 95% confidence. For the English-Arabic experiment, the agreement rate increased from 69.7% to 75.1% (error reduction of 17.8%); this higher agreement rate is statistically significant with a confidence rate of 90%.

We also performed data analyses on two subsets of the full study. First, we focused on sentence pairs that were verified to contain divergences; the results are reported in Table 8. They were not significantly different from the complete set. We then considered a smaller subset of sentence pairs containing only expansion/contraction divergences whose transformations altered the surface words as well as the syntactic structures; the results are reported in Table 9. In this case, the higher agreement-rate for the English'-Spanish annotators is statistically sig-

[11] $F = \frac{2 \times Precision \times Recall}{Precision + Recall}$

nificant with 90% confidence; the higher agreement-rate for the English'-Arabic annotators is statistically significant with 95% confidence.

Additional statistics also support our hypothesis that transforming divergent English sentences facilitates word-level alignment by reducing the number of unaligned and multiply-aligned words. In the English-Spanish experiment, both the appearances of unaligned words and multiply-aligned words decreased when aligning to the modified English sentences. The percentage of unaligned words decreased from 17% to 14% (18% fewer unaligned words), and the average number of links to a word is lowered from 1.35 to 1.16.[12] In the English-Arabic experiment, the number of unaligned words is significantly smaller when aligning Arabic sentences to the modified English sentences; however, on average multiple-alignment increased. This may be due to the big difference in sentence lengths (English sentences are typically twice as long as the Arabic ones); thus it is not surprising that the average number of alignments per word would be closer to two when most of the words are aligned. The reason for the lower number in the unmodified English case might be that the subjects only aligned words that had clear translations.

5 Conclusion and Future Work

In this paper, we examined the frequency of occurrence of six divergence types in English-Spanish and English-Arabic. By examining bitext corpora, we have established conservative lower-bounds, estimating that these divergences occur at least 10% of the time. A realistic sampling indicates that the percentage is actually significantly higher, approximately 35% in Spanish.

We have shown that divergence cases can be systematically handled by transforming the syntactic structures of the English sentences to bear a closer resemblance to those of the foreign language, using a small set of templates. The validity of the divergence handling has been verified through two word-level alignment experiments. In both cases, the human subjects consistently had higher agreement rate with each other on the task of performing word-level alignment when divergent English phrases were transformed.

The results of this work suggest several future research directions. First, we are actively working on automating the process of divergence detection and classification, with the goal of replacing our "common search terms" in Table 2 with automatic detection routines based on parameterization of the transformation rules in Table 4.[13] Once the process has been automated, we will be able to

[12] The relatively high overall percentage of unaligned words is due to the fact that the subjects did not align punctuation marks.

[13] For example, we will make use of lexical parameters such as LightVB, MotionV, and Oblique for our Type I rules. We already adopt the LightVB parameter in our current scheme—the current setting is {do, give, take, put, have} in English and {hacer, dar, tomar, tener, put} in Spanish. Settings for MotionV and Obliques are also available for English—and preliminary settings have been assigned in Spanish and Arabic. Three additional parameters will be used for Type II rules—Direction,

perform large-scaled experiments to study the effect of divergence handling on statistical word-alignment models.

Second, while we have focused on the effect of divergence handling on the word-alignment process in this work, we also need to evaluate the effect of divergence handling on the foreign parse trees. Our latest experiments involve projection of English-Chinese experiments; we will evaluate whether our transformation rules on the English structures result in better projected Chinese dependency structures by evaluating against Chinese Treebank data [24].

Finally, we plan to compare our approach with that of [11] in creating foreign language treebanks from projected English syntactic structures. Both approaches apply techniques to improve the accuracy of projected dependency trees, but ours occurs *prior* to statistical alignment, making corrections relevant to general divergence classes—whereas the latter occurs *after* statistical alignment, making corrections relevant to syntactic constraints of the foreign language. We will evaluate different orderings of the two different correction types to determine which ordering is most appropriate for optimal projection of foreign-language dependency trees.

Acknowledgments

This work has been supported, in part, by ONR MURI Contract FCPO.810548265 and Mitre Contract 010418-7712. We are grateful for the assistance of our Spanish aligners, Irma Amenero, Emily Ashcraft, Allison Bigelow, and Clara Cabezas; and also our Arabic aligners, Moustafa Al-Bassyiouni, Eiman Elnahrawy, Tamer Nadeem, and Musa Nasir.

References

[1] Al-Onaizan, Y., Curin, J., Jahr, M., Knight, K., Lafferty, J., Melamed, I.D., Och, F.J., Purdy, D., Smith, N.A., Yarowsky, D.: Statistical machine translation: Final report. In *Summer Workshop on Language Engineering*. John Hopkins University Center for Language and Speech Processing, 1999.

[2] Alshawi, H., Douglas, S.: Learning Dependency Transduction Models from Unannotated Examples. *Philosophical Transactions, Series A: Mathematical, Physical and Engineering Sciences*, 2000.

[3] Alshawi, H., Bangalore, S., Douglas, S.: Learning Dependency Translation Models as Collections of Finite State Head Transducers. *Computational Linguistics*, 26, 2000.

[4] Brown, P.F., Cocke, J., Della-Pietra, S., Della-Pietra, V.J., Jelinek, F., Lafferty, J.D., Mercer, R.L., Roossin, P.S.: A Statistical Approach to Machine Translation. *Computational Linguistics*, 16(2):79–85, June 1990.

Swap, and CatVar—the latter one associated with categorial variation which will be semi-automatically acquired using resources developed in the categorial variation work of [8]. All settings are small enough to be constructed in 1 person-day by a native speaker of the foreign language.

[5] Brown, P.F., Della-Pietra, S.A., Della-Pietra, V.J., Mercer, R.L.: The Mathematics of Machine Translation: Parameter Estimation. *Computational Linguistics*, 1993.

[6] Dorr, B.J., Pearl, L., Hwa, R., Habash, N.: Improved Word-Level Alignment: Injecting Knowledge about MT Divergences. Technical report, University of Maryland, College Park, MD, 2002. LAMP-TR-082, CS-TR-4333, UMIACS-TR-2002-15.

[7] Fellbaum, C., Palmer, M., Dang, H.T., Delfs, L., Wolff, S.: Manual and Automatic Semantic Annotation with WordNet. In *Proceedings of the NAACL Workshop on WordNet and Other Lexical Resources: Applications, Customizations*, Carnegie Mellon University, Pittsburg, PA, 2001.

[8] Habash, N., Dorr, B.J.: Generation-Heavy Machine Translation. In *Proceedings of the Fifth Conference of the Association for Machine Translation in the Americas, AMTA-2002*, Tiburon, CA, 2002 (this volume).

[9] Hermjakob, U., Mooney, R.J.: Learning Parse and Translation Decisions from Examples with Rich Context. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, pages 482–489, 1997.

[10] Hwa, R.: Sample selection for statistical grammar induction. In *Proceedings of the 2000 Joint SIGDAT Conference on EMNLP and VLC*, pages 45–52, Hong Kong, China, October 2000.

[11] Hwa, R., Resnik, P., Weinberg, A., Kolak, O.: Evaluating Translational Correspondence Using Annotation Projection. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, PA, 2002.

[12] Han, C.-H., Lavoie, B., Palmer, M., Rambow, O., Kittredge, R., Korelsky, T., Kim, N., Kim, M.: Handling Structural Divergences and Recovering Dropped Arguments in a Korean/English Machine Translation System. In *Proceedings of the Fourth Conference of the Association for Machine Translation in the Americas, AMTA-2000*, Cuernavaca, Mexico, 2000.

[13] Lavoie, B., Kittredge, R., Korelsky, T., Rambow, O.: A Framework for MT and Multilingual NLG Systems Based on Uniform Lexico-Structural Processing. In *Proceedings of the 1st Annual North American Association of Computational Linguistics, ANLP/NAACL-2000*, Seattle, WA, 2000.

[14] Lavoie, B., White, M., Korelsky, T.: Inducing Lexico-Structural Transfer Rules from Parsed Bi-texts. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics – DDMT Workshop*, Toulouse, France, 2001.

[15] Lin, D.: Government-Binding Theory and Principle-Based Parsing. Technical report, University of Maryland, 1995. Submitted to Computational Linguistics.

[16] Lin, D.: Dependency-Based Evaluation of MINIPAR. In *Proceedings of the Workshop on the Evaluation of Parsing Systems, First International Conference on Language Resources and Evaluation*, Granada, Spain, May 1998.

[17] Melamed, I.D.: Empirical Methods for MT Lexicon Development. In *Proceedings of the Third Conference of the Association for Machine Translation in the Americas, AMTA-98, in Lecture Notes in Artificial Intelligence, 1529*, pages 18–30, Langhorne, PA, October 28–31 1998.

[18] Menezes, A., Richardson, S.D.: A best-first alignment algorithm for automatic extraction of transfer mappings from bilingual corpora. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics – DDMT Workshop*, Toulouse, France, 2001.

[19] Meyers, A., Kosaka, M., Grishman, R.: Chart-Based Transfer Rule Application in Machine Translation. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING'2000)*, Saarbrucken, Germany, 2000.

[20] Och, F.J., Ney, H.: Improved Statistical Alignment Models. In *Proceedings of the 38th Annual Conference of the Association for Computational Linguistics*, pages 440–447, Hongkong, China, 2000.

[21] Slobin, D.I.: Two Ways to Travel: Verbs of Motion in English and Spanish. In M. Shibatani and S. A. Thompson, editors, *Grammatical Constructions: Their Form and Meaning*, pages 195–219. Oxford University Press, New York, 1996.

[22] Watanabe, H., Kurohashi, S., Aramaki, E.: Finding Structural Correspondences from Bilingual Parsed Corpus for Corpus-based Transaltion. In *Proceedings of COLING-2000*, Saarbrüken, Germany, 2000.

[23] Wu, D.: Stochastic Inversion Transduction Grammars and Bilingual Parsing of Parallel Corpora. *Computational Linguistics*, 23(3):377–400, 1997.

[24] Xia, F., Palmer, M., Xue, N., Okurowski, M.E., Kovarik, J., Huang, S., Kroch, T., Marcus, M.: Developing Guidelines and Ensuring Consistency for Chinese Text Annotation. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation (LREC-2000)*, Athens, Greece, 2000.

[25] Yamada, K., Knight, K.: A Syntax-Based Statistical Translation Model. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, pages 523–529, Toulouse, France, 2001.

[26] Yarowsky, D., Ngai, G.: Inducing Multilingual POS Taggers and NP Bracketers via Robust Projection across Aligned Corpora. In *Proc. of NAACL-2001*, pages 200–207, 2001.