Linguistic Knowledge for Statistical Machine Translation

Alexander Fraser

Visiting: ICL, Uni Heidelberg
Permanent: CIS, Uni München

22. April 2014 - Uni Heidelberg

Hauptseminar: Linguistic Knowledge for SMT

- Hauptseminar meeting on Tuesdays at 11:15
- Prerequisite: a course on statistical machine translation (SMT), talk to me otherwise, after class or at office hours (15:00-16:00 Tuesdays, i.e., today)
- Usual setup: Referat + Hausarbeit (6-10 pages + references) on a research topic
- If you strongly prefer, I may be able to give you a project - this will be more work (and also involve a Referat).
- Referat: English preferred.

Information about you

- Please download the text file from the course web page
- Contains a few questions about your background (courses, human languages)
- I will use this to put together a list of topics which will be presented next week, you will then email me with preferences
- You may also request specific topics (you might have an idea about this after today's lecture, or some use of Google Scholar)

Today

- I did my PhD in 2007 at the University of Southern California / Information Sciences Institute with Daniel Marcu, the topic was word alignment.
- After that I moved to the University of Stuttgart, and recently to LMU München
- I work on a number of things (including cross-language information retrieval and syntactic parsing), but this is now my primary research area
- Today I will lecture about my work on translating from and to morphologically rich languages (mostly German for this talk)
- This is one way (of very many!) to deal with integrating linguistic knowledge into SMT
- Feel free to ask questions during the talk

Next week

- Looking for a volunteer for next week, the Referat will cover two basic papers focusing on preprocessing:
- Empirical Methods for Compound Splitting. Philipp Koehn and Kevin Knight. EACL 2003. (German compound splitting)
- Improving Statistical MT Through Morphological Analysis. Sharon Goldwater and David McClosky. EMNLP 2005. (Reducing inflection)

- Questions before we start?

Linguistic Knowledge for Statistical Machine Translation

- 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

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

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:

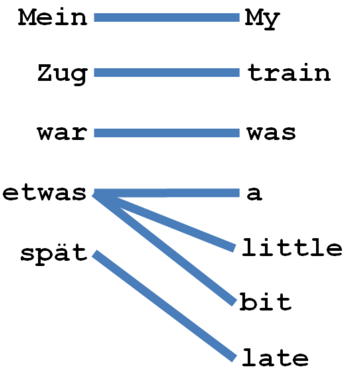
- Problem definition: translate sentences (independently!)
- Evaluation: use human judgements of quality (and correlated automatic metrics)
- Linguistic representation: subject of the rest of this talk
- Model: see next few slides
- Training: see next few slides
- Search

(Statistical) machine translation is a structured prediction problem

Structured prediction problems in computational linguistics are defined like this:

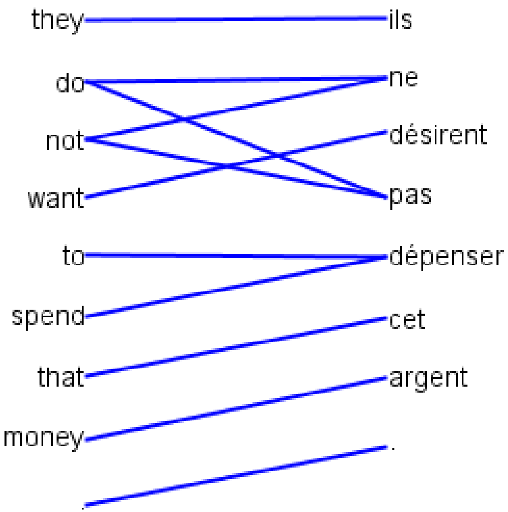
- Problem definition: translate sentences (independently!)
- Evaluation: use human judgements of quality (and correlated automatic metrics)
- Linguistic representation: subject of the rest of this talk
- Model: see next few slides
- Training: see next few slides
- Search: beyond the scope of this talk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

Basic non-linguistic representation - word alignment

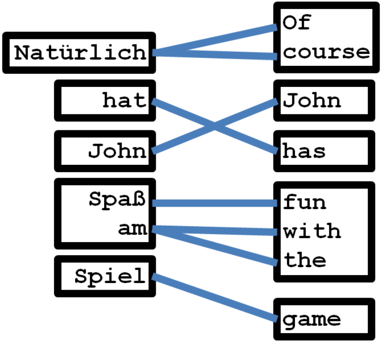


Word alignment: bigraph, connected components show "minimal translation units"

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)!

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

- DFG project: 3 year project (recently renewed for 3 more years)
- Combined with support from the FP7 TTC project:
  Terminology Extraction, Translation Tools and Comparable Corpora
- 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

Long-distance reordering during inference - N-gram system

- We have developed a system which uses n-grams of *minimal translation units* rather than phrases (this is called N-Gram SMT)
- Our contribution to N-Gram SMT is that we have added reordering (previous work models only lexical choice)
- One aspect of this work is that we have lexicalized target jump positions
  - Consider generating German in English order...
  - I - Ich, have - habe, INSERT-GERMAN-GAP, eaten - gegessen, JUMP-BACK-TO-GERMAN-GAP, cookies - Kekse
  - This is a lingustic-knowledge-free way of handling German syntax and is very powerful
  - Better performance than phrase-based model in German to English and French to English evaluations
  - See (Durrani, Schmid, Fraser ACL 2011) for details
- However, this does not solve the long-distance reordering problem completely, there is clearly the need for linguistic knowledge
- We are adding syntactic knowledge to this model

Long-distance reordering during inference - Hiero - 1 of 2

- We also work with decoders based on SCFG (Synchronous Context-Free Grammar)
- One framework is called Hiero (Chiang 2007)
- In Hiero, hierarchical phrase pairs are made by taking a larger phrase pair and subtracting out one or more smaller phrase pairs
- For instance:
  - "habe X_1 gegessen -> have eaten X_1"
  - X_1 will be replaced by another hierarchical phrase pair
- Popular framework, no explicit linguistic knowledge, only implicitly learned from word alignment

Long-distance reordering during inference - Hiero - 2 of 2

- We have worked on making Hiero able to deal with reorderings up to 50 words (not uncommon in German newspaper text!)
  - Found that Hiero requires knowledge of German clause structure and verbal complex
  - It seems likely that this result holds for all hierarchical systems (due to data sparsity)
  - Implemented improvements using hard constraints, see (Braune, Gojun, Fraser EAMT 2012)
- Current work involves tight integration of Vowpal Wabbit classifier (Langford) into Moses SMT toolkit
- Models hierarchical phrase pair choices given source language syntactic context obtained from a parser
  (Braune, Fraser, Daume, Carpuat, JHU Summer Workshop Team - in preparation)

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

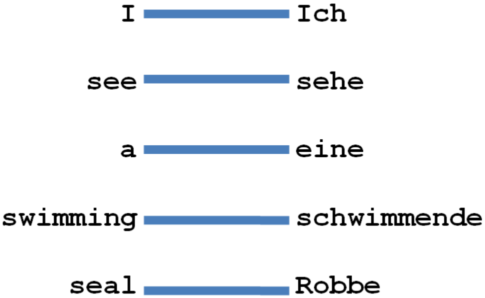
Example: baseline vs. two-step system

- 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`

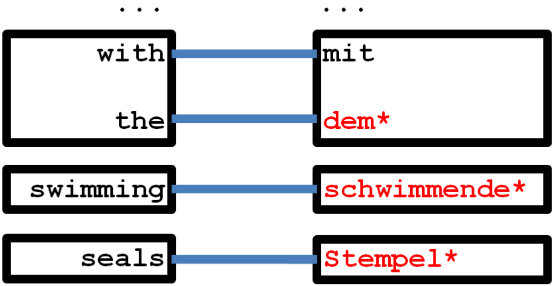
Inflection - example



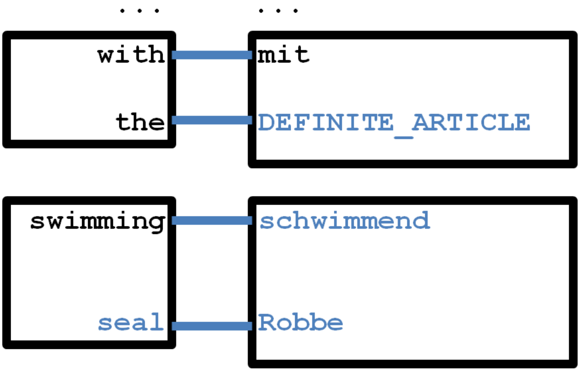
Suppose the training data is typical European Parliament material and this sentence pair is also in the training data.

We would like to translate: “... with the swimming seals”

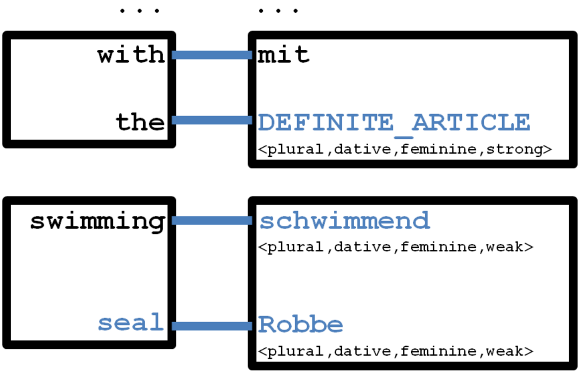
Inflection - problem in baseline



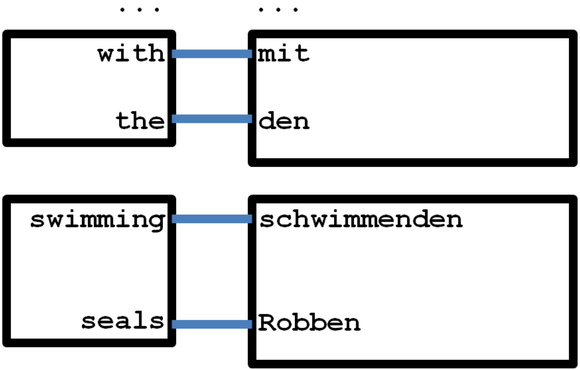
Dealing with inflection - translation to underspecified representation



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

Future work - text structure/pragmatics

First steps towards using text structure/pragmatics in SMT.
Breaking the sentence independence assumption:

- Using coreference models, initially for underspecified pronouns:
  - What is gender of "it" in English when translated to German? (we have completed an initial study)
  - Translating from Subject-Drop languages like Spanish/Italian/Czech/Arabic: which pronoun should be used in the translation? (we have some work on this already)
- Consistent lexical choice across sentences (Carpuat, others have initial work here)
- Many other problems here, many of which have never been solved (even in rule-based MT)
- Starting point: see (Hardmeier 2013) for a comprehensive survey

Lessons/questions for translating to other morphologically rich languages

- Should we directly predict surface forms (Toutanova et al 2008), or predict linguistic features and generate (or work)?
  - What role does language syncretism (irregular dropped inflection) play?
- German: rule-based morphological analysis/generation and statistical disambiguation. The right combination for other languages?
- How should we better model ambiguity in word formation? (Lattices?) What role does the compositionality assumption play?
- I mostly discussed nominal inflection (See (de Gispert and Mariño) for Spanish verbal inflection)
  - How to deal with reflexives and other complex verbal phenomena?
- El Kholy and Habash: for Arabic only number and gender should be predicted (translate determiners). Automate determining this?
- Where to enrich the source language rather than target? (Oflazer, others)
- Sequence models work for English (configurational) and somewhat for German (less-configurational)
  - What about non-configurational (Czech, Russian, etc)?

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, Anita Ramm, Aoife Cahill, Helmut Schmid, Sabine Schulte im Walde, Hinrich Schütze, others in my group...
- Funding: Deutsche Forschungsgemeinschaft, EU FP7 TTC project: Terminology Extraction, Translation Tools and Comparable Corpora

Next week

- Looking for a volunteer for next week, the Referat will cover two basic papers focusing on preprocessing:
- Empirical Methods for Compound Splitting. Philipp Koehn and Kevin Knight. EACL 2003. (German compound splitting)
- Improving Statistical MT Through Morphological Analysis. Sharon Goldwater and David McClosky. EMNLP 2005. (Reducing inflection)

Thank you!

Thank you!