Translation to Morphologically Rich Languages

Alexander Fraser

Centrum für Informations- und Sprachverarbeitung (CIS)
Ludwig-Maximilians-Universität München

EAMT 2016 - Riga, May 30th 2016

Translating to Morphologically Richer Languages with SMT

- Most research on statistical machine translation (SMT) is on translating into English, which is a **morphologically-not-at-all-rich** language
  - Usually significant interest in **morphological reduction**
- Recent interest in the other direction, translation to morphologically rich languages (MRLs)
  - Requires **morphological generation**

Research on machine translation - present situation

The situation up until recently

Research on machine translation - present situation

The situation up until recently

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

Research on machine translation - present situation

The situation up until recently

- 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 up until recently

- 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 up until recently

- 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 up until recently

- 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 up until recently

- 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

Challenges

- The challenges I am currently focusing on:
  - How to generate morphology (for German) 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

Outline

- History and motivation
- 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 and discussion

Our work

- Before: DFG (German National Science Foundation), FP7 TTC project
- Now: two new Horizon2020 projects (ERC Starting Grant and Health in my Language "HimL" (come to our poster!))
- **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 to German

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 (unpublished experiments on phrase-based)

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 a MRL 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
- We have a lot of work on this, for instance:
  - Discriminative rule selection in Hiero and String-to-Tree (Braune et al. EMNLP 2015; see Fabienne Braune presentation tomorrow)
  - Operation Sequence Model (new CL journal article: Durrani et al. 2015)
- Also, if you have a different alphabetic script to deal with, see:
  - Learning transliteration models through transliteration mining (Sajjad et al. 2012, Durrani et al. 2014)

Translating from English to German

English to German is a challenging problem - previous generation rule-based systems relatively competitive

Translating from English to German

English to German is a challenging problem - previous generation rule-based systems relatively competitive

- 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 relatively competitive

- 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 relatively competitive

- 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 relatively competitive

- 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 relatively competitive

- 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”)

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

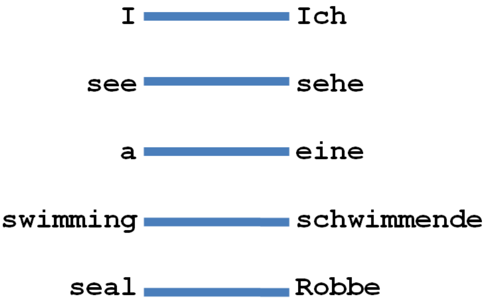
   a) predict inflection features using a sequence classifier
   b) generate inflected forms based on predicted features and lemmas

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`

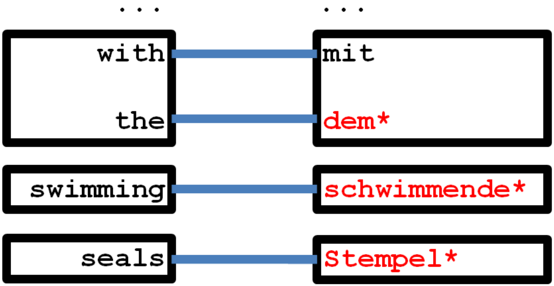
Inflection - example



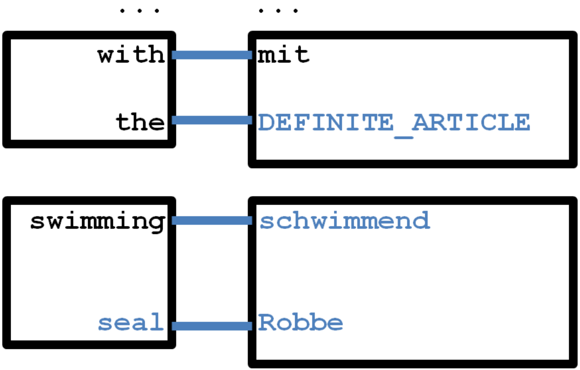
Suppose the training data is typical European Parliament material and this sentence pair is also in the training data.

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

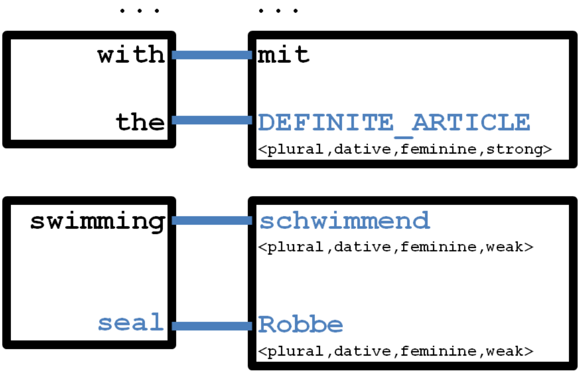
Inflection - problem in baseline



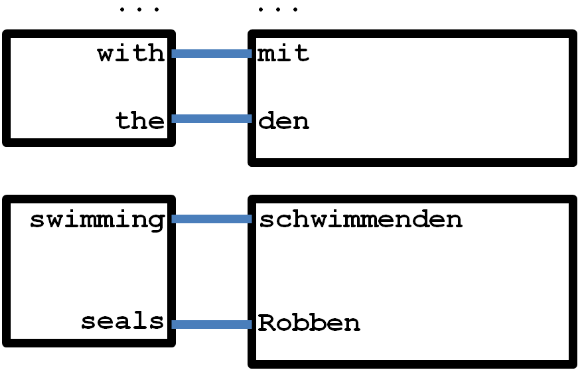
Dealing with inflection - translation to underspecified representation



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 (EN-DE)
- 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*

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>			
Bus<+NN><Masc><Pl>			
haben<VAFIN>			
dann<ADV>			
zwar<ADV>			
Zugang<+NN><Masc><Sg>			
zu<APPR><Dat>			
die<+ART><Def>			
betreffend<+ADJ><Pos>			
Land<+NN><Neut><Sg>			

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		
Bus<+NN><Masc><Pl>	NN-Masc Pl		
haben<VAFIN>	haben<V>		
dann<ADV>	ADV		
zwar<ADV>	ADV		
Zugang<+NN><Masc><Sg>	NN-Masc Sg		
zu<APPR><Dat>	APPR-Dat		
die<+ART><Def>	ART		
betreffend<+ADJ><Pos>	ADJA		
Land<+NN><Neut><Sg>	NN-Neut Sg		

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		
Bus<+NN><Masc><Pl>	NN-Masc.Nom.Pl.Wk		
haben<VAFIN>	haben<V>		
dann<ADV>	ADV		
zwar<ADV>	ADV		
Zugang<+NN><Masc><Sg>	NN-Masc.Acc.Sg.St		
zu<APPR><Dat>	APPR-Dat		
die<+ART><Def>	ART-Neut.Dat.Sg.St		
betreffend<+ADJ><Pos>	ADJA-Neut.Dat.Sg.Wk		
Land<+NN><Neut><Sg>	NN-Neut.Dat.Sg.Wk		

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

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>`

- I'll present an example, if there is time, some details
- See (Cap, Fraser, Weller, Cahill EACL 2014) and Cap's PhD thesis for more

Compound Merging for English to German SMT

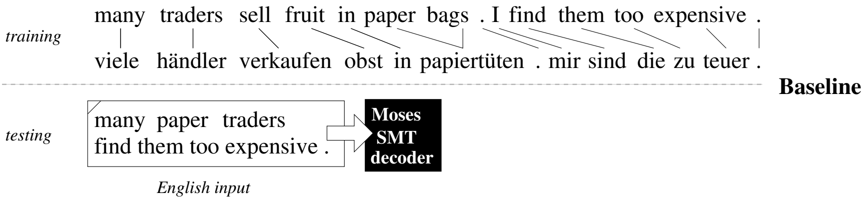
Compound Merging for English to German SMT

training

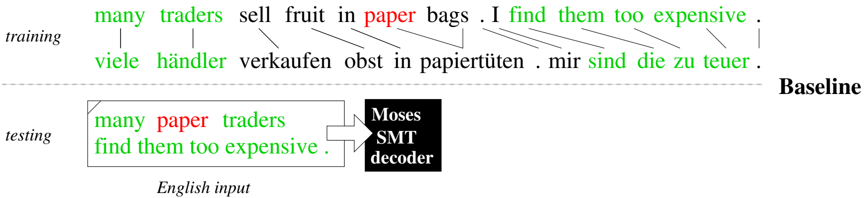
many traders sell fruit in paper bags . I find them too expensive .

viele händler verkaufen obst in papiertüten . mir sind die zu teuer .

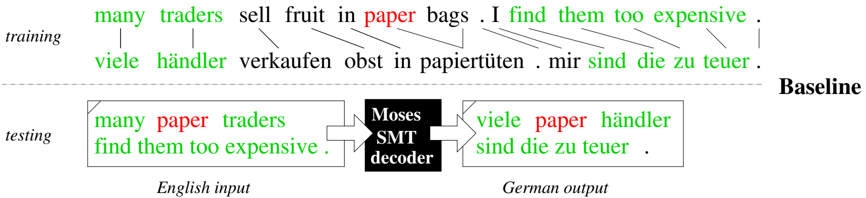
Compound Merging for English to German SMT



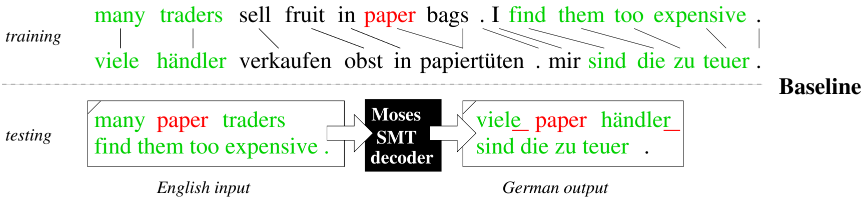
Compound Merging for English to German SMT



Compound Merging for English to German SMT



Compound Merging for English to German SMT



Compound Merging for English to German SMT

training

many traders sell fruit in paper bags . I find them too expensive .

viele händler verkaufen obst in papiertüten . mir sind die zu teuer .

Baseline

testing

many paper traders
find them too expensive .

English input

Moses SMT decoder

viele_ paper händler_
sind die zu teuer .

German output

---

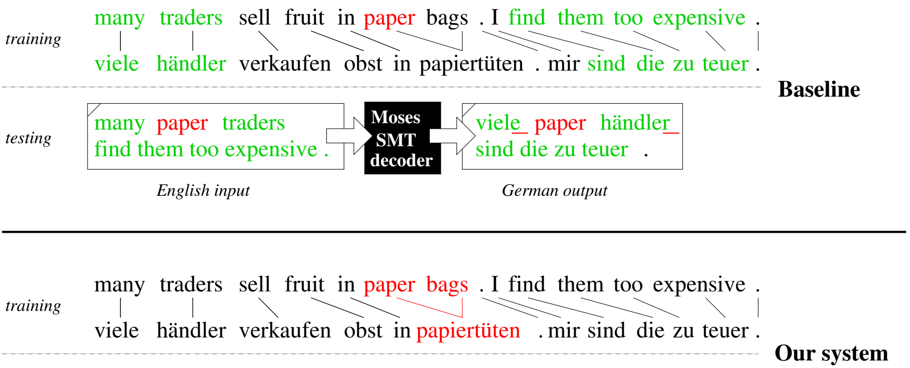
training

many traders sell fruit in paper bags . I find them too expensive .

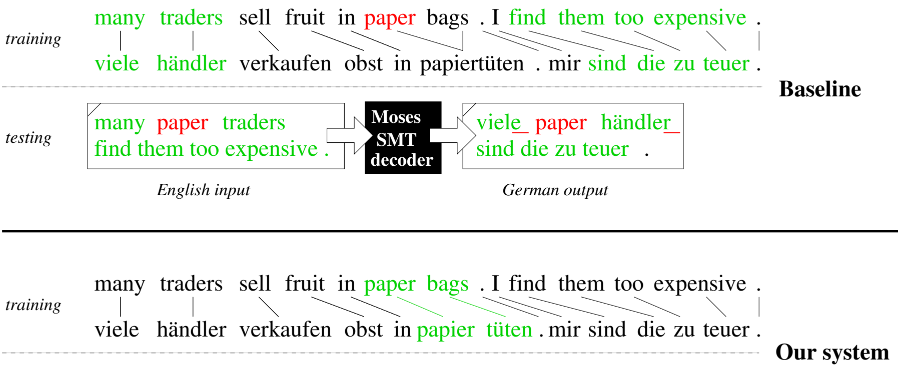
viele händler verkaufen obst in papiertüten . mir sind die zu teuer .

Our system

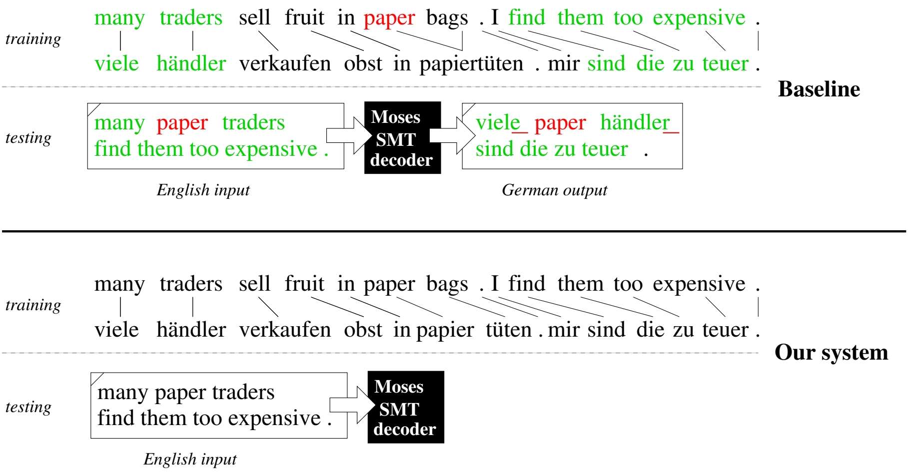
Compound Merging for English to German SMT



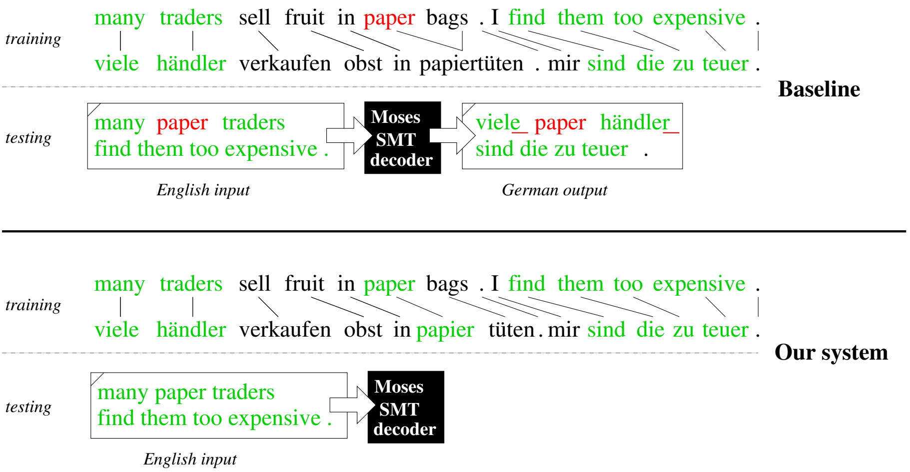
Compound Merging for English to German SMT



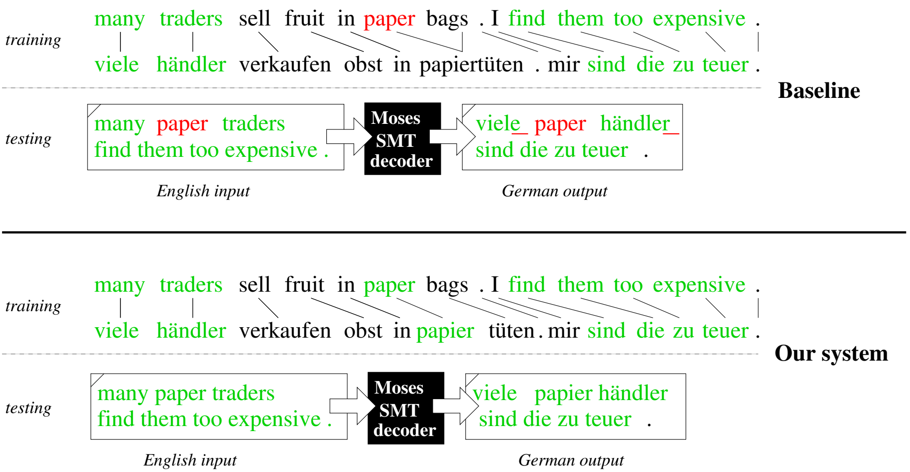
Compound Merging for English to German SMT



Compound Merging for English to German SMT

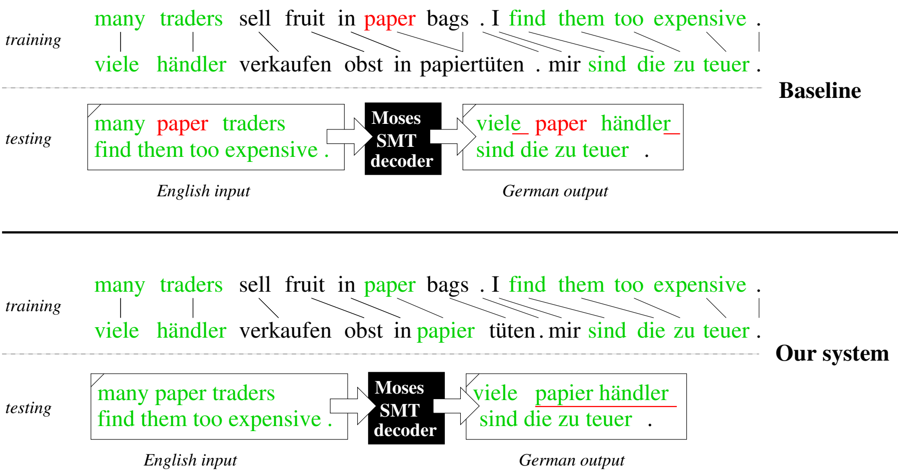


Compound Merging for English to German SMT



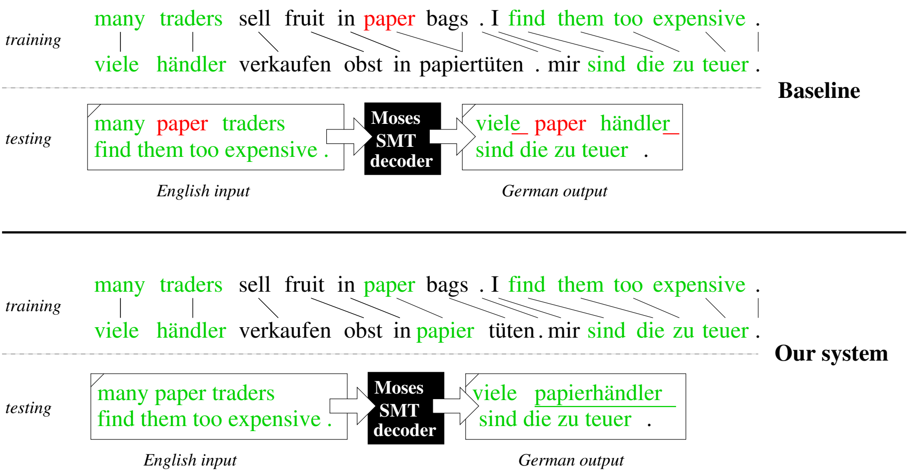
Translate into split and lemmatised German

Compound Merging for English to German SMT



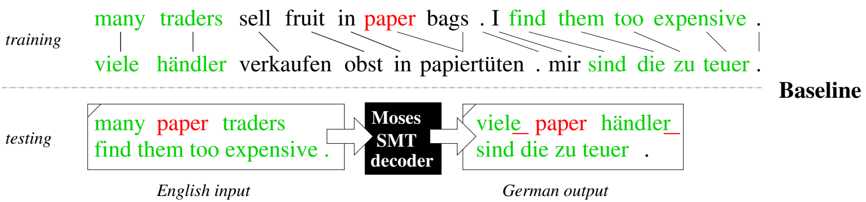
Step 1: Compound Merging

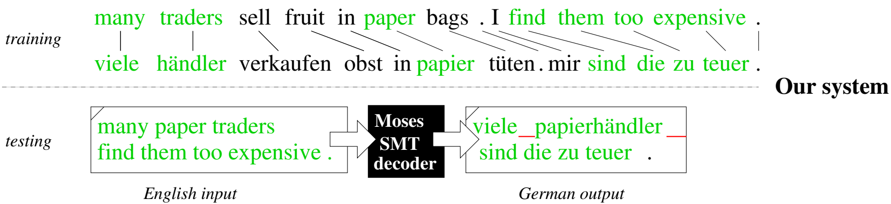
Compound Merging for English to German SMT



Step 1: Compound Merging

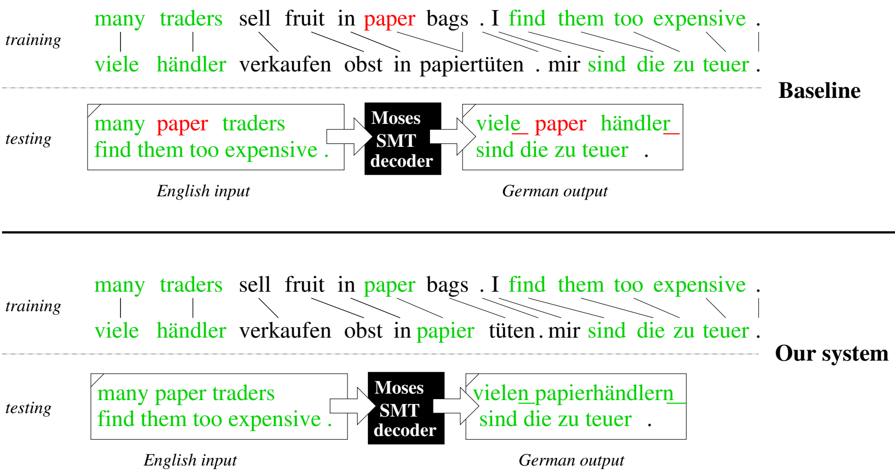
Compound Merging for English to German SMT





Step 1: Compound Merging   Step 2: Re-Inflection

Compound Merging for English to German SMT



Step 1: Compound Merging   Step 2: Re-Inflection

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

Compound merging is a challenging task:

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

Compound merging is a challenging task:

- not all two consecutive words that **could** be merged, **should** be merged:

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

Compound merging is a challenging task:

- not all two consecutive words that **could** be merged, **should** be merged:

  "kind" + *"punsch"* = *"kinderpunsch"* (punch for children)

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

Compound merging is a challenging task:

- not all two consecutive words that **could** be merged, **should** be merged:

  "kind" + *"punsch"* = *"kinderpunsch"* (punch for children)
  but: *"darf ein kind punsch trinken?"*
  (may a child drink punch?)

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

Compound merging is a challenging task:

- not all two consecutive words that **could** be merged, **should** be merged:

  "kind" + *"punsch"* = *"kinderpunsch"* (punch for children)
  but: *"darf ein kind punsch trinken?"*
  (may a child drink punch?)

- **solution:** use linear chain **C**onditional **R**andom **F**ields (Crfs).

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

Compound merging is a challenging task:

- not all two consecutive words that **could** be merged, **should** be merged:

  "kind" + *"punsch"* = *"kinderpunsch"* (punch for children)
  but: *"darf ein kind punsch trinken?"*
  (may a child drink punch?)

- **solution:** use linear chain **C**onditional **R**andom **F**ields (Crfs).
  - machine learning technique

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

Compound merging is a challenging task:

- not all two consecutive words that **could** be merged, **should** be merged:

  "kind" + *"punsch"* = *"kinderpunsch"* (punch for children)
  but: *"darf ein kind punsch trinken?"*
  (may a child drink punch?)

- **solution:** use linear chain **C**onditional **R**andom **F**ields (Crfs).
  - machine learning technique
  - learn **context-dependent merging decisions**

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

Compound merging is a challenging task:

- not all two consecutive words that **could** be merged, **should** be merged:

  "kind" + *"punsch"* = *"kinderpunsch"* (punch for children)
  but: *"darf ein kind punsch trinken?"*
  (may a child drink punch?)

- **solution:** use linear chain **C**onditional **R**andom **F**ields (Crfs).
  - machine learning technique
  - learn **context-dependent merging decisions** based on **features** assigned to each word

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

Compound merging is a challenging task:

- not all two consecutive words that **could** be merged, **should** be merged:

  "kind" + *"punsch"* = *"kinderpunsch"* (punch for children)
  but: *"darf ein kind punsch trinken?"*
  (may a child drink punch?)

- **solution:** use linear chain **C**onditional **R**andom **F**ields (Crfs).
  - machine learning technique
  - learn **context-dependent merging decisions** based on **features** assigned to each word
  - features can be derived from the **target** and/or the **source language** (here: German and English)

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

Examples of CRF features derived from the **target language**:

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

Examples of CRF features derived from the **target language**:

part of speech	some POS patterns are more likely to form compounds than others

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

Examples of CRF features derived from the **target language**:

part of speech	some POS patterns are more likely to form compounds than others
modifier vs. head position	some words occur much more often as modifiers than as heads (and vice versa)

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

Examples of CRF features derived from the **target language**:

part of speech	some POS patterns are more likely to form compounds than others
modifier vs. head position	some words occur much more often as modifiers than as heads (and vice versa)
productivity of a modifier	some words are more productive than others: for each modifier, count the number of different head types

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

Examples of CRF features derived from the **target language**:

part of speech	some POS patterns are more likely to form compounds than others
modifier vs. head position	some words occur much more often as modifiers than as heads (and vice versa)
productivity of a modifier	some words are more productive than others: for each modifier, count the number of different head types

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

Examples of CRF features derived from the **target language**:

part of speech	some POS patterns are more likely to form compounds than others
modifier vs. head position	some words occur much more often as modifiers than as heads (and vice versa)
productivity of a modifier	some words are more productive than others: for each modifier, count the number of different head types

However, as all these features are derived from the (often **disfluent**) target language, they might not be very reliable

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

In contrast, the source sentence is **fluent** language, and sometimes, the English **source sentence structure** may help the decision:

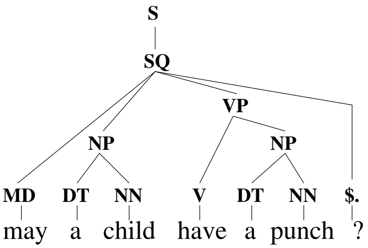
Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

In contrast, the source sentence is **fluent** language, and sometimes, the English **source sentence structure** may help the decision:

should **not** be merged:
darf ein **kind punsch** trinken?

In contrast, the source sentence is **fluent** language, and sometimes, the English **source sentence structure** may help the decision:

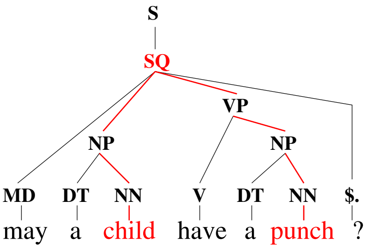


should **not** be merged:
darf ein **kind punsch** trinken?

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

In contrast, the source sentence is **fluent** language, and sometimes, the English **source sentence structure** may help the decision:

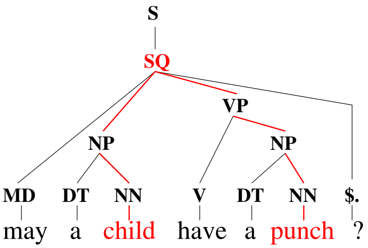


should **not** be merged:
darf ein **kind punsch** trinken?

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

In contrast, the source sentence is **fluent** language, and sometimes, the English **source sentence structure** may help the decision:



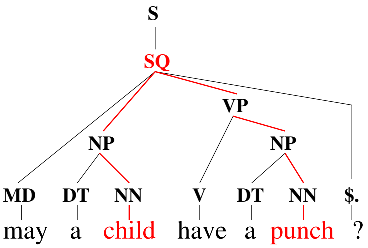
should **not** be merged:
darf ein **kind punsch** trinken?

should be merged:
jeder darf **kind punsch** haben!

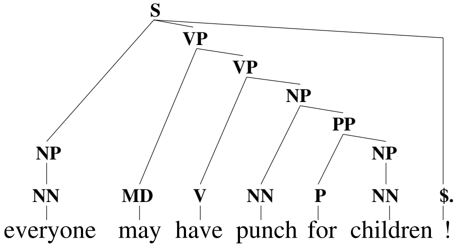
Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

In contrast, the source sentence is **fluent** language, and sometimes, the English **source sentence structure** may help the decision:



should **not** be merged:
darf ein **kind punsch** trinken?

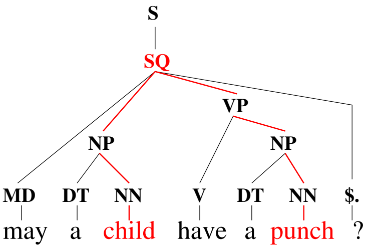


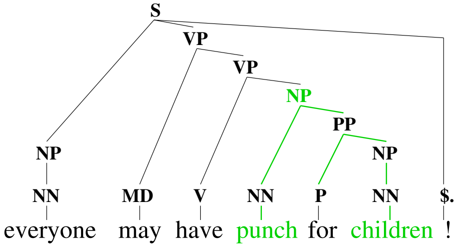
should be merged:
jeder darf **kind punsch** haben!

Compound Merging for English to German SMT

Post-Processing: Transform back into fluent German

In contrast, the source sentence is **fluent** language, and sometimes, the English **source sentence structure** may help the decision:





should **not** be merged:
darf ein **kind punsch** trinken?

should be merged:
jeder darf **kind punsch** haben!

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), which prepositions are subcategorized)
- Working on 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
  - Viewing PPs and NPs in a unified way (always headed by a preposition/case combination)
- One of the first lines of work on directly integrating lexical semantic research into statistical machine translation
- See papers by (Weller, Fraser, Schulte im Walde - ACL 2013, EAMT 2015, others) for more details

How will NMT change all of this? 1 of 2

- Neural machine translation (NMT) is a big deal!
  - The shared task at WMT this year will be almost completely dominated by Edinburgh NMT systems (Sennrich, Haddow, Birch) based on the **Bahdanau, Cho, Bengio 2015 model ("the attentional model")**
- Maybe you should buy some GPUs?
  - Possible first purchase: gamer-PC with Nvidia Titan-X GPU
  - Switch to buying Pascal 1080 GPUs as soon as tested
- **Downside:**
  - NMT training is really slow, and highly unstable
  - NMT model parameters are a matrix of floating point numbers (no phrase table!)

How will NMT change all of this? 2 of 2

- **Upside:** these models have the potential to learn how to do the kind of decisions we need
- Have the ability to condition decisions on arbitrary positions in the input and output sentences
  - Automatically learn to do this in training (no complex feature functions or pre-/post-processing)!
- **My view:** will significantly change the details of solving these problems
  - But will not magically solve them for us!
  - Still need linguistic resources/knowledge, still need to think about morphological generation

Lessons/questions for translating to other morphologically rich languages - 1 of 3

- Key ability is to produce things that are unseen in the training data. For instance:
  - Adjectives agreeing with nouns they have not previously appeared with
  - New inflections of seen lemmas (think of the inflectional richness of the Balto-Slavic languages, for instance!)
  - Noun phrases in a different case than was seen in the training data
  - Integration of terminology mined from comparable corpora (for preliminary work see Weller, Fraser, Heid EAMT 2014)
  - New compounds; similar techniques will hopefully generalize later to agglutinative languages such as Finnish, Estonian and Turkish

Lessons/questions for translating to other morphologically rich languages - 2 of 3

- Would be nice to generalize online in the decoder (rather than with pre-processing and post-processing), but difficult
- Should we directly predict surface forms (Toutanova et al 2008), or predict linguistic features and generate (as in our work)?
  - What role does language syncretism play here?
- 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?

Lessons/questions for translating to other morphologically rich languages - 3 of 3

- Mostly discussed nominal inflection (see, e.g., (de Gispert and Mariño) for Spanish verbal inflection))
  - Can we model tense and mood easily? (preliminary answer from work with Anita Ramm: no!)
  - 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 (Balto-Slavic languages, etc.)?
- Can we capture morphological phenomena dependent on text structure/pragmatics? (E.g., pronouns, many other phenomena)

Conclusion

- The key questions in data-driven 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 a little syntax, and talked a lot about morphological generation (skipping many details of things like portmanteaus)
  - We solve many interesting machine learning problems as well
  - Research on translating to MRLs is very interdisciplinary, plenty of room for collaboration
  - We need better linguistic resources and **more people** working on these problems!
- Credits: Fabienne Braune, Fabienne Cap, Nadir Durrani, Anita Ramm, Hassan Sajjad, Marion Weller, Helmut Schmid, Sabine Schulte im Walde, Aoife Cahill, Hinrich Schütze
- See my website for the slides with references

Thank you!

Thank you!

Work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 644402 and grant agreement No 640550, the DFG grants Distributional Approaches to Semantic Relatedness and Models of Morphosyntax for Statistical Machine Translation and a DFG Heisenberg Fellowship

Word formation: dealing with portmanteaus

- Portmanteau: combination of article and preposition
- split: `im<APPRART><Dat>` → `in<APPR><Dat> die<+ART><Def>`
- merging step: create portmanteaus based on the inflection decision (e.g. no merging if number= plural)

- Splitting portmanteaus allows for better generalization:

in das *internationale Rampenlicht* (*in the international spotlight*)	Acc	in parallel data
im *internationalen Rampenlicht* (*in-the international spotlight*)	Dat	not in parallel data

- With portmanteau-splitting:
  `die<+ART><Def> international<ADJ> Rampenlicht<+NN><Neut><Sg>`
- Generalize from the accusative example with no portmanteau
- Predict inflection features and then re-merge if necessary

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; Hardmeier et al is very nice)
  - 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

References I

- Fabienne Braune, Nina Seemann, Alexander Fraser (2015). Rule Selection with Soft Syntactic Features for String-to-Tree Statistical Machine Translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1095 to 1101, Lisbon, Portugal, September.
- Marion Weller, Alexander Fraser, Sabine Schulte im Walde (2015). Target-side Generation of Prepositions for SMT. In *Proceedings the 18th Annual Conference of the European Association for Machine Translation (EAMT)*, pages 177 to 184, Antalya, Turkey, May.
- Nadir Durrani, Helmut Schmid, Alexander Fraser, Philipp Koehn, Hinrich Schütze (2015). The Operation Sequence Model – Combining N-Gram-based and Phrase-based Statistical Machine Translation. *Computational Linguistics*. 41(2), pages 157 to 186.

References II

- Marion Weller, Fabienne Cap, Stefan Müller, Sabine Schulte im Walde and Alexander Fraser (2014). Distinguishing Degrees of Compositionality in Compound Splitting for Statistical Machine Translation. In *Proceedings of the First Workshop on Computational Approaches to Compound Analysis (ComaComa) at COLING*, Dublin, Ireland, August, to appear.
- Aleš Tamchyna, Fabienne Braune, Alexander Fraser, Marine Carpuat, Hal Daumé III, Chris Quirk (2014). Integrating a Discriminative Classifier into Phrase-based and Hierarchical Decoding. The Prague Bulletin of Mathematical Linguistics, No. 101, pages 29 to 41, April.
- Marion Weller, Alexander Fraser, Ulrich Heid (2014). Combining Bilingual Terminology Mining and Morphological Modeling for Domain Adaptation in SMT. In *Proceedings of the Seventeenth Annual Conference of the European Association for Machine Translation (EAMT)*, pages 11 to 18, Dubrovnik, Croatia, June.

References III

- Fabienne Cap, Alexander Fraser, Marion Weller, Aoife Cahill (2014). How to Produce Unseen Teddy Bears: Improved Morphological Processing of Compounds in SMT. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 579 to 587, Goteborg, Sweden, April.
- Nadir Durrani, Alexander Fraser, Helmut Schmid, Hieu Hoang, 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 (ACL)*, pages 399 to 405, Sofia, Bulgaria, August.
- Marion Weller, Alexander Fraser, Sabine Schulte im Walde (2013). Using Subcategorization Knowledge To Improve Case Prediction For Translation To German. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 593 to 603, Sofia, Bulgaria, August.

References IV

- Fabienne Braune, Anita Gojun, Alexander Fraser (2012). Long-distance Reordering During Search for Hierarchical Phrase-based SMT. In *Proceedings of the 16th Annual Conference of the European Association for Machine Translation (EAMT)*, pages 177 to 184, Trento, Italy, May.
- Alexander Fraser, Marion Weller, Aoife Cahill, 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 (EACL)*, pages 664 to 674, Avignon, France, April.
- Anita Gojun, Alexander Fraser (2012). Determining the Placement of German Verbs in English-to-German SMT. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 726 to 735, Avignon, France, April.
- Hassan Sajjad, Alexander Fraser, Helmut Schmid (2012). A Statistical Model for Unsupervised and Semi-supervised Transliteration Mining. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 469 to 477, Jeju Island, Korea, July.

References V

- Nadir Durrani, Helmut Schmid, Alexander Fraser (2011). A Joint Sequence Translation Model with Integrated Reordering. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1045 to 1054, Portland, Oregon, USA, June.
- Fabienne Fritzinger, Alexander Fraser (2010). How to Avoid Burning Ducks: Combining Linguistic Analysis and Corpus Statistics for German Compound Processing. In *Proceedings of the ACL 2010 Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pages 224 to 234, Uppsala, Sweden, July.
- Alexander Fraser (2009). Experiments in Morphosyntactic Processing for Translating to and from German. In *Proceedings of the EACL Fourth Workshop on Statistical Machine Translation*, pages 115 to 119, Athens, Greece, March.
- Alexander Fraser, Daniel Marcu (2005). ISI's Participation in the Romanian-English Alignment Task. In *Proceedings of the ACL Workshop on Building and Using Parallel Texts*, pages 91 to 94, Ann Arbor, Michigan, USA, June.