Machine Translation

Alexander Fraser & Liane Guillou
[fraser,liane@cis.uni-muenchen.de]

CIS, Ludwig-Maximilians-Universität München

Computational Morphology and Electronic Dictionaries
SoSe 2016
2016-06-20

Machine Translation

Today I'll present some slides on four topics in machine translation:

- Machine translation (history and present)
- Transfer-based machine translation (Apertium)
- Basics of statistical machine translation
- Modeling morphology in statistical machine translation

Research on machine translation - past

(1970-present) Previous generation: So-called "Rule-based"

- Parse source sentence with rule-based parser
- Transfer source syntactic structure using hand-written rules to obtain target language representation
- Generate text from target language representation
- Scattered using of machine learning, particularly in parsing (recently in generation as well)

Research on machine translation - current generation

About 2000: Start of current generation: “Statistical Machine Translation”

- Relies only on corpus statistics, no linguistic structure (this will be explained further)
- First commercial product in 2004: Language Weaver Arabic/English (I was the PI of this)
- Google Translate and Bing, others

A brief history

- Machine translation was one of the first applications envisioned for computers
- Warren Weaver (1949): “I have a text in front of me which is written in Russian but I am going to pretend that it is really written in English and that it has been coded in some strange symbols. All I need to do is strip off the code in order to retrieve the information contained in the text.”
- First demonstrated by IBM in 1954 with a basic word-for-word translation system

Modified from Callison-Burch, Koehn

Interest in machine translation

- Commercial interest:
  - U.S. has invested in machine translation (MT) for intelligence purposes
  - MT is popular on the web—it is the most used of Google's special features
  - EU spends more than $1 billion on translation costs each year.
  - (Semi-)automated translation could lead to huge savings

Modified from Callison-Burch, Koehn

Interest in machine translation

- Academic interest:
  - One of the most challenging problems in NLP research
  - Requires knowledge from many NLP sub-areas, e.g., lexical semantics, syntactic parsing, morphological analysis, statistical modeling,…
  - Being able to establish links between two languages allows for transferring resources from one language to another

Modified from Dorr, Monz

Machine translation

- Goals of machine translation (MT) are varied, everything from *gisting* to rough draft
- Largest known application of MT: Microsoft knowledge base
  - Documents (web pages) that would not otherwise be translated at all

Language Weaver Arabic to English

Description of the Iraqi President George Bush American elections– which will follow in the current month of the thirty–that they constitute a historic moment, recognizing that the organization of elections in .current circumstances difficult issue

It was considered bush in the press that the pronouncements of the possible organization of elections in most regions of the Iraqi punctually wish that the turnout where high. He added that "Iraqi 14 ."appear in the relative calm 18ال governorates

v.2.0 – October 2003

A description of the American president George W. Bush elections– Iraq, which will take place on the thirtieth session of the month–– as a historic moment, acknowledging that the organization of elections in the current difficult circumstances.

Bush said in press statements that it is possible to organize elections in most regions of Iraq to the deadline and I wish that the turnout are high. He added that "14 governorates of Iraq's 18 appeared in relative calm".

v.2.4 – October 2004

Iraqi troops had become a target always Iraqi gunmen (French)

US President George W. Bush described Iraq elections-which will take place on the 30th of this month-- as a historic moment, acknowledging that the elections in the current situation is difficult.

Bush said in a press statement that it be possible to organize elections in most regions of Iraq in time and hoped that the rate of participation in the high. He added that "Iraqi 14 of the provinces of 18 appears to be relatively calm."

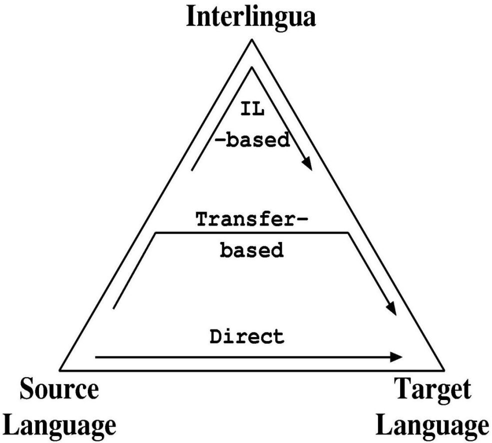
v.3.0 - February 2005

Document versus sentence

- MT problem: generate high quality translations of **documents**
- However, all current MT systems work only at **sentence** level!
- Translation of independent sentences is a difficult problem that is worth solving
- But remember that important discourse phenomena are ignored!
  - Example: How to translate English *it* to French (choice of feminine vs masculine *it*) or German (feminine/masculine/neuter *it*) if object referred to is in another sentence?

Machine Translation Approaches

- Grammar-based
  - Interlingua-based
  - Transfer-based
- Direct
  - Example-based
  - Statistical



Modified from Vogel

Apertium Shallow Transfer

Apertium is an open-source project to create "shallow transfer" systems for many language pairs. Four main components:

- Morphological analyzer for source language (with associated disambiguator)
- Dictionary for mapping words from source to target
- Transfer rules for:
  - Reordering words
  - Copying and modifying linguistic features (e.g., copying plural marker from English noun to German noun, copying gender from German noun to German article)
- Morphological generator for target language

Apertium Pros

- Rule-based MT is easy to understand (can trace through derivation if output is wrong)
- Executes quickly, based on finite-state-technology similar to two-level morphology
- Easy to add new vocabulary

Apertium Cons

- Slow and hard work to extend system
- Changing existing rules (often necessary) can have unpredictable effects, as rules are executed in sequence
- Difficult to model non-deterministic choices (for instance, word-sense disambiguation like “bank”); but these are very frequent
- In general: not robust to unexpected input

Status of Apertium EN-DE

EN-DE is in the Apertium "Nursery"

- Can get some basic sentences right currently
- But needs two-three more months before it works reasonably

EN-DE is a very difficult pair

- Apertium requires rules which have seen entire sequence of POS-tags
- But the German "mittelfeld" can have arbitrary sequences of POS-tags!

(If time: example)

Statistical versus Grammar-Based

- Often statistical and grammar-based MT are seen as alternatives, even opposing approaches – wrong !!!
- Dichotomies are:
  - Use probabilities – everything is equally likely (in between: heuristics)
  - Rich (deep) structure – no or only flat structure
- Both dimensions are continuous
- Examples
  - EBMT: flat structure and heuristics
  - SMT: flat structure and probabilities
  - XFER: deep(er) structure and heuristics

	No Probs	Probs
Flat Structure	EBMT	SMT
Deep Structure	XFER, Interlingua	Holy Grail

- Goal: structurally rich probabilistic models

Modified from Vogel

Statistical Approach

- Using statistical models
  - Create many alternatives, called hypotheses
  - Give a score to each hypothesis
  - Select the best -> search
- Advantages
  - Avoid hard decisions
  - Speed can be traded with quality, no all-or-nothing
  - Works better in the presence of unexpected input
- Disadvantages
  - Difficulties handling structurally rich models, mathematically and computationally
  - Need data to train the model parameters
  - Difficult to understand decision process made by system

Modified from Vogel

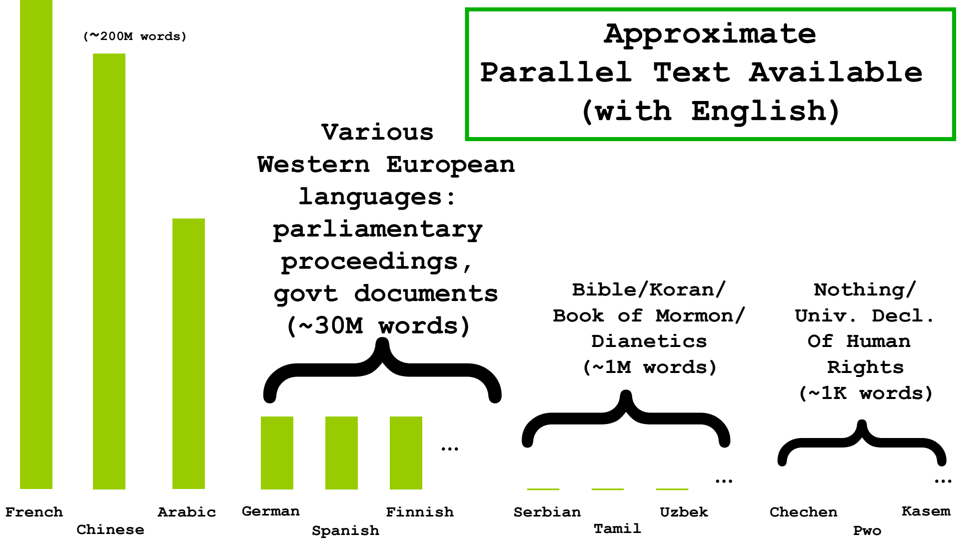
Parallel corpus

- Example from DE-News (8/1/1996)

English	German
Diverging opinions about planned tax reform	Unterschiedliche Meinungen zur geplanten Steuerreform
The discussion around the envisaged major tax reform continues .	Die Diskussion um die vorgesehene grosse Steuerreform dauert an .
The FDP economics expert , Graf Lambsdorff , today came out in favor of advancing the enactment of significant parts of the overhaul , currently planned for 1999 .	Der FDP - Wirtschaftsexperte Graf Lambsdorff sprach sich heute dafuer aus , wesentliche Teile der fuer 1999 geplanten Reform vorzuziehen .

Modified from Dorr, Monz

Most statistical machine translation research has focused on a few high-resource languages (European, Chinese, Japanese, Arabic).



How to Build an SMT System

- Start with a large parallel corpus
  - Consists of document pairs (document and its translation)
- Sentence alignment: in each document pair automatically find those sentences which are translations of one another
  - Results in sentence pairs (sentence and its translation)
- Word alignment: in each sentence pair automatically annotate those words which are translations of one another
  - Results in word-aligned sentence pairs
- Automatically estimate a statistical model from the word-aligned sentence pairs
  - Results in model parameters
- Given new text to translate, apply model to get most probable translation

Research on machine translation - present situation

The situation now (for this talk)

Research on machine translation - present situation

The situation now (for this talk)

- Realization that high quality machine translation requires modeling **linguistic structure** is widely held (even at Google)

Research on machine translation - present situation

The situation now (for this talk)

- Realization that high quality machine translation requires modeling **linguistic structure** is widely held (even at Google)
- Significant research efforts in Europe, USA, Asia on incorporating **syntax** in inference, but good results still limited to a handful of language pairs

Research on machine translation - present situation

The situation now (for this talk)

- Realization that high quality machine translation requires modeling **linguistic structure** is widely held (even at Google)
- Significant research efforts in Europe, USA, Asia on incorporating **syntax** in inference, but good results still limited to a handful of language pairs
- Only scattered work on incorporating **morphology** in inference

Research on machine translation - present situation

The situation now (for this talk)

- Realization that high quality machine translation requires modeling **linguistic structure** is widely held (even at Google)
- Significant research efforts in Europe, USA, Asia on incorporating **syntax** in inference, but good results still limited to a handful of language pairs
- Only scattered work on incorporating **morphology** in inference
- Difficulties to beat previous generation rule-based systems, particularly when the target language is **morphologically rich** and **less configurational** than English, e.g., German, Arabic, Czech and many more languages

Research on machine translation - present situation

The situation now (for this talk)

- Realization that high quality machine translation requires modeling **linguistic structure** is widely held (even at Google)
- Significant research efforts in Europe, USA, Asia on incorporating **syntax** in inference, but good results still limited to a handful of language pairs
- Only scattered work on incorporating **morphology** in inference
- Difficulties to beat previous generation rule-based systems, particularly when the target language is **morphologically rich** and **less configurational** than English, e.g., German, Arabic, Czech and many more languages
- First attempts to integrate **semantics**

Research on machine translation - present situation

The situation now (for this talk)

- Realization that high quality machine translation requires modeling **linguistic structure** is widely held (even at Google)
- Significant research efforts in Europe, USA, Asia on incorporating **syntax** in inference, but good results still limited to a handful of language pairs
- Only scattered work on incorporating **morphology** in inference
- Difficulties to beat previous generation rule-based systems, particularly when the target language is **morphologically rich** and **less configurational** than English, e.g., German, Arabic, Czech and many more languages
- First attempts to integrate **semantics**
- This progression (mostly) parallels the development of rule-based MT, with the noticeable exception of morphology

Research on statistical machine translation

Statistical machine translation (SMT) uses automatically learned rules.

SMT is following same progression as previous generation rule-based:

Research on statistical machine translation

Statistical machine translation (SMT) uses automatically learned rules.
SMT is following same progression as previous generation rule-based:

- Word-based

Research on statistical machine translation

Statistical machine translation (SMT) uses automatically learned rules.
SMT is following same progression as previous generation rule-based:

- Word-based
- Phrase-based (unique to SMT)

Research on statistical machine translation

Statistical machine translation (SMT) uses automatically learned rules.
SMT is following same progression as previous generation rule-based:

- Word-based
- Phrase-based (unique to SMT)
- Morphology

Research on statistical machine translation

Statistical machine translation (SMT) uses automatically learned rules.
SMT is following same progression as previous generation rule-based:

- Word-based
- Phrase-based (unique to SMT)
- Morphology
- Syntax

Research on statistical machine translation

Statistical machine translation (SMT) uses automatically learned rules.
SMT is following same progression as previous generation rule-based:

- Word-based
- Phrase-based (unique to SMT)
- Morphology
- Syntax
- Semantics

Research on statistical machine translation

Statistical machine translation (SMT) uses automatically learned rules.
SMT is following same progression as previous generation rule-based:

- Word-based
- Phrase-based (unique to SMT)
- Morphology
- Syntax
- Semantics
- Text Structure/Pragmatics

Research on statistical machine translation

Statistical machine translation (SMT) uses automatically learned rules.
SMT is following same progression as previous generation rule-based:

- Word-based
- Phrase-based (unique to SMT)
- Morphology
- Syntax
- Semantics
- Text Structure/Pragmatics
- World Knowledge

Research on statistical machine translation

Statistical machine translation (SMT) uses automatically learned rules.
SMT is following same progression as previous generation rule-based:

- Word-based
- Phrase-based (unique to SMT)
- Morphology
- Syntax
- Semantics
- Text Structure/Pragmatics
- World Knowledge

- We need all of these levels of representation to reach perfect machine translation!
- This talk will focus on integrating morphological and syntactic modeling into SMT
- We have also started integrating semantics, have some ideas about text structure/pragmatics

Outline

- History
- Basic statistical approach
- Word alignment (morphologically rich)
- Translating from morphologically rich to less rich
- Improved translation to morphologically rich languages
  - Translating English clause structure to German
  - Morphological generation
  - Adding lexical semantic knowledge to morphological generation
- Bigger picture: questions about adding more linguistic structure, dealing with ambiguity/underspecification

(Statistical) machine translation is a structured prediction problem

Structured prediction problems in computational linguistics are defined like this:

- Problem definition
- Evaluation

- Linguistic representation
- Model
- Training
- Search

(Statistical) machine translation is a structured prediction problem

Structured prediction problems in computational linguistics are defined like this:

- Problem definition: translate sentences (independently!)
- Evaluation

- Linguistic representation
- Model
- Training
- Search

(Statistical) machine translation is a structured prediction problem

Structured prediction problems in computational linguistics are defined like this:

- Problem definition: translate sentences (independently!)
- Evaluation: use human judgements of quality (and correlated automatic metrics)
- Linguistic representation
- Model
- Training
- Search

(Statistical) machine translation is a structured prediction problem

Structured prediction problems in computational linguistics are defined like this:

- Problem definition: translate sentences (independently!)
- Evaluation: use human judgements of quality (and correlated automatic metrics)
- Linguistic representation: subject of the rest of this talk
- Model
- Training
- Search

(Statistical) machine translation is a structured prediction problem

Structured prediction problems in computational linguistics are defined like this:

- Problem definition: translate sentences (independently!)
- Evaluation: use human judgements of quality (and correlated automatic metrics)
- Linguistic representation: subject of the rest of this talk
- Model: see next few slides
- Training
- Search

(Statistical) machine translation is a structured prediction problem

Structured prediction problems in computational linguistics are defined like this:

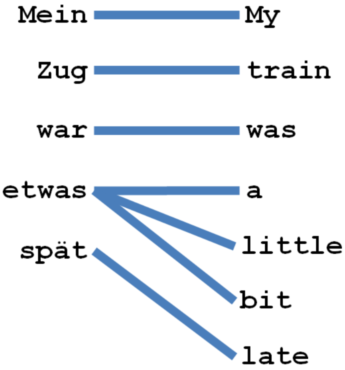
- Problem definition: translate sentences (independently!)
- Evaluation: use human judgements of quality (and correlated automatic metrics)
- Linguistic representation: subject of the rest of this talk
- Model: see next few slides
- Training: see next few slides
- Search

(Statistical) machine translation is a structured prediction problem

Structured prediction problems in computational linguistics are defined like this:

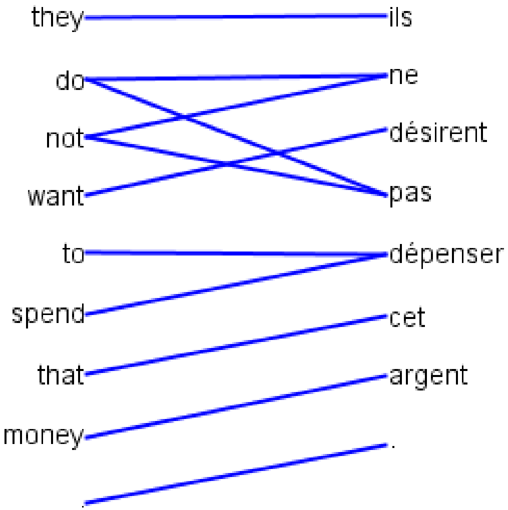
- Problem definition: translate sentences (independently!)
- Evaluation: use human judgements of quality (and correlated automatic metrics)
- Linguistic representation: subject of the rest of this talk
- Model: see next few slides
- Training: see next few slides
- Search: beyond the scope of this talk (think of beam search and CYK+)

Basic non-linguistic representation - word alignment

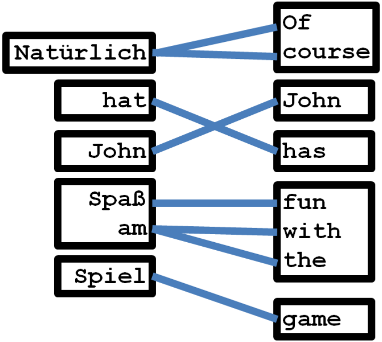


Word alignment: bigraph, connected components show "minimal translation units"

Introduction to SMT - Word Alignment



Phrase-based SMT (Koehn's example) - German to English



Phrase pairs are either minimal translation units or contiguous groups of them (e.g., spass -> fun, am -> with the). Often not linguistic phrases!

- German word sequence is segmented into German phrases seen in the word aligned training data
- German phrases are used to produce English phrases
- English phrases are reordered

Phrase-based SMT: training

Start with a large collection of parallel documents (for instance: Proceedings of the European Parliament)

Phrase-based SMT: training

Start with a large collection of parallel documents (for instance: Proceedings of the European Parliament)

Identify parallel sentences

Phrase-based SMT: training

Start with a large collection of parallel documents (for instance: Proceedings of the European Parliament)

Identify parallel sentences (Braune and Fraser COLING 2010), others

Phrase-based SMT: training

Start with a large collection of parallel documents (for instance: Proceedings of the European Parliament)

Identify parallel sentences (Braune and Fraser COLING 2010), others

Compute word alignments:

Phrase-based SMT: training

Start with a large collection of parallel documents (for instance: Proceedings of the European Parliament)

Identify parallel sentences (Braune and Fraser COLING 2010), others

Compute word alignments:

using an unsupervised generative model

Phrase-based SMT: training

Start with a large collection of parallel documents (for instance: Proceedings of the European Parliament)

Identify parallel sentences (Braune and Fraser COLING 2010), others

Compute word alignments:

- using an unsupervised generative model (Brown et al 1993, Fraser and Marcu EMNLP 2007)

Phrase-based SMT: training

Start with a large collection of parallel documents (for instance: Proceedings of the European Parliament)

Identify parallel sentences (Braune and Fraser COLING 2010), others

Compute word alignments:

- using an unsupervised generative model (Brown et al 1993, Fraser and Marcu EMNLP 2007)
- using a supervised discriminative model

Phrase-based SMT: training

Start with a large collection of parallel documents (for instance: Proceedings of the European Parliament)

Identify parallel sentences	(Braune and Fraser COLING 2010), others
Compute word alignments:	
using an unsupervised generative model	(Brown et al 1993, Fraser and Marcu EMNLP 2007)
using a supervised discriminative model	many models including (Fraser and Marcu ACL-WMT 2005), (Fraser and Marcu ACL 2006)

Phrase-based SMT: training

Start with a large collection of parallel documents (for instance: Proceedings of the European Parliament)

Identify parallel sentences (Braune and Fraser COLING 2010), others

Compute word alignments:

- using an unsupervised generative model (Brown et al 1993, Fraser and Marcu EMNLP 2007)
- using a supervised discriminative model many models including (Fraser and Marcu ACL-WMT 2005), (Fraser and Marcu ACL 2006)
- or a semi-supervised model combining these

Phrase-based SMT: training

Start with a large collection of parallel documents (for instance: Proceedings of the European Parliament)	
Identify parallel sentences	(Braune and Fraser COLING 2010), others
Compute word alignments:	
using an unsupervised generative model	(Brown et al 1993, Fraser and Marcu EMNLP 2007)
using a supervised discriminative model	many models including (Fraser and Marcu ACL-WMT 2005), (Fraser and Marcu ACL 2006)
or a semi-supervised model combining these	(Fraser and Marcu ACL 2006)

Phrase-based SMT: training

Given a word aligned parallel corpus, learn to translate unseen sentences (supervised structured learning)

- Learn a phrase lexical translation sub-model and a phrase reordering sub-model from the word alignment (Och and Ney 2004; Koehn, Och, Marcu 2003)
- Combine these with other knowledge sources to learn a full model of translation (Och and Ney 2004)
- Most important other knowledge source: monolingual n-gram language model in the target language
  - Models "fluency", good target language sentences
- IMPORTANT: no explicit linguistic knowledge (syntactic parses, morphology, etc)!

Translating to Morphologically Rich(er) Languages with SMT

- Most research on statistical machine translation (SMT) is on translating into English, which is a **morphologically-not-at-all-rich** language, with significant interest in **morphological reduction**
- Recent interest in the other direction - requires **morphological generation**
- We will start with a very brief review of MT and SMT

Challenges

- The challenges I am currently focusing on:
  - How to generate morphology (for German or French) which is more specified than in the source language (English)?
  - How to translate from a configurational language (English) to a less-configurational language (German)?
  - Which linguistic representation should we use and where should specification happen?

configurational roughly means "fixed word order" here

Our work

- Several projects funded by the EU, including an ERC Starting Grant, and by the DFG (German Research Foundation)
- Basic research question: can we integrate linguistic resources for morphology and syntax into (large scale) statistical machine translation?
- Will talk about German/English word alignment and translation from German to English briefly
- Primary focus: translation from English (and French) to German
- Secondary: translation to French, others (recently: Russian, not ready yet)

Lessons: word alignment

- My thesis was on word alignment...
- Our work in the project shows that word alignment involving morphologically rich languages is a task where:
  - One should throw away inflectional marking (Fraser ACL-WMT 2009)
  - One should deal with compounding by aligning split compounds (Fritzinger and Fraser ACL-WMT 2010)
  - Syntactic information doesn't seem to help much (at least for training phrase-based SMT models)

Lessons: translating from German to English

First, let's look at the morphologically rich to morphologically poor direction...

1. Parse the German, and deterministically reorder it to look like English "ich habe gegessen einen Erdbeerkuchen" (Collins, Koehn, Kucerova 2005; Fraser ACL-WMT 2009)
    - German main clause order: I have a strawberry cake **eaten**
    - Reordered (English order): I have **eaten** a strawberry cake

Lessons: translating from German to English

First, let's look at the morphologically rich to morphologically poor direction...

1. Parse the German, and deterministically reorder it to look like English "ich habe gegessen einen Erdbeerkuchen" (Collins, Koehn, Kucerova 2005; Fraser ACL-WMT 2009)
   - German main clause order: I have a strawberry cake **eaten**
   - Reordered (English order): I have **eaten** a strawberry cake
2. Split German compounds ("Erdbeerkuchen" -> "Erdbeer Kuchen") (Koehn and Knight 2003; Fritzinger and Fraser ACL-WMT 2010)

Lessons: translating from German to English

First, let's look at the morphologically rich to morphologically poor direction...

1. Parse the German, and deterministically reorder it to look like English "ich habe gegessen einen Erdbeerkuchen" (Collins, Koehn, Kucerova 2005; Fraser ACL-WMT 2009)
   - German main clause order: I have a strawberry cake **eaten**
   - Reordered (English order): I have **eaten** a strawberry cake
2. Split German compounds ("Erdbeerkuchen" -> "Erdbeer Kuchen") (Koehn and Knight 2003; Fritzinger and Fraser ACL-WMT 2010)
3. Keep German noun inflection marking plural/singular, throw away most other inflection ("einen" -> "ein") (Fraser ACL-WMT 2009)

Lessons: translating from German to English

First, let's look at the morphologically rich to morphologically poor direction...

1. Parse the German, and deterministically reorder it to look like English "ich habe gegessen einen Erdbeerkuchen" (Collins, Koehn, Kucerova 2005; Fraser ACL-WMT 2009)
    - German main clause order: I have a strawberry cake **eaten**
    - Reordered (English order): I have **eaten** a strawberry cake
2. Split German compounds ("Erdbeerkuchen" -> "Erdbeer Kuchen") (Koehn and Knight 2003; Fritzinger and Fraser ACL-WMT 2010)
3. Keep German noun inflection marking plural/singular, throw away most other inflection ("einen" -> "ein") (Fraser ACL-WMT 2009)
4. Apply standard phrase-based techniques to this representation

Lessons: translating from German to English

- I described how to integrate syntax and morphology deterministically for this task
- We don't see the need for modeling morphology in the translation model for German to English: simply preprocess
- But for getting the target language word order right, we should be using reordering models, not deterministic rules
  - This allows us to use target language context (modeled by the language model)
  - Critical to obtaining well-formed target language sentences

Translating from English to German

English to German is a challenging problem - previous generation rule-based systems still superior, but gap rapidly narrowing as we generalize better and better

Translating from English to German

English to German is a challenging problem - previous generation rule-based systems still superior, but gap rapidly narrowing as we generalize better and better

- Use classifiers to classify English clauses with their German word order

Translating from English to German

English to German is a challenging problem - previous generation rule-based systems still superior, but gap rapidly narrowing as we generalize better and better

- Use classifiers to classify English clauses with their German word order
- Predict German verbal features like person, number, tense, aspect

Translating from English to German

English to German is a challenging problem - previous generation rule-based systems still superior, but gap rapidly narrowing as we generalize better and better

- Use classifiers to classify English clauses with their German word order
- Predict German verbal features like person, number, tense, aspect
- Use phrase-based model to translate English words to German lemmas (with split compounds)

Translating from English to German

English to German is a challenging problem - previous generation rule-based systems still superior, but gap rapidly narrowing as we generalize better and better

- Use classifiers to classify English clauses with their German word order
- Predict German verbal features like person, number, tense, aspect
- Use phrase-based model to translate English words to German lemmas (with split compounds)
- Create compounds by merging adjacent lemmas
  - Use a sequence classifier to decide where and how to merge lemmas to create compounds

Translating from English to German

English to German is a challenging problem - previous generation rule-based systems still superior, but gap rapidly narrowing as we generalize better and better

- Use classifiers to classify English clauses with their German word order
- Predict German verbal features like person, number, tense, aspect
- Use phrase-based model to translate English words to German lemmas (with split compounds)
- Create compounds by merging adjacent lemmas
  - Use a sequence classifier to decide where and how to merge lemmas to create compounds
- Determine how to inflect German noun phrases (and prepositional phrases)
  - Use a sequence classifier to predict nominal features

Reordering for English to German translation

(SL) [Yesterday I **read** a book][which I **bought** last week]

⇩

(SL reordered) [Yesterday **read** I a book][which I last week **bought**]

(TL) [Gestern **las** ich ein Buch][das ich letzte Woche **kaufte**]

Reordering for English to German translation

(SL) [Yesterday I **read** a book][which I **bought** last week]

⇩

(SL reordered) [Yesterday **read** I a book][which I last week **bought**]

(TL) [Gestern **las** ich ein Buch][das ich letzte Woche **kaufte**]

- Use deterministic reordering rules to reorder English parse tree to German word order, see (Gojun, Fraser EACL 2012)

Reordering for English to German translation

(SL) [Yesterday I **read** a book][which I **bought** last week]

(SL reordered) [Yesterday **read** I a book][which I last week **bought**]

(TL) [Gestern **las** ich ein Buch][das ich letzte Woche **kaufte**]

- Use deterministic reordering rules to reorder English parse tree to German word order, see (Gojun, Fraser EACL 2012)
- Translate this representation to obtain German lemmas

Reordering for English to German translation

(SL) [Yesterday I **read** a book][which I **bought** last week]

(SL reordered) [Yesterday **read** I a book][which I last week **bought**]

(TL) [Gestern **las** ich ein Buch][das ich letzte Woche **kaufte**]

- Use deterministic reordering rules to reorder English parse tree to German word order, see (Gojun, Fraser EACL 2012)
- Translate this representation to obtain German lemmas
- Finally, predict German verb features. Both verbs in example: <first person, singular, past, indicative>

Reordering for English to German translation

(SL) [Yesterday I **read** a book][which I **bought** last week]

⇩

(SL reordered) [Yesterday **read** I a book][which I last week **bought**]

(TL) [Gestern **las** ich ein Buch][das ich letzte Woche **kaufte**]

- Use deterministic reordering rules to reorder English parse tree to German word order, see (Gojun, Fraser EACL 2012)
- Translate this representation to obtain German lemmas
- Finally, predict German verb features. Both verbs in example: <first person, singular, past, indicative>
- New work on this uses lattices to represent alternative clausal orderings (e.g., “las”, “habe ... gelesen”)

Word formation: dealing with compounds

- German compounds are highly productive and lead to data sparsity. We split them in the training data using corpus/linguistic knowledge techniques (Fritzinger and Fraser ACL-WMT 2010)
- At test time, we translate English test sentence to the German split lemma representation

  split `Inflation<+NN><Fem><Sg> Rate<+NN><Fem><Sg>`
- Determine whether to merge adjacent words to create a compound (Stymne & Cancedda 2011)
  - Classifier is a linear-chain CRF using German lemmas (in split representation) as input

  compound `Inflationsrate<+NN><Fem><Sg>`
- Initial implementation documented in (Fraser, Weller, Cahill, Cap EACL 2012)
- New approach additionally using machine learning features on the syntax of the aligned English (Cap, Fraser, Weller, Cahill EACL 2014)

Predicting nominal inflection

Idea: separate the translation into two steps:

(1) Build a translation system with non-inflected forms (lemmas)

(2) Inflect the output of the translation system

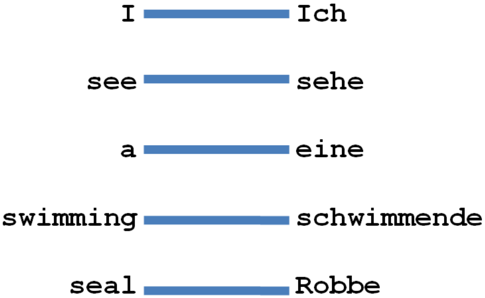
   a) predict inflection features using a sequence classifier
   b) generate inflected forms based on predicted features and lemmas

Example: baseline vs. two-step system

- A standard system using inflected forms needs to decide on one of the possible inflected forms:
  `blue` → `blau`, `blaue`, `blauer`, `blaues`, `blauen`, `blauem`
- A translation system built on lemmas, followed by inflection prediction and inflection generation:

  (1) `blue` → `blau<ADJECTIVE>`
  (2) `blau<ADJECTIVE><nominative><feminine><singular>`
  `<weak-inflection>` → `blaue`

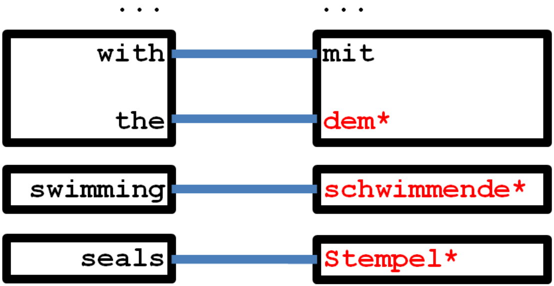
Inflection - example



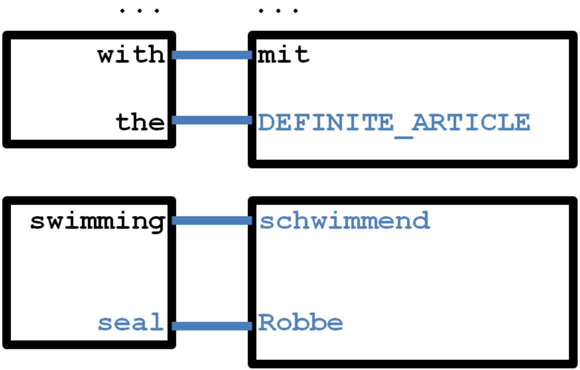
Suppose the training data is typical European Parliament material and this sentence pair is also in the training data.

We would like to translate: "... with the swimming seals"

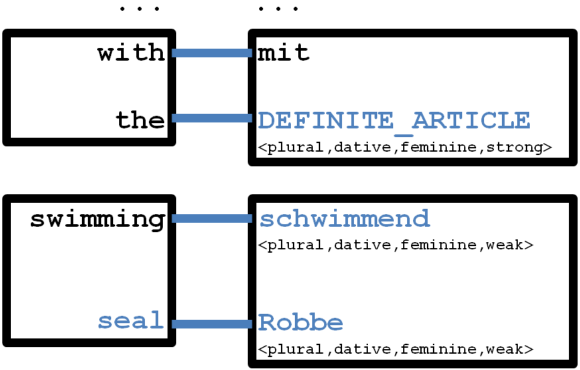
Inflection - problem in baseline



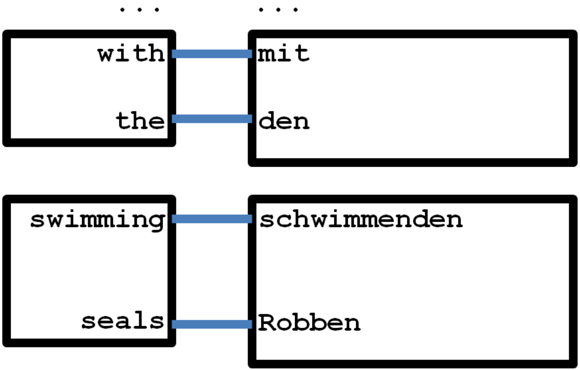
Dealing with inflection - translation to underspecified representation



Dealing with inflection - nominal inflection features prediction



Dealing with inflection - surface form generation



Sequence classification

- Initially implemented using simple language models (input = underspecified, output = fully specified)
- Linear-chain CRFs work much better
- We use the Wapiti Toolkit (Lavergne et al., 2010)
- We use a huge feature space
  - 6-grams on German lemmas
  - 8-grams on German POS-tag sequences
  - various other features including features on aligned English
  - L1 regularization is used to obtain a sparse model
- See (Fraser, Weller, Cahill, Cap EACL 2012) for more details
- We'd like to integrate this into the Moses SMT toolkit in future work (however, tractability will be a challenge!)
- Here are two examples (French first)...

DE-FR inflection prediction system

Overview of the inflection process

stemmed SMT-output	predicted features	inflected forms	after post-processing	gloss
le[DET]				*the*
plus[ADV]				*most*
grand[ADJ]				*large*
démocratie<Fem>[NOM]				*democracy*
musulman[ADJ]				*muslim*
dans[PRP]				*in*
le[DET]				*the*
histoire<Fem>[NOM]				*history*

DE-FR inflection prediction system

Overview of the inflection process

stemmed SMT-output	predicted features	inflected forms	after post-processing	gloss
le[DET]	DET-Fem.Sg			*the*
plus[ADV]	ADV			*most*
grand[ADJ]	ADJ-Fem.Sg			*large*
démocratie<Fem>[NOM]	NOM-Fem.Sg			*democracy*
musulman[ADJ]	ADJ-Fem.Sg			*muslim*
dans[PRP]	PRP			*in*
le[DET]	DET-Fem.Sg			*the*
histoire<Fem>[NOM]	NOM-Fem.Sg			*history*

DE-FR inflection prediction system

Overview of the inflection process

stemmed SMT-output	predicted features	inflected forms	after post-processing	gloss
le[DET]	DET-Fem.Sg	la		*the*
plus[ADV]	ADV	plus		*most*
grand[ADJ]	ADJ-Fem.Sg	grande		*large*
démocratie<Fem>[NOM]	NOM-Fem.Sg	démocratie		*democracy*
musulman[ADJ]	ADJ-Fem.Sg	musulmane		*muslim*
dans[PRP]	PRP	dans		*in*
le[DET]	DET-Fem.Sg	la		*the*
histoire<Fem>[NOM]	NOM-Fem.Sg	histoire		*history*

DE-FR inflection prediction system

Overview of the inflection process

stemmed SMT-output	predicted features	inflected forms	after post-processing	gloss
le[DET]	DET-Fem.Sg	la	la	*the*
plus[ADV]	ADV	plus	plus	*most*
grand[ADJ]	ADJ-Fem.Sg	grande	grande	*large*
démocratie<Fem>[NOM]	NOM-Fem.Sg	démocratie	démocratie	*democracy*
musulman[ADJ]	ADJ-Fem.Sg	musulmane	musulmane	*muslim*
dans[PRP]	PRP	dans	dans	*in*
le[DET]	DET-Fem.Sg	la	l'	*the*
histoire<Fem>[NOM]	NOM-Fem.Sg	histoire	histoire	*history*

DE-FR inflection prediction system

Overview of the inflection process

stemmed SMT-output	predicted features	inflected forms	after post-processing	gloss
le[DET]	DET-Fem.Sg	la	la	*the*
plus[ADV]	ADV	plus	plus	*most*
grand[ADJ]	ADJ-Fem.Sg	grande	grande	*large*
démocratie<Fem>[NOM]	NOM-Fem.Sg	démocratie	démocratie	*democracy*
musulman[ADJ]	ADJ-Fem.Sg	musulmane	musulmane	*muslim*
dans[PRP]	PRP	dans	dans	*in*
le[DET]	DET-Fem.Sg	la	l'	*the*
histoire<Fem>[NOM]	NOM-Fem.Sg	histoire	histoire	*history*

Next step

Combine this with bilingual terminology mining (Daille and Morin 2008), this enables **context-dependent** inflection of mined terminology in translation

Feature prediction and inflection: example

English input these buses may have access to that country [...]

SMT output	predicted features	inflected forms	gloss
solche<+INDEF><Pro>	PIAT-Masc.Nom.Pl.St	**solche**	such
Bus<+NN><Masc><Pl>	NN-Masc.Nom.Pl.Wk	**Busse**	buses
haben<VAFIN>	haben<V>	haben	have
dann<ADV>	ADV	dann	then
zwar<ADV>	ADV	zwar	though
Zugang<+NN><Masc><Sg>	NN-Masc.Acc.Sg.St	**Zugang**	access
zu<APPR><Dat>	APPR-Dat	zu	to
die<+ART><Def>	ART-Neut.Dat.Sg.St	**dem**	the
betreffend<+ADJ><Pos>	ADJA-Neut.Dat.Sg.Wk	**betreffenden**	respective
Land<+NN><Neut><Sg>	NN-Neut.Dat.Sg.Wk	**Land**	country

Extending inflection prediction using knowledge of subcategorization

- Subcategorization knowledge captures information about arguments of a verb (e.g., what sort of direct objects are allowed (if any))
- Working on two approaches for adding subcategorization information into SMT
  - Using German subcategorization information extracted from web corpora to improve case prediction for contexts suffering from sparsity
  - Using machine learning features based on semantic roles from the English source sentence (obtained through dependency parsing
- Unique in terms of directly integrating lexical semantic research into statistical machine translation
- See (Weller, Fraser, Schulte im Walde - ACL 2013) for more details

Conclusion

- In my opinion: the key questions in statistical machine translation are about linguistic representation and learning from data
- I presented what we have done so far and how we plan to continue
  - I focused mostly on linguistic representation
  - I discussed some syntax, and talked a lot about morphological generation (skipping many details of things like portmanteaus)
  - We solve interesting machine learning and computer science problems as well (most details were skipped)
  - In future work, we will integrate these techniques more tightly into inference (decoding)
- Credits: Marion Weller, Fabienne Cap, Fabienne Braune, Nadir Durrani, Anita Ramm, Hassan Sajjad, Aoife Cahill, Helmut Schmid, Sabine Schulte im Walde, Hinrich Schütze

Thank you!

Thank you!