Multilingual Coreference Resolution

Sanda M. Harabagiu
Southern Methodist University
Dallas, TX 75275-0122
sanda@seas.smu.edu

Steven J. Maiorano
IPO
Washington, D.C. 20505
maiorano@cais.com

Abstract

In this paper we present a new, multilingual data-driven method for coreference resolution as implemented in the SWIZZLE system. The results obtained after training this system on a bilingual corpus of English and Romanian tagged texts, outperformed coreference resolution in each of the individual languages.

1 Introduction

The recent availability of large bilingual corpora has spawned interest in several areas of multilingual text processing. Most of the research has focused on bilingual terminology identification, either as parallel multiwords forms (e.g. the *Champollion* system (Smadja et al.1996)), technical terminology (e.g. the *Termight* system (Dagan and Church, 1994) or broad-coverage translation lexicons (e.g. the *SABLE* system (Resnik and Melamed, 1997)). In addition, the Multilingual Entity Task (MET) from the TIPSTER program[1] (*http://www-nlpir.nist.gov/related-projects/tipster/met.htm*) challenged the participants in the Message Understanding Conference (MUC) to extract named entities across several foreign language corpora, such as Chinese, Japanese and Spanish.

In this paper we present a new application of aligned multilingual texts. Since coreference resolution is a pervasive discourse phenomenon causing performance impediments in current IE systems, we considered a corpus of aligned English and Romanian texts to identify coreferring expressions. Our task focused on the same kind of coreference as considered in the past MUC competitions, namely the *identity* coreference. Identity coreference links nouns, pronouns and noun phrases (including proper names) to their corresponding antecedents.

We created our bilingual collection by translating the MUC-6 and MUC-7 coreference training texts into Romanian using native speakers. The training data set for Romanian coreference used, wherever possible, the same coreference identifiers as the English data and incorporated additional tags as needed. Our claim is that by adding the wealth of coreferential features provided by multilingual data, new powerful heuristics for coreference resolution can be developed that outperform monolingual coreference resolution systems.

For both languages, we resolved coreference by using SWIZZLE, our implementation of a bilingual coreference resolver. SWIZZLE is a multilingual enhancement of COCKTAIL (Harabagiu and Maiorano, 1999), a coreference resolution system that operates on a mixture of heuristics that combine semantic and textual cohesive information[2]. When COCKTAIL was applied separately on the English and the Romanian texts, coreferring links were identified for each English and Romanian document respectively. When aligned referential expressions corefer with non-aligned anaphors, SWIZZLE derived new heuristics for coreference. Our experiments show that SWIZZLE outperformed COCKTAIL on both English and Romanian test documents.

The rest of the paper is organized as follows. Section 2 presents COCKTAIL, a monolingual coreference resolution system used separately on both the English and Romanian texts. Section 3 details the data-driven approach used in SWIZZLE and presents some of its resources. Section 4 reports and discusses the experimental results. Section 5 summarizes the

---

[1]The TIPSTER Text Program was a DARPA-led government effort to advance the state of the art in text processing technologies.

[2]The name of COCKTAIL is a pun on CogNIAC because COCKTAIL combines a larger number of heuristics than those reported in (Baldwin, 1997). SWIZZLE, moreover, adds new heuristics, discovered from the bilingual aligned corpus.

conclusions.

2 COCKTAIL

Currently, some of the best-performing and most robust coreference resolution systems employ knowledge-based techniques. Traditionally, these techniques have combined extensive syntactic, semantic, and discourse knowledge. The acquisition of such knowledge is time-consuming, difficult, and error-prone. Nevertheless, recent results show that knowledge-poor methods perform with amazing accuracy (cf. (Mitkov, 1998), (Kennedy and Boguraev, 1996) (Kameyama, 1997)). For example, CogNIAC (Baldwin, 1997), a system based on seven ordered heuristics, generates high-precision resolution (over 90%) for some cases of pronominal reference. For this research, we used a coreference resolution system ((Harabagiu and Maiorano, 1999)) that implements different sets of heuristics corresponding to various forms of coreference. This system, called COCKTAIL, resolves coreference by exploiting several textual cohesion constraints (e.g. term repetition) combined with lexical and textual coherence cues (e.g. subjects of communication verbs are more likely to refer to the last person mentioned in the text). These constraints are implemented as a set of heuristics ordered by their priority. Moreover, the COCKTAIL framework uniformly addresses the problem of interaction between different forms of coreference, thus making the extension to multilingual coreference very natural.

2.1 Data-Driven Coreference Resolution

In general, we define a data-driven methodology as a sequence of actions that captures the data patterns capable of resolving a problem with both a high degree of precision and recall. Our data-driven methodology reported here generated *sets of heuristics* for the coreference resolution problem. *Precision* is the number of correct references out of the total number of coreferences resolved, whereas the *recall* measures the number of resolved references out of the total number of keys, i.e., the annotated coreference data.

The data-driven methodology used in COCKTAIL is centered around the notion of a *coreference chain*. Due to the *transitivity* of coreference relations, k coreference relations having at least one common argument generate $k + 1$ *coreferring expressions*. The text position induces an order among coreferring expressions. A *coreference structure* is created when a set of coreferring expressions are connected in an oriented graph such that each node is related only to one of its preceding nodes. In turn, a *corefer-*

ence chain is the coreference structure in which every node is connected to its immediately preceding node. Clearly, multiple coreference structures for the same set of coreferring expressions can be mapped to a single coreference chain. As an example, both coreference structures illustrated in Figure 1(a) and (c) are cast into the coreference chain illustrated in Figure 1(b).
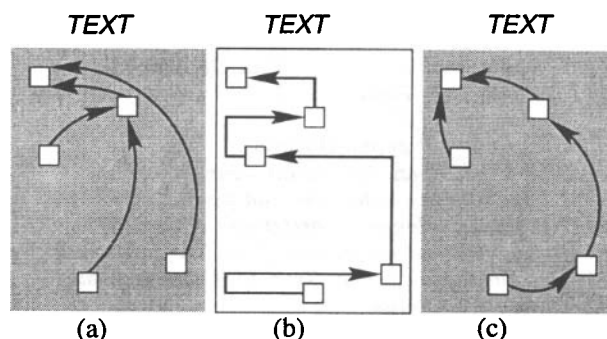

Figure 1: Three coreference structures.

Given a corpus annotated with coreference data, the data-driven methodology first generates all coreference chains in the data set and then considers all possible combinations of coreference relations that would generate the same coreference chains. For a coreference chain of length l with nodes n_1, n_2, ... n_{l+1}, each node n_k $(1 \leq k \leq l)$ can be connected to any of the $l - k$ nodes preceding it. From this observation, we find that a number of $1 \times 2 \times ... \times (l - k)... \times l = l!$ coreference structures can generate the same coreference chain. This result is very important, since it allows for the automatic generation of coreference data. For each coreference relation $\mathcal{R}$ from an annotated corpus we created a median of $(l - 1)!$ new coreference relations, where l is the length of the coreference chain containing relation $\mathcal{R}$. This observation gave us the possibility of expanding the test data provided by the coreference keys available in the MUC-6 and MUC-7 competitions (MUC-6 1996), (MUC-7 1998). The MUC-6 coreference annotated corpus contains 1626 coreference relations, while the MUC-7 corpus has 2245 relations. The average length of a coreference chain is 7.21 for the MUC-6 data, and 8.57 for the MUC-7 data. We were able to expand the number of annotated coreference relations to 6,095,142 for the MUC-6 corpus and to 8,269,403 relations for the MUC-7 corpus; this represents an expansion factor of 3,710. We are not aware of any other automated way of creating coreference annotated data, and we believe that much of the COCKTAIL's impressive performance is due to the plethora of data provided by this method.

Heuristics for 3rd person pronouns	Heuristics for nominal reference

o_Heuristic 1-Pronoun_(H1Pron)
Search in the same sentence for the **same**
   3rd person pronoun Pron'
   if *(Pron' belongs to coreference chain CC)*
    and *there is an element from CC which is*
     closest to Pron in Text, Pick that element.
   else *Pick Pron'.*
o_Heuristic 2-Pronoun_(H2Pron)
Search for PN, the closest proper name from Pron
if *(PN agrees in number and gender with Pron)*
   if *(PN belongs to coreference chain CC)*
    then *Pick the element from CC which is*
     closest to Pron in Text.
   else *Pick PN.*
o_Heuristic 3-Pronoun_(H3Pron)
Search for Noun, the closest noun from Pron
if *(Noun agrees in number and gender with Pron)*
   if *(Noun belongs to coreference chain CC)*
    and *there is an element from CC which is*
     closest to Pron in Text, Pick that element.
   else *Pick Noun*
	o_Heuristic 1-Nominal_(H1Nom)
	if *(Noun is the head of an appositive)*
	   then Pick the preceding NP.
	o_Heuristic 2-Nominal_(H2Nom)
	if *(Noun belongs to an NP, Search for NP'*
	   such that Noun'=same_name(head(NP),head(NP'))
	   or
	   Noun'=same_name(adjunct(NP),adjunct(NP')))
	   then if *(Noun' belongs to coreference chain CC)*
	    then *Pick the element from CC which is*
	     closest to Noun in Text.
	    else *Pick Noun'.*
	o_Heuristic 3-Nominal_(H3Nom)
	if *Noun is the head of an NP*
	   then *Search for proper name PN*
	   such that head(PN)=Noun
	   if *(PN belongs to coreference chain CC)*
	    and *there is an element from CC which is*
	     closest to Noun in Text, Pick that element.
	   else *Pick PN.*

Table 1: Best performing heuristics implemented in COCKTAIL

2.2 Knowledge-Poor Coreference Resolution

The result of our data-driven methodology is the set of heuristics implemented in COCKTAIL which cover both nominal and pronoun coreference. Each heuristic represents a pattern of coreference that was mined from the large set of coreference data. COCKTAIL uses knowledge-poor methods because (a) it is based only on a limited number of heuristics and (b) text processing is limited to part-of-speech tagging, named-entity recognition, and approximate phrasal parsing. The heuristics from COCKTAIL can be classified along two directions. First of all, they can be grouped according to the type of coreference they resolve, e.g., heuristics that resolve the anaphors of reflexive pronouns operate differently than those resolving bare nominals. Currently, in COCKTAIL there are heuristics that resolve five types of pronouns (personal, possessive, reflexive, demonstrative and relative) and three forms of nominals (definite, bare and indefinite).

Secondly, for each type of coreference, there are three classes of heuristics categorized according to their suitability to resolve coreference. The first class is comprised of strong indicators of coreference. This class resulted from the analysis of the distribution of the antecedents in the MUC annotated data. For example, repetitions of named entities and appositives account for the majority of the nominal coreferences, and, therefore, represent anchors for the first class of heuristics.

The second class of coreference covers cases in which the arguments are recognized to be semantically consistent. COCKTAIL's test of semantic consistency blends together information available from WordNet and statistics gathered from Treebank. Different consistency checks are modeled for each of the heuristics.

Example of the application of heuristic H2Pron
Mr. Adams$_1$, 69 years old, is the retired chairman of Canadian-based Emco Ltd., a maker of plumbing and petroleum equipment; he$_1$ has served on the Woolworth board since 1981.
Example of the application of heuristic H3Pron
"We have got to stop pointing our fingers at *these kids*$_2$ who have no future," he said, "and reach our hands out to *them*$_2$.
Example of the application of heuristic H2Nom
The chairman and *the chief executive officer*$_3$ of Woolworth Corp. have temporarily relinquished their posts while the retailer conducts its investigation into *alleged accounting irregularities*$_4$. Woolworth's board named John W. Adams, an outsider, to serve as interim chairman and *executive officer*$_3$, while a special committee, appointed by the board last week and led by Mr. Adams, investigates *the alleged irregularities*$_4$.

Table 2: Examples of coreference resolution. The same annotated index indicates coreference.

The third class of heuristics resolves coreference by coercing nominals. Sometimes coercions involve only derivational morphology - linking verbs with their nominalizations. On other occasions, coercions are obtained as paths of meronyms (e.g. is-part relations) and hypernyms (e.g. is-a relations). Con-

sistency checks implemented for this class of coreference are conservative: either the adjuncts must be identical or the adjunct of the referent must be less specific than the antecedent. Table 1 lists the top performing heuristics of COCKTAIL for pronominal and nominal coreference. Examples of the heuristics operation on the MUC data are presented presented in Table 2. Details of the top performing heuristics of COCKTAIL were reported in (Harabagiu and Maiorano, 1999).

2.3 Bootstrapping for Coreference Resolution

One of the major drawbacks of existing coreference resolution systems is their inability to recognize many forms of coreference displayed by many real-world texts. Recall measures of current systems range between 36% and 59% for both knowledge-based and statistical techniques. Knowledge based-systems would perform better if more coreference constraints were available whereas statistical methods would be improved if more annotated data were available. Since knowledge-based techniques outperform inductive methods, we used high-precision coreference heuristics as knowledge seeds for machine learning techniques that operate on large amounts of unlabeled data. One such technique is *bootstrapping*, which was recently presented in (Riloff and Jones 1999), (Jones et al.1999) as an ideal framework for text learning tasks that have knowledge seeds. The method does not require large training sets. We extended COCKTAIL by using *meta-bootstrapping* of both new heuristics and clusters of nouns that display semantic consistency for coreference.

The coreference heuristics are the seeds of our bootstrapping framework for coreference resolution. When applied to large collections of texts, the heuristics determine classes of coreferring expressions. By generating coreference chains out of all these coreferring expressions, often new heuristics are uncovered. For example, Figure 2 illustrates the application of three heuristics and the generation of data for a new heuristic rule. In COCKTAIL, after a heuristic is applied, a new coreference chain is calculated. For the example illustrated in Figure 2, if the reference of expression A is sought, heuristic H1 indicates expression B to be the antecedent. When the coreference chain is built, expression A is directly linked to expression D, thus uncovering a new heuristic H0.

As a rule of thumb, we do not consider a new heuristic unless there is massive evidence of its coverage in the data. To measure the coverage we use

the *FOIL_Gain measure*, as introduced by the FOIL inductive algorithm (Cameron-Jones and Quinlan 1993). Let H_0 be the new heuristic and H_1 a heuristic that is already in the seed set. Let p_0 be the number of positive coreference examples of H_{new} (i.e. the number of coreference relations produced by the heuristic that can be found in the test data) and n_0 the number of negative examples of H_{new} (i.e. the number of relations generated by the heuristic which cannot be found in the test data). Similarly, p_1 and n_1 are the positive and negative examples of H_1. The new heuristics are scored by their *FOIL_Gain* distance to the existing set of heuristics, and the best scoring one is added to the COCKTAIL system. The *FOIL_Gain* formula is:

$$FOIL\_Gain(H_1, H_0) = k(log_2\frac{p_1}{p_1 + n_1} - log_2\frac{p_0}{p_0 + n_0})$$

where k is the number of positive examples covered by both H_1 and H_0. Heuristic H_0 is added to the seed set if there is no other heuristic providing larger *FOIL_Gain* to any of the seed heuristics.
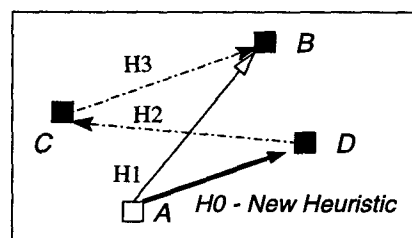

Figure 2: Bootstrapping new heuristics.

Since in COCKTAIL, semantic consistency of coreferring expressions is checked by comparing the similarity of noun classes, each new heuristic determines the adjustment of the similarity threshold of all known coreferring noun classes. The steps of the bootstrapping algorithm that learns both new heuristics and adjusts the similarity threshold of coreferential expressions is:

MUTUAL BOOTSTRAPPING LOOP
1. *Score all candidate heuristics with FOIL_Gain*
2. *Best_h=closest candidate to heuristics(COCKTAIL)*
3. *Add Best_h to heuristics(COCKTAIL)*
4. *Adjust semantic similarity threshold for semantic consistency of coreferring nouns*
5. *Goto step 1 if the precision and recall did not degrade under minimal performance.*

(Riloff and Jones 1999) note that the bootstrapping algorithm works well but its performance can deteriorate rapidly when non-coreferring data enter as candidate heuristics. To make the algorithm more robust, a second level of bootstrapping can be introduced. The outer bootstrapping mechanism, called

meta-bootstrapping compiles the results of the inner (mutual) bootstrapping process and identifies the k most reliable heuristics, where k is a number determined experimentally. These k heuristics are retained and the rest of them are discarded.

3 SWIZZLE

3.1 Multilingual Coreference Data

To study the performance of a data-driven multilingual coreference resolution system, we prepared a corpus of Romanian texts by translating the MUC-6 and MUC-7 coreference training texts. The translations were performed by a group of four Romanian native speakers, and were checked for style by a certified translator from Romania. In addition, the Romanian texts were annotated with coreference keys. Two rules were followed when the annotations were done:

o*1:* Whenever an expression E_R represents a translation of an expression E_E from the corresponding English text, if E_E is tagged as a coreference key with identification number *ID*, then the Romanian expression E_R is also tagged with the same *ID* number. This rule allows for translations in which the textual position of the referent and the antecedent have been swapped.

o*2:* Since the translations often introduce new coreferring expressions in the same chain, the new expressions are given new, unused *ID* numbers. For example, Table 3 lists corresponding English and Romanian fragments of coreference chains from the original MUC-6 Wall Street Journal document DOCNO: 930729-0143.

Table 3 also shows the original MUC coreference SGML annotations. Whenever present, the *REF* tag indicates the *ID* of the antecedent, whereas the *MIN* tag indicates the minimal reference expression.

3.2 Lexical Resources

The multilingual coreference resolution method implemented in SWIZZLE incorporates the heuristics derived from COKCTAIL's monolingual coreference resolution processing in both languages. To this end, COCKTAIL required both sets of texts to be tagged for part-of-speech and to recognize the noun phrases. The English texts were parsed with Brill's part-of-speech tagger (Brill 1992) and the noun phrases were identified by the grammar rules implemented in the phrasal parser of FASTUS (Appelt et al., 1993). Corresponding resources are not available in Romanian.

To minimize COCKTAIL's configuration for processing Romanian texts, we implemented a Romanian part-of-speech rule-based tagger that used the same

Economic adviser Gene Sperling described
<COREF ID="29" TYPE="IDENT" REF="30">
it< /COREF> as "a true full-court press" to pass
<COREF ID="31" TYPE="IDENT" REF="26"
MIN="<u>bill</u>">the <COREF ID="32"
TYPE="IDENT" REF="10" MIN="reduction">
<COREF ID="33" TYPE="IDENT" REF="12">
deficit< /COREF>-**reduction**< /COREF>
bill, the final version of which is now being
hammered out by <COREF ID="43">**House**
< /COREF> and <COREF ID="41">**Senate**
< /COREF>**negotiators**< /COREF>.
....
<COREF ID="34" TYPE="IDENT" REF="2">
The executives< /COREF>' backing – however tepid
– gives the administration a way to counter
<COREF ID="35" TYPE="IDENT" REF="36">
business< /COREF> critics of <COREF ID="500"
TYPE="IDENT" REF="31" MIN="<u>package</u>"
STATUS="OPT">**the overall package**
< /COREF>,...

Consilierul cu probleme economice Gene Sperling a
descris-<COREF ID="29" TYPE="IDENT"
REF="30">o< /COREF> ca pe un efort de
avengură menit să promoveze <COREF ID="1125"
TYPE="IDENT" REF="26" MIN="legea">**legea**
< /COREF> pentru <COREF TYPE="IDENT"
REF="10" MIN="reducerea"> reducerea
< /COREF><COREF ID="33" TYPE="IDENT"
REF="12"> deficitului in bugetul SUA< /COREF>.
Versiunea finala a acestei <COREF ID="1126"
TYPE="IDENT" REF="1125" MIN="legi">**legi**
< /COREF> este desfiinţată chiar în aceste
zile în cadrul dezbaterilor ce au loc în
<COREF ID="43">**Camera Reprezentativilor**
< /COREF> şi în <COREF ID="41">
Senat< /COREF>< /COREF>.
....
Sprijinirea <COREF ID="127" TYPE="IDENT"
REF="1126" MIN="legii">**legii**> /COREF>
de către specialişti în economie – deşi
în manieră moderată – oferă administraţiei o
modalitate de a contrabalansa criticile aduse
<COREF ID="500" TYPE="IDENT" REF="31"
MIN="legii" STATUS="OPT">**legii**< /COREF>
de către companiile americane,...

Table 3: Example of parallel English and Romanian text annotated for coreference. The elements from a coreference chain in the respective texts are underlined. The English text has only two elements in the coreference chain, whereas the Romanian text contains four different elements. The two additional elements of the Romanian coreference chain are derived due to (1) the need to translate the relative clause from the English fragment into a separate sentence in Romanian; and (2) the reordering of words in the second sentence.

tags as generated by the Brill tagger. In addition, we implemented rules that identify noun phrases in Romanian.

To take advantage of the aligned corpus, SWIZZLE also relied on bilingual lexical resources that help translate the referential expressions. For this purpose, we used a core Romanian WordNet (Harabagiu, 1999) which encoded, wherever possible, links between the English synsets and their Romanian counterparts. This resource also incorporated knowledge derived from several bilingual dictionaries (e.g. (Bantaş, 1969)).

Having the parallel coreference annotations, we can easily identify their translations because they have the same identification coreference key. Looking at the example given in Table 3, the expression "legii", with ID=500 is the translation of the expression "package", having the same ID in the English text. However, in the test set, the REF fields are intentionally voided, entrusting COCKTAIL to identify the antecedents. The bilingual coreference resolution performed in SWIZZLE, however, requires the translations of the English and Romanian antecedents. The principles guiding the translations of the English and Romanian antecedents (A^{E-R} and A^{R-E}, respectively) are:

• *Circularity:* Given an English antecedent, due to semantic ambiguity, it can belong to several English WordNet sysnsets. For each such sysnset S_i^E we consider the Romanian corresponding sysnet(s) S_j^R. We filter out all S_j^R that do not contain A^{E-R}. If only one Romanian sysnset is left, then we identified a translation. Otherwise, we start from the Romanian antecedent, find all synsets S_k^R to which it belongs, and obtain the corresponding English sysnets S_l^E. Similarly, all English synsets not containing the English antecedent are filtered out. If only one synset remains, we have again identified a translation. Finally, in the last case, the intersection of the multiple synsets in either language generates a legal translation. For example, the English synset $S^E = \{bill, measure\}$ translates into the Romanian synset $S^R = \{lege\}$. First, none of the dictionary translations of *bill* into Romanian (e.g. *poliţă, bacnotă, afiş*) translate back into any of the elements of S^E. However the translation of *measure* into the Romanian *lege* translates back into *bill*, its synonym.

• *Semantic density:* Given an English and a Romanian antecedent, to establish whether they are translations of one another, we disambiguate them by first collapsing all sysnsets that have common elements. Then we apply the circularity principle, relying on the semantic alignment encoded in the Romanian WordNet. When this core lexical database was first

implemented, several other principles were applied. In our experiment, we were satisfied with the quality of the translations recognized by following only these two principles.

3.3 Multilingual Coreference Resolution

The SWIZZLE system was run on a corpus of 2335 referential expressions in English (927 from MUC-6 and 1408 from MUC-7) and 2851 Romanian expressions (1219 from MUC-6 and 1632 from MUC-7). Initially, the heuristics implemented in COCKTAIL were applied separately to the two textual collections. Several special cases arose.
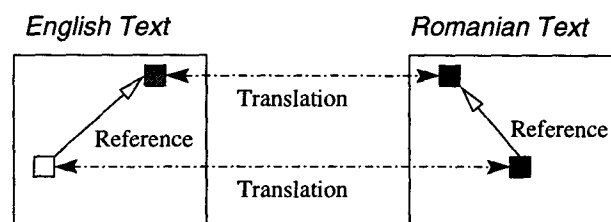

Figure 3: Case 1 of multilingual coreference

Case 1, which is the ideal case, is shown in Figure 3. It occurs when two referential expressions have antecedents that are translations of one another. This situation occurred in 63.3% of the referential expressions from MUC-6 and in 58.7% of the MUC-7 references. Over 50% of these are pronouns or named entities. However, all the non-ideal cases are more interesting for SWIZZLE, since they port knowledge that enhances system performance.
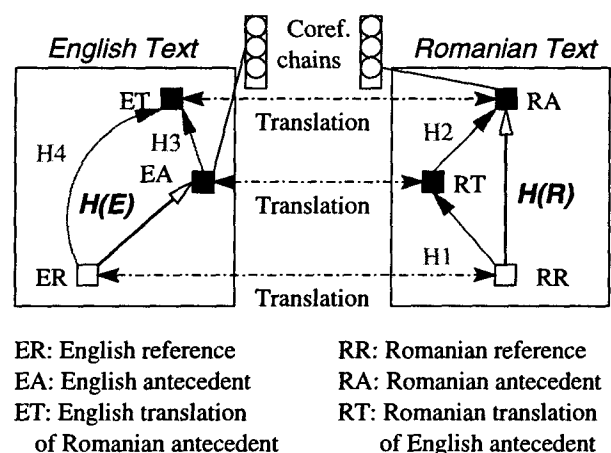

ER: English reference   RR: Romanian reference
EA: English antecedent   RA: Romanian antecedent
ET: English translation   RT: Romanian translation
  of Romanian antecedent   of English antecedent

Figure 4: Case 2 of multilingual coreference

Case 2 occurs when the antecedents are not translations, but belong to or corefer with elements of some coreference chains that were already established. Moreover, one of the antecedents is textually

147

closer to its referent. Figure 4 illustrates the case when the English antecedent is closer to the referent than the Romanian one.

SWIZZLE Solutions: (1) If the heuristic *H(E)* used to resolve the reference in the English text has higher priority than *H(R)*, which was used to resolve the reference from the Romanian text, then we first search for *RT*, the Romanian translation of *EA*, the English antecedent. In the next step, we add heuristic *H1* that resolves *RR* into *RT*, and give it a higher priority than *H(R)*. Finally, we also add heuristic *H2* that links *RT* to *RA* when there is at least one translation between the elements of the coreference chains containing *EA* and *ET* respectively.

(2) If *H(R)* has higher priority than *H(E)*, heuristic *H3* is added while *H(E)* is removed. We also add *H4* that relates *ER* to *ET*, the English translation of *RA*.

Case 3 occurs when at least one of the antecedents starts a new coreference chain (i.e., no coreferring antecedent can be found in the current chains).

SWIZZLE Solution: If one of the antecedents corefers with an element from a coreference chain, then the antecedent in the opposite language is its translation. Otherwise, SWIZZLE chooses the antecedent returned by the heuristic with highest priority.

4 Results

The foremost contribution of SWIZZLE was that it improved coreference resolution over both English and Romanian texts when compared to monolingual coreference resolution performance in terms of precision and recall. Also relevant was the contribution of SWIZZLE to the process of understanding the cultural differences expressed in language and the way these differences influence coreference resolution. Because we do not have sufficient space to discuss this issue in detail here, let us state, in short, that English is more economical than Romanian in terms of referential expressions. However the referential expressions in Romanian contribute to the resolution of some of the most difficult forms of coreference in English.

4.1 Precision and Recall

Table 4 summarizes the precision results for both English and Romanian coreference. The results indicate that the English coreference is more precise than the Romanian coreference, but SWIZZLE improves coreference resolution in both languages. There were 64% cases when the English coreference was resolved by a heuristic with higher priority than the corresponding heuristic for the Romanian coun-

terpart. This result explains why there is better precision enhancement for the English coreference.

	Nominal	Pronominal	Total
English	73%	89%	84%
Romanian	66%	78%	72%
SWIZZLE on English	76%	93%	87%
SWIZZLE on Romanian	71%	82%	76%

Table 4: Coreference precision

	Nominal	Pronominal	Total
English	69%	89%	78%
Romanian	63%	83%	72%
SWIZZLE on English	66%	87%	77%
SWIZZLE on Romanian	61%	80%	70%

Table 5: Coreference recall

Table 5 also illustrates the recall results. The advantage of the data-driven coreference resolution over other methods is based on its better recall performance. This is explained by the fact that this method captures a larger variety of coreference patterns. Even though other coreference resolution systems perform better for some specific forms of reference, their recall results are surpassed by the data-driven approach. Multilingual coreference in turn improves more the precision than the recall of the monolingual data-driven coreference systems.

In addition, Table 5 shows that the English coreference results in better recall than Romanian coreference. However, the recall shows a decrease for both languages for SWIZZLE because imprecise coreference links are deleted. As is usually the case, deleting data lowers the *recall*. All results were obtained by using the automatic scorer program developed for the MUC evaluations.

5 Conclusions

We have introduced a new data-driven method for multilingual coreference resolution, implemented in the SWIZZLE system. The results of this method are encouraging since they show clear improvements over monolingual coreference resolution. Currently, we are also considering the effects of a bootstrapping algorithm for multilingual coreference resolution. Through this procedure we would learn concurrently semantic consistency knowledge and better performing heuristic rules. To be able to develop such a learning approach, we must first develop a method for automatic recognition of multilingual referential expressions.

We also believe that a better performance evaluation of SWIZZLE can be achieved by measuring its impact on several complex applications. We intend to analyze the performance of SWIZZLE when it is used as a module in an IE system, and separately in a Question/Answering system.

Acknowledgements This paper is dedicated to the memory of our friend Megumi Kameyama, who inspired this work.

References

Douglas E. Appelt, Jerry R. Hobbs, John Bear, David Israel, Megumi Kameyama and Mabry Tyson. 1993. The SRI MUC-5 JV-FASTUS Information Extraction System. In *Proceedings of the Fifth Message Understanding Conference (MUC-5)*.

Brack Baldwin. 1997. CogNIAC: high precision coreference with limited knowledge and linguistic resources. In *Proceedings of the ACL'97/EACL'97 Workshop on Operational factors in practical, robust anaphora resolution*, pages 38–45, Madrid, Spain.

Andrei Bantaş. 1969. Dicţionar Român-Englez, Enlgez-Român. Editura Ştiinţifică, Bucureşti.

David Bean and Ellen Riloff. 1999. Corpus-Based Identification of Non-Anaphoric Noun Phrases. In *Proceedings of the 37th Conference of the Assosiation for Computatioanl Linguistics (ACL-99)*, pages 373–380.

Eric Brill. A simple rule-based part of speech tagger. In *Proceedings of the Third Conference on Applied Natural Language Processing*, pages 152–155, 1992.

Joseph F. Cameron-Jones and Ross Quinlan. 1993. Avoiding Pitfalls When Learning Recursive Theories. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI-93)*, pages 1050-1055.

Claire Cardie and Kiri Wagstaff. 1999. Noun phrase coreference as clustering. In *Proceedings of the Joint Conference on Empirical Methods in NLP and Very Large Corpora*, pages 82–89.

Niyu Ge, John Gale and Eugene Charniak. 1998. Anaphora Resolution: A Multi-Strategy Approach. In *Proceedings of the 6th Workshop on Very Large Corpora, (COLING/ACL'98)*.

Ido Dagan and Ken W. Church. 1994. TERMIGHT: Identifying and translating technical terminology. In *Proceedings of the 4th ACL Conference on Applied Natural Language Processing (ANLP-94)*.

Sanda M. Harabagiu. 1999. Lexical Acquisition for a Romanian WordNet. *Proceeding of the 3rd European Summer School on Computational Linguistics*.

Sanda M. Harabagiu and Steve J. Maiorano. 1999. Knowledge-Lean Coreference Resolution and its Relation to Textual Cohesion and Coherence. In *Proceedings of the Workshop on the Relation of Discourse/Dialogue Structure and Reference, ACL'98*, pages 29–38.

Jerry R. Hobbs. Resolving pronoun references. *Lingua*, 44:311–338.

Andrew Kehler. 1997. Probabilistic Coreference in Information Extraction. In *Proceedings of the Second Conference on Empirical Methods in Natural Language Processing (SIGDAT)*, pages 163–173.

Shalom Lappin and Herbert Leass. 1994. An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20(4):535–562.

Rosie Jones, Andrew McCallum, Kevin Nigam and Ellen Riloff. 1999. Bootstrapping for Text Learning Tasks. In *Proceedings of the IJCAI-99 Workshop on Text Mining: Foundations, Techniques, and Applications*.

Megumi Kameyama. 1997. Recognizing Referential Links: An Information Extraction Perspective. In *Proceedings of the Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts, (ACL-97/EACL-97)*, pages 46–53, Madrid, Spain.

Christopher Kennedy and Branimir Bogureav. 1996. Anaphora for everyone: Pronominal anaphora resolution without a parser. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING-96)*.

George A. Miller. 1995. WordNet: A Lexical Database. *Communication of the ACM*, 38(11):39–41.

Ruslan Mitkov. 1998. Robust pronoun resolution with limited knowledge. In *Proceedings of COLING-ACL'98*, pages 869–875.

1996. *Proceedings of the Sixth Message Understanding Conference (MUC-6)*,Morgan Kaufmann, San Mateo, CA.

1998. *Proceedings of the Seventh Message Understanding Conference (MUC-7)* ,Morgan Kaufmann, San Mateo, CA.

Philip Resnik and I. Dan Melamed. 1997. Semi-Automatic Acquisition of Domain-Specific Translation Lexicons. In *Proceedings of the 5th ACL Conference on Applied Natural Language Processing (ANLP-97)*.

Ellen Riloff and Rosie Jones. 1999. Learning Dictionaries for Information Extraction by Multi-Level Bootstrapping. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)*.

Frank Smadja, Katheleen R. McKeown and Vasileios Hatzivassiloglou. 1996. Translating collocations for bilingual lexicons: A statistical approach. *Computational Linguistics* , 21(1):1–38.