Morphological Processing for Statistical Machine Translation

Fabienne Cap

UPPSALA
UNIVERSITY
SWEDEN

Goals for Today

Why Morphological Processing?

Morphological Processing in SMT

A closer look at Compound Merging

Goals for Today

Why Morphological Processing?

Morphological Processing in SMT

A closer look at Compound Merging

Data Sparsity

What is data sparsity?

Data Sparsity

What is data sparsity?

$\rightarrow$ rarely occurring words cause problems in statistical applications

Data Sparsity

What is data sparsity?

$\rightarrow$ rarely occurring words cause problems in statistical applications

Why is this problematic for SMT?

Data Sparsity

What is data sparsity?

$\rightarrow$ rarely occurring words cause problems in statistical applications

Why is this problematic for SMT?

$\rightarrow$ less occurrences $\rightarrow$ less reliable translations

Data Sparsity

What is data sparsity?

→ rarely occurring words cause problems in statistical applications

Why is this problematic for SMT?

→ less occurrences → less reliable translations

→ unseen words cannot be translated

Data Sparsity

What is data sparsity?

→ rarely occurring words cause problems in statistical applications

Why is this problematic for SMT?

→ less occurrences → less reliable translations

→ unseen words cannot be translated

What can we do about it?

Data Sparsity

What is data sparsity?

→ rarely occurring words cause problems in statistical applications

Why is this problematic for SMT?

→ less occurrences → less reliable translations

→ unseen words cannot be translated

What can we do about it?

→ make the most out of the available training data!

Hands on Data Sparsity

There are two kinds of sparse data in parallel corpora for SMT:

1. unseen/rarely seen **simplex** words
2. unseen/rarely seen **complex** words

Hands on Data Sparsity

There are two kinds of sparse data in parallel corpora for SMT:

1. unseen/rarely seen **simplex** words
2. unseen/rarely seen **complex** words

Hands on Data Sparsity

There are two kinds of sparse data in parallel corpora for SMT:

1. unseen/rarely seen **simplex** words
2. unseen/rarely seen **complex** words

Hands on Data Sparsity

There are two kinds of sparse data in parallel corpora for SMT:

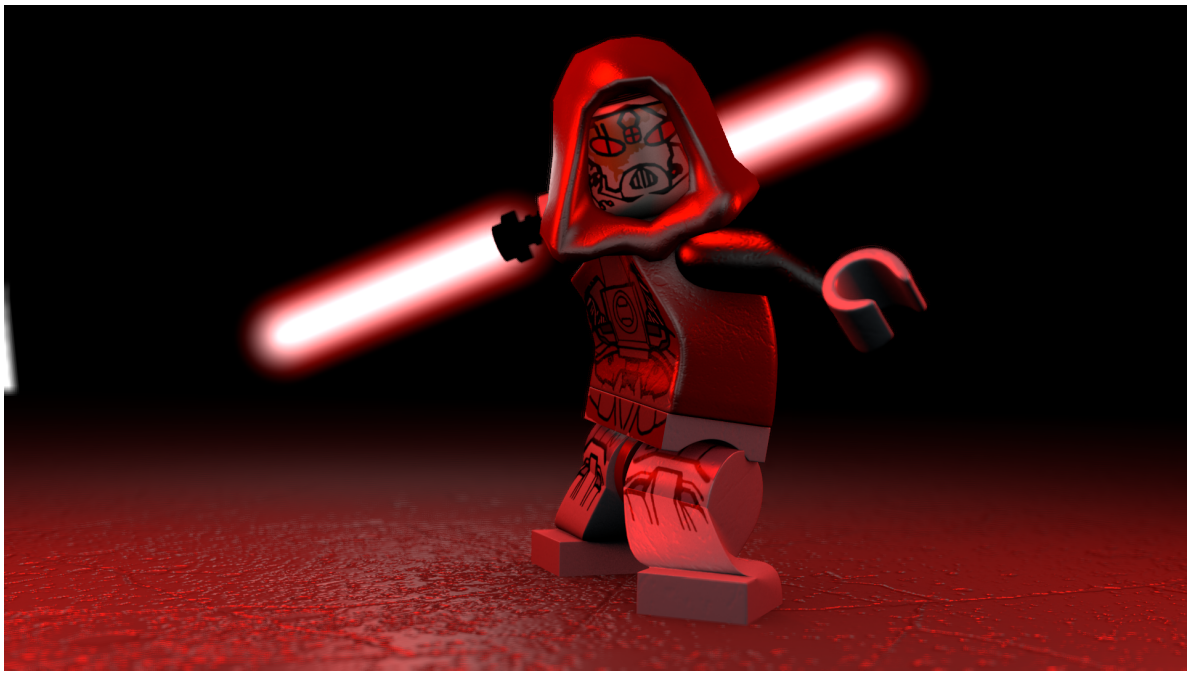
1. unseen/rarely seen **simplex** words
   $\rightarrow$ use more data
2. unseen/rarely seen **complex** words

Hands on Data Sparsity

There are two kinds of sparse data in parallel corpora for SMT:

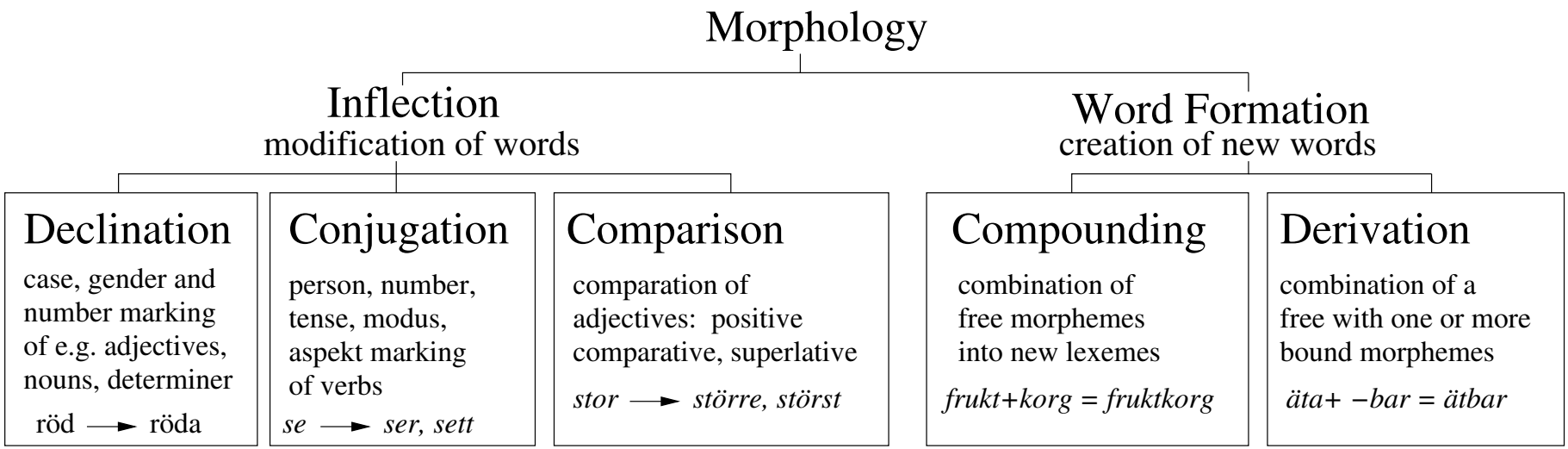
1. unseen/rarely seen **simplex** words
   $\rightarrow$ use more data
2. unseen/rarely seen **complex** words
   $\rightarrow$ decomposition into seen words and word parts

The Revenge of the Sith:

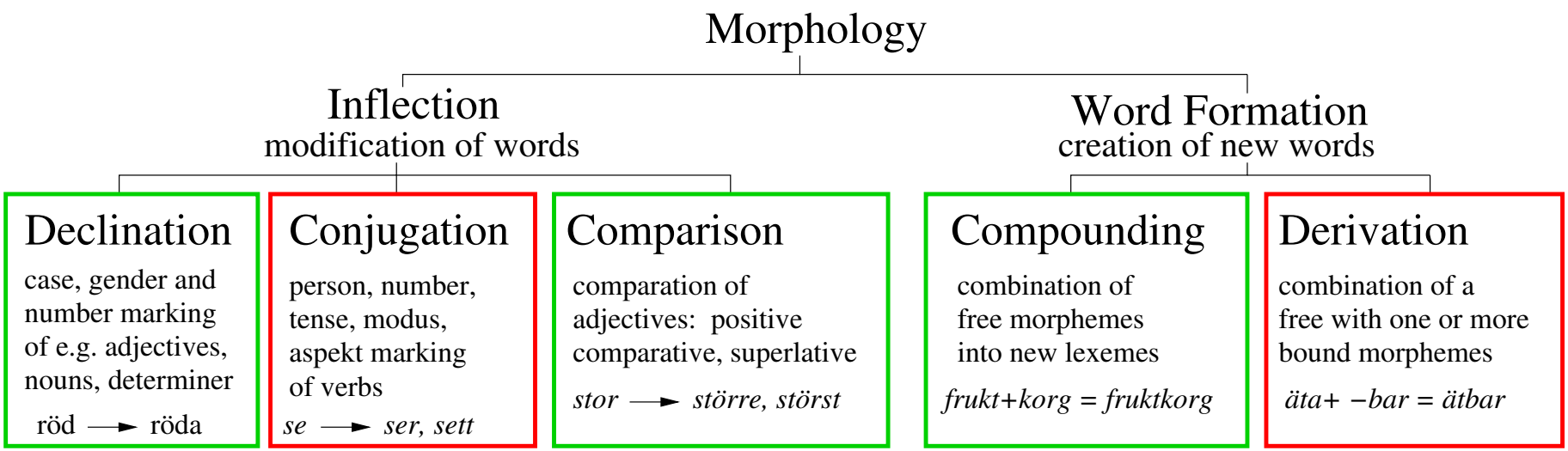


Sith language is morphologically richer than you thought!

Morphology

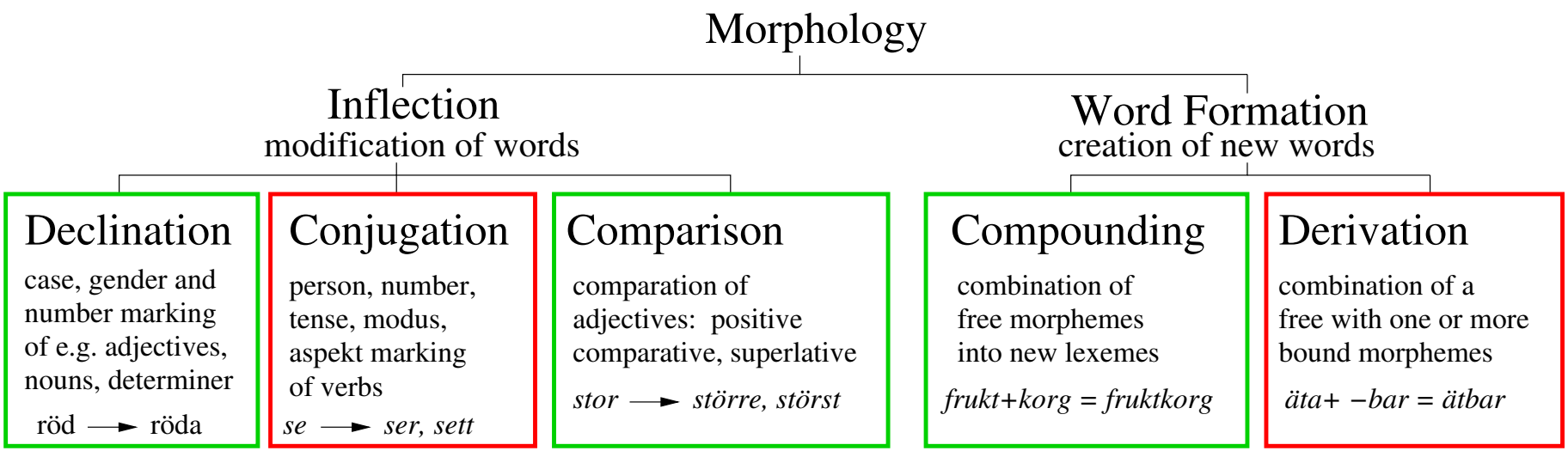

Morphology
Inflection
modification of words
Word Formation
creation of new words
Declination
case, gender and number marking of e.g. adjectives, nouns, determiner
röd → röda
Conjugation
person, number, tense, modus, aspekt marking of verbs
se → ser, sett
Comparison
comparation of adjectives: positive comparative, superlative
stor → större, störst
Compounding
combination of free morphemes into new lexemes
frukt+korg = fruktkorg
Derivation
combination of a free with one or more bound morphemes
äta+ −bar = ätbar

Morphology



Previous work on morphological processing for SMT
has mostly dealt with Deklination, Comparation and Compounding

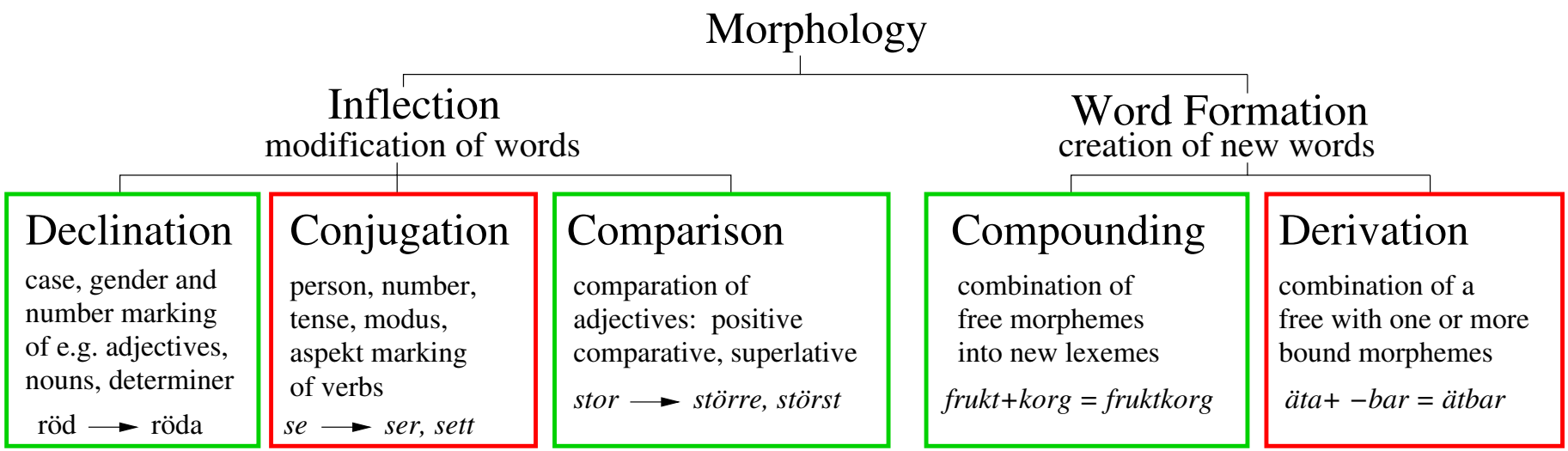
Morphology



Previous work on morphological processing for SMT
has mostly dealt with Deklination, Comparation and Compounding

The goal is to make the source and the target language
as similar as possible prior to word alignment

Morphology



Previous work on morphological processing for SMT
has mostly dealt with Deklination, Comparation and Compounding

The goal is to make the source and the target language
as similar as possible prior to word alignment
e.g. through lemmatisation or compound splitting

Morphological Processing: Lemmatisation

Das Haus ist blau - The house is blue

Morphological Processing: Lemmatisation

Das Haus ist blau - The house is blue

Number	Case	Definite	Indefinite
Singular	Nominativ	das blau**e** Haus	ein blau**es** Haus
	Genitiv	d**es** blau**en** Haus**es**	ein**es** blau**en** Haus**es**
	Akkusativ	in das blau**e** Haus	in ein blau**es** Haus
	Dativ	in d**em** blau**en** Haus	in ein**em** blau**en** Haus
Plural	Nominativ	die blau**en** Häuser	einige blau**e** Häuser
	Genitiv	d**er** blau**en** Häuser	einig**er** blau**er** Häuser
	Akkusativ	in die blau**en** Häuser	in einige blau**e** Häuser
	Dativ	in d**en** blau**en** Häuser**n**	in einige**n** blau**en** Häuser**n**

Morphological Processing: Lemmatisation

Das Haus ist blau - The house is blue

Number	Case	Definite	Indefinite
Singular	Nominativ	das blau**e** Haus	ein blau**es** Haus
	Genitiv	d**es** blau**en** Haus**es**	ein**es** blau**en** Haus**es**
	Akkusativ	in das blau**e** Haus	in ein blau**es** Haus
	Dativ	in d**em** blau**en** Haus	in ein**em** blau**en** Haus
Plural	Nominativ	die blau**en** Häuser	einige blau**e** Häuser
	Genitiv	d**er** blau**en** Häuser	einig**er** blau**er** Häuser
	Akkusativ	in die blau**en** Häuser	in einige blau**e** Häuser
	Dativ	in d**en** blau**en** Häuser**n**	in einig**en** blau**en** Häuser**n**

blau, blaue, blaues, blauen, blauer $\rightarrow$ blue

Morphological Processing: Lemmatisation

Das Haus ist blau - The house is blue

Number	Case	Definite	Indefinite
Singular	Nominativ	das blau**e** Haus	ein blau**es** Haus
	Genitiv	d**es** blau**en** Haus**es**	ein**es** blau**en** Haus**es**
	Akkusativ	in das blau**e** Haus	in ein blau**es** Haus
	Dativ	in d**em** blau**en** Haus	in ein**em** blau**en** Haus
Plural	Nominativ	die blau**en** Häuser	einige blau**e** Häuser
	Genitiv	d**er** blau**en** Häuser	einig**er** blau**er** Häuser
	Akkusativ	in die blau**en** Häuser	in einige blau**e** Häuser
	Dativ	in d**en** blau**en** Häuser**n**	in einige**n** blau**en** Häuser**n**

blau, blaue, blaues, blauen, blauer $\rightarrow$ blue

but: keep differences that are made in both languages!

Haus, Hauses $\rightarrow$ house
Häuser, Häusern $\rightarrow$ house**s**

Morphological Processing: Compound Splitting

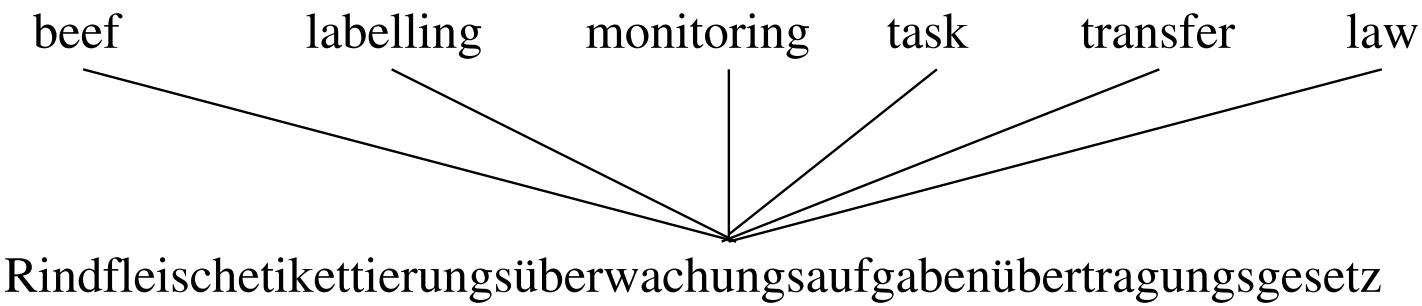
Rindfleischetikettierungsüberwachungsaufgabenübertragungsgesetz

Morphological Processing: Compound Splitting

Rindfleischetikettierungsüberwachungsaufgabenübertragungsgesetz

This is a real example!

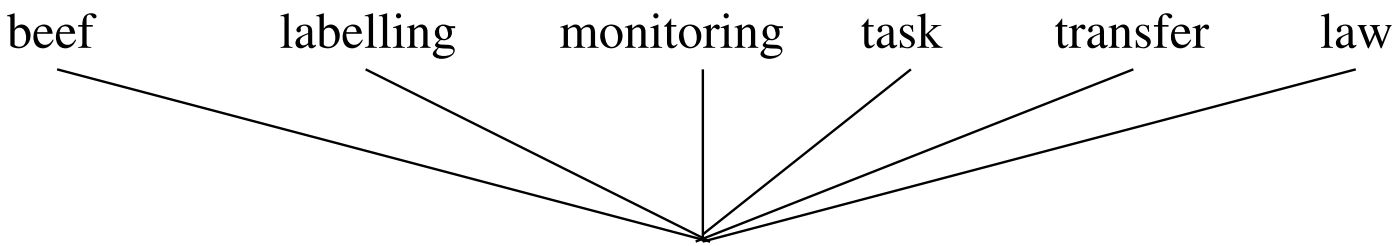
Morphological Processing: Compound Splitting



1:6

This is a real example!

Morphological Processing: Compound Splitting



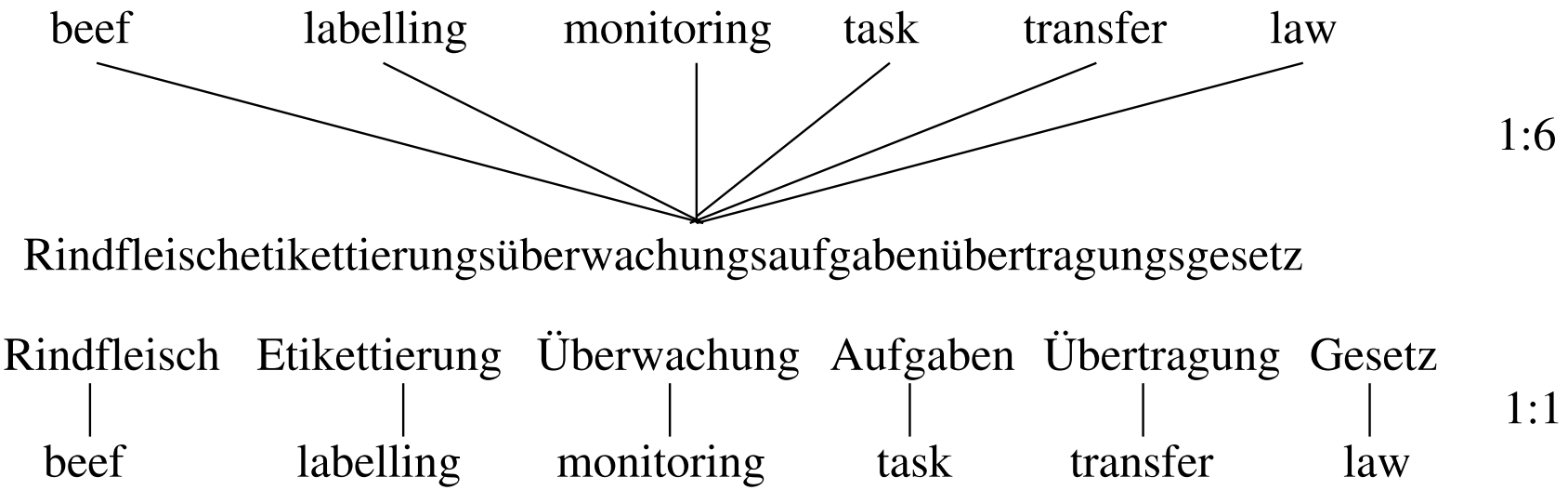
1:6

Rindfleischetikettierungsüberwachungsaufgabenübertragungsgesetz

Rindfleisch Etikettierung Überwachung Aufgaben Übertragung Gesetz

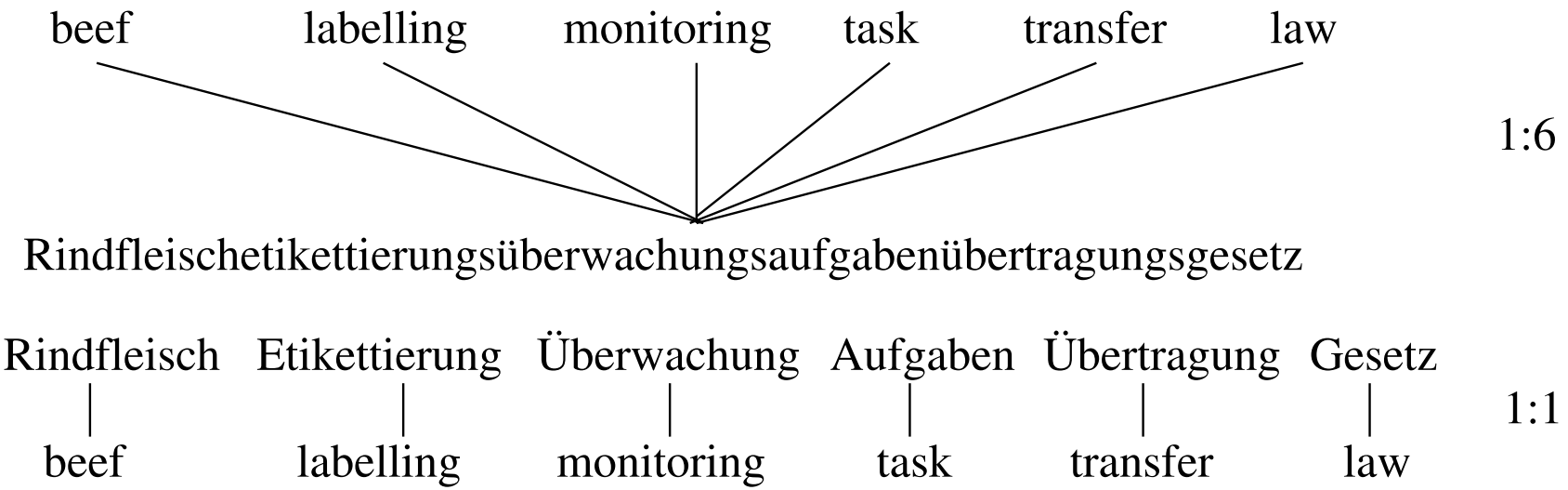
This is a real example!

Morphological Processing: Compound Splitting



This is a real example!

Morphological Processing: Compound Splitting



This is a real example!

→ more compound splitting for SMT in a student project!

Goals for Today

Why Morphological Processing?

Morphological Processing in SMT

A closer look at Compound Merging

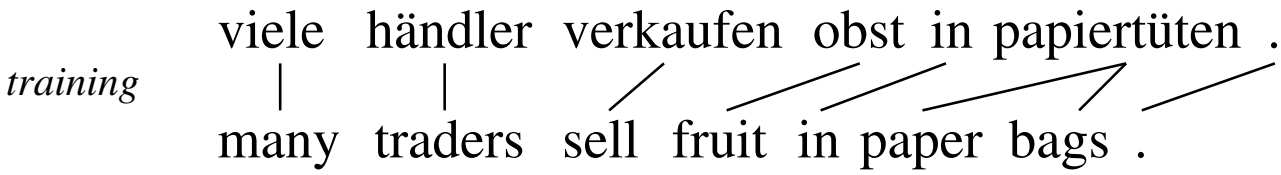
Goals for Today

Why Morphological Processing?

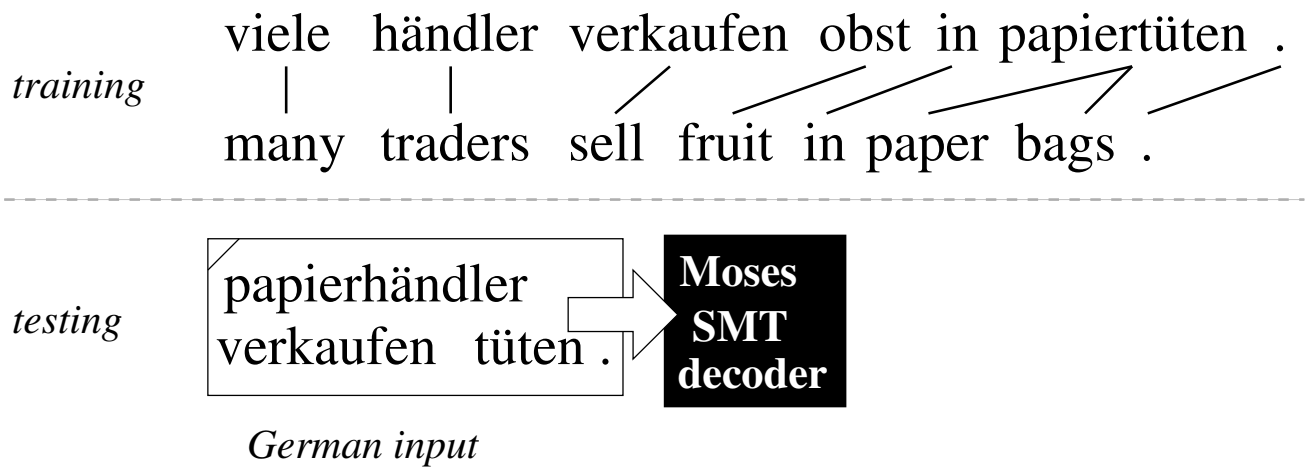
Morphological Processing in SMT

A closer look at Compound Merging

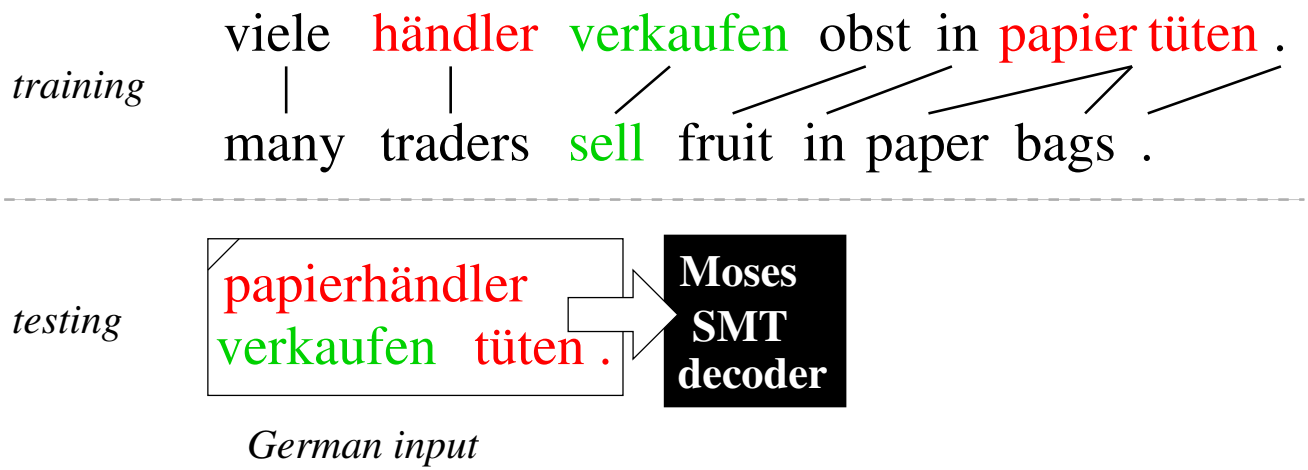
German to English SMT Example



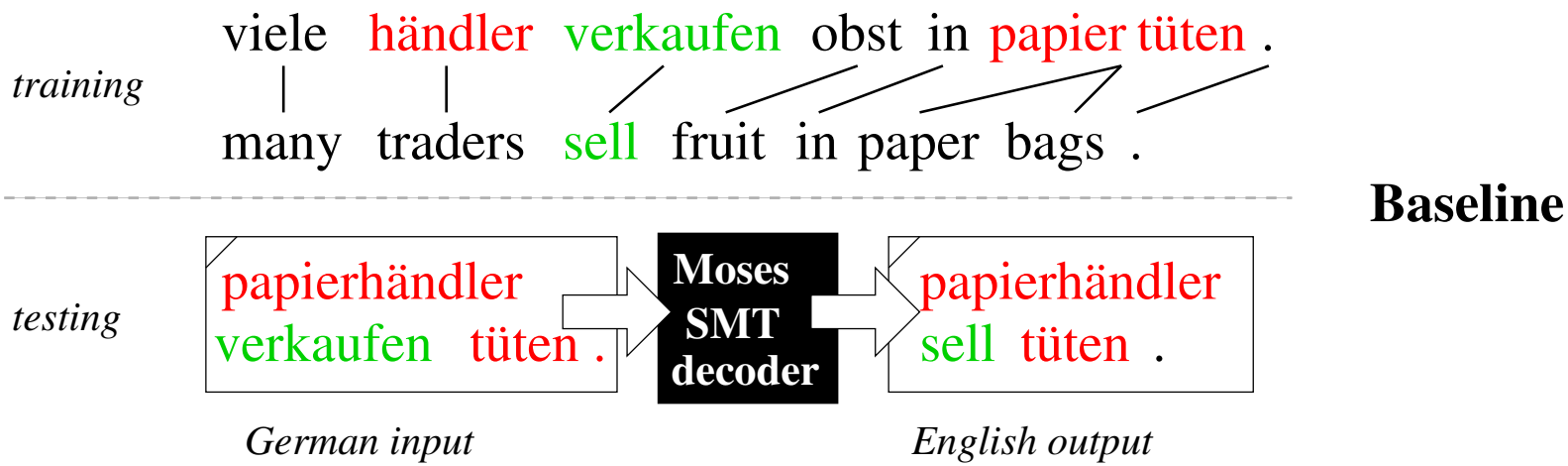
German to English SMT Example



German to English SMT Example

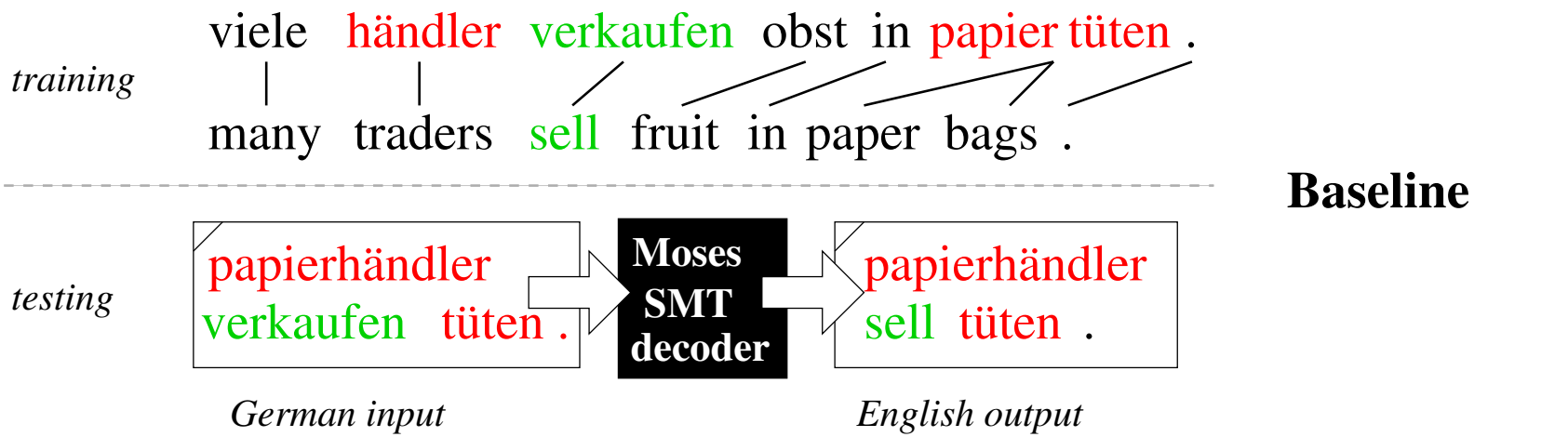


German to English SMT Example

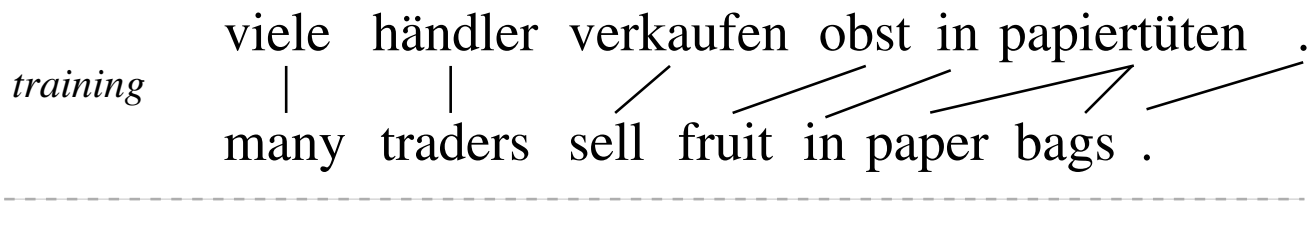


Baseline

German to English SMT Example

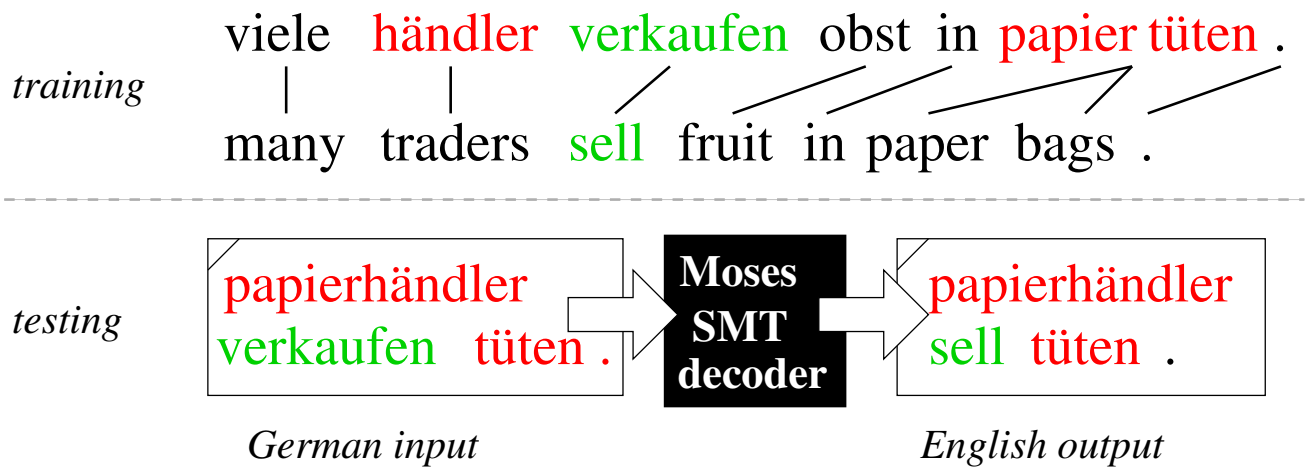


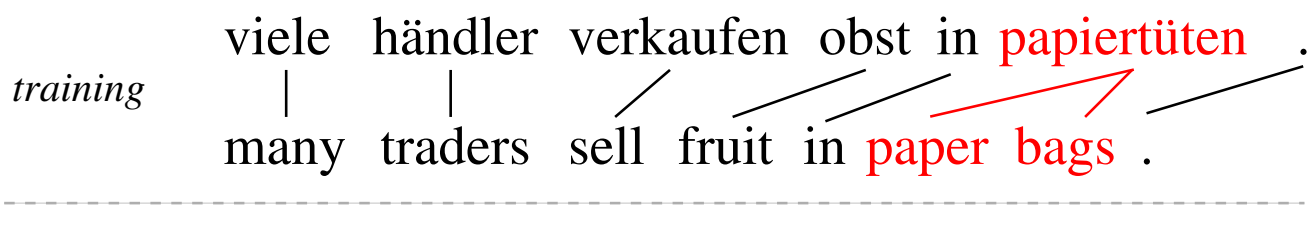
Baseline



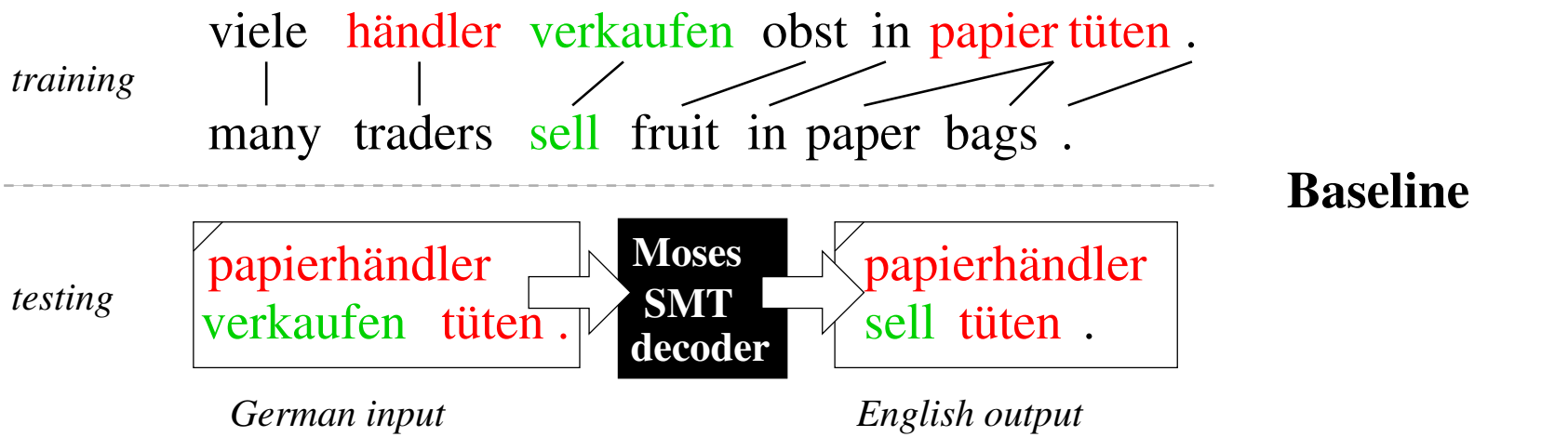
Our system

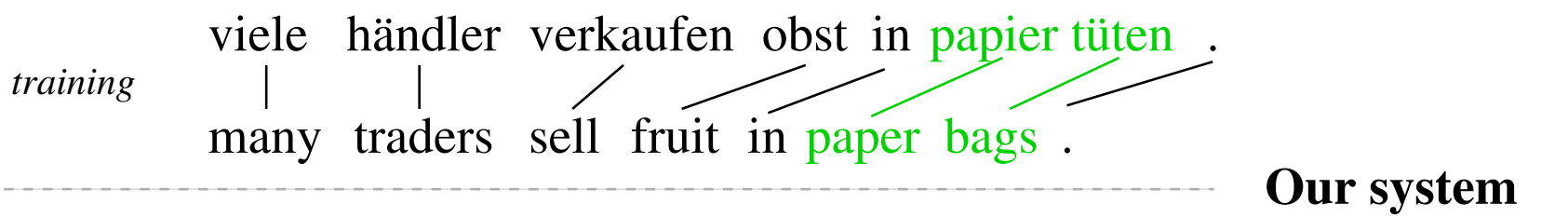
German to English SMT Example



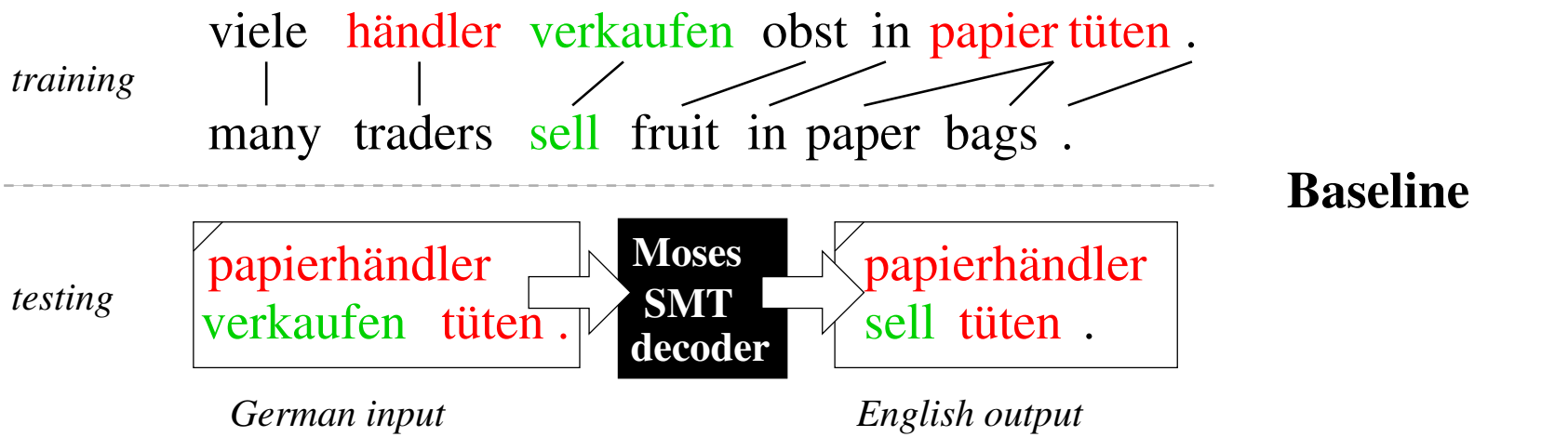


German to English SMT Example

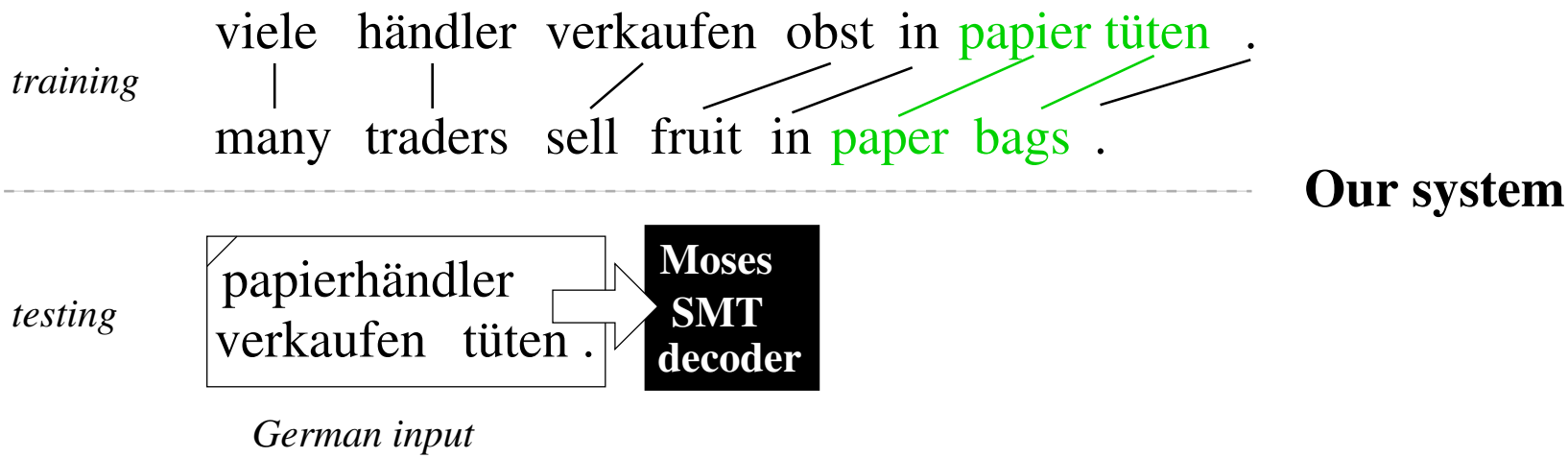




German to English SMT Example

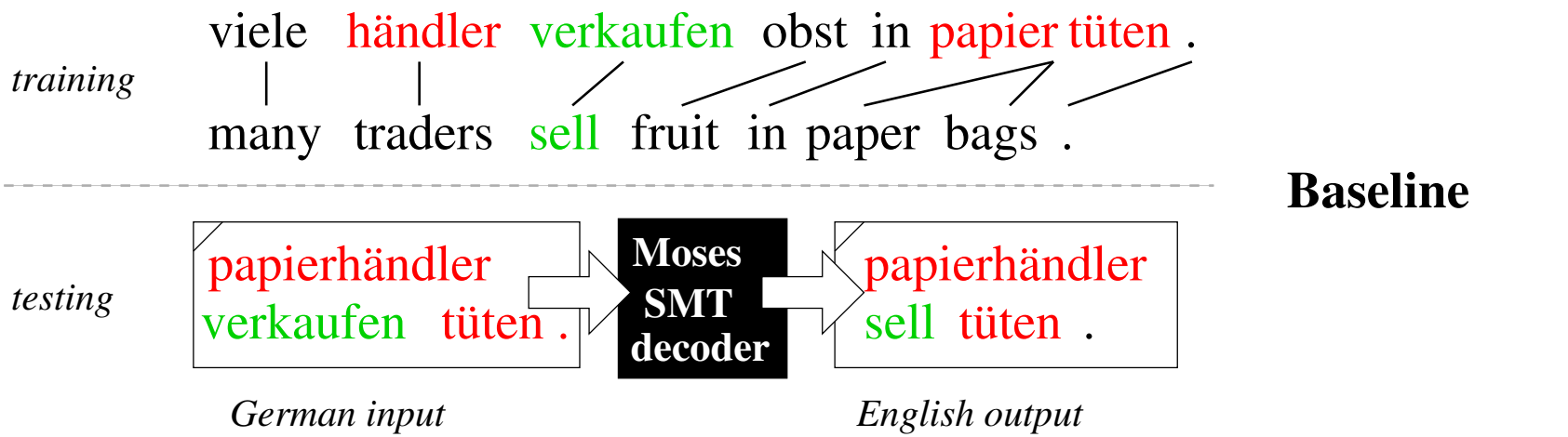


Baseline

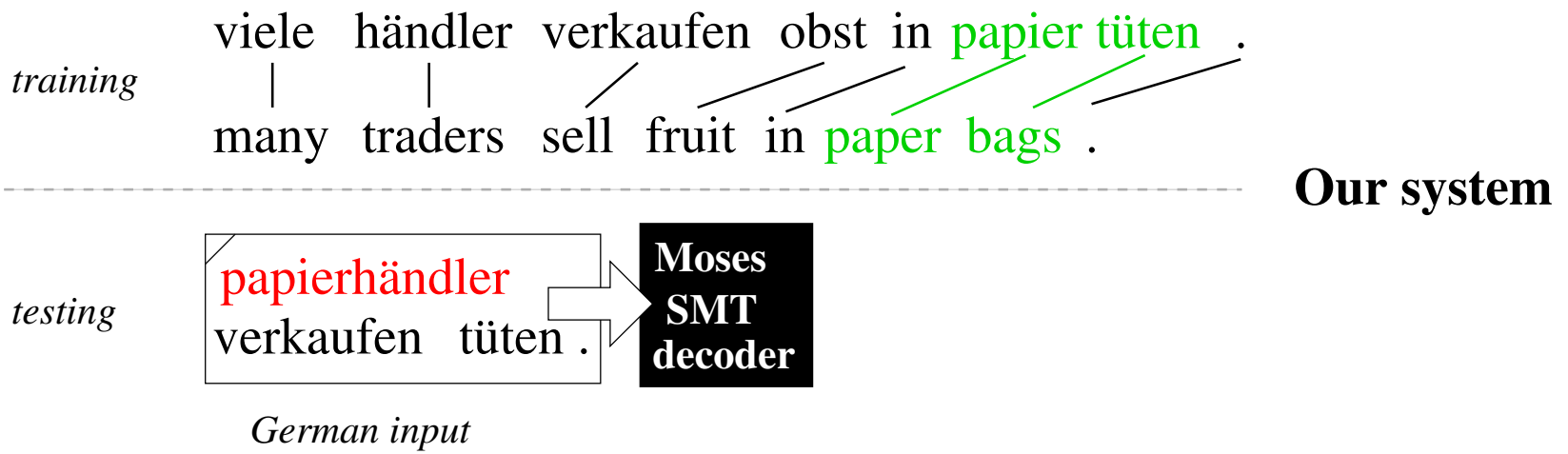


Our system

German to English SMT Example

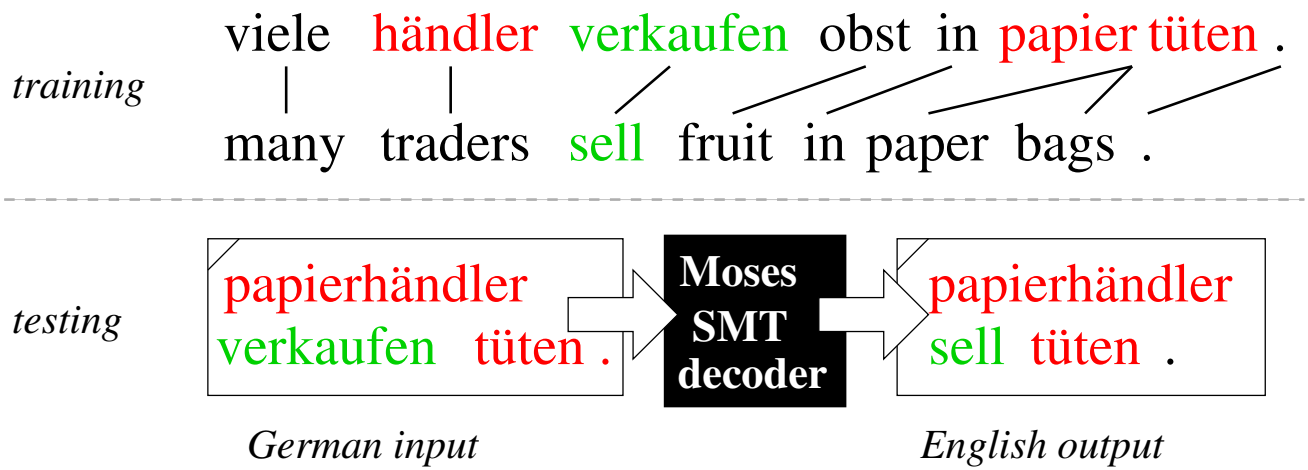


Baseline

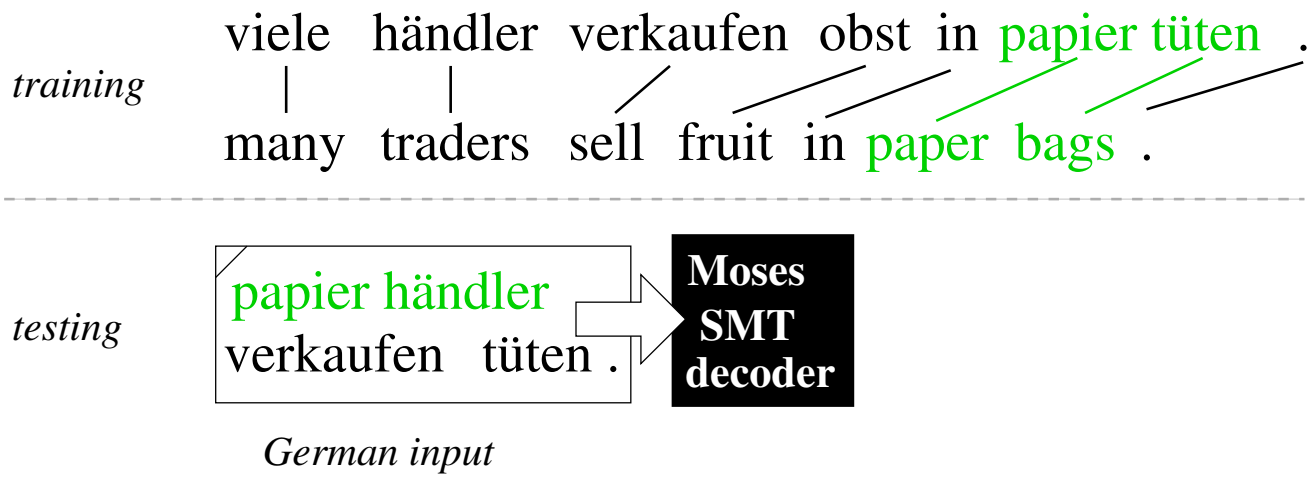


Our system

German to English SMT Example

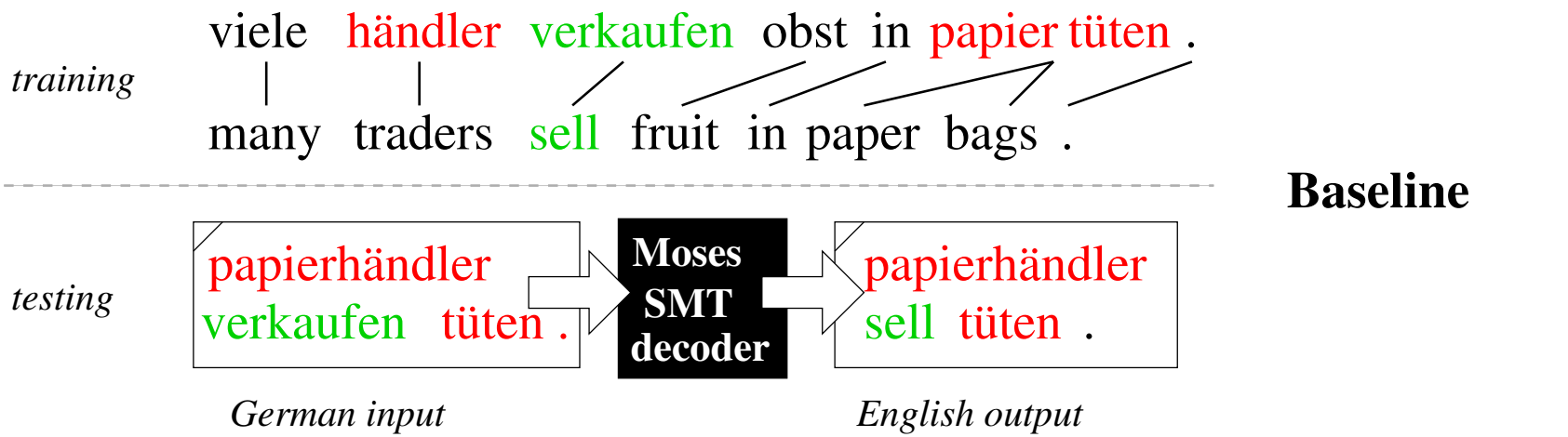


Baseline

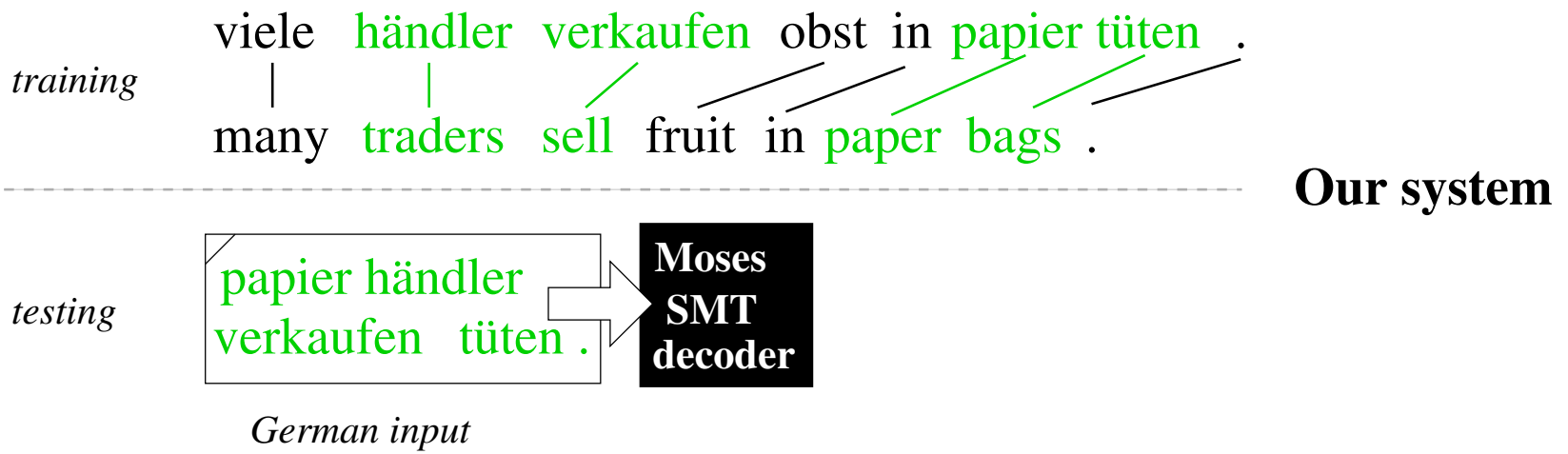


Our system

German to English SMT Example

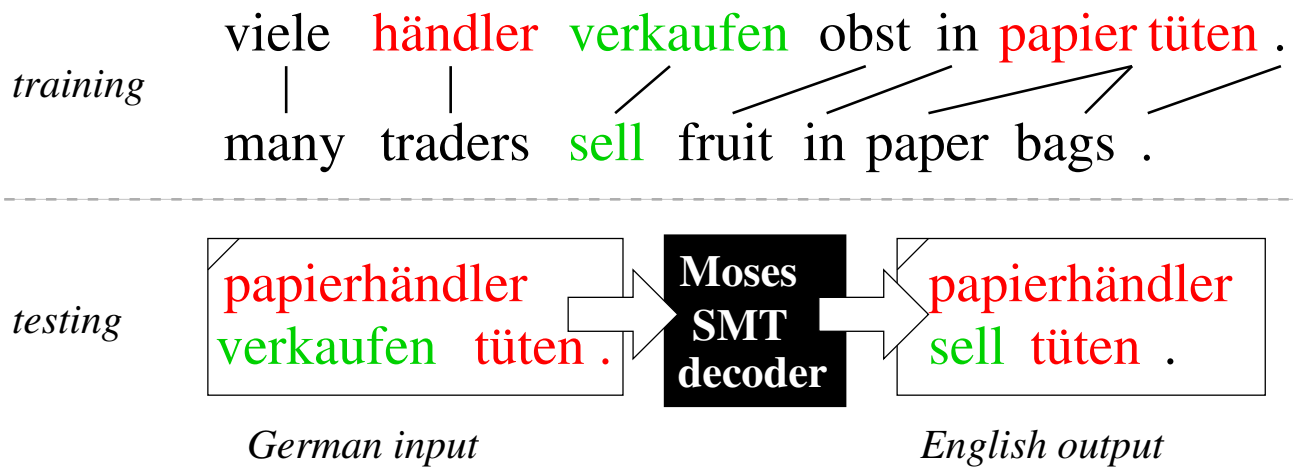


Baseline

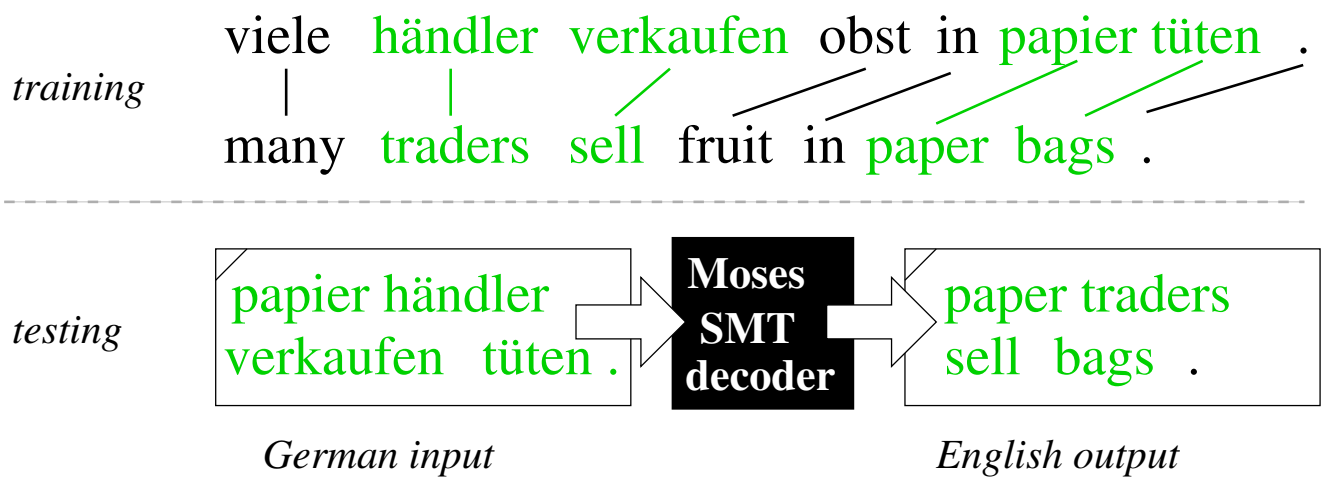


Our system

German to English SMT Example



Baseline

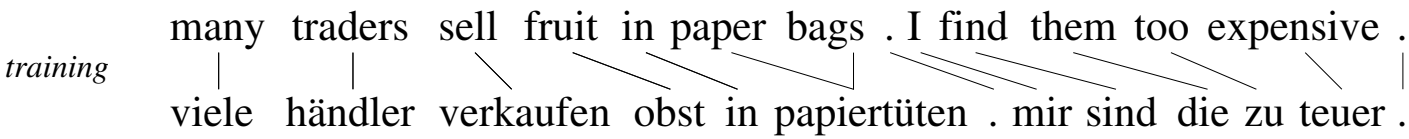


Our system

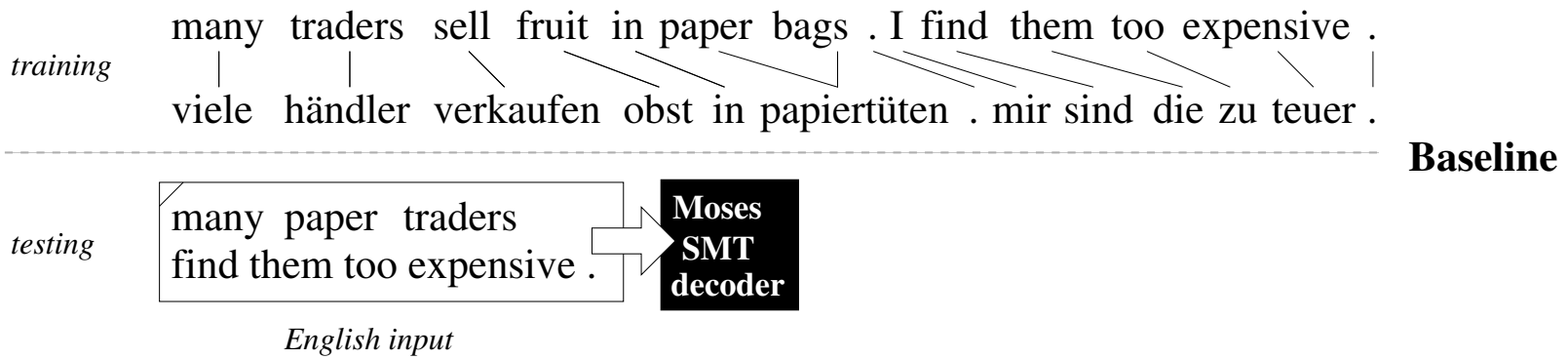
Now: opposite translation direction!!!

Pay Attention You Must!!

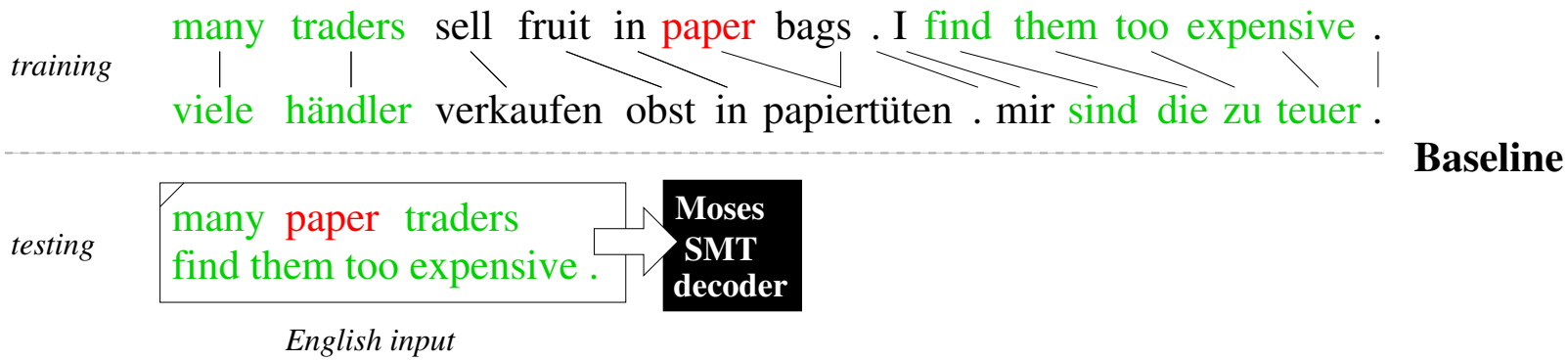
English to German SMT Example



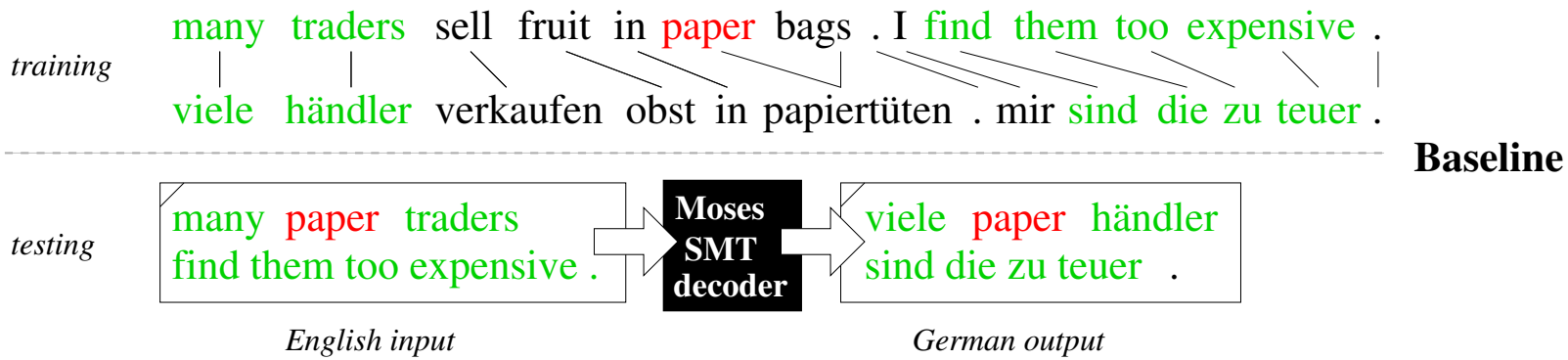
English to German SMT Example



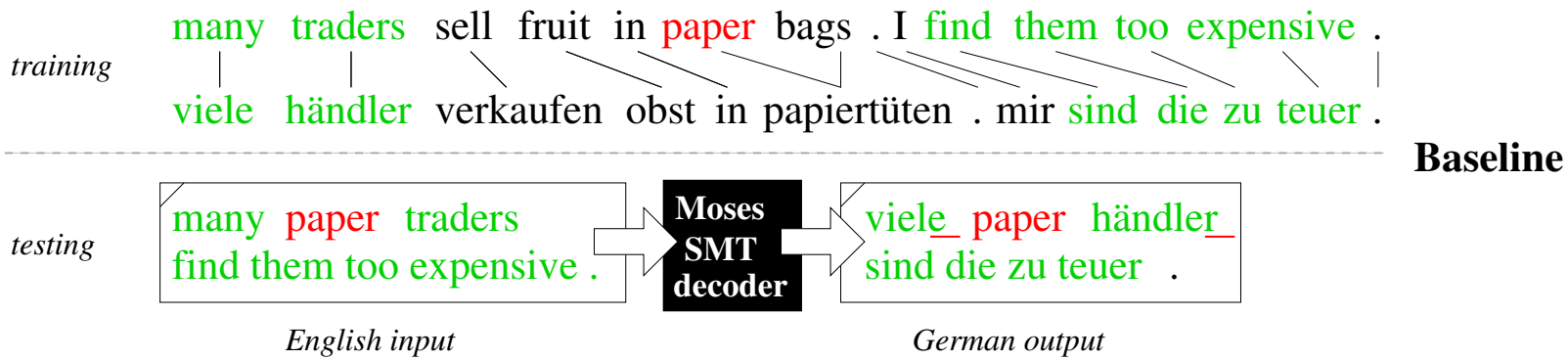
English to German SMT Example



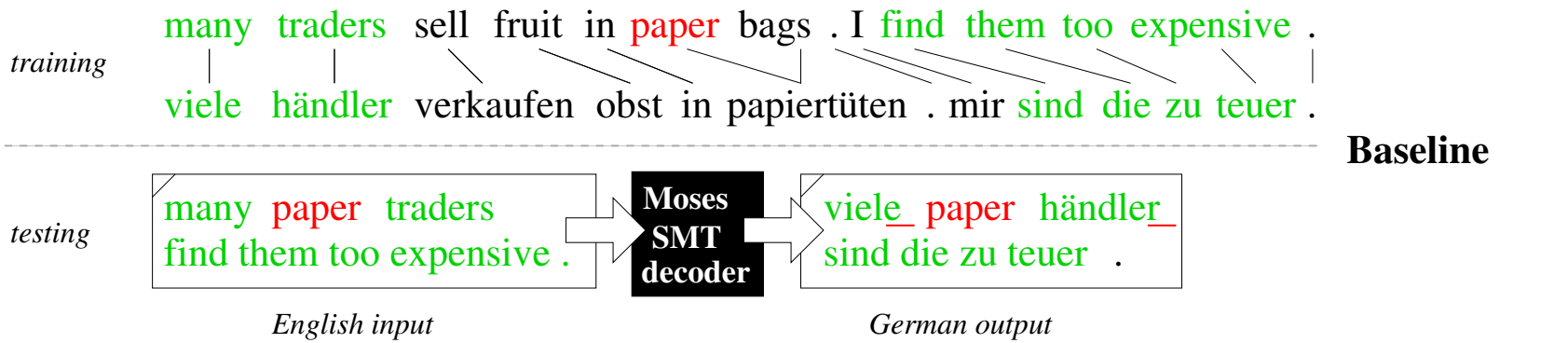
English to German SMT Example

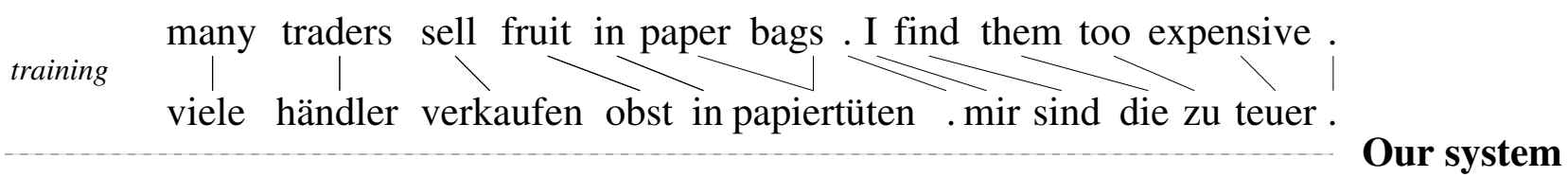


English to German SMT Example

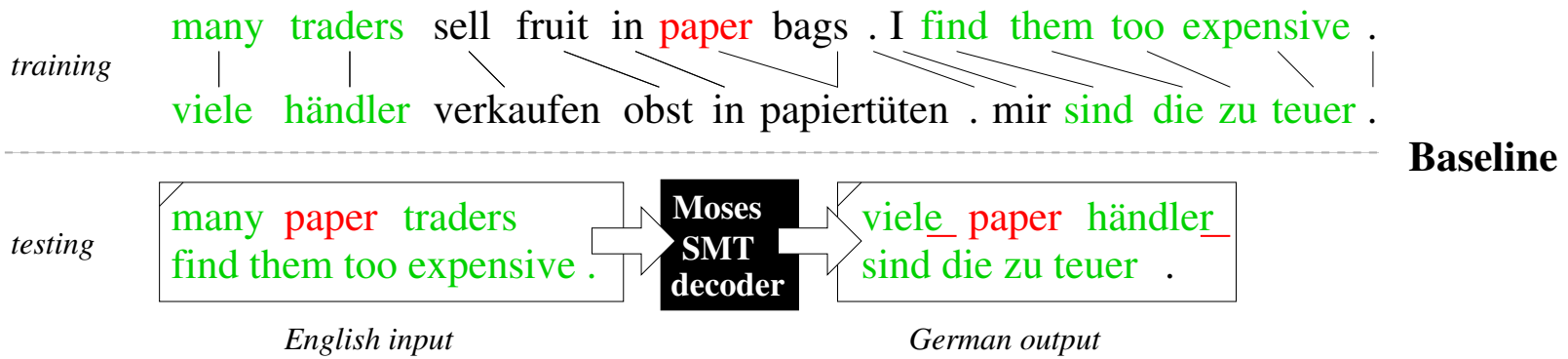


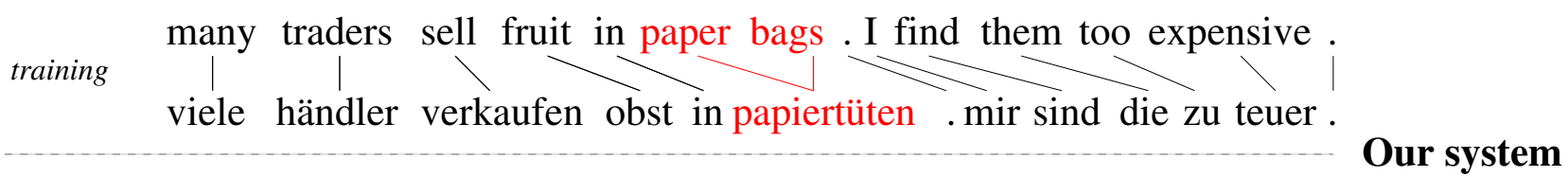
English to German SMT Example



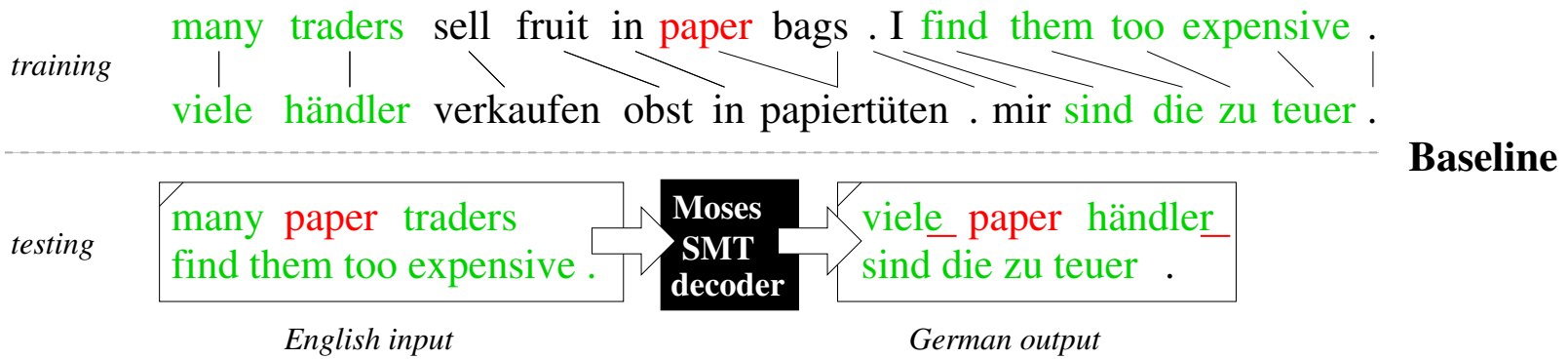


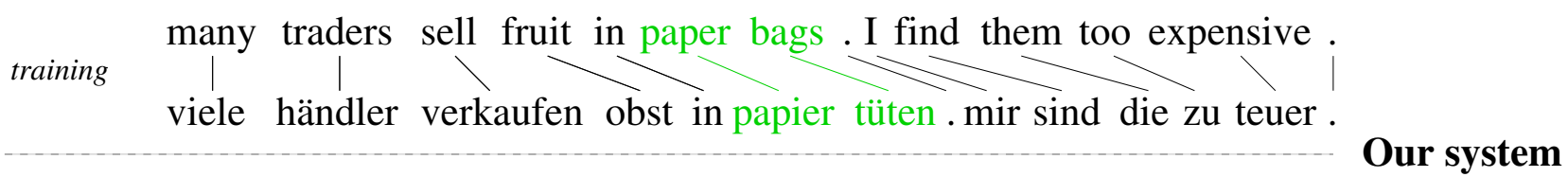
English to German SMT Example



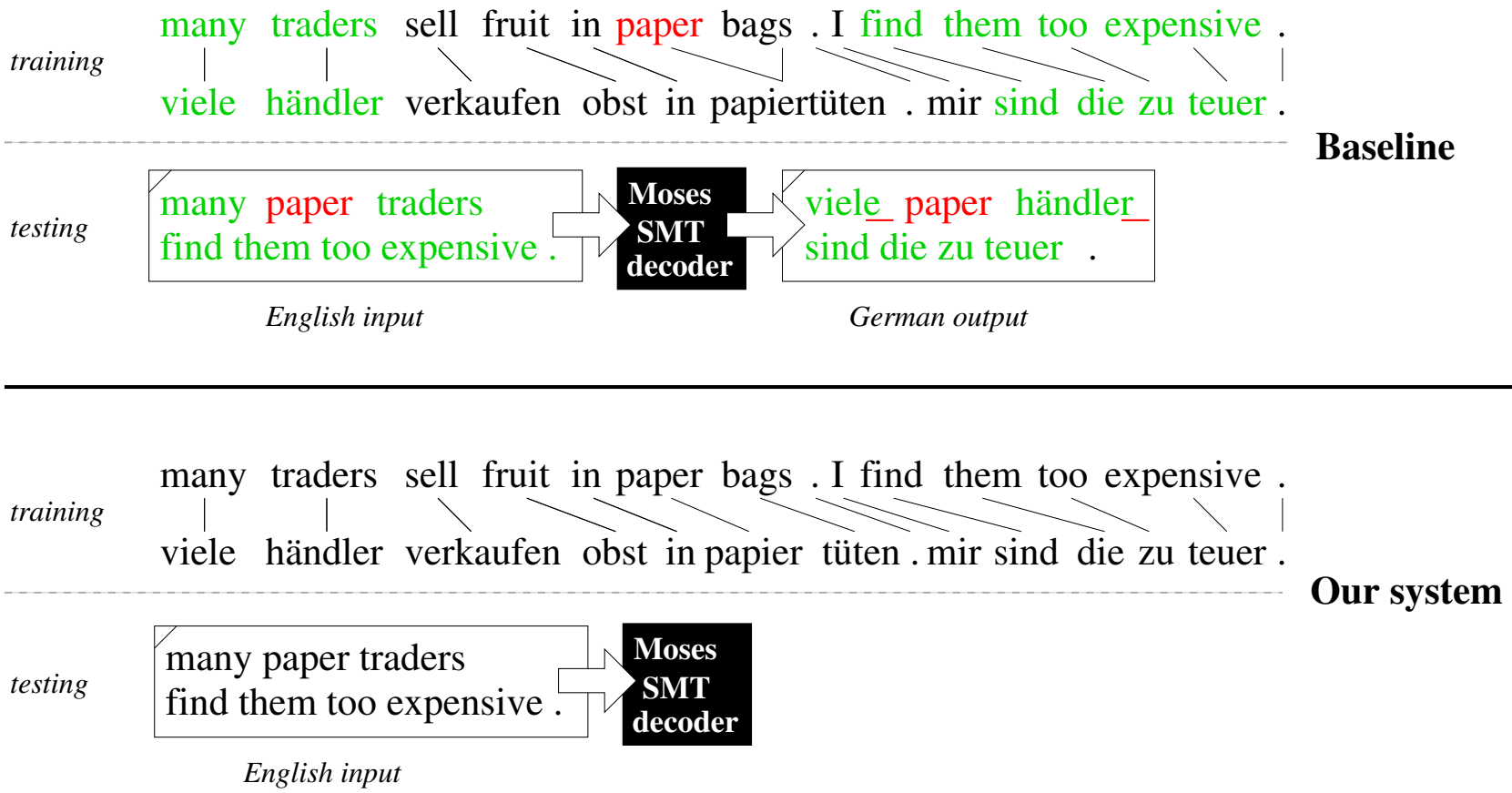


English to German SMT Example

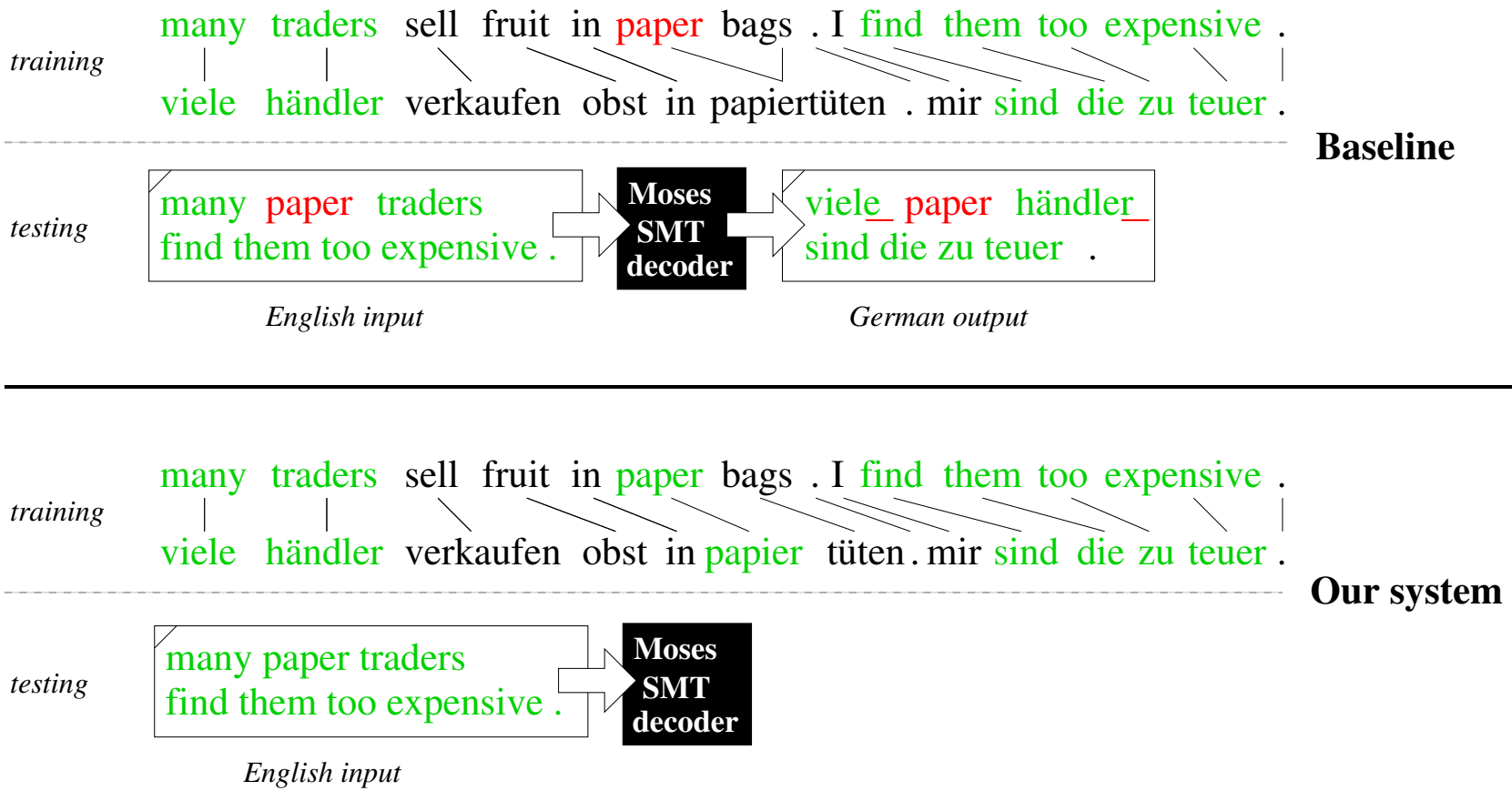




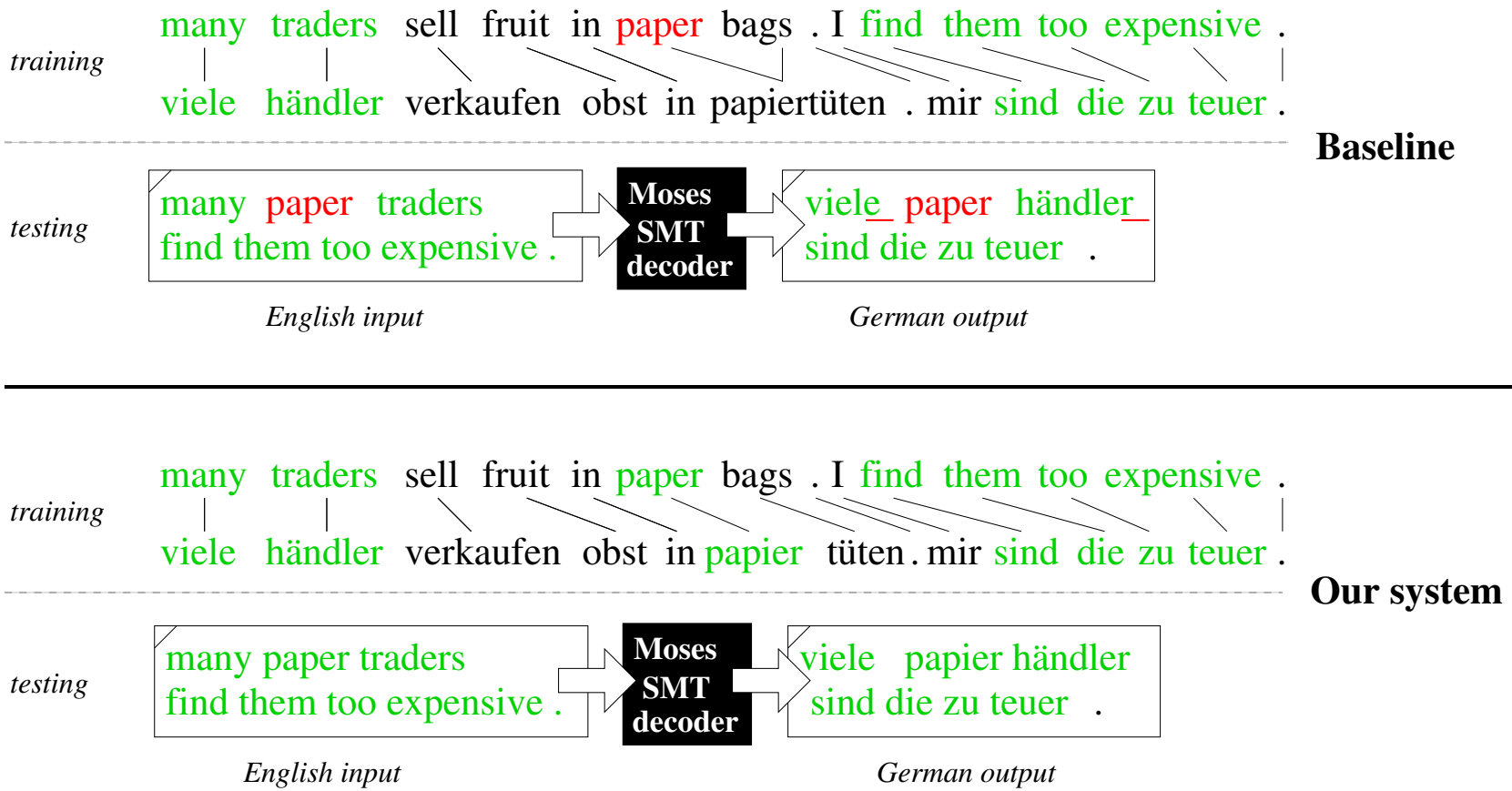
English to German SMT Example



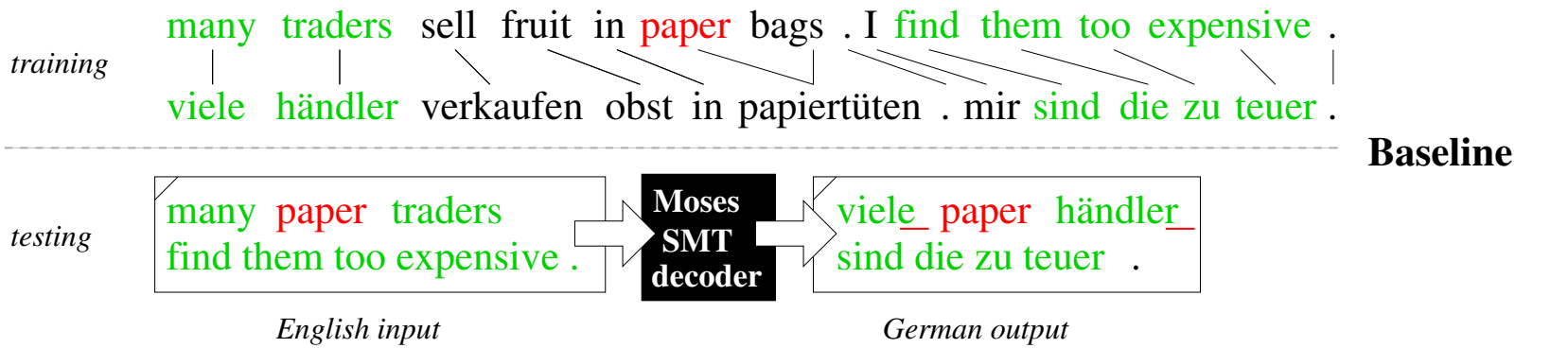
English to German SMT Example

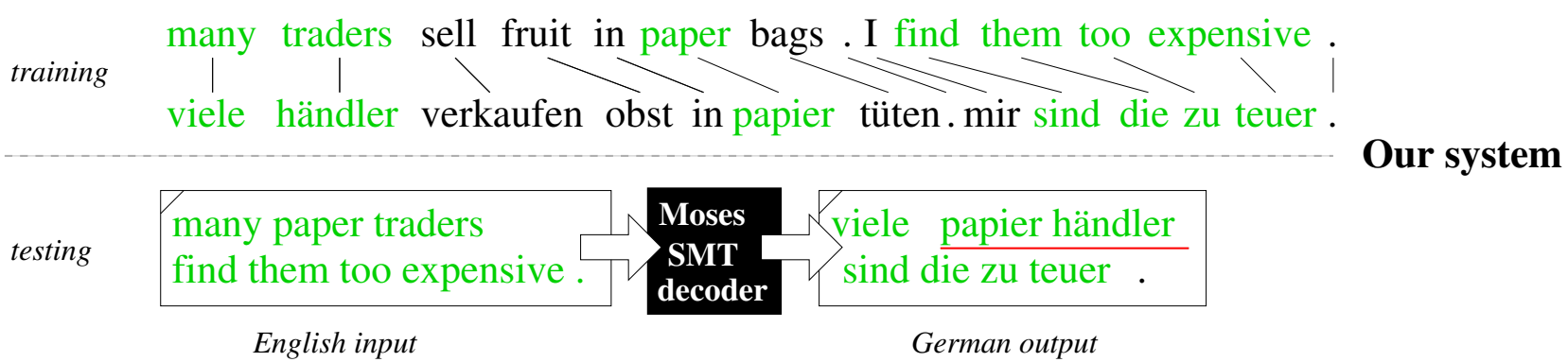


English to German SMT Example

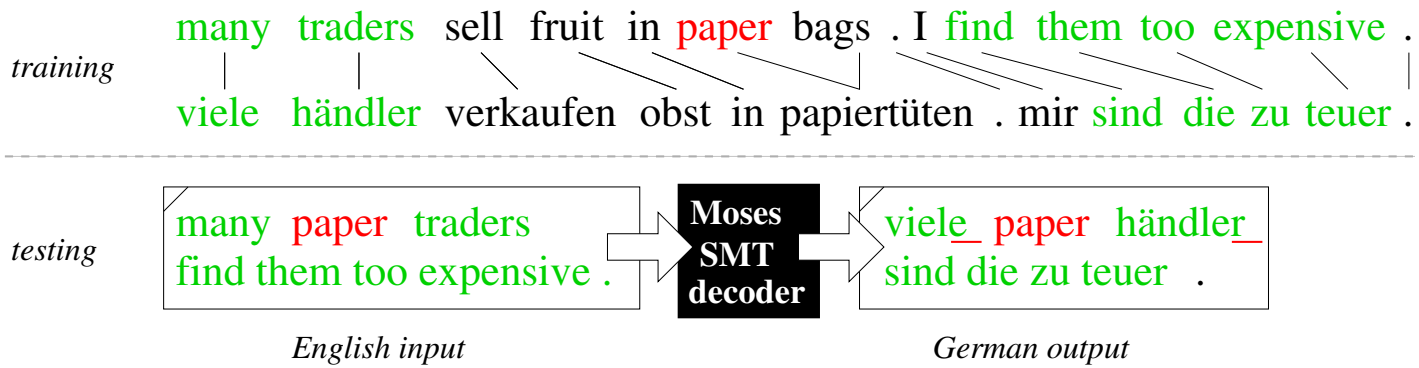


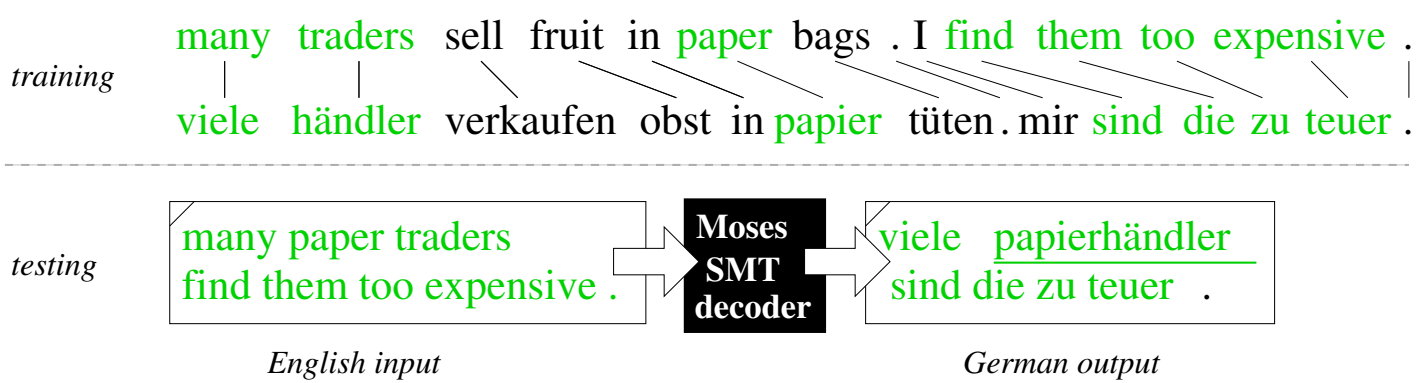
English to German SMT Example



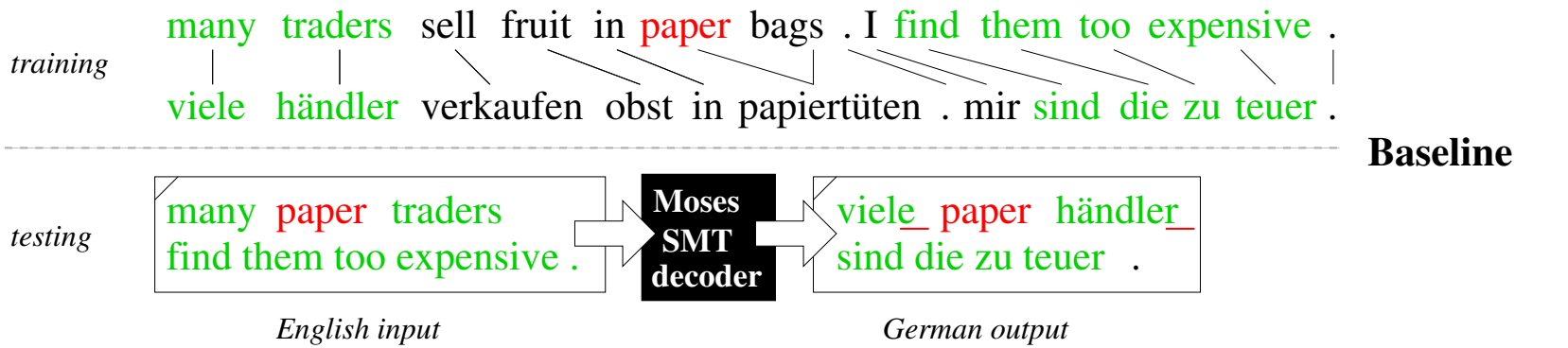


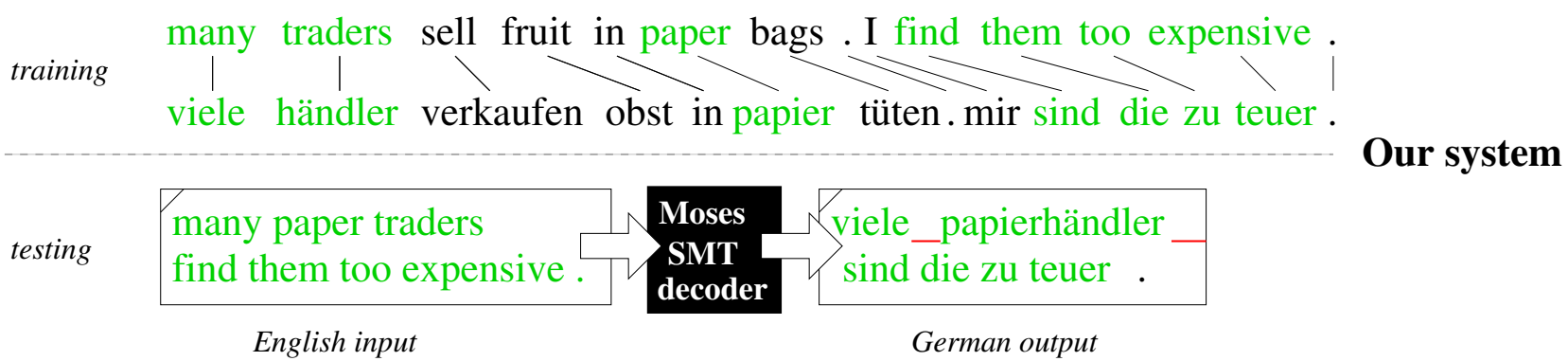
English to German SMT Example



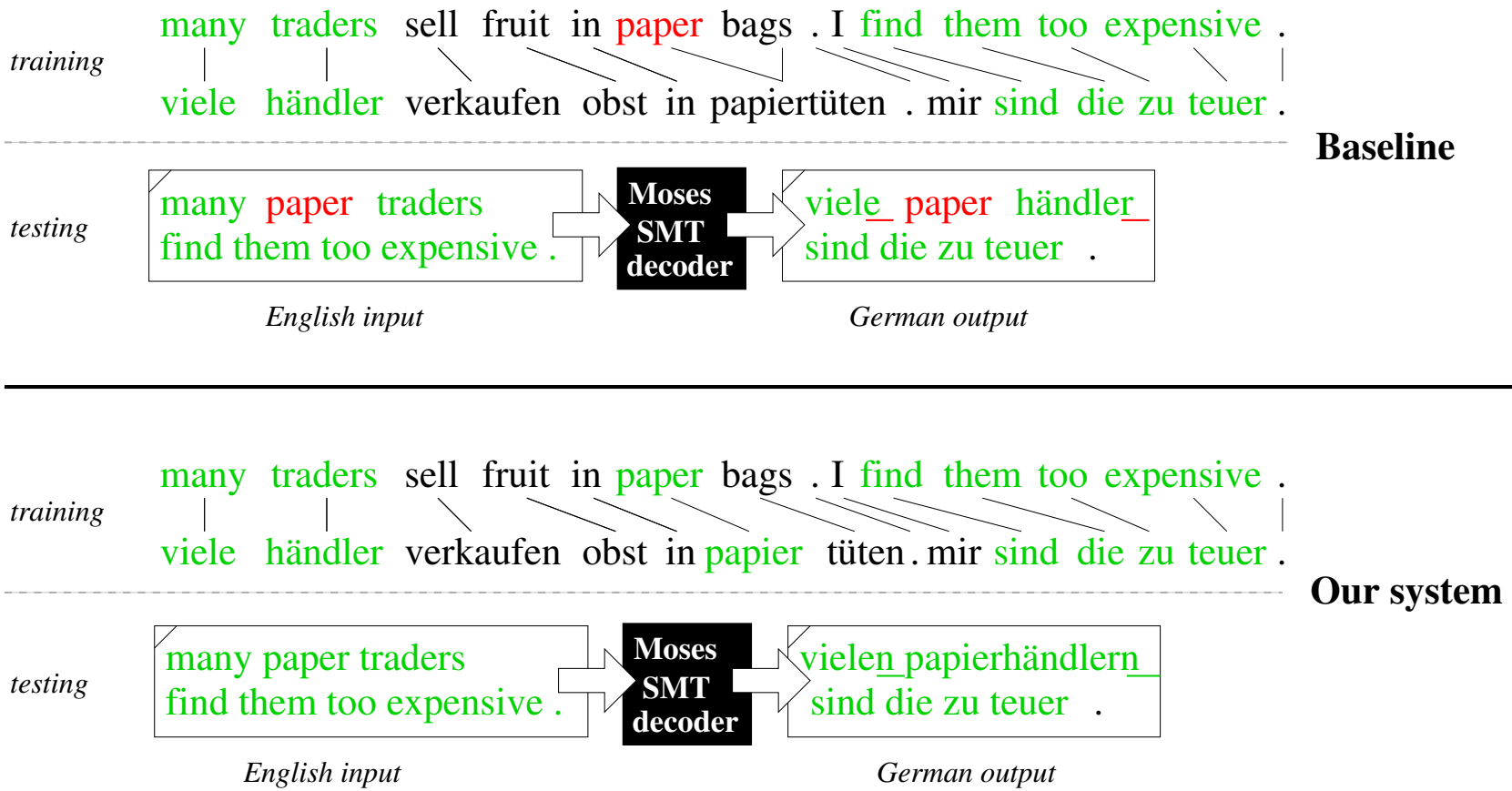


English to German SMT Example





English to German SMT Example



German to English SMT Example

Morphological Processing....

- allows to translate compounds that have not occurred in the training data:
  - provided that they have been properly split
  - their parts must have occurred in the training data
  - it is irrelevant how the parts occurred:
    as simplex words, compound modifiers or heads
- enhances the word counts of simplex words and thus makes their translations more reliable as well
- can produce unseen inflectional variants of seen words
- can produce coherent inflected sequences of words

Goals for Today

Why Morphological Processing?

Morphological Processing in SMT

A closer look at Compound Merging

Goals for Today

Why Morphological Processing?

Morphological Processing in SMT

A closer look at Compound Merging

Compound Merging Approaches

1) List-based approach

- store compounds and their parts after splitting
- only words that are on this list are merged into compounds

- POS-markup for compound modifiers:
  Inflations|N-Part + Rate|N = Inflationsrate|N
  restricts the POS of candidate heads for merging
- use CRFs for the merging decision

- use a rule-based morphological analyser for analysis and generation of compounds
- use CRFs for the merging decision

Compound Merging Approaches

1) List-based approach

- store compounds and their parts after splitting
- only words that are on this list are merged into compounds
- POS-markup for compound modifiers:
  Inflations|N-Part + Rate|N = Inflationsrate|N
  restricts the POS of candidate heads for merging
- use CRFs for the merging decision
- use a rule-based morphological analyser for analysis and generation of compounds
- use CRFs for the merging decision

Compound Merging Approaches

1) List-based approach

- store compounds and their parts after splitting
- only words that are on this list are merged into compounds

- POS-markup for compound modifiers:
  Inflations|N-Part + Rate|N = Inflationsrate|N
  restricts the POS of candidate heads for merging
- use CRFs for the merging decision

- use a rule-based morphological analyser for analysis and generation of compounds
- use CRFs for the merging decision

Compound Merging Approaches

1) List-based approach

- store compounds and their parts after splitting
- only words that are on this list are merged into compounds

2) POS-based approach

- POS-markup for compound modifiers:
  Inflations|N-Part + Rate|N = Inflationsrate|N
  restricts the POS of candidate heads for merging
- use CRFs for the merging decision

- use a rule-based morphological analyser for analysis and generation of compounds
- use CRFs for the merging decision

Compound Merging Approaches

1) List-based approach

- store compounds and their parts after splitting
- only words that are on this list are merged into compounds

2) POS-based approach

- POS-markup for compound modifiers:
  Inflations|N-Part + Rate|N = Inflationsrate|N
  restricts the POS of candidate heads for merging
- use CRFs for the merging decision

- use a rule-based morphological analyser for analysis and generation of compounds
- use CRFs for the merging decision

Compound Merging Approaches

1) List-based approach

- store compounds and their parts after splitting
- only words that are on this list are merged into compounds

2) POS-based approach

- POS-markup for compound modifiers:
  Inflations|N-Part + Rate|N = Inflationsrate|N
  restricts the POS of candidate heads for merging
- use CRFs for the merging decision

- use a rule-based morphological analyser for analysis and generation of compounds
- use CRFs for the merging decision

Compound Merging Approaches

1) List-based approach

- store compounds and their parts after splitting
- only words that are on this list are merged into compounds

2) POS-based approach

- POS-markup for compound modifiers:
  Inflations|N-Part + Rate|N = Inflationsrate|N
  restricts the POS of candidate heads for merging
- use CRFs for the merging decision

3) Morphological approach

- use a rule-based morphological analyser for analysis and generation of compounds
- use CRFs for the merging decision

Compound Merging Approaches

1) List-based approach

- store compounds and their parts after splitting
- only words that are on this list are merged into compounds

2) POS-based approach

- POS-markup for compound modifiers:
  Inflations|N-Part + Rate|N = Inflationsrate|N
  restricts the POS of candidate heads for merging
- use CRFs for the merging decision

3) Morphological approach

- use a rule-based morphological analyser for analysis and generation of compounds
- use CRFs for the merging decision

Compound Merging Approaches

1) List-based approach

- store compounds and their parts after splitting
- only words that are on this list are merged into compounds

2) POS-based approach

- POS-markup for compound modifiers:
  Inflations|N-Part + Rate|N = Inflationsrate|N
  restricts the POS of candidate heads for merging
- use CRFs for the merging decision

3) Morphological approach

- use a rule-based morphological analyser for analysis and generation of compounds
- use CRFs for the merging decision

Compound Merging for English to German SMT

Compound merging is a challenging task:

- not all two consecutive words that **could** be merged, **should** be merged:
- **solution:** use linear chain **C**onditional **R**andom **F**ields (CRFs).
  - machine learning technique
  - learn context-dependent merging decisions
  - features can be derived from the target and/or the source language

Compound Merging for English to German SMT

Compound merging is a challenging task:

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

- **solution:** use linear chain **C**onditional **R**andom **F**ields (CRFs).
  - machine learning technique
  - learn context-dependent merging decisions
  - features can be derived from the target and/or the source language

Compound Merging for English to German SMT

Compound merging is a challenging task:

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

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

- **solution:** use linear chain **C**onditional **R**andom **F**ields (CRFs).
  - machine learning technique
  - learn context-dependent merging decisions
  - features can be derived from the target and/or the source language

Compound Merging for English to German SMT

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
  - features can be derived from the target and/or the source language

Compound Merging for English to German SMT

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**
  - features can be derived from the **target** and/or the **source language**

Compound Merging for English to German SMT

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**
  - features can be derived from the **target** and/or the **source language**

Compound Merging for English to German SMT

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**
  - features can be derived from the **target** and/or the **source language**

Compound Merging for English to German SMT

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**

Compound Merging for English to German SMT

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**

Compound Merging for English to German SMT

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

Compound Merging for English to German SMT

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

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

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, I count the number of different head types

Compound Merging for English to German SMT

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, I count the number of different head types

Compound Merging for English to German SMT

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, I 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

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

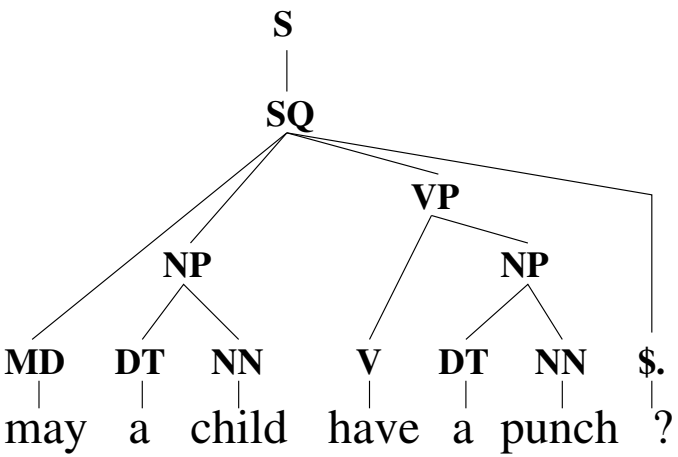
Compound Merging for English to German SMT

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

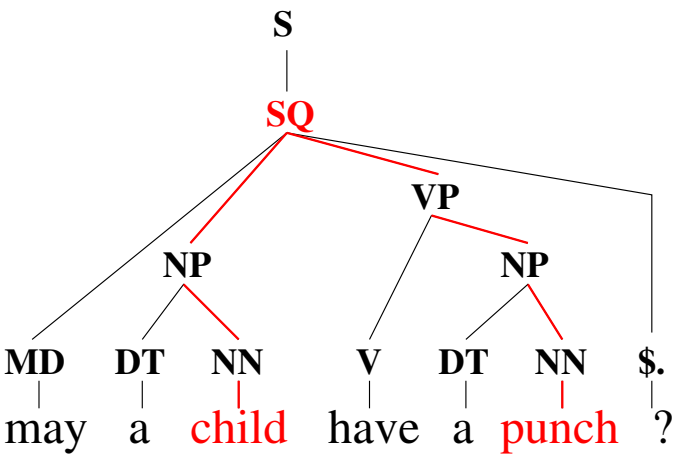
In contrast, the source sentence is **fluent** language, and sometimes, the English **source sentence structure** may help the decision:



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

Compound Merging for English to German SMT

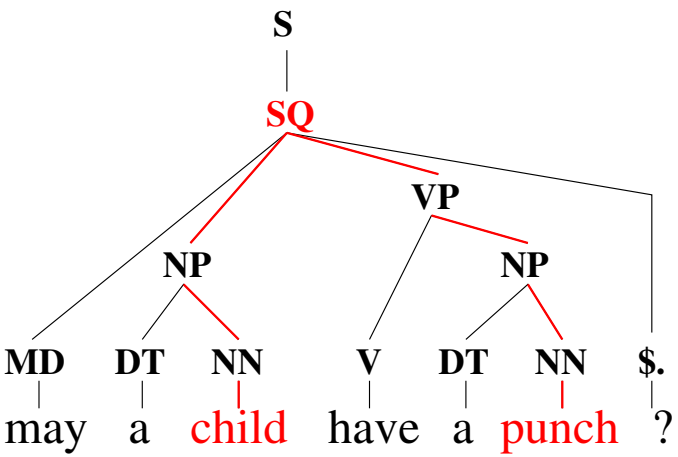
In contrast, the source sentence is **fluent** language, and sometimes, the English **source sentence structure** may help the decision:



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

Compound Merging for English to German SMT

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

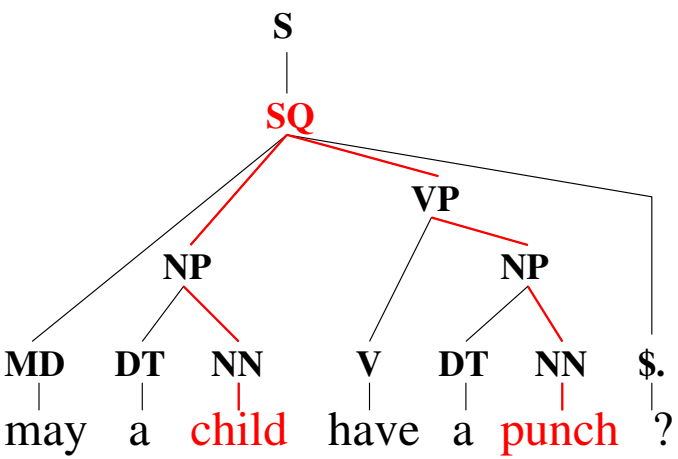


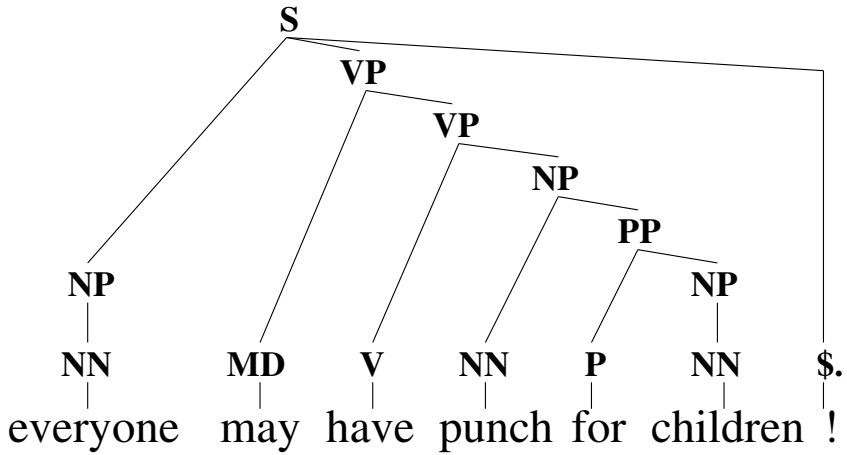
should **not** be merged:
darf ein **kind punsch** trinken?

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

Compound Merging for English to German SMT

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



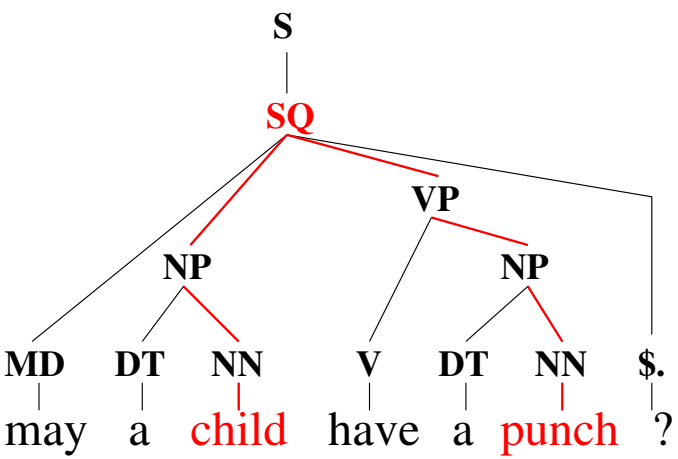


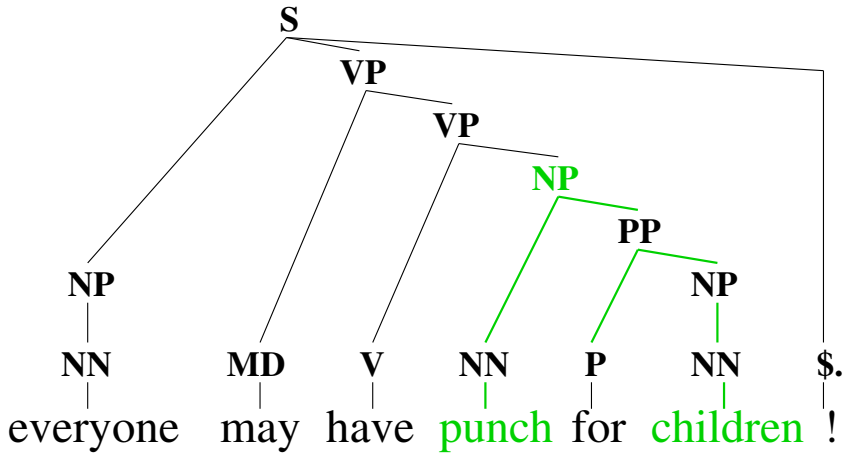
should **not** be merged:
darf ein **kind punsch** trinken?

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

Compound Merging for English to German SMT

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

CRF Training: learn binary merging decisions

German					English	MERGE?
word	POS	MOD	HEAD	PROD	EN:NP	
darf	VM	0	0	0	0	0
ein	DET	0	0	0	0	0
kind	NN	16,126	1,195	1,824	0	0
punsch	NN	2	13	2	0	0
trinken	VV	0	0	0	0	0
?	?	0	0	0	0	0
jeder	PRO	0	0	0	0	0
darf	VM	0	0	0	0	0
kind	NN	16,126	1,195	1,824	1	1
punsch	NN	2	13	2	0	0
haben	VV	0	0	0	0	0
!	!	0	0	0	0	0

Compound Merging for English to German SMT

CRF Training: learn binary merging decisions

German					English	MERGE?
word	**POS**	**MOD**	**HEAD**	**PROD**	**EN:NP**	
darf	VM	0	0	0	0	0
ein	DET	0	0	0	0	0
kind	NN	16,126	1,195	1,824	0	0
punsch	NN	2	13	2	0	0
trinken	VV	0	0	0	0	0
?	?	0	0	0	0	0
jeder	PRO	0	0	0	0	0
darf	VM	0	0	0	0	0
kind	NN	16,126	1,195	1,824	1	1
punsch	NN	2	13	2	0	0
haben	VV	0	0	0	0	0
!	!	0	0	0	0	0

In the **training** data...

→ "Kind" occurred more often as a **modifier** than as a **head**

Compound Merging for English to German SMT

CRF Training: learn binary merging decisions

German					English	MERGE?
word	**POS**	**MOD**	**HEAD**	**PROD**	**EN:NP**	
darf	VM	0	0	0	0	0
ein	DET	0	0	0	0	0
kind	NN	16,126	1,195	1,824	0	0
punsch	NN	2	13	2	0	0
trinken	VV	0	0	0	0	0
?	?	0	0	0	0	0
jeder	PRO	0	0	0	0	0
darf	VM	0	0	0	0	0
kind	NN	16,126	1,195	1,824	1	1
punsch	NN	2	13	2	0	0
haben	VV	0	0	0	0	0
!	!	0	0	0	0	0

In the **training** data...

→ "Kind" occurred more often as a **modifier** than as a **head**

→ the **opposite** applies to "Punsch"!

Compound Merging for English to German SMT

CRF Training: learn binary merging decisions

German					English	MERGE?
word	POS	MOD	HEAD	PROD	EN:NP	
darf	VM	0	0	0	0	0
ein	DET	0	0	0	0	0
kind	NN	16,126	1,195	1,824	0	0
punsch	NN	2	13	2	0	0
trinken	VV	0	0	0	0	0
?	?	0	0	0	0	0
jeder	PRO	0	0	0	0	0
darf	VM	0	0	0	0	0
kind	NN	16,126	1,195	1,824	1	1
punsch	NN	2	13	2	0	0
haben	VV	0	0	0	0	0
!	!	0	0	0	0	0

English feature determines merging decision!

To thank you I want

Where are we?

Type	Date	Time	Place	Topic	Reading / Assignments
F	2016-03-30	10-12	6-K1031	Introduction (SS)	Koehn 1; JM 25.1-2; Hutchins; CFMF
F	2016-03-30	14-16	6-K1031	MT evaluation (SS)	Koehn 8; JM 25.9
F	2016-04-04	10-12	2-0076	MT in practice (Convertus) - guest lecture	
L	2016-04-06	10-12	Chomsky	MT in practice (AS)	lab report 1
F	2016-04-11	10-12	2-0076	Introduction to SMT (FC)	Koehn Ch 4, Ch 7, KK97
L	2016-04-13	10-12	Chomsky	Word-based SMT (SS)	lab report 2
L	2016-04-18	10-12	Chomsky	Word-based SMT (SS)	lab report 2
F	2016-04-18	14-16	2-1077	Machine translation at Semantix, a translation provider - guest lecture	
F	2016-04-20	10-12	6-K1031	Parallel Corpora, Alignment (AS)	Koehn 2-4, JT 3-4, KK97, KK99
L	2016-04-25	10-12	Chomsky	Parallel corpora & alignment (AS)	lab report 3
F	2016-04-27	10-12	2-0076	Phrase-based SMT (FC)	Koehn Ch 5
L	2016-05-02	10-12	Chomsky	Phrase-based SMT (AS)	lab report 4
F	2016-05-04	10-12	6-K1031	Decoding (CH)	Koehn Ch 6
L	2016-05-09	10-12	Chomsky	Phrase-based SMT (AS)	lab report 4
F	2016-05-11	10-12	2-0076	Tree-based SMT & MT for morphologically rich languages (SS, FC)	Koehn 10.2, 11
F	2016-05-16	10-12	2-0076	Document-wide decoding & Neural MT (CH)	
L	2016-05-18	10-12	Chomsky	Document-wide decoding lab (AS)	oral lab report 5
S	2016-05-23	10-12	2-0076	Seminar - master student presentations	
S	2016-05-25	10-12	6-K1031	Seminar - master student presentations	

Where are we?

Type	Date	Time	Place	Topic	Reading / Assignments
F	2016-03-30	10-12	6-K1031	Introduction (SS)	Koehn 1; JM 25.1-2; Hutchins; CFMF
F	2016-03-30	14-16	6-K1031	MT evaluation (SS)	Koehn 8; JM 25.9
F	2016-04-04	10-12	2-0076	MT in practice (Convertus) - guest lecture	
L	2016-04-06	10-12	Chomsky	MT in practice (AS)	lab report 1
F	2016-04-11	10-12	2-0076	Introduction to SMT (FC)	Koehn Ch 4, Ch 7, KK97
L	2016-04-13	10-12	Chomsky	Word-based SMT (SS)	lab report 2
L	2016-04-18	10-12	Chomsky	Word-based SMT (SS)	lab report 2
F	2016-04-18	14-16	2-1077	Machine translation at Semantix, a translation provider - guest lecture	
F	2016-04-20	10-12	6-K1031	Parallel Corpora, Alignment (AS)	Koehn 2-4, JT 3-4, KK97, KK99
L	2016-04-25	10-12	Chomsky	Parallel corpora & alignment (AS)	lab report 3
F	2016-04-27	10-12	2-0076	Phrase-based SMT (FC)	Koehn Ch 5
L	2016-05-02	10-12	Chomsky	Phrase-based SMT (AS)	lab report 4
F	2016-05-04	10-12	6-K1031	Decoding (CH)	Koehn Ch 6
L	2016-05-09	10-12	Chomsky	Phrase-based SMT (AS)	lab report 4
F	2016-05-11	10-12	2-0076	Tree-based SMT & MT for morphologically rich languages (SS, FC)	Koehn 10.2, 11
F	2016-05-16	10-12	2-0076	Document-wide decoding & Neural MT (CH)	
L	2016-05-18	10-12	Chomsky	Document-wide decoding lab (AS)	oral lab report 5
S	2016-05-23	10-12	2-0076	Seminar - master student presentations	
S	2016-05-25	10-12	6-K1031	Seminar - master student presentations	