Inducing Information Extraction Systems for New Languages via Cross-Language Projection

Ellen Riloff
School of Computing
University of Utah
Salt Lake City, UT 84112
riloff@cs.utah.edu

Charles Schafer and David Yarowsky
Department of Computer Science
Johns Hopkins University
Baltimore, MD 21218
{cschafer,yarowsky}@cs.jhu.edu

Abstract

Information extraction (IE) systems are costly to build because they require development texts, parsing tools, and specialized dictionaries for each application domain and each natural language that needs to be processed. We present a novel method for rapidly creating IE systems for new languages by exploiting existing IE systems via cross-language projection. Given an IE system for a source language (e.g., English), we can transfer its annotations to corresponding texts in a target language (e.g., French) and learn information extraction rules for the new language automatically. In this paper, we explore several ways of realizing both the transfer and learning processes using off-the-shelf machine translation systems, induced word alignment, attribute projection, and transformation-based learning. We present a variety of experiments that show how an English IE system for a plane crash domain can be leveraged to automatically create a French IE system for the same domain.

1 Introduction

Information extraction (IE) is an important application for natural language processing, and recent research has made great strides toward making IE systems easily portable across domains. However, IE systems depend on parsing tools and specialized dictionaries that are language specific, so they are not easily portable across languages. In this research, we explore the idea of using an information extraction system designed for one language to automatically create a comparable information extraction system for a different language.

To achieve this goal, we rely on the idea of *cross-language projection*. The basic approach is the following. First, we create an artificial parallel corpus by applying an off-the-shelf machine translation (MT) system to source language text (here, English) to produce target language text (here, French). Or conversely, in some experiments we generate a parallel corpus by applying MT to a French corpus to produce artificial English. We then run a word alignment algorithm over the parallel corpus. Next, we apply an English IE system to the English texts and project the IE annotations over to the corresponding French words via the induced word alignments. In effect, this produces an automatically annotated French corpus. We explore several strategies for transferring the English IE annotations to the target language, including evaluation of the French annotations produced by the direct projection alone, as well as the use of transformation-based learning to create French extraction rules from the French annotations.

2 Information Extraction

The goal of information extraction systems is to identify and extract facts from natural language text. IE systems are usually designed for a specific domain, and the types of facts to be extracted are defined in advance. In this paper, we will focus on the domain of plane crashes and will try to extract descriptions of the vehicle involved in the crash, victims of the crash, and the location of the crash.

Most IE systems use some form of *extraction patterns* to recognize and extract relevant information. Many techniques have been developed to generate extraction patterns for a new domain automatically, including PALKA (Kim & Moldovan, 1993), AutoSlog (Riloff, 1993), CRYSTAL (Soderland et al., 1995), RAPIER (Califf, 1998), SRV (Freitag, 1998), meta-bootstrapping (Riloff & Jones, 1999), and ExDisco (Yangarber et al., 2000). For this work, we will use AutoSlog-TS (Riloff, 1996b) to generate IE patterns for the plane crash domain. AutoSlog-TS is a derivative of AutoSlog that automatically generates extraction patterns by gathering statistics from a corpus of relevant texts (within the domain) and irrelevant texts (outside the domain). Each extraction pattern represents a linguistic ex-

pression that can extract noun phrases from one of three syntactic positions: subject, direct object, or object of a prepositional phrase. For example, the following patterns could extract vehicles involved in a plane crash: "<*subject*> *crashed*", "*hijacked* <*direct-object*>", and "*wreckage of* <*np*>".

We trained AutoSlog-TS using AP news stories about plane crashes as the relevant text, and AP news stories that do not mention plane crashes as the irrelevant texts. AutoSlog-TS generates a list of extraction patterns, ranked according to their association with the domain. A human must review this list to decide which patterns are useful for the IE task and which ones are not. We manually reviewed the top patterns and used the accepted patterns for the experiments described in this paper. To apply the extraction patterns to new text, we used a shallow parser called Sundance that also performs information extraction.

3 Cross-Language Projection

3.1 Motivation and Previous Projection Work

Not all languages have received equal investment in linguistic resources and tool development. For a select few, resource-rich languages such as English, annotated corpora and text analysis tools are readily available. However, for the large majority of the world's languages, resources such as treebanks, part-of-speech taggers, and parsers do not exist. And even for many of the better-supported languages, cutting edge analysis tools in areas such as information extraction are not readily available.

One solution to this NLP-resource disparity is to transfer linguistic resources, tools, and domain knowledge from resource-rich languages to resource-impoverished ones. In recent years, there has been a burst of projects based on this paradigm. Yarowsky et al. (2001) developed cross-language projection models for part-of-speech tags, base noun phrases, named-entity tags, and morphological analysis (lemmatization) for four languages. Resnik et al. (2001) developed related models for projecting dependency parsers from English to Chinese. There has also been extensive work on the cross-language transfer and development of ontologies and WordNets (e.g., (Atserias et al., 1997)).

3.2 Mechanics of Projection

The cross-language projection methodology employed in this paper is based on Yarowsky et al. (2001), with one important exception. Given the absence of available naturally occurring bilingual

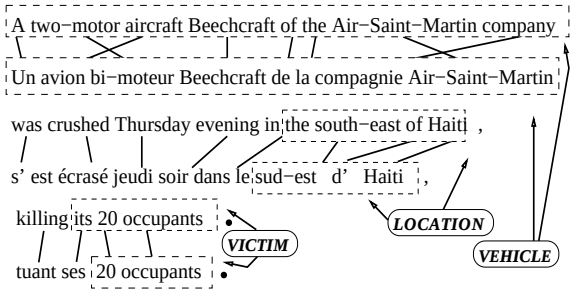

Figure 1: French text word aligned with its English machine translation (extractions highlighted)

corpora in our target domain, we employ commercial, off-the-shelf machine translation to generate an artificial parallel corpus. While machine translation errors present substantial problems, MT offers great opportunities because it frees cross-language projection research from the relatively few large existing bilingual corpora (such as the Canadian Hansards). MT allows projection to be performed on any corpus, such as the domain-specific plane-crash news stories employed here. Section 5 gives the details of the MT system and corpora that we used.

Once the artificial parallel corpus has been created, we apply an English IE system to the English texts and transfer the IE annotations to the target language as follows:

1. Sentence align the parallel corpus.[1]
2. Word-align the parallel corpus using the Giza++ system (Och and Ney, 2000).
3. Transfer English IE annotations and noun-phrase boundaries to French via the mechanism described in Yarowsky et al. (2001), yielding annotated sentence pairs as illustrated in Figure 1.
4. Train a stand-alone IE tagger on these projected annotations (described in Section 4).

4 Transformation-Based Learning

We used *transformation-based learning (TBL)* (Brill, 1995) to learn information extraction rules for French. TBL is well-suited for this task because it uses rule templates as the basis for learning, which can be easily modeled after English extraction patterns. However, information extraction systems typically rely on a shallow parser to identify syntactic elements (e.g., subjects and direct objects) and verb

[1]This is trivial because each sentence has a numbered anchor preserved by the MT system.

constructions (e.g., passive vs. active voice). Our hope was that the rules learned by TBL would be applicable to new French texts without the need for a French parser. One of our challenges was to design rule templates that could approximate the recognition of syntactic structures well enough to duplicate most of the functionality of a French shallow parser.

When our TBL training begins, the initial state is that no words are annotated. We experimented with two sets of "truth" values: Sundance's annotations and human annotations. We defined 56 language-independent rule templates, which can be broken down into four sets designed to produce different types of behavior. *Lexical N-gram rule templates* change the annotation of a word if the word(s) immediately surrounding it exactly match the rule. We defined rule templates for 1, 2, and 3-grams. In Table 1, Rules 1-3 are examples of learned Lexical N-gram rules. *Lexical+POS N-gram rule templates* can match exact words or part-of-speech tags. Rules 4-5 are Lexical+POS N-gram rules. Rule 5 will match verb phrases such as "went down in", "shot down in", and "came down in".

One of the most important functions of a parser is to identify the subject of a sentence, which may be several words away from the main verb phrase. This is one of the trickest behaviors to duplicate without the benefit of syntactic parsing. We designed *Subject Capture rule templates* to identify words that are likely to be a syntactic subject. As an example, Rule 6 looks for an article at the beginning of a sentence and the word "crashed" a few words ahead[2], and infers that the article belongs to a vehicle noun phrase. (The NP Chaining rules described next will extend the annotation to include the rest of the noun phrase.) Rule 7 attempts relative pronoun disambiguation when it finds the three tokens "COMMA which crashed" and infers that the word preceding the comma is a vehicle.

Without the benefit of a parser, another challenge is identifying noun phrase boundaries. We designed *NP Chaining rule templates* to look at words that have already been labelled and extend the boundaries of the annotation to cover a complete noun phrase. As examples, Rules 8 and 9 extend location and victim annotations to the right, and Rule 10 extends a vehicle annotation to the left.

[2] ϕ is a start-of-sentence token. w_{4-7} means that the item occurs in the range of $word_4$ through $word_7$.

	Rule Condition	Rule Effect
1.	w_1=crashed w_2=in	w_3 is LOC.
2.	w_1=wreckage w_2=of	w_3 is VEH.
3.	w_1=injuring	w_2 is VIC.
4.	w_1=NOUN w_2=crashed	w_1 is VEH.
5.	w_1=VERB w_2=down w_3=in	w_4 is LOC.
6.	w_1=ϕ w_2=ART w_{4-7}=crashed	w_2 is VEH.
7.	w_2=COMMA w_3=which w_4=crashed	w_1 is VEH.
8.	w_1=in w_2=LOCATION w_3=NOUN	w_3 is LOC.
9.	w_1=VERB w_2=VICTIM w_3=NOUN	w_3 is VIC.
10.	w_1=ART w_2=VEHICLE	w_1 is VEH.

Table 1: Examples of Learned TBL Rules (LOC.=location, VEH.=vehicle, VIC.=victim)

5 Resources

The corpora used in these experiments were extracted from English and French AP news stories. We created the corpora automatically by searching for articles that contain plane crash keywords. The news streams for the two languages came from different years, so the specific plane crash events described in the two corpora are disjoint. The English corpus contains roughly 420,000 words, and the French corpus contains about 150,000 words.

For each language, we hired 3 fluent university students to do annotation. We instructed the annotators to read each story and mark relevant entities with SGML-style tags. Possible labels were *location* of a plane crash, *vehicle* involved in a crash, and *victim* (any persons killed, injured, or surviving a crash). We asked the annotators to align their annotations with noun phrase boundaries. The annotators marked up 1/3 of the English corpus and about 1/2 of the French corpus.

We used a high-quality commercial machine translation (MT) program (Systran Professional Edition) to generate a translated parallel corpus for each of our English and French corpora. These will henceforth be referred to as MT-French (the Systran translation of the English text) and MT-English (the Systran translation of our French text).

6 Experiments and Evaluation

6.1 Scoring and Annotator Agreement

We explored two ways of measuring annotator agreement and system performance. (1) The *exact-word-match* measure considers annotations to match if their start and end positions are exactly the same. (2) The *exact-NP-match* measure is more forgiving and considers annotations to match if they both include the head noun of the same noun phrase. The *exact-word-match* criterion is very conservative

because annotators may disagree about equally acceptable alternatives (e.g., "Boeing 727" vs. "new Boeing 727"). Using the *exact-NP-match* measure, "Boeing 727" and "new Boeing 727" would constitute a match. We used different tools to identify noun phrases in English and French. For English, we applied the base noun phrase chunker supplied with the fnTBL toolkit (Ngai & Florian, 2001). In French, we ran a part-of-speech tagger (Cucerzan & Yarowsky, 2000) and applied regular-expression heuristics to detect the heads of noun phrases.

We measured agreement rates among our human annotators to assess the difficulty of the IE task. We computed pairwise agreement scores among our 3 English annotators and among our 3 French annotators. The *exact-word-match* scores ranged from 16-31% for French and 24-27% for English. These relatively low numbers suggest that the *exact-word-match* criterion is too strict. The *exact-NP-match* agreement scores were much higher, ranging from 43-54% for French and 51-59% for English[3].

These agreement numbers are still relatively low, however, which partly reflects the fact that IE is a subjective and difficult task. Inspection of the data revealed some systematic differences of approach among annotators. For example, one of the French annotators marked 4.5 times as many locations as another. On the English side, the largest disparity was a factor of 1.4 in the tagging of victims.

6.2 Monolingual English & French Evaluation

As a key baseline for our cross-language projection studies, we first evaluated the AutoSlog-TS and TBL training approaches on monolingual English and French data. Figure 2 shows (1) English training by running AutoSlog-TS on unannotated texts and then applying its patterns to the human-annotated English test data, (2) English training and testing by applying TBL to the human-annotated English data with 5-fold cross-validation, (3) English training by applying TBL to annotations produced by Sundance (using AutoSlog-TS patterns) and then testing the TBL rules on the human-annotated English data, and (4) French training and testing by applying TBL to human annotated French data with 5-fold cross-validation.

Table 2 shows the performance in terms of Precision (P), Recall (R) and F-measure (F). Throughout our experiments, AutoSlog-TS training achieves higher precision but lower recall than TBL training. This may be due to the exhaustive coverage provided by the human annotations used by TBL, compared to the more labor-efficient but less-complete AutoSlog-TS training that used only unannotated data.

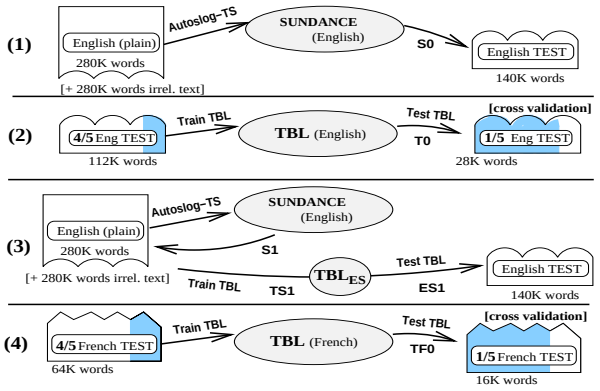

Figure 2: Monolingual IE Evaluation pathways[4]

Monolingual Training Route	P	R	F
English			
(1) Train AutoSlog-TS on English-plain (AS_E) $S0$: Apply AS_E to English Test	**.44**	.42	.43
(2) Train TBL on 4/5 of English-Test (TBL_E) $T0$: Apply TBL_E to 1/5 of English Test (perform in 5-fold cross-validation)	.35	**.62**	**.45**
(3) Train AutoSlog-TS on English-plain (AS_E) $S1$: Apply AS_E to English-plain $TS1$ Train TBL on Sundance annotations $ES1$: Apply TBL_{ES} to English Test	.31	.40	.35
French			
(4) Train TBL on 4/5 of French-Test (TBL_F) $TF0$: Apply TBL_F to 1/5 of French Test (perform in 5-fold cross-validation)	**.47**	**.66**	**.54**

Table 2: Monolingual IE Baseline Performance

6.3 TBL-based IE Projection and Induction

As noted in Section 5, both the English and French corpora were divided into unannotated ("plain") and annotated ("antd" or "Tst") sections. Figure 3 illustrates these native-language data subsets in white. Each native-language data subset also has a machine-translated mirror in French/English respectively (shown in black), with an identical number of sentences to the original. By word-aligning these 4 native/MT pairs, each becomes a potential vehicle for cross-language information projection.

Consider the pathway $T_E1 \Rightarrow P1 \Rightarrow T_F1$ as a representative example pathway for projection. Here

[3] Agreement rates were computed on a subset of the data annotated by multiple people; systems were scored against the full corpus, of which each annotator provided the standard for one third.

an English TBL classifier is trained on the 140K-word human annotated data and the learned TBL rules are applied to the unannotated English sub-corpus. The annotations are then projected across the Giza++ word alignments to their MT-French mirror. Next, a French TBL classifier (TBL1) is trained on the projected MT-French annotations and the learned French TBL rules are subsequently applied to the native-French test data.

An alternative path ($T_{\rm E}4 \Rightarrow P4 \Rightarrow$ *French-Test*) is more direct, in that the English TBL classifier is applied immediately to the word-aligned MT-English translation of the French test data. The MT-English annotations can then be directly projected to the French test data, so no additional training is necessary. Another short direct projection path ($P_{\rm HA}2 \Rightarrow T_{\rm HA}2 \Rightarrow$ *French-test*) skips the need to train an English TBL model by projecting the English human annotations directly onto MT-French texts, which can then be used to train a French TBL system which can be applied to the French test data.

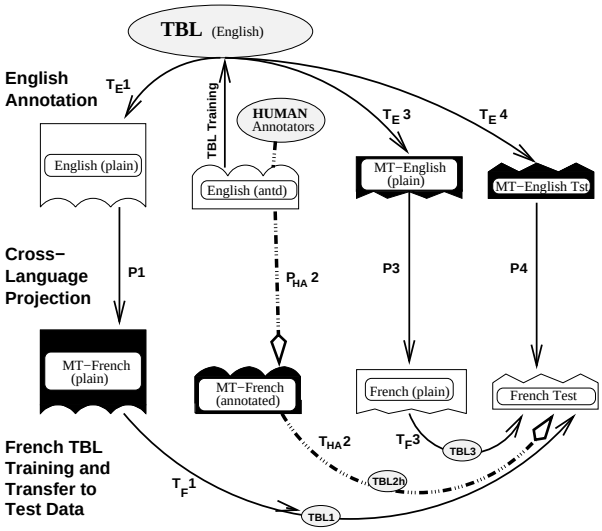

Figure 3: TBL-based IE projection pathways

Table 3 shows the results of our TBL-based experiments. The top performing pathway is the $T_E4 \Rightarrow P4$ two-step projection pathway shown in Figure 3. Note the F-measure of the best pathway is .45, which is equal to the highest F-measure for monolingual English and only 9% lower than the F-measure for monolingual French.

[4]The irrelevant texts are needed to train AutoSlog-TS, but not TBL.

Projection and Training Route	P	R	F
T_E1: Apply TBL_E to English-plain $P1$: Project to MT-French(English-Plain) T_F1: Train TBL & Apply to FrTest	**.69**	.24	.36
–: Use human Annos from Eng_Antd $P_{ha}2$: Project to MT-French(English_Antd) $T_{ha}2$: Train TBL & Apply to FrTest	.56	.29	.39
T_E3: Apply TBL_E to MT-Eng(FrenchPlain) $P3$: Project to French-Plain T_F3: Train TBL & Apply to FrTest	.49	.34	.40
T_E4: Apply TBL_E to MT-Eng(FrenchTest) $P4$: Direct Project to French-Test	.49	**.41**	**.45**

Table 3: TBL-based IE projection performance

6.4 Sundance-based IE Projection and Induction

Figure 4 shows the projection and induction model using Sundance for English IE annotation, which is almost isomorphic to that using TBL. One notable difference is that Sundance was trained by applying AutoSlog-TS to the unannotated English text rather than the human-annotated data. Figure 4 also shows an additional set of experiments ($S_{MT}3$ and $S_{MT}4$) in which AutoSlog-TS was trained on the English MT translations of the unannotated French data. The motivation was that native-English extraction patterns tend to achieve low recall when applied to MT-English text (given frequent mistranslations such as "to crush" a plane rather than "to crash" a plane). By training AutoSlog-TS on the sentences generated by an MT system (seen in the $S_{MT}3$ and $S_{MT}4$ pathways), the F-measure increases.[5]

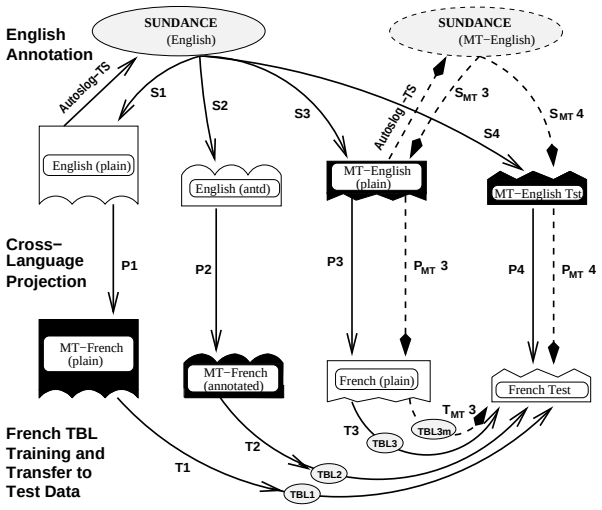

Figure 4: Sundance-based projection pathways

[5]This is a "fair" gain, in that the MT-trained AutoSlog-TS patterns didn't use translations of any of the French test data.

Projection and Training Route		P	R	F
AutoSlog-TS trained on native English (AS_E)				
$S2$:	Apply AS_E to English-Antd	.39	.24	.29
$P2$:	Project to MT-French(English-Antd)			
$T2$:	Train TBL_{FP2} & Apply to FrTest			
$S(1+2)$:	Apply AS_E to English_Antd+Plain	.43	.23	.30
$P(1+2)$:	Project to MT-French(Eng-Ant+Pl)			
$T(1+2)$:	Train TBL_{FP1+2} & Apply to FrTest			
$S3$:	Apply AS_E to MT-Eng(FrenchPlain)	.45	.04	.07
$P3$:	Project to French-Plain			
$T3$:	Train TBL_{FP3} & Apply to FrTest			
$S4$:	Apply AS_E to MT-Eng(FrenchTest)	.48	.07	.13
$P4$:	Direct Project to French-Test			
AutoSlog-TS trained on MT English (AS_{MTE})				
$S_{MT}3$:	Apply AS_{MTE} to MT-Eng(FrPlain)	.46	.25	.32
$P_{MT}3$:	Project to French-Plain			
$T_{MT}3$:	Train TBL_{FMT3} & Apply to FrTest			
$S_{MT}4$:	Apply AS_{MTE} to MT-Eng(FrTest)	**.55**	**.28**	**.37**
$P_{MT}4$:	Direct Project to French-Test			

Table 4: Sundance-based IE projection performance [6]

Table 4 shows that the best Sundance pathway achieved an F-measure of .37. Overall, Sundance averaged 7% lower F-measures than TBL on comparable projection pathways. However, AutoSlog-TS training required only 3-4 person hours to review the learned extraction patterns while TBL training required about 150 person-hours of manual IE annotations, so this may be a viable cost-reward tradeoff. However, the investment in manual English IE annotations can be reused for projection to new foreign languages, so the larger time investment is a fixed cost per-domain rather than per-language.

6.5 Analysis and Implications

• For both TBL and Sundance, the $P1$, $P2$ and $P3$-family of projection paths all yield stand-alone monolingual French IE taggers not specialized for any particular test set. In contrast, the $P4$ series of pathways (e.g. $P_{MT}4$ for Sundance), were trained specifically on the MT output of the target test data. Running an MT system on test data can be done automatically and requires no additional human language knowledge, but it requires additional time (which can be substantial for MT). Thus, the higher performance of the $P4$ pathways has some cost.

• The significant performance gains shown by Sundance when AutoSlog-TS is trained on MT-English rather than native-English are not free because the MT data must be generated for each new language and/or MT system to optimally tune to its peculiar language variants. No target-language knowledge is needed in this process, however, and reviewing AutoSlog-TS' patterns can be done successfully by imaginative English-only speakers.

• In general, recall and F-measure drop as the number of experimental steps increases. Averaged over TBL and Sundance pathways, when comparing 2 and 3-step projections, mean recall decreases from 26.8 to 21.8 (5 points), and mean F-measure drops from 32.6 to 28.8 (3.8 points). Viable extraction patterns may simply be lost or corrupted via too many projection and retraining phases.

• One advantage of the projection path families of $P1$ and $P2$ is that no domain-specific documents in the foreign language are required (as they are in the $P3$ family). A collection of domain-specific English texts can be used to project and induce new IE systems even when no domain-specific documents exist in the foreign language.

6.6 Multipath Projection

Finally, we explored the use of classifier combination to produce a premium system. We considered a simple voting scheme over sets of individual IE systems. Every annotation of a head noun was considered a vote. We tried 4 voting combinations: (1) the systems that used Sundance with English extraction patterns, (2) the systems that used Sundance with MT-English extraction patterns, (3) the systems that used TBL trained on English human annotations, (4) all systems. For each combination of n systems, n answer sets were produced using the voting thresholds $T_v = 1..n$. For example, for $T_v = 2$ every annotation receiving $>= 2$ votes (picked by at least 2 individual systems) was output in the answer set. This allowed us to explore a precision/recall tradeoff based on varying levels of consensus.

Figure 5 shows the precision/recall curves. Voting yields some improvement in F-measure and provides a way to tune the system for higher precision or higher recall by choosing the T_v threshold. When using all English knowledge sources, the F-measure at T_v=1 (.48) is nearly 3% higher than the strongest individual system. Figure 5 also shows the performance of a 5th system (5), which is a TBL system trained directly from the French annotations under 5-fold cross-validation. It is remarkable that the most effective voting-based projection system from English to French comes within 6% F-measure of the monolingually trained system, given that this cross-validated French monolingual system was trained directly on data in the same language

[6] $S(1+2)$ combines the training data in $S1$ (280K) and $S2$ (140K), yielding a 420K-word sample.

and source as the test data. This suggests that cross-language projection of IE analysis capabilities can successfully approach the performance of dedicated systems in the target language.

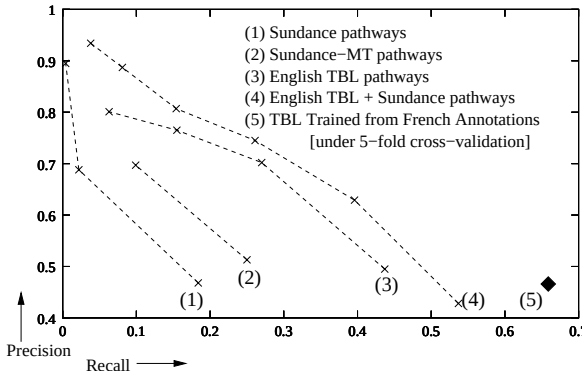

Figure 5: Precision/Recall curves for voting systems. Each point represents performance for a particular voting threshold. In all cases, precision increases and recall decreases as the threshold is raised.

French Test-Set Performance	P	R	F
Multipath projection from all English resources	**.43**	**.54**	**.48**

Table 5: Best multipath English-French Projection Performance (from English TBL and Sundance pathways)

7 Conclusions

We have used IE systems for English to automatically derive IE systems for a second language. Even with the quality of MT available today, our results demonstrate that we can exploit translation tools to transfer information extraction expertise from one language to another. Given an IE system for a source language, an MT system that can translate between the source and target languages, and a word alignment algorithm, our approach allows a user to create a functionally comparable IE system for the target language with very little human effort. Our experiments demonstrated that the new IE system can achieve roughly the same level of performance as the source-language IE system. French and English are relatively close languages, however, so how well these techniques will work for more distant language pairs is still an open question.

Additional performance benefits could be achieved in two ways: (1) put more effort into obtaining better resources for English, or (2) implement (minor) specializations per language. While it is expensive to advance the state of the art in English IE or to buy annotated data for a new domain, these additions will improve performance not only in English but for other languages as well. On the other hand, with minimal effort (hours) it is possible to custom-train a system such as Autoslog/Sundance to work relatively well on noisy MT-English, providing a substantial performance boost for the IE system learned for the target language, and further gains are achieved via voting-based classifier combination.

References

J. Atserias, S. Climent, X. Farreres, G. Rigau and H. Rodriguez. 1997. Combining multiple methods for the automatic construction of multilingual WordNets. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*.

E. Brill. 1995. Transformation-based error-driven learning and natural language processing: A case study in part of speech tagging. *Computational Linguistics*, 21(4):543–565.

M. E. Califf. 1998. *Relational learning techniques for natural language information extraction.* Ph.D. thesis, Tech. Rept. AI98-276, Artificial Intelligence Laboratory, The University of Texas at Austin.

S. Cucerzan and D. Yarowsky. 2000. Language independent minimally supervised induction of lexical probabilities. In *Proceedings of ACL-2000*, pages 270-277.

D. Freitag. 1998. Toward general-purpose learning for information extraction. In *Proceedings of COLING-ACL'98*, pages 404-408.

J. Kim and D. Moldovan. 1993. Acquisition of semantic patterns for information extraction from corpora. In *Proceedings of the Ninth IEEE Conference on Artificial Intelligence for Applications*, pages 171–176.

G. Ngai and R. Florian. 2001. Transformation-based learning in the fast lane. In *Proceedings of NAACL*, pages 40-47.

F. J. Och and H. Ney. 2000. Improved statistical alignment models. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pages 440–447.

E. Riloff. 1993. Automatically Constructing a dictionary for information extraction tasks. In *Proceedings of the Eleventh National Conference on Artificial Intelligence*, pages 811–816.

E. Riloff. 1996b. Automatically generating extraction patterns from untagged text. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 1044–1049. AAAI Press/MIT Press.

E. Riloff and R. Jones. 1999. Learning dictionaries for information extraction by multi-level bootstrapping. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence*, pages 474–479.

S. Soderland, D. Fisher, J. Aseltine, and W. Lehnert. 1995. CRYSTAL: Inducing a conceptual dictionary. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1314–1319.

R. Yangarber, R. Grishman, P. Tapanainen, and S. Huttunen. 2000. Automatic acquisiton of domain knowledge for information extraction. In *Proceedings of COLING-2000*, pages 940-946.

Yarowsky, D., G. Ngai and R. Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of HLT-01*, pages 161–168.