AER: Do we need to “improve” our alignments?

David Vilar, Maja Popović, Hermann Ney

International Workshop on Spoken Language Translation 2006
Kyoto, Japan
November 28, 2006

Human Language Technology and Pattern Recognition
Lehrstuhl für Informatik 6
Computer Science Department
RWTH Aachen University, Germany

Outline

1 Introduction

- Alignments are a key concept for statistical machine translation:
  - Represent the correspondence between the words of the source and target sentences.
  - Introduced in [Brown & Della Pietra+ 93] as a hidden variable (framework: EM Algorithm).
- Most widely used alignment models: IBM1–IBM5 (single word-based models).
- State-of-the-art translation systems: phrase based models (or variations).
  - Most approaches make use of the Viterbi alignment in the training process.
  - Alignment used as an additional knowledge source.

- **It is expected that an increase in alignment quality leads to an increase in translation quality.**
- **Alignment Error Rate (AER): Standard measure for alignment quality:**

$$\mathrm{AER}(S, P; A) = 1 - \frac{|A \cap S| + |A \cap P|}{|A| + |S|} .$$

- **We will present empirical evidence that a worse AER can result in an improvement in translation quality.**

2 Related Work

- [Fraser & Marcu 06] present an experimental study on the correlation of AER and BLEU.
- They propose to use the F-measure directly as an indicator of alignment quality:

$$\text{F-measure}(A,P,S,\alpha) = \frac{1}{\frac{\alpha}{\text{Precision}(A,P)} + \frac{1-\alpha}{\text{Recall}(A,S)}}$$

with

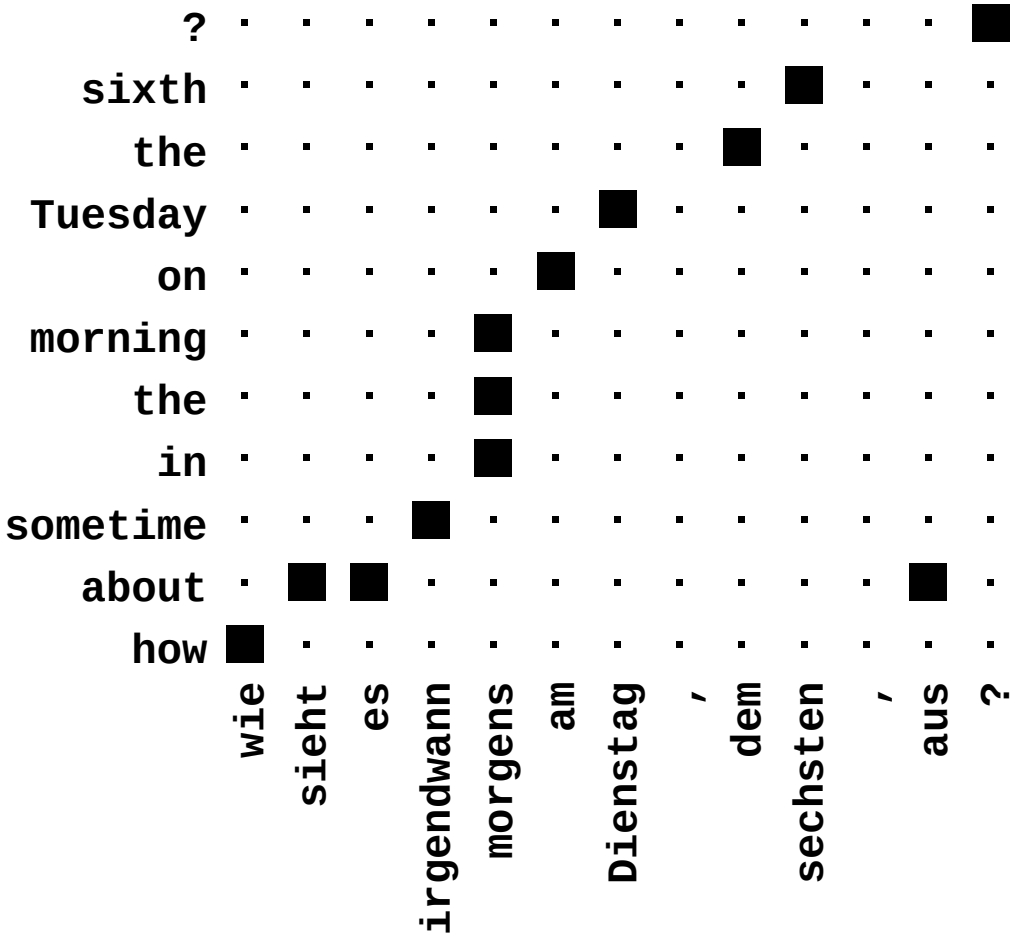
$$\text{Precision}(A,P) = \frac{|A \cap P|}{|A|} \qquad \text{Recall}(A,S) = \frac{|A \cap S|}{|S|}.$$

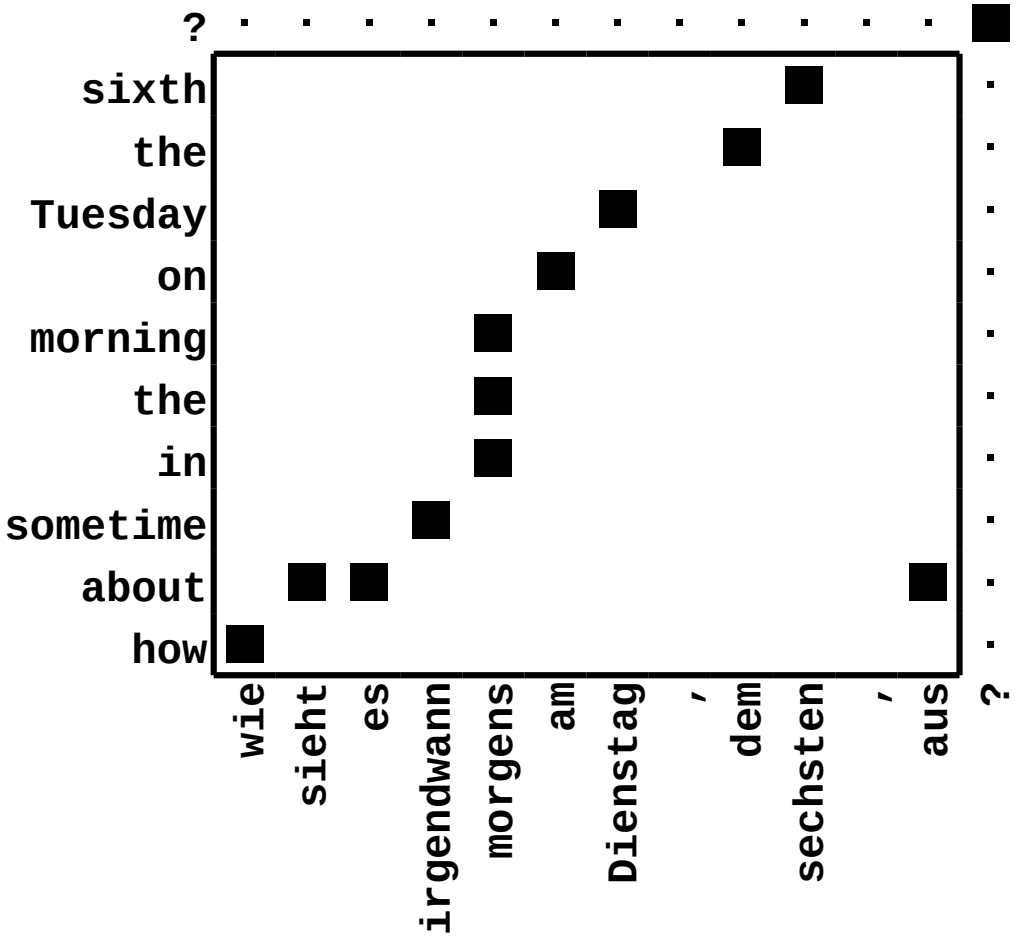
- We will also investigate this measure in this work.

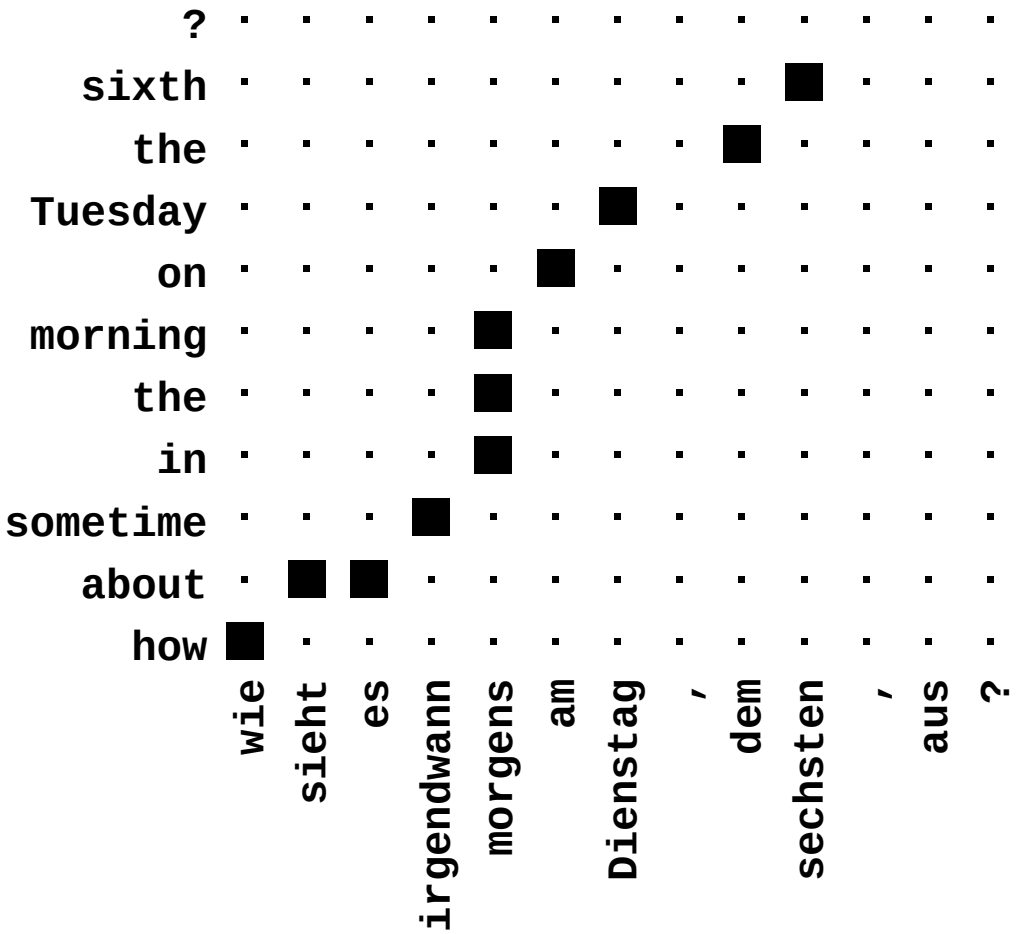
3 Phrase-Based Translation

- **Most important model(s) in a log-linear framework.**
- **Given a word-aligned parallel corpus, all phrases are extracted such that:**
  1. **all source words within the phrase are aligned only to target words within the phrase and**
  2. **all target words within the phrase are aligned only to source words within the phrase.**
- **More formally:**

$$\mathcal{BP}(f_1^J, e_1^I, A) = \{(f_j^{j+m}, e_i^{i+n}) \mid \forall (i', j') \in A : j \leq j' \leq j+m \Leftrightarrow i \leq i' \leq i+n\}.$$


?
sixth
the
Tuesday
on
morning
the
in
sometime
about
how
wie
sieht
es
irgendwann
morgens
am
Dienstag
,
dem
sechsten
,
aus
?


?
sixth
the
Tuesday
on
morning
the
in
sometime
about
how
wie
sieht
es
irgendwann
morgens
am
Dienstag
,
dem
sechsten
,
aus
?

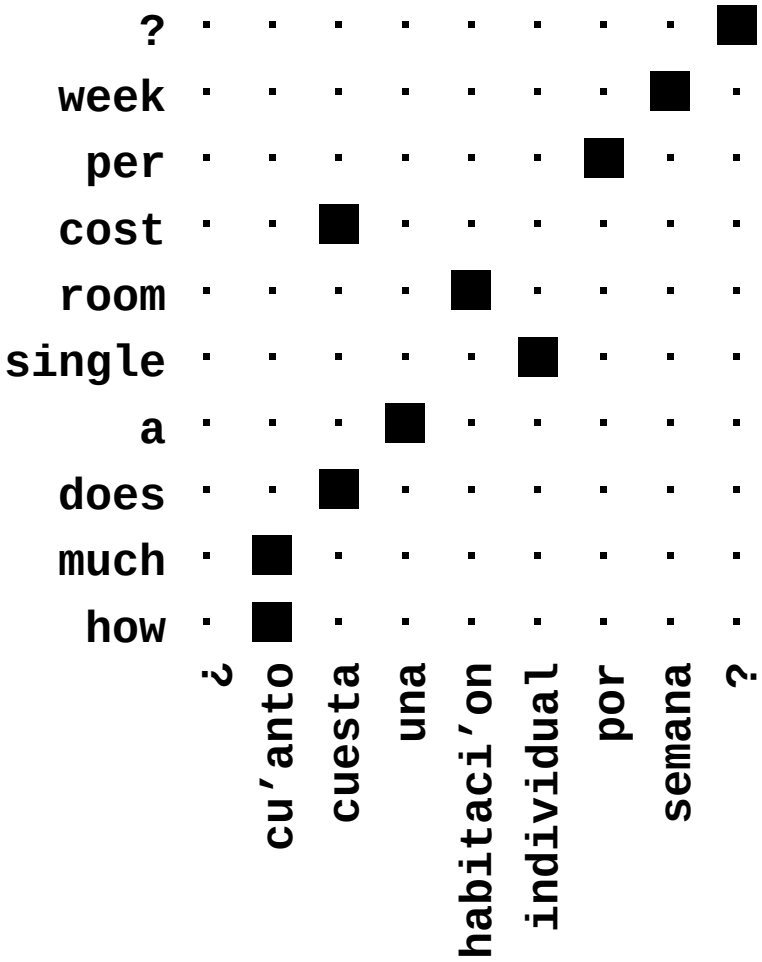

?
sixth
the
Tuesday
on
morning
the
in
sometime
about
how
wie
sieht
es
irgendwann
morgens
am
Dienstag
,
dem
sechsten
,
aus
?

4 Tuple-Based Translation

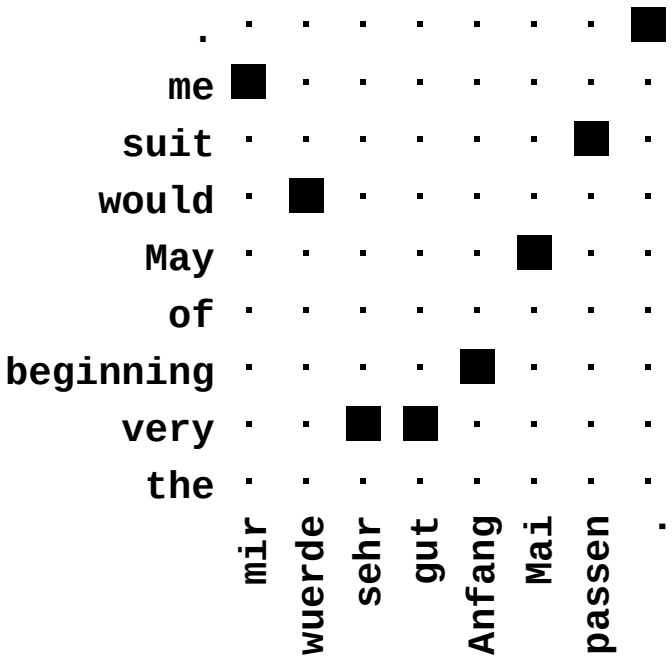
- Use joint probability of source and target string.
- Estimate an m-gram language model over the corpus of tuples $(\tilde{f}_k, \tilde{e}_k)$:

$$\hat{e}_1^I = \underset{\tilde{e}_1^K, A, K}{\operatorname{argmax}} \prod_{k=1}^{K} p(\tilde{f}_k, \tilde{e}_k | \tilde{f}_{k-m}^{k-1}, \tilde{e}_{k-m}^{k-1}, A, K) \, .$$

- RWTH system:
  - $\tilde{f}$ is always a single source word, $\tilde{e}$ is a phrase of 0 or more target words.
  - Use an alignment that is a function of target words.
- See also GIATI method proposed in [Casacuberta & Vidal 04].



¿|$ cuánto|how_much cuesta|does una|a habitación|$ individual|single_room_cost por|per semana|week ?|?



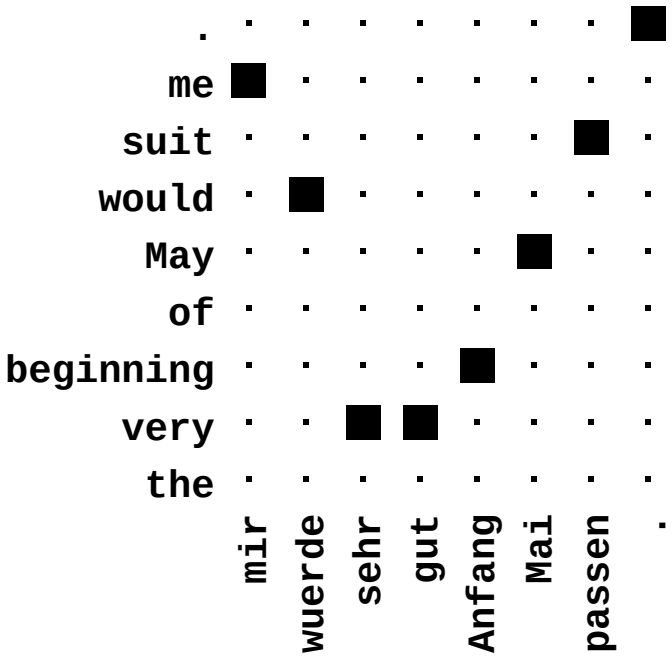

```
mir|$ wuerde|$ sehr|$ gut|$ Anfang|$ Mai|$
passen|the_very_beginning_of_May_would_suit_me
.|.
```

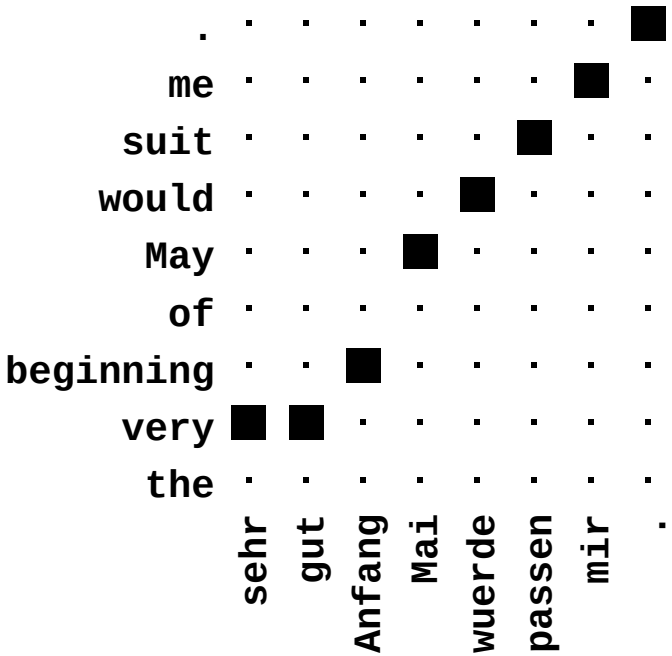
Alignment as a cost minimization problem:

- **Introduce costs $c(i, j)$ of aligning source word f_j to target word e_i.**
- **Define appropriate alignment constraints (full source or target sentence coverage, monotonicity, etc.).**
- **Find an alignment with minimal costs under these constraints.**

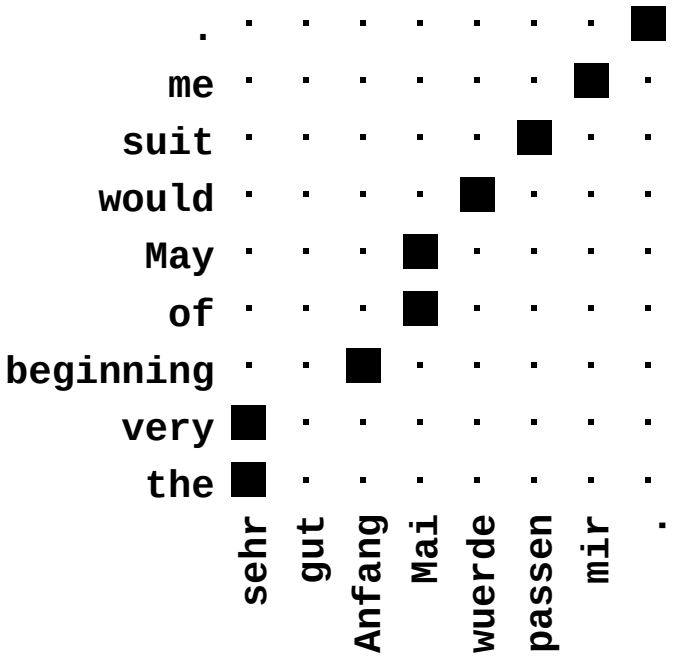
In our case the alignment procedure is:

1. **Find an alignment that is a function of the source words.**
2. **Reorder source sentence guided by this alignment.**
3. **Find an alignment that is a function of the target words.**


.
me
suit
would
May
of
beginning
very
the
mir
wuerde
sehr
gut
Anfang
Mai
passen
.

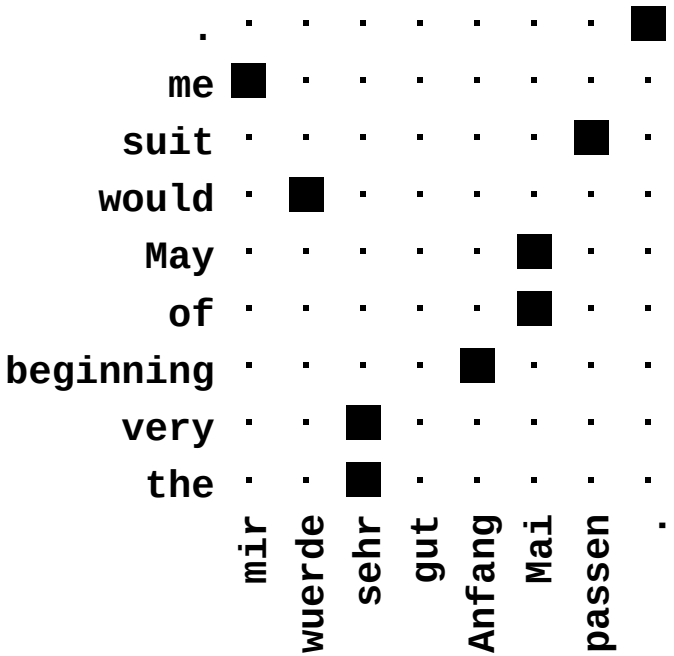


sehr gut Anfang Mai wuerde passen mir .
the very beginning of May would suit me .




```
sehr gut Anfang Mai wuerde passen mir .
the very beginning of May would suit me .

sehr|the_very gut|$ Anfang|beginning
Mai|of_May wuerde|would passen|suit mir|me .|.
```



sehr gut Anfang Mai wuerde passen mir .
the very beginning of May would suit me .

sehr|the_very gut|$ Anfang|beginning
Mai|of_May wuerde|would passen|suit mir|me .|.

5 Experimental Results

- Europarl Corpus (German to English)
- ACL2005 Shared Task

		German	English
Train	Sentences	751 088	
	Words	15 256 793	16 052 269
	Vocabulary	195 291	65 889
Test	Sentences	2 000	
	Words	54 247	57 945

- Reference alignment on a subset of 508 randomly selected sentences.
- Both sure and possible alignments.

System	Alignment	AER[%]	F[%]	BLEU[%]
Phrase Based	Baseline	20.8	77.5	24.6
	Phrases	24.2	71.8	24.8
	Tuples	26.4	73.6	24.5
Tuple Based	Baseline	20.8	77.5	18.2
	Phrases	24.2	71.8	14.8
	Tuples	26.4	73.6	19.4

Example for the Phrase-Based System

Original	**Es wird ein ganzes Kapitel über Wissenschaft, Gesellschaft und Bürger geben.**
Baseline	**It is a chapter on science, society and citizens.**
Phrases	**It will be a whole chapter on science, society and citizens.**
Reference	**There will be an entire chapter on science, society and the citizens.**

Example for the Tuple-Based System

Original	**Litauen verfügt über ein beträchtliches Potential für ein langfristiges Wirtschaftswachstum.**
Baseline	**Has a considerable potential for a long-term Lithuania, although economic growth.**
Tuples	**Lithuania has a considerable potential for a long-term economic growth.**
Reference	**Lithuania has considerable potential for long-term economic growth.**

6 Discussion

- AER and F-measure are appropriate for measuring alignment quality.
- "Inconsistency" between alignment models and translation models:
  - Improvements in alignment quality can lead to decreased translation performance (and vice-versa).
- [Ayan & Dorr 06] presents the "consistent phrase error rate".
- Best solution would be that alignments remain a hidden variable in the statistical modelling process.
- Future work on alignments (oriented to machine translation) should always report results on translation quality.

References

[Ayan & Dorr 06] N. F. Ayan, B. J. Dorr. Going Beyond AER: An Extensive Analysis of Word Alignments and Their Impact on MT. In *Proceedings of the Joint Conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics (COLING-ACL'2006)*, pp. 9–16, Sydney, Australia, July 2006.

[Brown & Della Pietra+ 93] P. F. Brown, S. A. Della Pietra, V. J. Della Pietra, R. L. Mercer. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, Vol. 19, No. 2, pp. 263–311, June 1993.

[Casacuberta & Vidal 04] F. Casacuberta, E. Vidal. Machine Translation with Inferred Stochastic Finite-State Transducers. *Computational Linguistics*, Vol. 30, No. 2, pp. 205–225, 2004.

[Chiang 05] D. Chiang. A hierarchical phrase-based model for statistical machine translation. In *Proc. of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 263–270, Ann Arbor, MI, June 2005.

[Fraser & Marcu 06] A. Fraser, D. Marcu. Measuring word alignment quality for statistical machine translation. Technical report, ISI-University of Southern California, May 2006.

[Kanthak & Vilar+ 05] S. Kanthak, D. Vilar, E. Matusov, R. Zens, H. Ney. Novel reordering approaches in phrase-based statistical machine translation. In *43rd Annual Meeting of the Assoc. for Computational Linguistics: Proc. Workshop on Building and Using Parallel Texts: Data-Driven Machine Translation and Beyond*, pp. 167–174, Ann Arbor, Michigan, June 2005.

[Koehn & Monz 05] P. Koehn, C. Monz. Shared task: Statistical machine translation between european languages. In *Proceedings of the ACL 2005 Workshop on Parallel Text*, pp. 119–124, Ann Arbor, MI, USA, June 2005.

[Marcu & Wong 02] D. Marcu, W. Wong. A phrase-based, joint probability model for statistical machine translation. In *Proc. of the Conf. on Empirical Methods for Natural Language Processing (EMNLP)*, pp. 133–139, Philadelphia, PA, July 2002.

[Mariño & Banchs+ 05] J. Mariño, R. Banchs, P. Lambert, M. Ruiz, J. Crego, J. Fonollosa. Bilingual N-gram Statistical Machine Translation. In *Proceedings of MT Summit X*, pp. 275–282, Phuket, Thailand, September 2005. Asia-Pacific Association for Machine Translation (AAMT).

[Matusov & Zens+ 04] E. Matusov, R. Zens, H. Ney. Symmetric word alignments for statistical machine translation. In *COLING '04: The 20th Int. Conf. on Computational Linguistics*, pp. 219–225, Geneva, Switzerland, August 2004.

[Och & Ney 02] F. J. Och, H. Ney. Discriminative training and maximum entropy models for statistical machine translation. In *Proc. of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 295–302, Philadelphia, PA, July 2002.

[Och & Ney 03] F. J. Och, H. Ney. A systematic comparison of various statistical alignment models. *Computational Linguistics*, Vol. 29, No. 1, pp. 19–51, March 2003.

[Venugopal & Vogel+ 03] A. Venugopal, S. Vogel, A. Waibel. Effective phrase translation extraction from alignment models. In *ACL '03: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pp. 319–326, Morristown, NJ, USA, 2003. Association for Computational Linguistics.

[Vilar & Matusov+ 05] D. Vilar, E. Matusov, S. Hasan, R. Zens, H. Ney. Statistical Machine Translation of European Parliamentary Speeches. In *Proceedings of MT Summit X*, pp. 259–266, Phuket, Thailand, September 2005. Asia-Pacific Association for Machine Translation (AAMT).

[Vilar 00] J. M. Vilar. Improve the learning of subsequential transducers by using alignments and dictionaries. In A. de Oliveira, editor, *Grammatical Inference: Algorithms and Applications*, Vol. 1891 of *Lecture Notes in Artificial Intelligence*, pp. 298–311, Lisbon, Portugal, Sept. 2000. Springer-Verlag.

[Zens & Bender+ 05] R. Zens, O. Bender, S. Hasan, S. Khadivi, E. Matusov, J. Xu, Y. Zhang, H. Ney. The RWTH phrase-based statistical machine translation system. In *Proceedings of the International Workshop on Spoken Language Translation (IWSLT)*, pp. 155–162, Pittsburgh, PA, October 2005.

[Zens & Och+ 02] R. Zens, F. J. Och, H. Ney. Phrase-based statistical machine translation. In M. Jarke, J. Koehler, G. Lakemeyer, editors, *25th German Conf. on Artificial Intelligence (KI2002)*, Vol. 2479 of *Lecture Notes in Artificial Intelligence (LNAI)*, pp. 18–32, Aachen, Germany, September 2002. Springer Verlag.