Statistical Machine Translation

Nadir Durrani

21-November-2014

Machine Translation

- Problem: Automatic translation the foreign text:

འགྲོ་བ་མི་རིགས་ག་ར་དབང་ཆ་འདྲ་མཉམ་འབད་སྐྱེཝ་ལས་ག་ར་
གིས་གཅིག་གིས་གཅིག་ལུ་སྤུན་ཆའི་དམ་ཚིག་བསྟན་དགོ།

Open Problems in Machine Translation

- Ambiguity in translation
  - He deposited money in a bank account with a high interest rate
  - Sitting on the bank of the Mississippi, a passing ship piqued his interest
  - How do we find the right meaning and thus translation?
  - Context should be helpful

- Phrase translation problem

It’s raining cats and dogs

موسلادھار بارش ہو رہی ہے

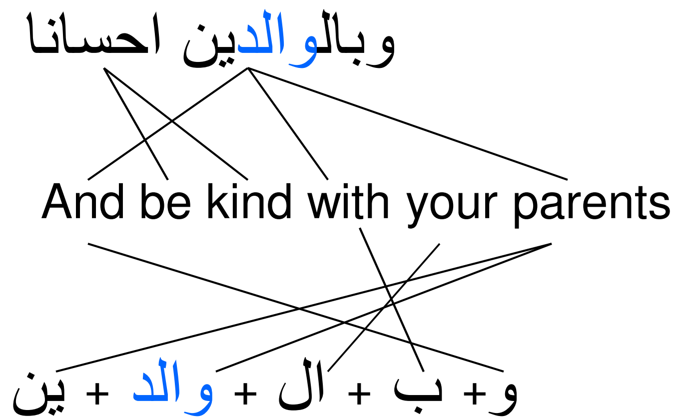
Open Problems in Machine Translation

- Morphological Differences

  Collins et. al (2005)

  Koehn and Hoang (2007)

  Fraser et. al (2012)



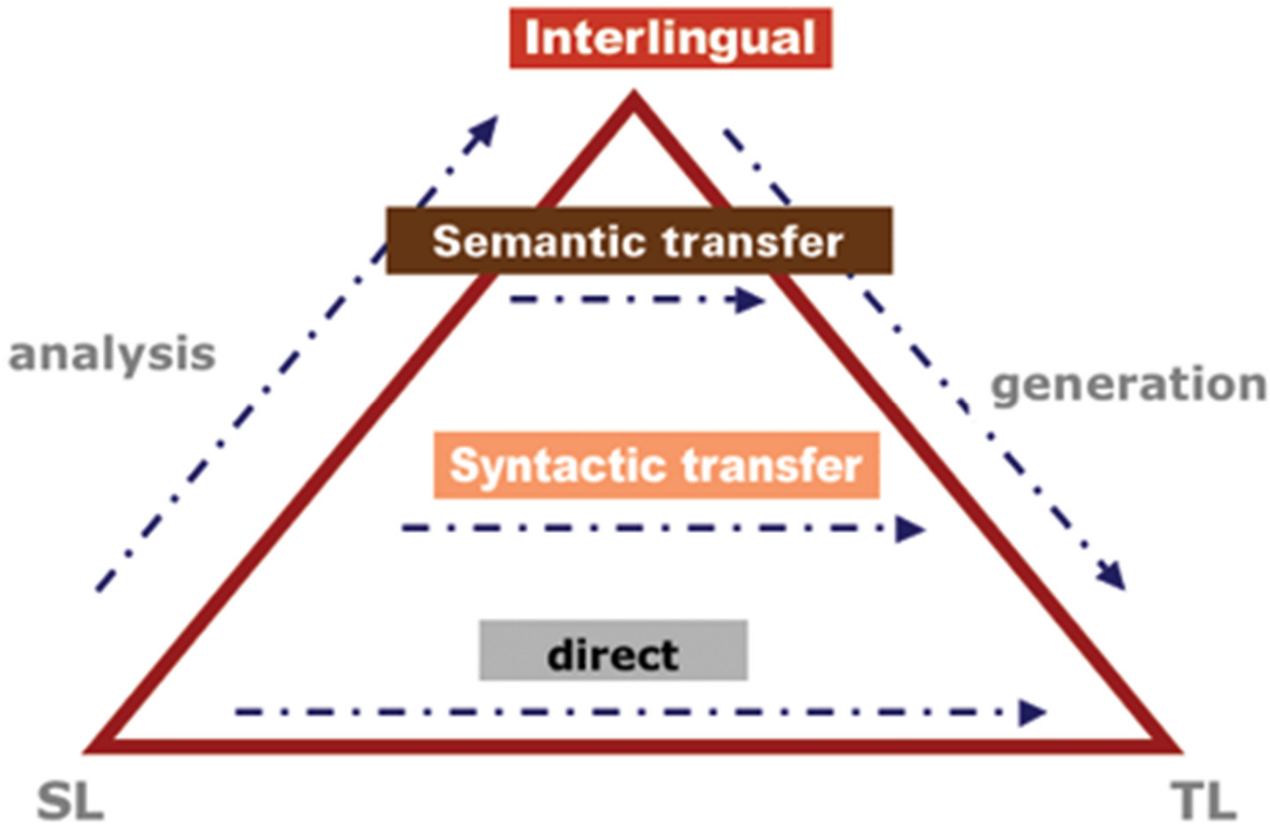
- Structural Differences

  Galley and Manning (2008)

  Green et. al (2010)

  Durrani et al (2011)

The Grand Plan



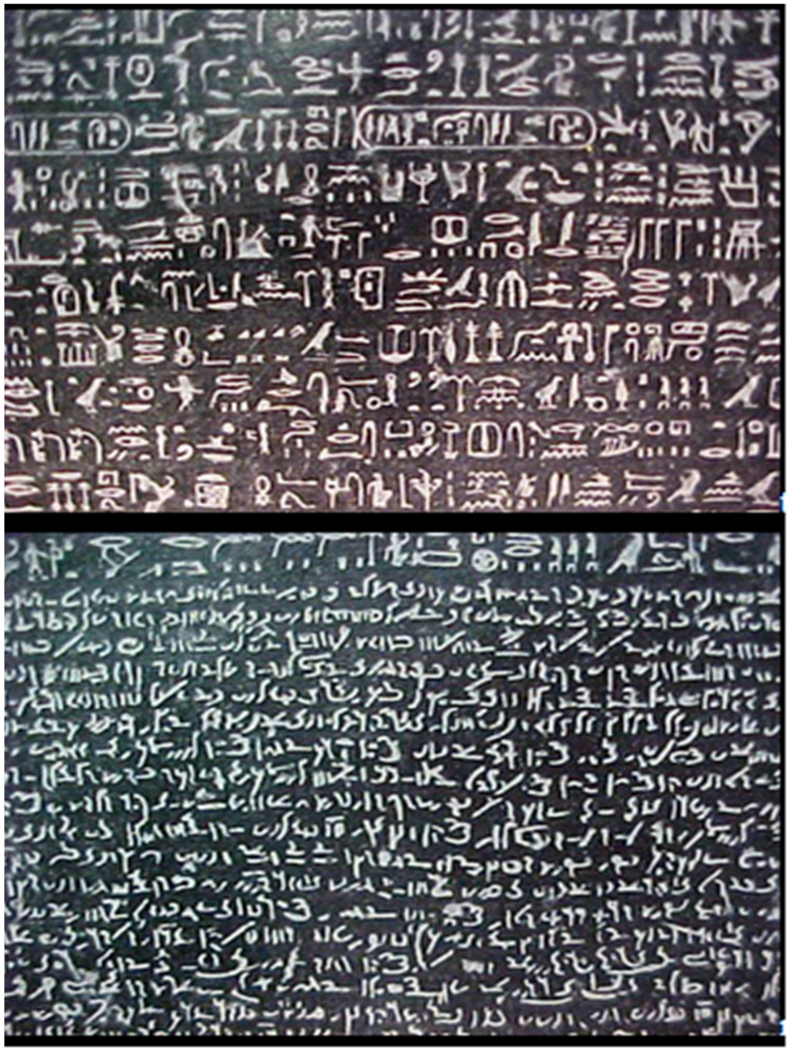
Different Machine Translation Frameworks

- Rule-based
- Empirical
  - Example-based machine translation
  - Statistical machine translation
- Hybrid Machine Translation

Rosetta Stone

- Egyptian language was a mystery for centuries
- The Rosetta stone is written in three scripts
  - Hieroglyphic (used for religious documents)
  - Demotic (common script of Egypt)
  - Greek (language of rulers of Egypt at that time)

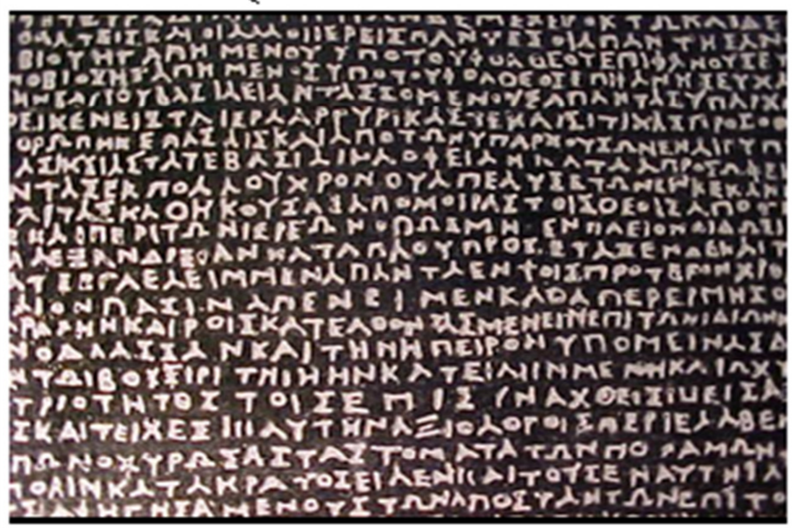
Parallel Data



Heiroglyphic

Demotic

Greek



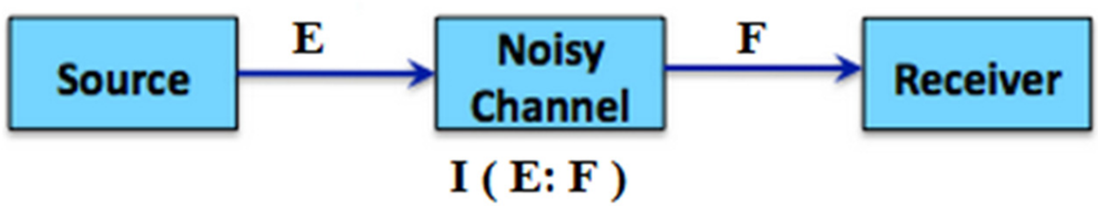
Parallel Data

- UN and European Parliamentary Proceedings
  - German, French, Spanish etc.
- News Corpus and Common Crawl Data
- NIST Data (Arabic, Chinese)

Noisy Channel Model

- Decipherment problem

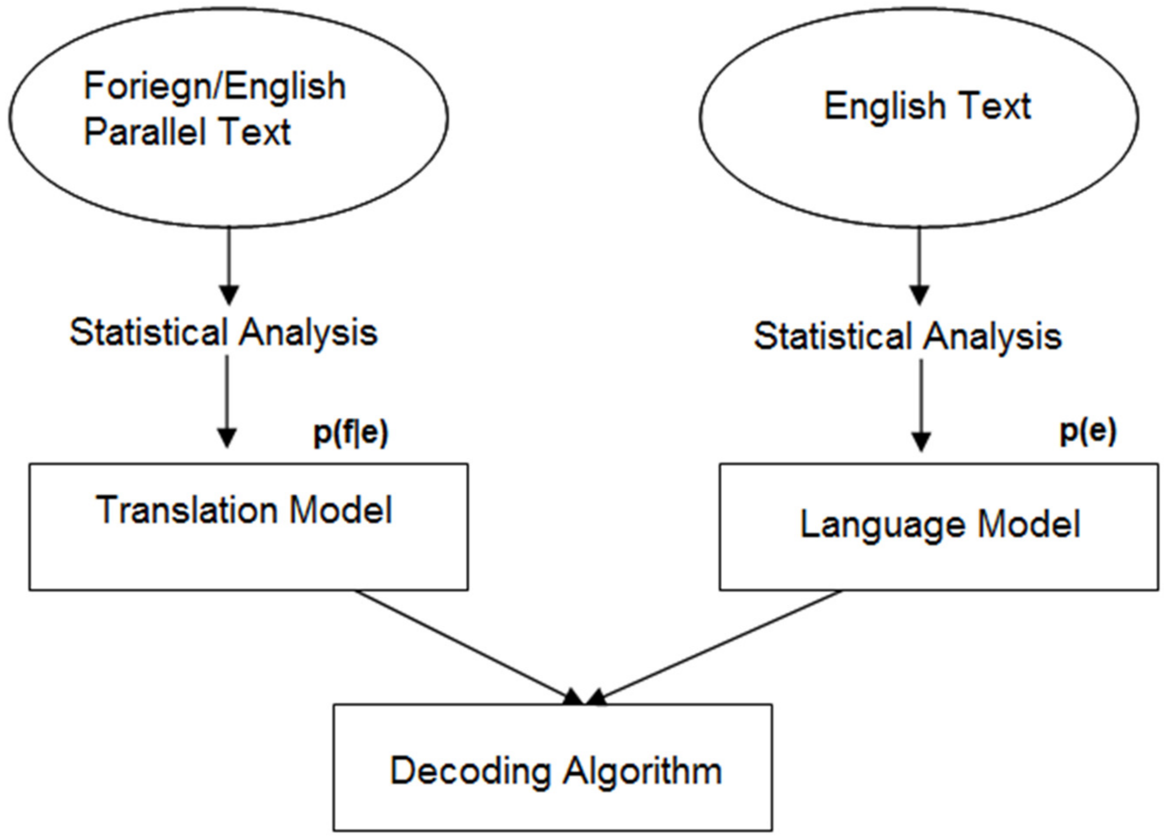
  Warren Weaver: “When I look at an article in Russian, I say: This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode”



- Bayes Rule: p (E | F) = p (F | E) x p(E) / p(F)

  e_{best} = argmax p (E | F) = argmax p (F | E) x p(E)

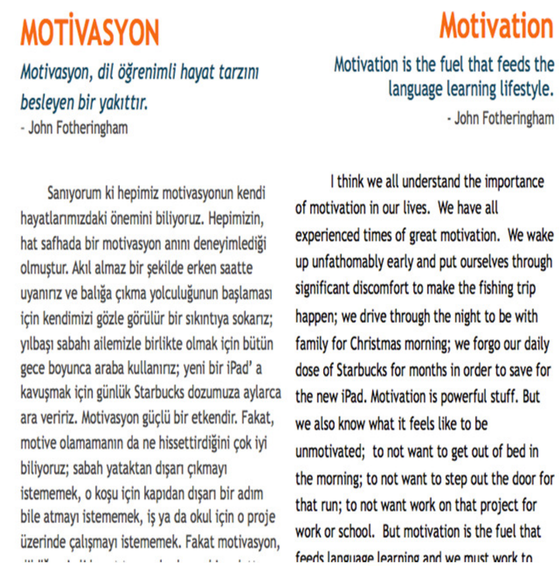
Statistical Machine Translation



From Koehn 2008. University of Edinburgh

Word-based Models (Brown et. al 1992)

- Word alignments
  - If we had word alignment we can learn translation model
  - If we knew model parameters we can learn word alignments
  - Chicken and Egg problem: EM-algorithm



MOTİVASYON

Motivasyon, dil öğrenimli hayat tarzını besleyen bir yakıttır.

- John Fotheringham

Sanıyorum ki hepimiz motivasyonun kendi hayatlarımızdaki önemini biliyoruz. Hepimizin, hat safhada bir motivasyon anını deneyimlediği olmuştur. Akıl almaz bir şekilde erken saatte uyanırız ve balığa çıkma yolculuğunun başlaması için kendimizi gözle görülür bir sıkıntıya sokarız; yılbaşı sabahı ailemizle birlikte olmak için bütün gece boyunca araba kullanırız; yeni bir iPad' a kavuşmak için günlük Starbucks dozumuza aylarca ara veririz. Motivasyon güçlü bir etkendir. Fakat, motive olamamanın da ne hissettirdiğini çok iyi biliyoruz; sabah yataktan dışarı çıkmayı istememek, o koşu için kapıdan dışarı bir adım bile atmayı istememek, iş ya da okul için o proje üzerinde çalışmayı istememek. Fakat motivasyon,

Motivation

Motivation is the fuel that feeds the language learning lifestyle.

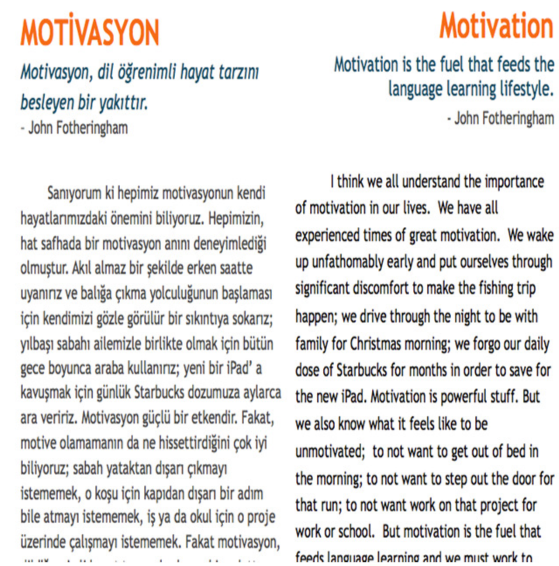
- John Fotheringham

I think we all understand the importance of motivation in our lives. We have all experienced times of great motivation. We wake up unfathomably early and put ourselves through significant discomfort to make the fishing trip happen; we drive through the night to be with family for Christmas morning; we forgo our daily dose of Starbucks for months in order to save for the new iPad. Motivation is powerful stuff. But we also know what it feels like to be unmotivated; to not want to get out of bed in the morning; to not want to step out the door for that run; to not want work on that project for work or school. But motivation is the fuel that feeds language learning and we must work to

Word-based Models (Brown et. al 1992)

- Word alignments
  - If we had word alignment we can learn translation model
  - If we knew model parameters we can learn word alignments
  - Chicken and Egg problem: EM-algorithm

- IBM Models
  - Model 1 (Word-to-word translation)
  - Model 2 (+additional distortion model)
  - Model 3 (+fertility: insertions, deletions)
  - Model 4 (+improved distortion model)
  - Model 5 (+non-deficient Model 4)


MOTİVASYON

Motivasyon, dil öğrenimli hayat tarzını besleyen bir yakıttır.

- John Fotheringham

Sanıyorum ki hepimiz motivasyonun kendi hayatlarımızdaki önemini biliyoruz. Hepimizin, hat safhada bir motivasyon anını deneyimlediği olmuştur. Akıl almaz bir şekilde erken saatte uyanırız ve balığa çıkma yolculuğunun başlaması için kendimizi gözle görülür bir sıkıntıya sokarız; yılbaşı sabahı ailemizle birlikte olmak için bütün gece boyunca araba kullanırız; yeni bir iPad' a kavuşmak için günlük Starbucks dozumuza aylarca ara veririz. Motivasyon güçlü bir etkendir. Fakat, motive olamamanın da ne hissettirdiğini çok iyi biliyoruz; sabah yataktan dışarı çıkmayı istememek, o koşu için kapıdan dışarı bir adım bile atmayı istememek, iş ya da okul için o proje üzerinde çalışmayı istememek. Fakat motivasyon,

Motivation

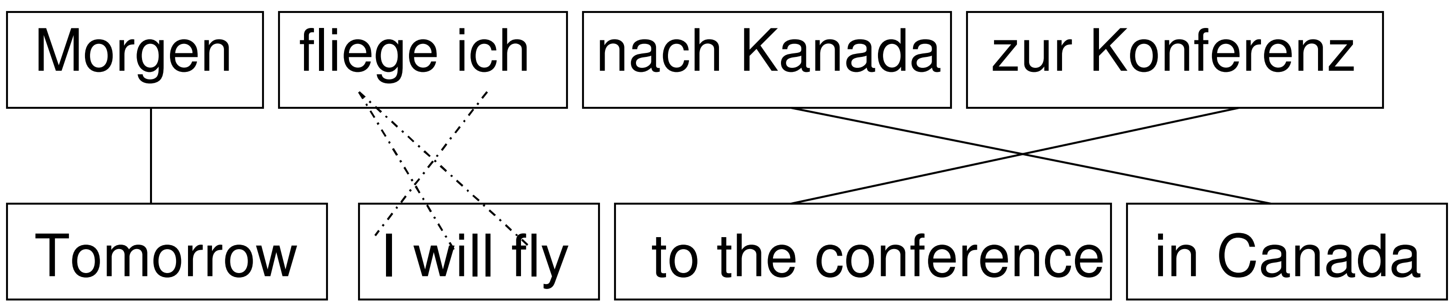
Motivation is the fuel that feeds the language learning lifestyle.

- John Fotheringham

I think we all understand the importance of motivation in our lives. We have all experienced times of great motivation. We wake up unfathomably early and put ourselves through significant discomfort to make the fishing trip happen; we drive through the night to be with family for Christmas morning; we forgo our daily dose of Starbucks for months in order to save for the new iPad. Motivation is powerful stuff. But we also know what it feels like to be unmotivated; to not want to get out of bed in the morning; to not want to step out the door for that run; to not want work on that project for work or school. But motivation is the fuel that feeds language learning and we must work to

Phrase-based Model (Och/Koehn et. al 2003)

- State-of-the-art for many language pairs



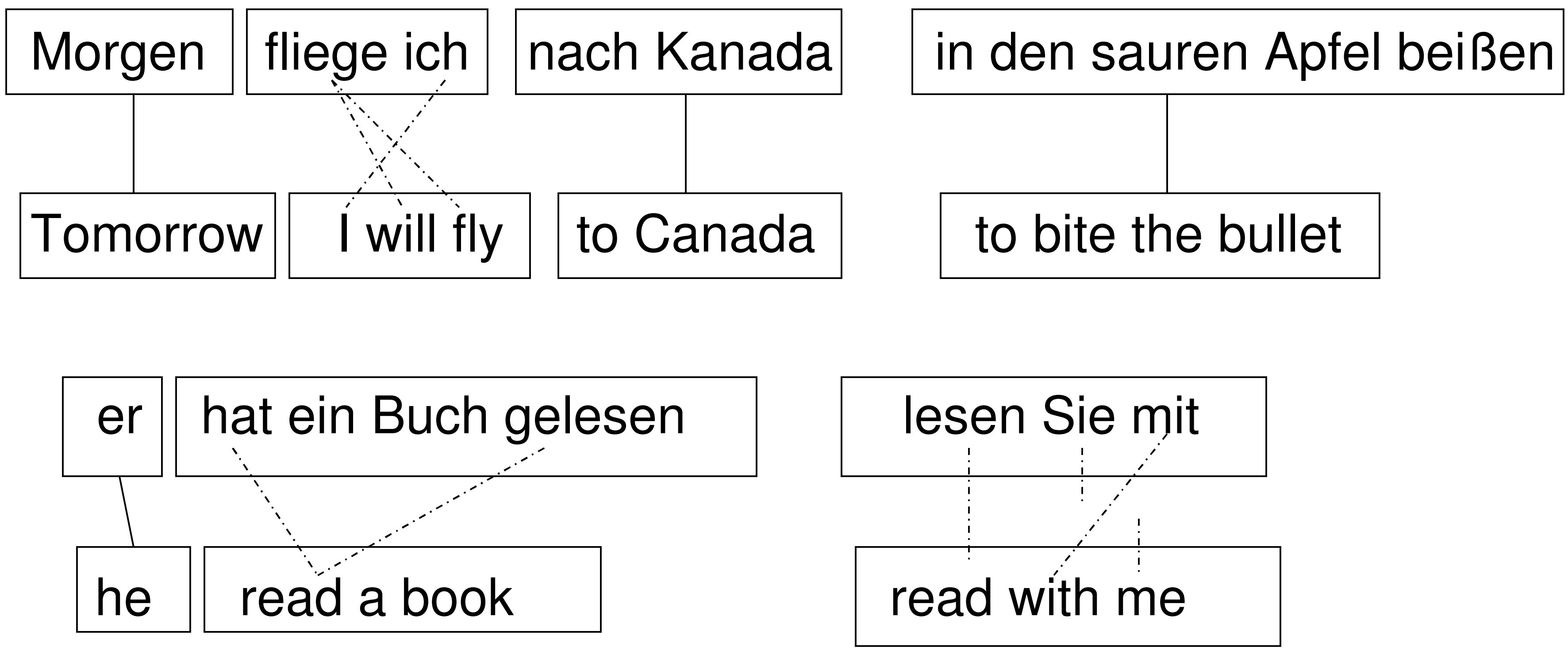
- Translation p(f|e) is estimated through phrases instead of words

From Koehn 2008

Benefits of phrase-based SMT

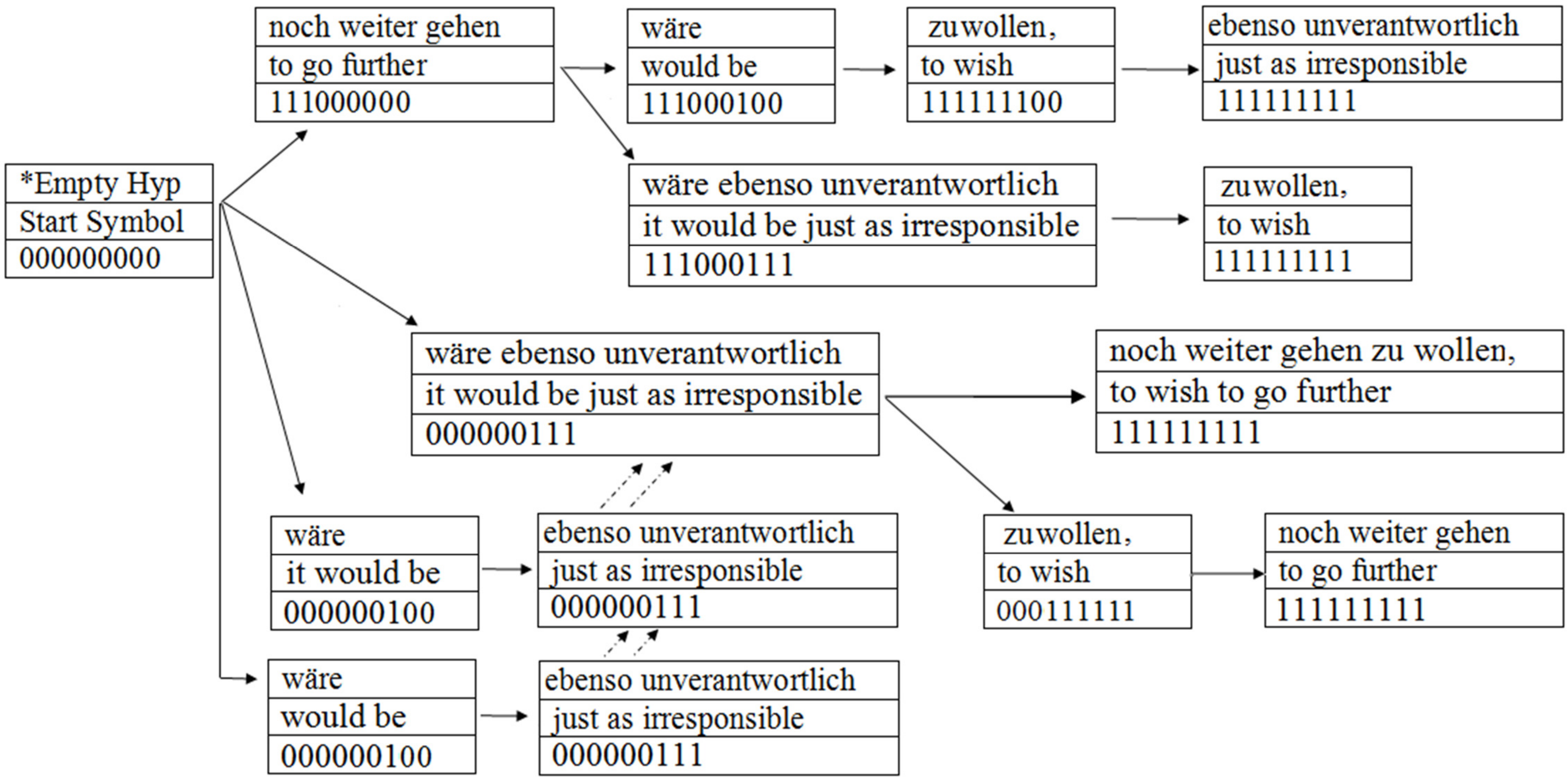
1. Local reordering

2. Idioms



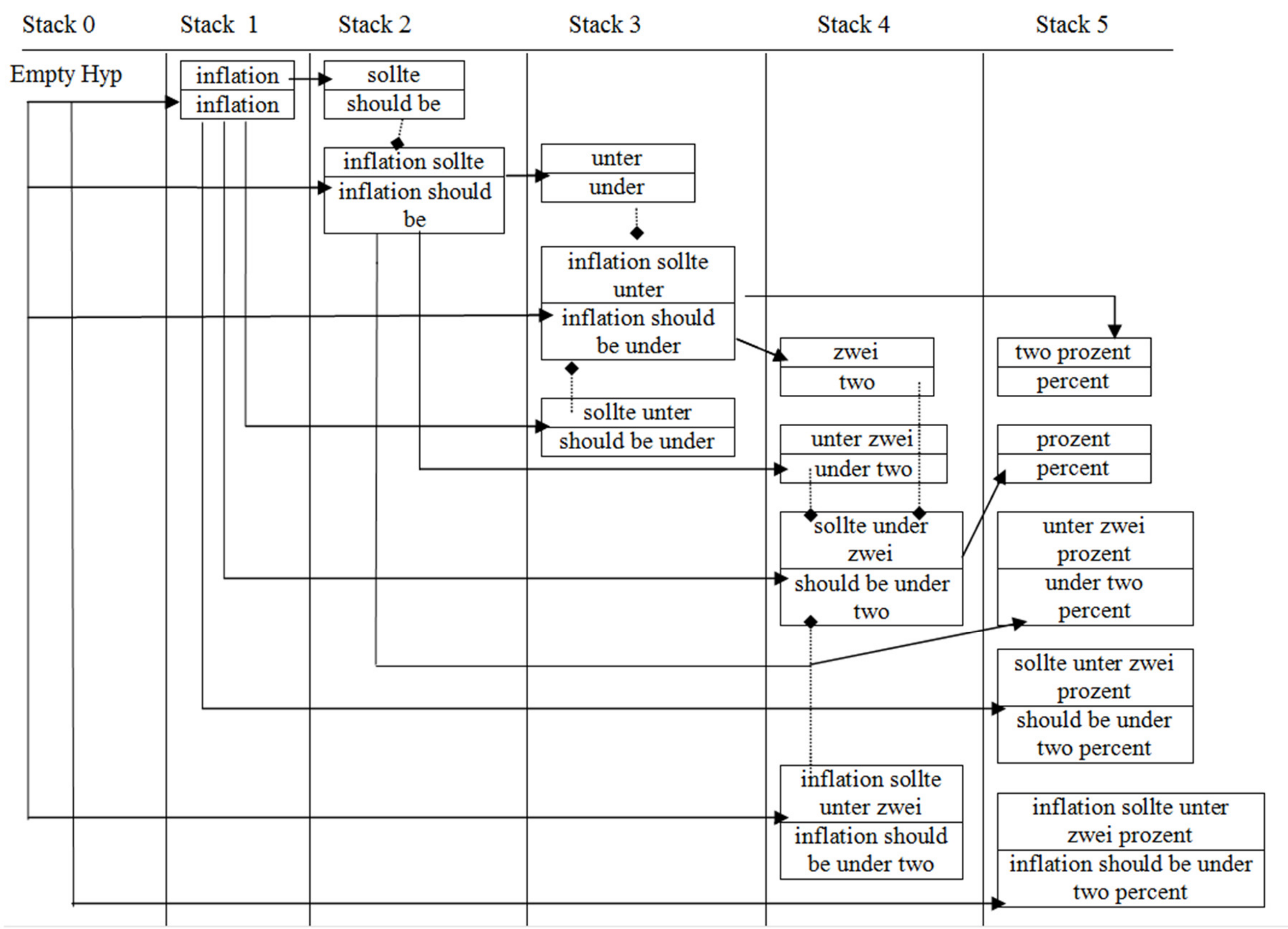
3. Discontinuities in phrases

4. Insertions and deletions



noch weiter gehen
to go further
111000000
wäre
would be
111000100
zuwollen,
to wish
111111100
ebenso unverantwortlich
just as irresponsible
111111111
*Empty Hyp
Start Symbol
000000000
wäre ebenso unverantwortlich
it would be just as irresponsible
111000111
zuwollen,
to wish
111111111
wäre ebenso unverantwortlich
it would be just as irresponsible
000000111
noch weiter gehen zu wollen,
to wish to go further
111111111
wäre
it would be
000000100
ebenso unverantwortlich
just as irresponsible
000000111
zuwollen,
to wish
000111111
noch weiter gehen
to go further
111111111
wäre
would be
000000100
ebenso unverantwortlich
just as irresponsible
000000111

Left-to-Right Stack Decoding



Phrasal Extraction

	er	hat	schon	sehr	viel	schoko	gegessen
he	x						
has		x					
eaten							x
so				x			
much					x		
chocolate						x	
already			x				

Reordering Sub-Model (Koehn et. al 2005)

	Morgan	fleige	ich	nach	Kanada	zur	Konferenz
Tomorrow	X	**M**					
I			X				
will							
fly		X					
to					**D**	X	
the							
conference					**S**		X
in				X			
Canada					X		

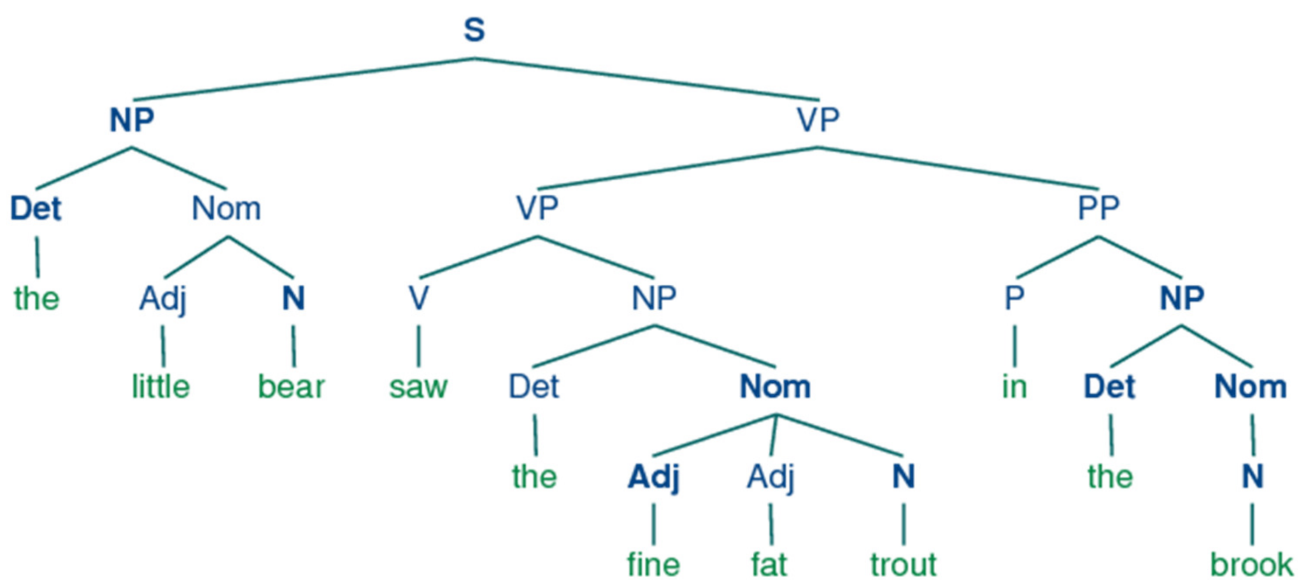
- Orientation-based model

  Monotonic (M), Swap (S), Discontinuous (D)

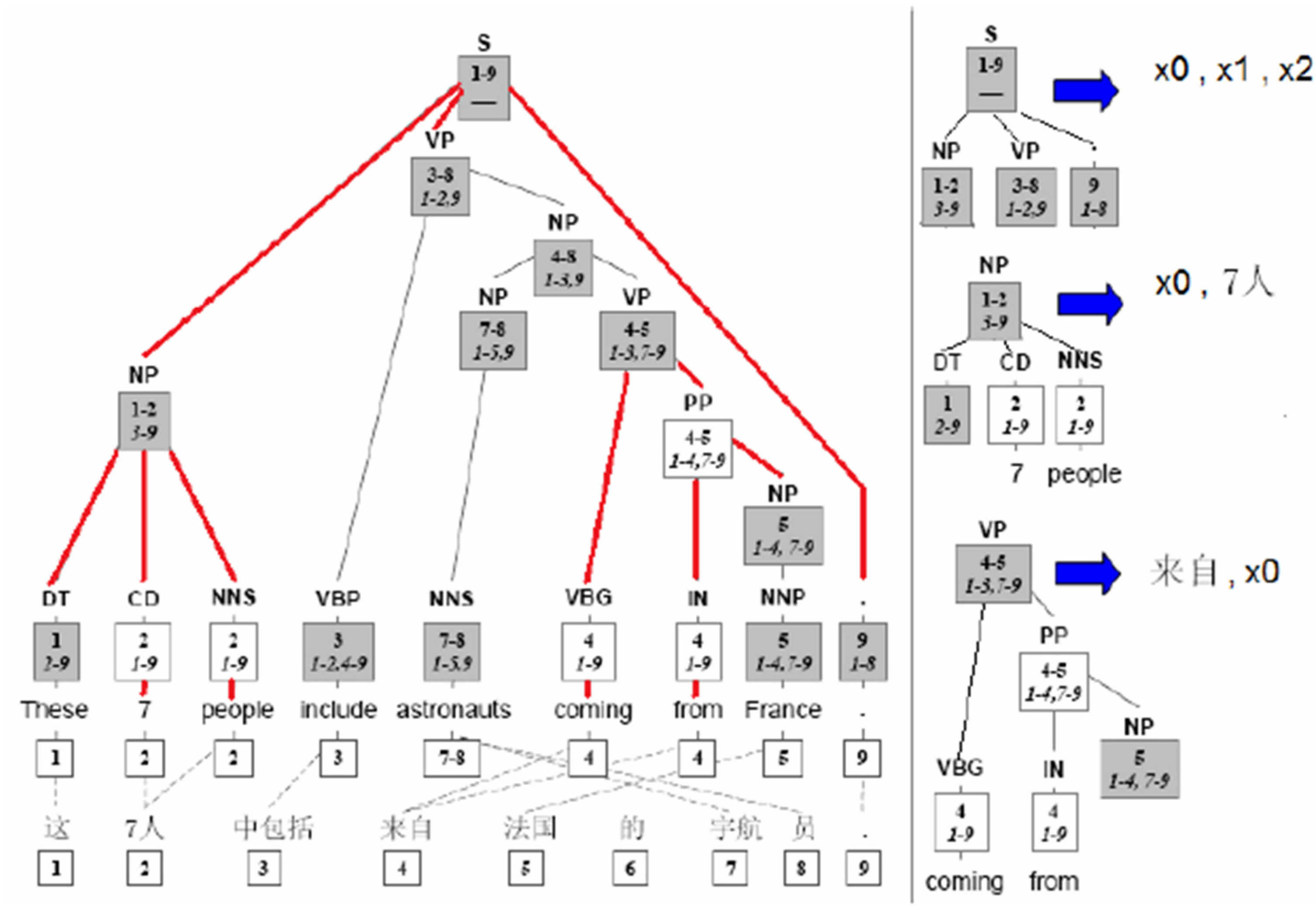
Syntax-based Models

- Phrase-based model can not capture long distance dependencies

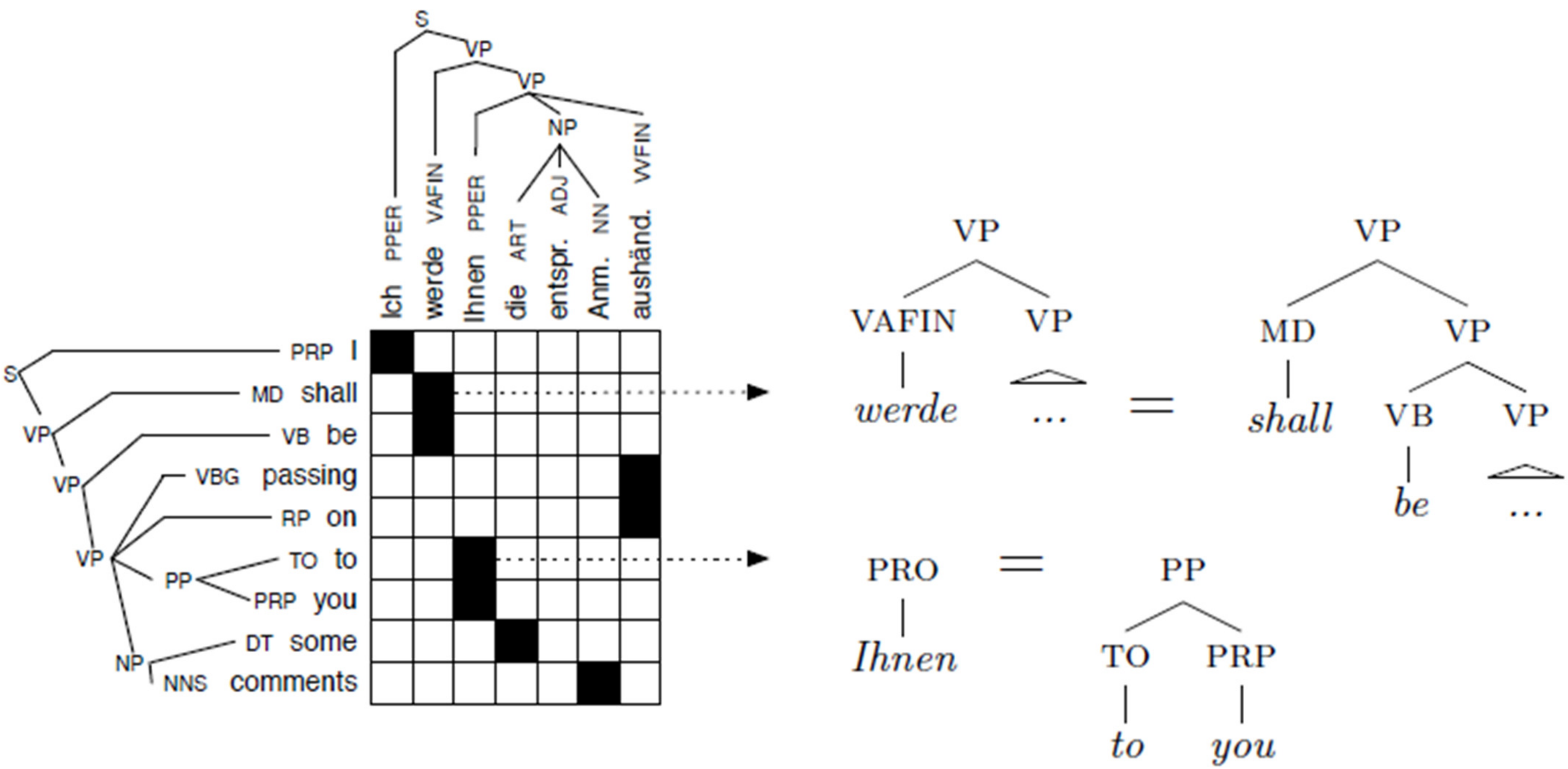
- Language is hierarchal and not flat



String-to-Tree Model (Galley et. al 2004, 2006)

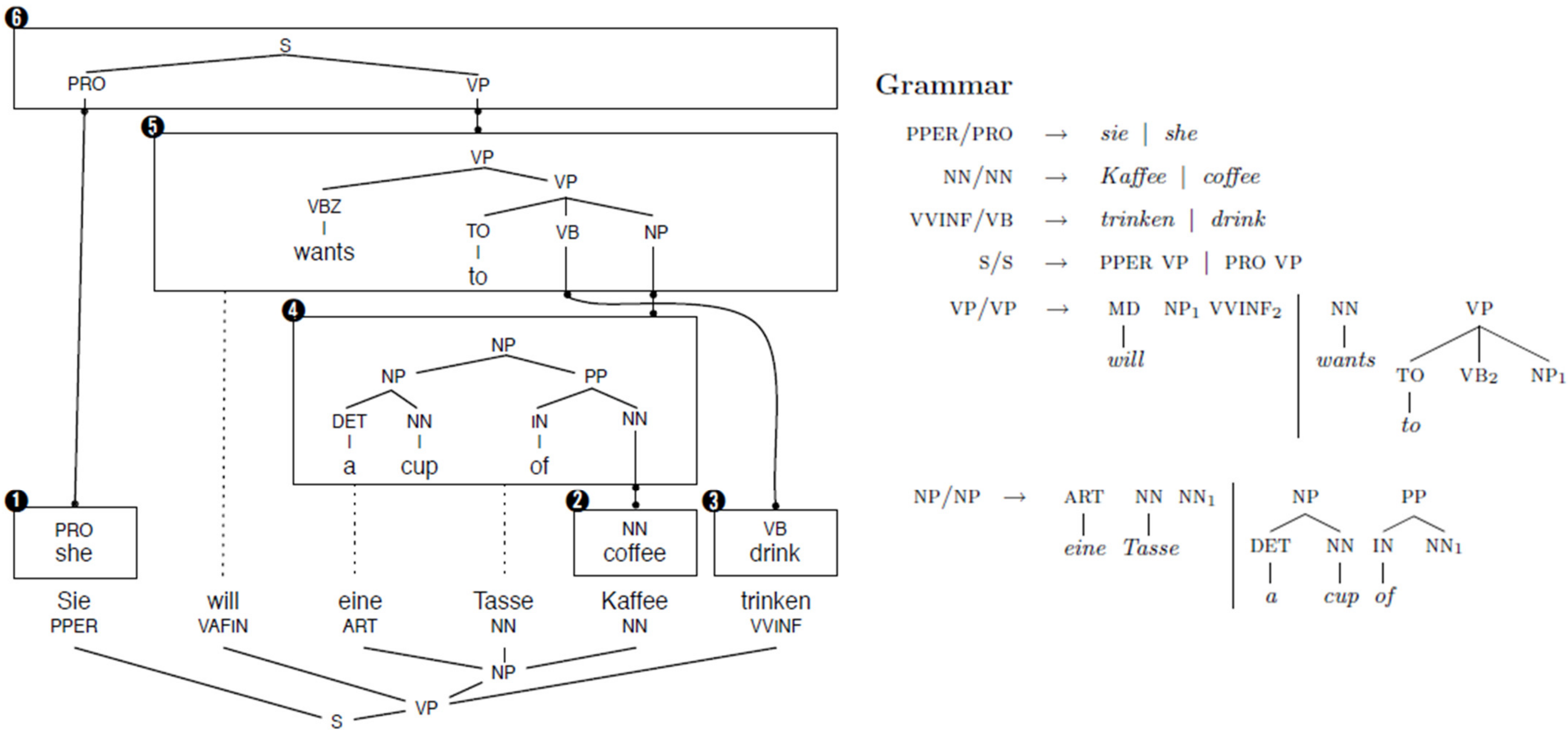


Tree-to-tree Model (Zhang et. al 2008)



From Koehn 2010. University of Edinburgh

Chart-based Decoding



Syntax-based Models

- Much progress, but success only for some language pairs

- Many open questions
  - Syntax on source/target/both?
  - Can we learn syntax unsupervised?
  - Phrase structure or dependency structure?
  - What grammar rules should be extracted?
  - Soft or hard constraints?
  - Feature design

Semantic-based Model

- What do existing models don't capture
  - Who did what to whom
  - Preservation of meaning can be more important than grammaticality/fluency
- ISI (Kevin Knight's Group)
  - Using semantic role labeling
  - Jones et. al (2012)

Log-linear Model (Och and Ney 2004)

- Typical features in Phrase-based Model
  - 4 Translation model features
  - 6 Reordering model features
  - Length Bonus
  - Phrase Bonus
  - Language Model

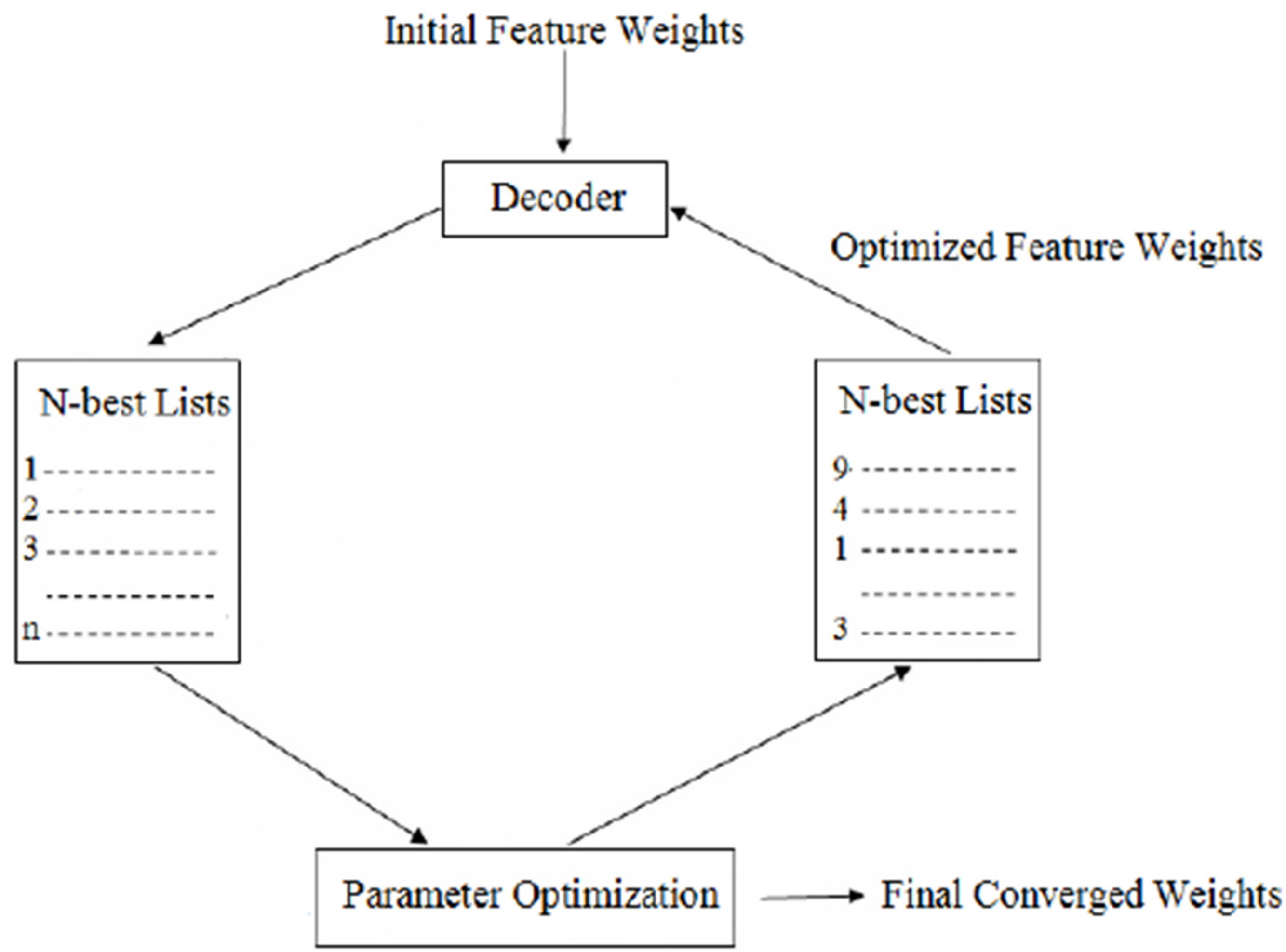
e_{best} = argmax p (E | F) = argmax p (F | E) x p(E)

$$\operatorname*{argmax}_e p(e|f) = \operatorname*{argmax}_e \left\{ \sum_{j=1}^{J} \lambda_j h_j(f, e) \right\}$$

- Tuning Algorithms
  - MERT (Och and Ney, 2004)
  - PRO (Hopkins and May, 2011)
  - MIRA (Chiang, 2012)

- 11,001 New Features for Statistical Machine Translation (Chiang et. al 2009)

Log-linear Model (Och and Ney 2004)



Open Problems in Machine Translation

- Evaluation
  - How good is a given machine translation system?
  - Hard problem, since many different translations acceptable
  - Evaluation metrics
    - Subjective judgments by human evaluators
    - Automatic evaluation metrics
- Automatic Evaluation Metrics
  - BLEU (Papineni et. al 2002)
  - METEOR (Banerjee and Lavie 2005)
  - WER/TER (Error rate)

Open Problems in Machine Translation

这个 机场 的 安全 工作 由 以色列 方面 负责 .

Israeli officials are responsible for airport security.
Israel is in charge of the security at this airport.
The security work for this airport is the responsibility of the Israel government.
Israeli side was in charge of the security of this airport.
Israel is responsible for the airport's security.
Israel is responsible for safety work at this airport.
Israel presides over the security of the airport.
Israel took charge of the airport security.
The safety of this airport is taken charge of by Israel.
This airport's security is the responsibility of the Israeli security officials.

(a typical example from the 2001 NIST evaluation set)

Open Problems in Machine Translation

- Human judgment
  - given: machine translation output
  - given: source and/or reference translation
  - task: asses the quality of machine translation output

- Metrics
  - Adequacy: Does the output convey the same meaning as the input sentence? Is part of the message lost, added, or distorted?
  - Fluency: Is the output good fluent English?

Open Problems in Machine Translation

- Domain Adaptation
  - Training data (News corpus, Europarl, Common Crawl Data)
  - Test data (Education domain, Medical domain)
  - Interpolation Models (Foster and Kuhn 2007)
  - MML Filter (Axelrod et. al 2011)
  - Domain Features (Hasler et. al 2012)

- OOV word translation
  - NE translation (Onaizan and Knight 2002)
  - NE disambiguation (Hermjakob et. al 2008)
  - Unsupervised Transliteration (Sajjad et. al 2012, Durrani et. al 2014)
    - Closely related languages (Durrani et. al 2011, Durrani and Koehn 2014)

Open Problems in Machine Translation

- Decoding Algorithms
  - Stack Decoding (Tillmann et. al 1997)
  - Efficient A* Decoding (Och et. al 2001)
  - Pruning Methods (Moore and Quirk 2007)

- Language Model
  - The house is big (good)
  - The house is xxl (worse)
  - House big is the (bad)
  - Markov-based language models with Kneser-Ney Smoothing
    - Considers history of 4 previous words
  - Syntax-based Language Models (Charniak et. al 2003)

Open Problems in Machine Translation

- Big Data and Scaling to Big Data
  - Parallel data (Billions of words) (Smith et. al 2013)
  - English monolingual data (trillions of words)
  - Randomized data structures (Talbot and Osborne 2007)
    - Developed at Edinburgh now used at Google
  - Distributed Systems
    - Distribute models over 100 machines
  - Efficient data-structures
    - Compact Phrase-tables (Junczys-Dowmunt 2012)
    - Scalable Language Model estimation (Heafield 2013)
      - Prefixes, back-off links in language models, binarization

Open Problems in Machine Translation

- Computer Assisted Translation
  - Machine Translation makes inroads in human translation industry
  - CASMACAT/MateCat Projects in Edinburgh

Why Do Machine Translation?

- Assimilation – reader initiates translation, wants to know the content (Gistable)
- Translation in Hand-held devices
- Post-editing (editable)
- User manuals in different languages, high quality translation (publishable)
- Integration with other NLP applications
  - Speech Technologies
  - Cross lingual information retrieval
- US Defense
  - Arabic-English post 9/11
  - Urdu-English, Pashto-English 2008
  - Dialectal Arabic (Egyptian, Labenese, Iraqi 2009-present)
  - Russian-English (2013-2014)

Open Source Resources

- Toolkits
  - Moses (Koehn et. al 2007), Phrasal (Cerr et. al 2010), NCode (Crego et. al 2011)
  - GIZA++ (Word Alignments)
  - SRILM, IRSTLM, KENLM, LMPLZ (Language Model)
- Data
  - French-English 39M
  - Chinese-English Spanish-English, Czech-English 15M
  - Arabic-English
  - German-English 5.5M
  - Urdu-English/Hindi-English ~300K
- Parsers
  - English, French, German

Thank you !!!

- Most of the slides are borrowed from Philipp Koehn

References

Michael Collins, Philipp Koehn, and Ivona Kucerova. 2005. Clause Restructuring for Statistical Machine Translation. In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05), pages 531–540, Ann Arbor, MI.

Philipp Koehn and Hieu Hoang. 2007. Factored Translation Models. In Proceedings of the 2007 Joint Conferenceon Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), pages 868–876, Prague, Czech Republic, June. Association for Computational Linguistics.

Alexander Fraser, Marion Weller, Aoife Cahill, and Fabienne Cap. 2012. Modeling Inflection and Word-Formation in SMT. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, pages 664–674, Avignon, France, April. Association for Computational Linguistics.

Galley, Michel, & Manning, Christopher D. (2008). A Simple and Effective Hierarchical Phrase Reordering Model. Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (pp. 848–856). Honolulu, Hawaii: Association for Computational Linguistics.

Green, Spence, Galley, Michel, and Manning, Christopher D. (2010). Improved Models of Distortion Cost for Statistical Machine Translation. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (pp. 867–875). Los Angeles, California: Association for Computational Linguistics.

Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and R. L. Mercer. 1993. The Mathematics of Statistical Machine Translation: Parameter Estimation. Computational Linguistics, 19(2):263–311.

Philipp Koehn, Franz J. Och, and Daniel Marcu. 2003. Statistical Phrase-Based Translation. In Proceedings of HLT-NAACL, pages 127–133, Edmonton, Canada.

Franz J. Och and Hermann Ney. 2004. The Alignment Template Approach to Statistical Machine Translation. Computational Linguistics, 30(1):417–449.

Franz J. Och. 2003. Minimum Error Rate Training in Statistical Machine Translation. In Proceedings of ACL, pages 160–167, Sapporo, Japan.

Colin Cherry and George Foster. 2012. Batch Tuning Strategies for Statistical Machine Translation. In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 427–436, Montréal, Canada, June. Association for Computational Linguistics.

Nadir Durrani, Helmut Schmid, and Alexander Fraser. 2011. A Joint Sequence Translation Model with Integrated Reordering. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pages 1045–1054, Portland, Oregon, USA, June.

Nadir Durrani, Alexander Fraser, Helmut Schmid, Hieu Hoang, and Philipp Koehn. 2013. Can Markov Models Over Minimal Translation Units Help Phrase-Based SMT? In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, August. Association for Computational Linguistics.

Nadir Durrani, Philipp Koehn, Helmut Schmid, and Alexander Fraser (2014). Investigating the Usefulness of Generalized Word Representations in SMT. In Proceedings of the 25th Annual Conference on Computational Linguistics (COLING), Dublin, Ireland.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL ’02, pages 311–318, Morristown, NJ, USA.

Satanjeev Banerjee and Alon Lavie. 2005. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In 43rd Annual Meeting of the Assoc. for Computational Linguistics: Proc.Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization, pages 65–72, Ann Arbor, MI, USA, June.

Michel Galley, Jonathan Graehl, Kevin Knight, Daniel Marcu, Steve DeNeefe, Wei Wang, and Ignacio Thayer. 2006. Scalable Inference and Training of Context-Rich Syntactic Translation Models. In Proceedings of COLING-ACL, pages 961–968, Sydney, Australia. Association for Computational Linguistics.

Min Zhang, Hongfei Jiang, Aiti Aw, Jun Sun, Sheng Li, and Chew Lim Tan. 2007. A tree-to-tree alignment-based model for statistical machine translation. In *Proceedings of MT-Summit*.

Chiang, D., Knight, K., and Wang, W. (2009). 11,001 new features for statistical machine translation. In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pages 218–226, Boulder, Colorado. Association for Computational Linguistics.

Bevan Jones, Jacob Andreas, Daniel Bauer, Karl Moritz Hermann, and Kevin Knight. 2012. Semantics-based machine translation with hyperedge replacement grammars. In Proc. COLING.

Foster, George and Roland Kuhn. 2007. Mixturemodel adaptation for SMT. In Proceedings of the Second Workshop on Statistical Machine Translation, pages 128–135, Prague, Czech Republic, June. Association for Computational Linguistics.

Axelrod, Amittai, Xiaodong He, and Jianfeng Gao. 2011. Domain adaptation via pseudo in-domain data selection. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pages 355–362, Edinburgh, Scotland, UK.

Franz J. Och and Hermann Ney. 2003. A Systematic Comparison of Various Statistical Alignment Models. Computational Linguistics, 29(1):19–51.

Eva Hasler, Barry Haddow, and Philipp Koehn. 2012. Sparse Lexicalised features and Topic Adaptation for SMT. In Proceedings of the seventh International Workshop on Spoken Language Translation (IWSLT), pages 268–275.

Al-Onaizan, Y. and Knight, K. (2002). Translating named entities using monolingual and bilingual resources. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics.

Nadir Durrani, Hassan Sajjad, Alexander Fraser and Helmut Schmid. (2010). Hindi-to-urdu machine translation through transliteration. In Proceedings of the 48th Annual Conference of the Association for Computational Linguistics, Uppsala, Sweden.

Nadir Durrani, Hassan Sajjad, Hieu Hoang, and Philipp Koehn. (2014). Integrating an Unsupervised Transliteration Model into Statistical Machine Translation. In Proceedings of the 15th Conference of the European Chapter of the ACL (EACL 2014), Gothenburg, Sweden. Association for Computational Linguistics.

Nadir Durrani and Philipp Koehn. (2014). Improving Machine Translation via Triangulation and Transliteration. In Proceedings of the 17th Annual Conference of the European Association for Machine Translation (EAMT), Dubrovnik, Croatia.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan,Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In ACL 2007 Demonstrations, Prague, Czech Republic.

Josep M. Crego, François Yvon, and José B. Mariño. 2011. Ncode: an Open Source Bilingual N-gram SMT Toolkit. The Prague Bulletin of Mathematical Linguistics, (96):49–58.

Daniel Cer, Michel Galley, Daniel Jurafsky, and Christopher D. Manning. 2010. Phrasal: A Statistical Machine Translation Toolkit for Exploring New model Features. In Proceedings of the NAACL HLT 2010 Demonstration Session, pages 9–12, Los Angeles, California, June.