Word Alignment for Languages with Scarce Resources

Joel Martin
National Research Council
Ottawa, ON, K1A 0R6
Joel.Martin@cnrc-nrc.gc.ca

Rada Mihalcea
University of North Texas
Denton, TX 76203
rada@cs.unt.edu

Ted Pedersen
University of Minnesota
Duluth, MN 55812
tpederse@umn.edu

Abstract

This paper presents the task definition, resources, participating systems, and comparative results for the shared task on word alignment, which was organized as part of the ACL 2005 Workshop on Building and Using Parallel Texts. The shared task included English–Inuktitut, Romanian–English, and English–Hindi sub-tasks, and drew the participation of ten teams from around the world with a total of 50 systems.

1 Defining a Word Alignment Shared Task

The task of word alignment consists of finding correspondences between words and phrases in parallel texts. Assuming a sentence aligned bilingual corpus in languages L1 and L2, the task of a word alignment system is to indicate which word token in the corpus of language L1 corresponds to which word token in the corpus of language L2.

This year's shared task follows on the success of the previous word alignment evaluation that was organized during the HLT/NAACL 2003 workshop on "Building and Using Parallel Texts: Data Driven Machine Translation and Beyond" (Mihalcea and Pedersen, 2003). However, the current edition is distinct in that it has a focus on languages with scarce resources. Participating teams were provided with training and test data for three language pairs, accounting for different levels of data scarceness: (1) *English–Inuktitut* (2 million words training data), (2) *Romanian–English* (1 million words), and (3) *English–Hindi* (60,000 words).

Similar to the previous word alignment evaluation and with the Machine Translation evaluation exercises organized by NIST, two different subtasks were defined: (1) *Limited resources*, where systems were allowed to use only the resources provided. (2) *Unlimited resources*, where systems were allowed to use any resources in addition to those provided. Such resources had to be explicitly mentioned in the system description.

Test data were released one week prior to the deadline for result submissions. Participating teams were asked to produce word alignments, following a common format as specified below, and submit their output by a certain deadline. Results were returned to each team within three days of submission.

1.1 Word Alignment Output Format

The word alignment result files had to include one line for each word-to-word alignment. Additionally, they had to follow the format specified in Figure 1. Note that the $S|P$ and confidence fields overlap in their meaning. The intent of having both fields available was to enable participating teams to draw their own line on what they considered to be a Sure or Probable alignment. Both these fields were optional, with some standard values assigned by default.

1.1.1 A Running Word Alignment Example

Consider the following two aligned sentences:
[English] *<s snum=18> They had gone . </s>*
[French] *<s snum=18> Ils étaient allés . </s>*

A correct word alignment for this sentence is:

```
18 1 1
18 2 2
18 3 3
18 4 4
```


Proceedings of the ACL Workshop on Building and Using Parallel Texts, pages 65–74,
Ann Arbor, June 2005.

sentence_no position_L1 position_L2 [$S|P$] [confidence]

where:
∘ **sentence_no** represents the id of the sentence within the test file. Sentences in the test data already have an id assigned. (see the examples below)
∘ **position_L1** represents the position of the token that is aligned from the text in language L1; the first token in each sentence is token 1. (not 0)
∘ **position_L2** represents the position of the token that is aligned from the text in language L2; again, the first token is token 1.
∘ **S|P** can be either S or P, representing a Sure or Probable alignment. All alignments that are tagged as S are also considered to be part of the P alignments set (that is, all alignments that are considered "Sure" alignments are also part of the "Probable" alignments set). If the $S|P$ field is missing, a value of S will be assumed by default.
∘ **confidence** is a real number, in the range (0-1] (1 meaning highly confident, 0 meaning not confident); this field is optional, and by default confidence number of 1 was assumed.

Figure 1: Word Alignment file format

stating that: all the word alignments pertain to sentence 18, the English token 1 *They* aligns with the French token 1 *Ils*, the English token 2 *had* aligns with the French token 2 *étaient*, and so on. Note that punctuation is also aligned (English token 4 aligned with French token 4), and counts toward the final evaluation figures.

Alternatively, systems could also provide an $S|P$ marker and/or a confidence score, as shown in the following example:

```
18 1 1 1
18 2 2 P 0.7
18 3 3 S
18 4 4 S 1
```

with missing $S|P$ fields considered by default S, and missing confidence scores considered by default 1.

1.2 Annotation Guide for Word Alignments

The word alignment annotation guidelines are similar to those used in the 2003 evaluation.

1. All items separated by a white space are considered to be a word (or token), and therefore have to be aligned (punctuation included).
2. Omissions in translation use the NULL token, i.e. token with id 0.
3. Phrasal correspondences produce multiple word-to-word alignments.

2 Resources

The shared task included three different language pairs, accounting for different language and data characteristics. Specifically, the three subtasks addressed the alignment of words in English–Inuktitut, Romanian–English, and English–Hindi parallel texts. For each language pair, training data were provided to participants. Systems relying only on these resources were considered part of the *Limited Resources* subtask. Systems making use of any additional resources (e.g. bilingual dictionaries, additional parallel corpora, and others) were classified under the *Unlimited Resources* category.

2.1 Training Data

Three sets of training data were made available. All data sets were sentence-aligned, and pre-processed (i.e. tokenized and lower-cased), with identical preprocessing procedures used for training, trial, and test data.

English–Inuktitut. A collection of sentence-aligned English–Inuktitut parallel texts from the Legislative Assembly of Nunavut (Martin et al., 2003). This collection consists of approximately 2 million Inuktitut tokens (1.6 million words) and 4 million English tokens (3.4 million words). The Inuktitut data was originally encoded in Unicode representing a syllabics orthography (qaniujaaqpait), but was transliterated to an ASCII encoding of the standardized roman orthography (qaliujaaqpait) for this evaluation.

Romanian–English. A set of Romanian–English parallel texts, consisting of about 1 million Romanian words, and about the same number of English words. This is the same training data set as used in the 2003 word alignment evaluation (Mihalcea and Pedersen, 2003). The data consists of:

- Parallel texts collected from the Web using a semi-supervised approach. The URLs format for pages containing potential parallel translations were manually identified (mainly from the archives of Romanian newspapers). Next, texts were automatically downloaded and sentence aligned. A manual verification of the alignment was also performed. These data collection process resulted in a corpus of about 850,000 Romanian words, and about 900,000 English words.

- Orwell's 1984, aligned within the MULTEXT-EAST project (Erjavec et al., 1997), with about 130,000 Romanian words, and a similar number of English words.

- The Romanian Constitution, for about 13,000 Romanian words and 13,000 English words.

English–Hindi. A collection of sentence aligned English–Hindi parallel texts, from the Emille project (Baker et al., 2004), consisting of approximately English 60,000 words and about 70,000 Hindi words. The Hindi data was encoded in Unicode Devangari script, and used the UTF–8 encoding. The English–Hindi data were provided by Niraj Aswani and Robert Gaizauskas from University of Sheffield (Aswani and Gaizauskas, 2005b).

2.2 Trial Data

Three sets of trial data were made available at the same time training data became available. Trial sets consisted of sentence aligned texts, provided together with manually determined word alignments. The main purpose of these data was to enable participants to better understand the format required for the word alignment result files. For some systems, the trial data has also played the role of a validation data set used for system parameter tuning. Trial sets consisted of 25 English–Inuktitut and English–Hindi aligned sentences, and a larger set of 248 Romanian–English aligned sentences (the same as the test data used in the 2003 word alignment evaluation).

2.3 Test Data

A total of 75 English–Inuktitut, 90 English–Hindi, and 200 Romanian–English aligned sentences were released one week prior to the deadline. Participants were required to run their word alignment systems on one or more of these data sets, and submit word alignments. Teams were allowed to submit an unlimited number of results sets for each language pair.

2.3.1 Gold Standard Word Aligned Data

The gold standard for the three language pair alignments were produced using slightly different alignment procedures.

For English–Inuktitut, annotators were instructed to align Inuktitut words or phrases with English phrases. Their goal was to identify the smallest phrases that permit one-to-one alignments between English and Inuktitut. These phrase alignments were converted into word-to-word alignments in the following manner. If the aligned English and Inuktitut phrases each consisted of a single word, that word pair was assigned a Sure alignment. Otherwise, all possible word-pairs for the aligned English and Inuktitut phrases were assigned a Probable alignment. Disagreements between the two annotators were decided by discussion.

For Romanian–English and English–Hindi, annotators were instructed to assign an alignment to *all* words, with specific instructions as to when to assign a NULL alignment. Annotators were not asked to assign a Sure or Probable label. Instead, we had an arbitration phase, where a third annotator judged the cases where the first two annotators disagreed. Since an inter-annotator agreement was reached for all word alignments, the final resulting alignments were considered to be Sure alignments.

3 Evaluation Measures

Evaluations were performed with respect to four different measures. Three of them – precision, recall, and F-measure – represent traditional measures in Information Retrieval, and were also frequently used in previous word alignment literature. The fourth measure was originally introduced by (Och and Ney, 2000), and proposes the notion of *quality of word alignment*.

Given an alignment $\mathcal{A}$, and a gold standard alignment $\mathcal{G}$, each such alignment set eventually consisting of two sets $\mathcal{A}_\mathcal{S}$, $\mathcal{A}_\mathcal{P}$, and $\mathcal{G}_\mathcal{S}$, $\mathcal{G}_\mathcal{P}$ corresponding to Sure and Probable alignments, the following measures are defined (where T is the alignment type, and can be set to either S or P).

$$P_T = \frac{|A_T \cap G_T|}{|A_T|} \tag{1}$$

$$R_T = \frac{|A_T \cap G_T|}{|G_T|} \tag{2}$$

$$F_T = \frac{2P_T R_T}{P_T + R_T} \tag{3}$$

$$AER = 1 - \frac{|A_P \cap G_S| + |A_P \cap G_P|}{|A_P| + |G_S|} \tag{4}$$

Each word alignment submission was evaluated in terms of the above measures. Given numerous (constructive) debates held during the previous word alignment evaluation, which questioned the informativeness of the NULL alignment evaluations, we decided

Team	System name	Description
Carnegie Mellon University	SPA	(Brown et al., 2005)
Information Sciences Institute / USC	ISI	(Fraser and Marcu, 2005)
Johns Hopkins University	JHU	(Schafer and Drabek, 2005)
Microsoft Research	MSR	(Moore, 2005)
Romanian Academy Institute of Artificial Intelligence	TREQ-AL, MEBA, COWAL	(Tufis et al., 2005)
University of Maryland / UMIACS	UMIACS	(Lopez and Resnik, 2005)
University of Sheffield	Sheffield	(Aswani and Gaizauskas, 2005a)
University of Montreal	JAPA, NUKTI	(Langlais et al., 2005)
University of Sao Paulo, University of Alicante	LIHLA	(Caseli et al., 2005)
University Jaume I	MAR	(Vilar, 2005)

Table 1: Teams participating in the word alignment shared task

to evaluate only no-NULL alignments, and thus the NULL alignments were removed from both submissions and gold standard data. We conducted therefore 7 evaluations for each submission file: AER, Sure/Probable Precision, Sure/Probable Recall, and Sure/Probable F-measure, all of them measured on no-NULL alignments.

4 Participating Systems

Ten teams from around the world participated in the word alignment shared task. Table 1 lists the names of the participating systems, the corresponding institutions, and references to papers in this volume that provide detailed descriptions of the systems and additional analysis of their results.

Seven teams participated in the Romanian–English subtask, four teams participated in the English–Inuktitut subtask, and two teams participated in the English–Hindi subtask. There were no restrictions placed on the number of submissions each team could make. This resulted in a total of 50 submissions from the ten teams, where 37 sets of results were submitted for the Romanian–English subtask, 10 for the English–Inuktitut subtask, and 3 for the English–Hindi subtask. Of the 50 total submissions, there were 45 in the *Limited resources* subtask, and 5 in the *Unlimited resources* subtask. Tables 2, 4 and 6 show all of the submissions for each team in the three subtasks, and provide a brief description of their approaches.

Results for all participating systems, including precision, recall, F-measure, and alignment error rate are listed in Tables 3, 5 and 7. Ranked results for all systems are plotted in Figures 2, 3 and 4. In the graphs, systems are ordered based on their AER scores. System names are preceded by a marker to indicate the system type: L stands for *Limited Resources*, and U stands for *Unlimited Resources*.

While each participating system was unique, there were a few unifying themes. Several teams had approaches that relied (to varying degrees) on an IBM model of statistical machine translation (Brown et al., 1993), with different improvements brought by different teams, consisting of new submodels, improvements in the HMM model, model combination for optimal alignment, etc. Se-veral teams used symmetrization metrics, as introduced in (Och and Ney, 2003) (union, intersection, refined), most of the times applied on the alignments produced for the two directions source–target and target–source, but also as a way to combine different word alignment systems. Significant improvements with respect to baseline word alignment systems were observed when the vocabulary was reduced using simple stemming techniques, which seems to be a particularly effective technique given the data sparseness problems associated with the relatively small amounts of training data.

In the *unlimited resources* subtask, systems made use of bilingual dictionaries, human–contributed word alignments, or syntactic constraints derived from a dependency parse tree applied on the English side of the corpus.

When only small amounts of parallel corpora were available (i.e. the English–Hindi subtask), the use of additional resources resulted in absolute improvements of up to 20% as compared to the case when the word alignment systems were based exclusively on the parallel texts. Interestingly, this was not the case for the language pairs that had larger training corpora (i.e. Romanian–English, English–Inuktitut), where the *limited resources* systems seemed to lead to comparable or sometime even better results than those that relied on *unlimited resources*. This suggests that the use of additional resources does not seem to contribute to improvements in word alignment quality when enough parallel corpora are available, but they can make a big difference when only small amounts of parallel texts are available.

Finally, in a comparison across language pairs, the best results are obtained in the English–Inuktitut task, followed by Romanian–English, and by English–Hindi, which corresponds to the ordering of the sizes of the training data sets. This is not surprising since, like many other NLP tasks, word alignment seems to highly benefit from large amounts of training data, and thus better results are obtained when larger training data sets are available.

5 Conclusion

A shared task on word alignment was organized as part of the ACL 2005 Workshop on Building and Using Parallel Texts. The focus of the task was on languages with scarce resources, with evaluations of alignments for three different language pairs: English–Inuktitut, English–Hindi, and Romanian–English. The task drew the participation of ten teams from around the world, with a total of 50 systems. In this paper, we presented the task definition, resources involved, and shortly described the participating systems. Comparative evaluations of results led to insights regarding the development of word alignment algorithms for languages with scarce resources, with performance evaluations of (1) various algorithms, (2) different amounts of training data, and (3) different additional resources. Data and evaluation software used in this exercise are available online at http://www.cs.unt.edu/˜rada/wpt05.

Acknowledgments

There are many people who contributed greatly to making this word alignment evaluation task possible. We are grateful to all the participants in the shared task, for their hard work and involvement in this evaluation exercise. Without them, all these comparative analyses of word alignment techniques would not be possible. In particular, we would like to thank Dan Tufiş and Bob Moore for their helpful comments concerning the Romanian–English data. We would also like to thank Benoit Farley for his valuable assistance with the English–Inuktitut data.

We are very thankful to Niraj Aswani and Rob Gaizauskas from University of Sheffield for making possible the English–Hindi word alignment evaluation. They provided sentence aligned training data from the Emille project, as well as word aligned trial and test data sets.

We are also grateful to all the Program Committee members for their comments and suggestions, which helped us improve the definition of this shared task.

References

N. Aswani and R. Gaizauskas. 2005a. Aligning words in english-hindi parallel corpora. In *(this volume)*.

N. Aswani and R. Gaizauskas. 2005b. A hybrid approach to align sentences and words in English-Hindi parallel corpora. In *Proceedings of the ACL Workshop on "Building and Exploiting Parallel Texts"*, Ann Arbor, MI.

P. Baker, K. Bontcheva, H. Cunningham, R. Gaizauskas, O. Hamza, A. Hardie, B. Jayaram, M. Leisher, A McEnery, D Maynard, V. Tablan, C. Ursu, and Z. Xiao. 2004. Corpus linguistics and south asian languages: Corpus creation and tool development. *Literary and Linguistic Computing*, 19(4).

P. Brown, S. della Pietra, V. della Pietra, and R. Mercer. 1993. The mathematics of statistical machine translation: parameter estimation. *Computational Linguistics*, 19(2).

R. D. Brown, J.D. Kim, P. J. Jansen, and J. G. Carbonell. 2005. Symmetric probabilistic alignment. In *(this volume)*.

H. Caseli, M. G. V. Nunes, and M. L. Forcada. 2005. Lihla: Shared task system description. In *(this volume)*.

T. Erjavec, N. Ide, and D. Tufis. 1997. Encoding and parallel alignment of linguistic corpora in six central and Eastern European languages. In *Proceedings of the Joint ACH/ALL Conference*, Queen's University, Kingston, Ontario, June.

A. Fraser and D. Marcu. 2005. Isi's participation in the romanian-english alignment task. In *(this volume)*.

P. Langlais, F. Gotti, and G. Cao. 2005. Nukti: English-inuktitut word alignment system description. In *(this volume)*.

A. Lopez and P. Resnik. 2005. Improved hmm alignment models for languages with scarce resources. In *(this volume)*.

J. Martin, H. Johnson, B. Farley, and A. Maclachlan. 2003. Aligning and using an english-inuktitut parallel corpus. In *Proceedings of the HLT-NAACL Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*, Edmonton, Canada.

R. Mihalcea and T. Pedersen. 2003. An evaluation exercise for word alignment. In *HLT-NAACL 2003 Workshop: Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*, Edmonton, Canada, May.

R. Moore. 2005. Association-based bilingual word alignment. In *(this volume)*.

F. Och and H. Ney. 2000. A comparison of alignment models for statistical machine translation. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000)*, Saarbrucken, Germany, August.

F.J. Och and H. Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1).

C. Schafer and E. Drabek. 2005. Models for inuktitut-english word alignment. In *(this volume)*.

D. Tufis, R. Ion, A. Ceausu, and D. Stefanescu. 2005. Combined word alignments. In *(this volume)*.

J.M. Vilar. 2005. Experiments using mar for aligning corpora. In *(this volume)*.

System	Resources	Description
JHU.AER.Emphasis.I	Limited	A word alignment system optimized for the characteristics of English–Inuktitut, exploiting cross-lingual affinities at sublexical level and regular patterns of transliteration. The system is based on classifier combination, performed under an AER target evaluation metric.
JHU.AER.Emphasis.II	Limited	Same as JHU.AER.Emphasis.I, but with a different minimum required votes for classifier combination.
JHU.F-meas.Emphasis	Limited	Same as JHU.AER.Emphasis.I, with classifier combination performed under an F-measure target evaluation metric.
JHU.AER.F-meas.AER DualEmphasis	Limited	Same as JHU.AER.Emphasis.I, with a dual emphasis on AER and F-measure.
JHU.Recall.Emphasis	Limited	Same as JHU.AER.Emphasis.I, with an emphasis on recall.
LIHLA	Limited	A word alignment tool based on language-independent heuristics. Starts with two bilingual probabilistic lexicons (source-target and target-source) generated by NATools (http://natura.di.uminho.pt/natura/natura/), which are combined with some language-independent heuristics that try to find the best alignment.
UMIACS.limited	Limited	A system using IBM Model 4 with improvements brought in the HMM model.
UMontreal.NUKTI	Limited	A system based on computation of log-likelihood ratios between all Inuktitut substrings and English words. Alignment with a greedy strategy trying to optimize this association score.
UMontreal.Japa-cart	Limited	A system based on alignment with a sentence aligner where Inuktitut and English words are considered to be sentences. In case a n-m alignment is produced, its cartesian product is output as the final alignment.
UMontreal.Japa-nukti	Limited	Same as UMontreal.Japa-cart except for the treatment of the n-m pairs (n,m $\geq$ 1). Instead of generating the cartesian product, this method uses the NUKTI approach to figure out which words should be aligned.

Table 2: Short description for English–Inuktitut systems

System	P_S	R_S	F_S	P_P	R_P	F_P	AER
Limited Resources							
JHU.AER.Emphasis.II	34.19%	76.79%	47.32%	**96.66%**	32.35%	48.37%	**9.46%**
JHU.AER.Emphasis.I	28.15%	82.25%	41.95%	90.65%	39.35%	54.88%	11.49%
JHU.F-measure.AER.DualEmphasis	19.71%	92.15%	32.47%	84.38%	58.62%	69.18%	14.25%
UMIACS.limited	**49.86%**	62.80%	55.59%	89.16%	16.68%	28.11%	22.51%
LIHLA	46.55%	73.72%	**57.07%**	79.53%	18.71%	30.30%	22.72%
JHU.F-measure.Emphasis	13.06%	91.81%	22.87%	70.67%	73.78%	**72.19%**	26.70%
UMontreal.nukti	12.24%	86.01%	21.43%	63.09%	65.87%	64.45%	34.06%
JHU.Recall.Emphasis	10.68%	**93.86%**	19.18%	62.63%	**81.74%**	70.92%	34.18%
UMontreal.Japa-nukti	9.62%	67.58%	16.84%	51.34%	53.60%	52.44%	46.64%
UMontreal.Japa-cart	0.00%	0.00%	0.00%	26.17%	74.49%	38.73%	71.27%

Table 3: Results for English–Inuktitut

System	Resources	Description
CMU.SPA contiguous	Limited	A tool based on Symmetric Probabilistic Alignment (SPA), which maximizes bi-directional translation probabilities of words in a selected source-language n-gram and every possible target-language n-gram. Probabilities are derived from a pair of probabilistic lexicons (source-to-target and target-to-source). Only contiguous target-language n-grams are considered as possible alignments.
CMU.SPA non-contiguous	Limited	Same as CMU.SPA.contiguous, but both contiguous and non-contiguous target-language n-grams are considered as possible alignments
CMU.SPA human-augmented	Unlimited	Same as CMU.SPA.contiguous, but the probabilistic dictionaries were modified with word and phrasal translations extracted from a human alignment of 204 sentences in the training corpus.
ISI.RUN1	Limited	A baseline word-based system using IBM Model 4 as implemented in Giza++. Different subruns include the two separate direction En–Ro, Ro–En, as well as the "union", "intersection", and "refined" symmetrization metrics, as defined in (Och and Ney, 2003)
ISI.RUN2	Limited	Same as ISI.RUN1, but uses stems of size 4 (instead of words) for both English and Romanian.
ISI.RUN4	Limited	A system using IBM Model 4 and a new submodel based on the intersection of two starting alignments. The submodels are grouped into a log-linear model, with optimal weights found through a search algorithm.
ISI.RUN5	Limited	Same as ISI.RUN4, but with 5 additional submodels, using translation tables for En–Ro, Ro–En, backoff fertility, zero or non-zero fertility English word penalty
UJaume.MAR	Limited	A new alignment model based on a recursive approach. Due to its high computational cost, heuristics have been used to split training and test data in smaller chunks.
USaoPaulo.LIHLA	Limited	A word alignment tool based on language-independent heuristics. Starts with two bilingual probabilistic lexicons (source-target and target-source) generated by NATools (http://natura.di.uminho.pt/natura/natura/), which are combined with some language-independent heuristics that try to find the best alignment.
MSR.word-align	Limited	A system based on competitive linking, first by log-likelihood-ratio association score, then by probability of link given joint occurrence; constrained by measuring monontonicity of alignment, and augmented with 1-2 and 2-1 alignments also derived by competitive linking.
RACAI.MEBA-V1	Limited	A system based on GIZA++, with a translation model constructed using seven major parameters that control the contribution of various heuristics (cognates, relative distance, fertility, displacement, etc.)
RACAI.MEBA-V2	Limited	Same as RACAI.MEBA-V1, but with a different set of parameters.
RACAI.TREQ-AL	Unlimited	Same as RACAI.MEBA-V1, but with an additional resource consisting of a translation dictionary extracted from the alignment of the Romanian and English WordNet.
RACAI.COWAL	Unlimited	A combination (union) of RACAI.MEBA and RACAI.TREQ-AL.
UMIACS.limited	Limited	A system using IBM Model 4 with improvements brought in the HMM model.
UMIACS.unlimited	Unlimited	Same as UMIACS.limited, but also integrating a distortion model based on a dependency parse built on the English side of the parallel corpus.

Table 4: Short description for Romanian–English systems

System	P_S	R_S	F_S	P_P	R_P	F_P	AER
Limited Resources							
ISI.Run5.vocab.grow	87.90%	63.08%	**73.45%**	87.90%	63.08%	**73.45%**	**26.55%**
ISI.Run3.vocab.grow	87.93%	62.98%	73.40%	87.93%	62.98%	73.40%	26.60%
ISI.Run4.vocab.grow	88.31%	62.75%	73.37%	88.31%	62.75%	73.37%	26.63%
ISI.Run2.vocab.grow	81.84%	66.28%	73.25%	81.84%	66.28%	73.25%	26.75%
ISI.Run5.simple.union	81.78%	65.35%	72.64%	81.78%	65.35%	72.64%	27.36%
ISI.Run5.simple.normal	87.09%	61.93%	72.39%	87.09%	61.93%	72.39%	27.61%
ISI.Run4.simple.union	81.85%	64.69%	72.27%	81.85%	64.69%	72.27%	27.73%
ISI.Run5.simple.inverse	86.96%	61.75%	72.22%	86.96%	61.75%	72.22%	27.78%
ISI.Run3.simple.normal	87.11%	61.63%	72.19%	87.11%	61.63%	72.19%	27.81%
ISI.Run3.simple.union	81.00%	65.05%	72.15%	81.00%	65.05%	72.15%	27.85%
ISI.Run4.simple.normal	87.20%	61.34%	72.02%	87.20%	61.34%	72.02%	27.98%
ISI.Run5.simple.intersect	93.77%	58.33%	71.93%	93.77%	58.33%	71.93%	28.07%
ISI.Run3.simple.intersect	93.92%	57.96%	71.68%	93.92%	57.96%	71.68%	28.32%
ISI.Run3.simple.inverse	86.12%	61.37%	71.67%	86.12%	61.37%	71.67%	28.33%
ISI.Run4.simple.inverse	87.33%	60.78%	71.67%	87.33%	60.78%	71.67%	28.33%
ISI.Run4.simple.intersect	**94.29%**	57.42%	71.38%	**94.29%**	57.42%	71.38%	28.62%
ISI.Run2.simple.inverse	81.32%	63.32%	71.20%	81.32%	63.32%	71.20%	28.80%
ISI.Run2.simple.union	70.46%	**71.31%**	70.88%	70.46%	**71.31%**	70.88%	29.12%
RACAI MEBA-V1	83.21%	60.54%	70.09%	83.21%	60.54%	70.09%	29.91%
ISI.Run2.simple.intersect	94.08%	55.22%	69.59%	94.08%	55.22%	69.59%	30.41%
ISI.Run2.simple.normal	77.04%	63.20%	69.44%	77.04%	63.20%	69.44%	30.56%
RACAI MEBA-V2	77.90%	61.85%	68.96%	77.90%	61.85%	68.96%	31.04%
ISI.Run1.simple.grow	75.82%	62.23%	68.35%	75.82%	62.23%	68.35%	31.65%
UMIACS.limited	73.77%	61.69%	67.19%	73.77%	61.69%	67.19%	32.81%
ISI.Run1.simple.inverse	72.70%	57.34%	64.11%	72.70%	57.34%	64.11%	35.89%
ISI.Run1.simple.union	59.96%	68.85%	64.10%	59.96%	68.85%	64.10%	35.90%
MSR.word-align	79.54%	53.13%	63.70%	79.54%	53.13%	63.70%	36.30%
CMU.SPA.contiguous	64.96%	61.34%	63.10%	64.96%	61.34%	63.10%	36.90%
CMU.SPA.noncontiguous	64.91%	61.34%	63.07%	64.91%	61.34%	63.07%	36.93%
ISI.Run1.simple.normal	67.41%	56.81%	61.66%	67.41%	56.81%	61.66%	38.34%
ISI.Run1.simple.intersect	93.75%	45.30%	61.09%	93.75%	45.30%	61.09%	38.91%
UJaume.MAR	54.04%	64.65%	58.87%	54.04%	64.65%	58.87%	41.13%
USaoPaulo.LIHLA	57.68%	53.51%	55.51%	57.68%	53.51%	55.51%	44.49%
Unlimited Resources							
RACAI.COWAL	71.24%	**76.77%**	**73.90%**	71.24%	**76.77%**	**73.90%**	**26.10%**
RACAI.TREQ-AL	**82.08%**	60.62%	69.74%	**82.08%**	60.62%	69.74%	30.26%
UMIACS.unlimited	72.41%	**62.15%**	66.89%	72.41%	**62.15%**	66.89%	33.11%
CMU.SPA.human-augmented	64.60%	60.54%	62.50%	64.60%	60.54%	62.50%	37.50%

Table 5: Results for Romanian–English

System	Resources	Description
USheffield	Unlimited	A multi-feature approach for many-to-many word alignment. Prior to word alignment, a pattern-based local word grouping is performed for both English and Hindi. Various methods such as dictionary lookup, transliteration similarity, expected English word(s) and nearest aligned neighbors are used.
UMIACS.limited	Limited	A system using IBM Model 4 with improvements brought in the HMM model.
UMIACS.unlimited	Unlimited	Same as UMIACS.limited, but also integrating a distortion model based on a dependency parse built on the English side of the parallel corpus.

Table 6: Short description for English–Hindi systems

System	P_S	R_S	F_S	P_P	R_P	F_P	AER
Limited Resources							
UMIACS.limited	**42.90%**	**56.00%**	**48.58%**	**42.90%**	**56.00%**	**48.58%**	**51.42%**
Unlimited Resources							
USheffield	**77.03%**	**60.68%**	**67.88%**	**77.03%**	**60.68%**	**67.88%**	**32.12%**
UMIACS.unlimited	43.65%	56.14%	49.12%	43.65%	56.14%	49.12%	50.88%

Table 7: Results for English–Hindi

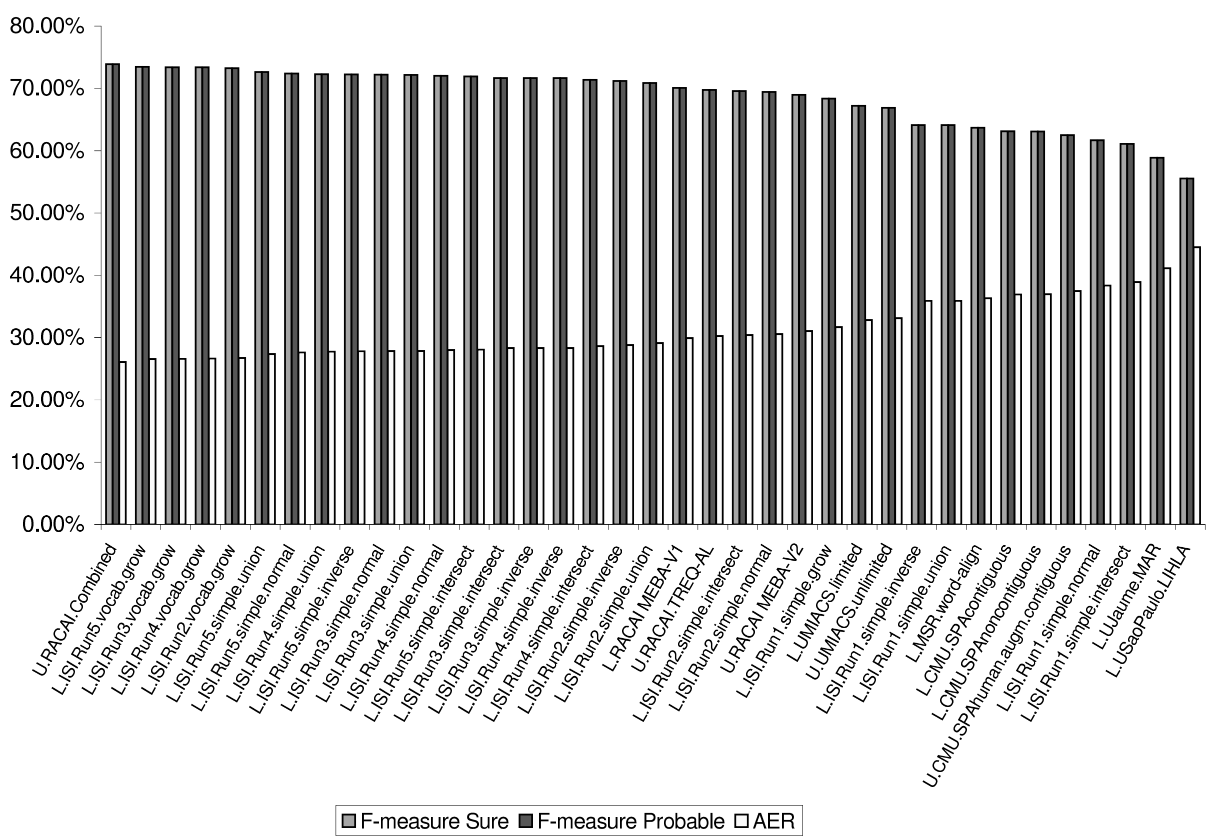

Figure 2: Ranked results for Romanian–English data

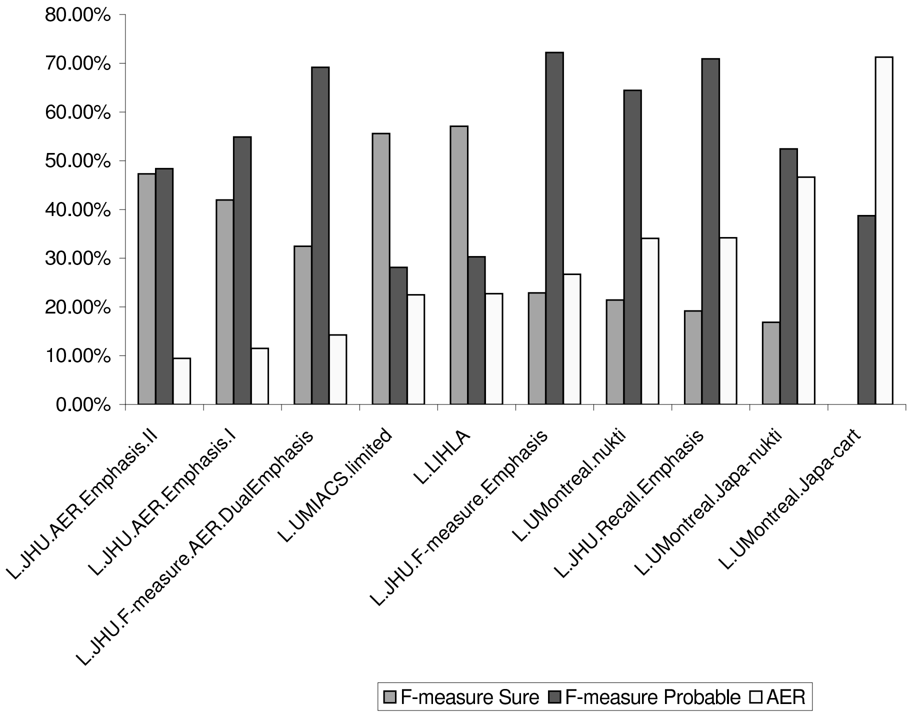

Figure 3: Ranked results for English–Inuktitut data

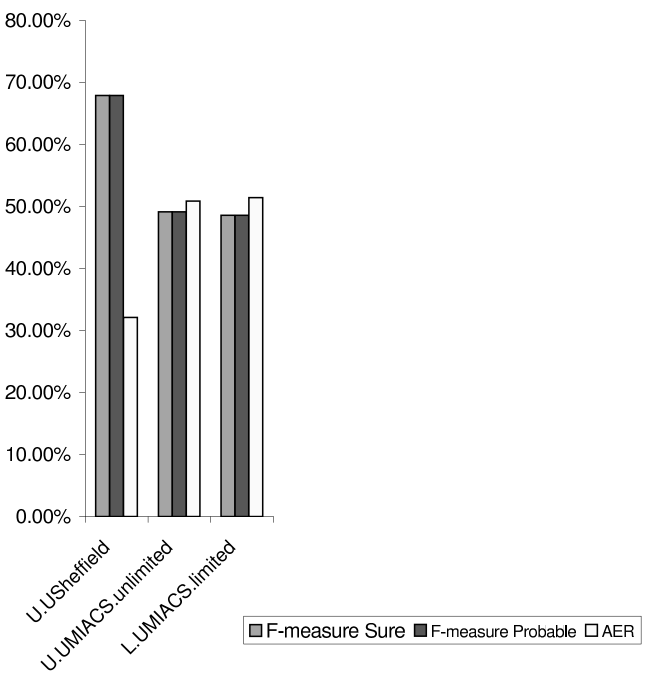

Figure 4: Ranked results for English–Hindi data