Machine Translation

Jordan Boyd-Graber

University of Maryland

Phrase-Based Models

Adapted from material by Philipp Koehn

Motivation

- Word-Based Models translate words as atomic units
- Phrase-Based Models translate phrases as atomic units
- Advantages:
  - many-to-many translation can handle non-compositional phrases
  - use of local context in translation
  - the more data, the longer phrases can be learned
- “Standard Model” before neural revolution

Motivation

- Word-Based Models translate words as atomic units
- Phrase-Based Models translate phrases as atomic units
- Advantages:
  - many-to-many translation can handle non-compositional phrases
  - use of local context in translation
  - the more data, the longer phrases can be learned
- “Standard Model” before neural revolution
- Led to much of the decoding black magic of today

Anstellen

English	German

Anstellen

English	German
turn on the AC	stell die Klimaanlage an

Anstellen

English	German
turn on the AC	stell die Klimaanlage an
We want to hire more teaching faculty	Wir wollen mehr Lehrkrafte anstellen

Anstellen

English	German
turn on the AC	stell die Klimaanlage an
We want to hire more teaching faculty	Wir wollen mehr Lehrkrafte anstellen
What mess have you made this time?	Was für ein Unsinn has du jetzt wieder angestellt?

Anstellen

English	German
turn on the AC	stell die Klimaanlage an
We want to hire more teaching faculty	Wir wollen mehr Lehrkrafte anstellen
What mess have you made this time?	Was für ein Unsinn has du jetzt wieder angestellt?
put the ladder against the wall	stell die Leiter an die Mauer an

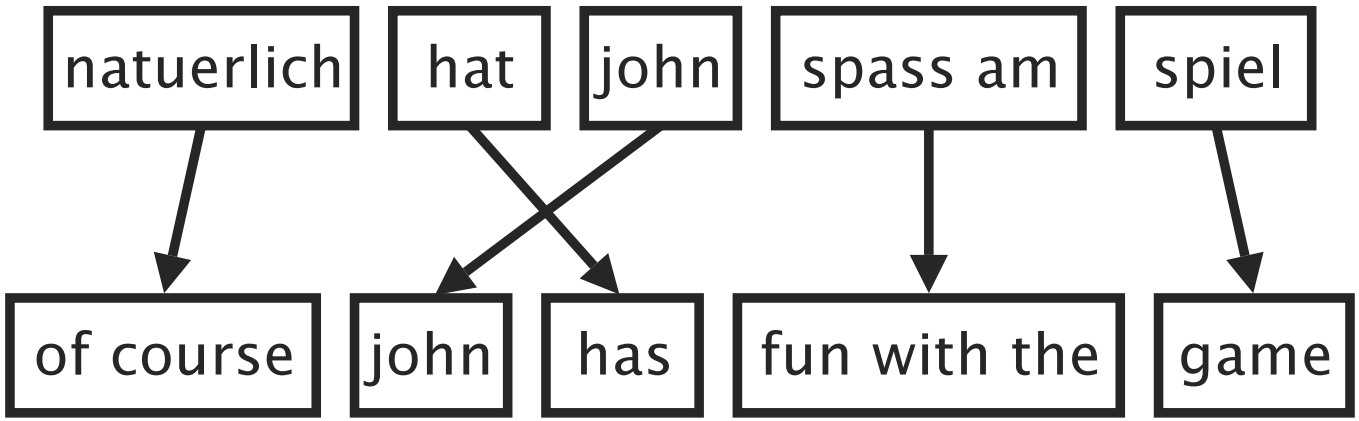
Anstellen

English	German
turn on the AC	stell die Klimaanlage an
We want to hire more teaching faculty	Wir wollen mehr Lehrkrafte anstellen
What mess have you made this time?	Was für ein Unsinn has du jetzt wieder angestellt?
put the ladder against the wall	stell die Leiter an die Mauer an

Anstellen

English	German
turn on the AC	stell die Klimaanlage an
We want to hire more teaching faculty	Wir wollen mehr Lehrkrafte anstellen
What mess have you made this time?	Was für ein Unsinn has du jetzt wieder angestellt?
put the ladder against the wall	stell die Leiter an die Mauer an

Phrase-Based Model



- Foreign input is segmented in phrases
- Each phrase is translated into English
- Phrases are reordered

Phrase Translation Table

- Main knowledge source: table with phrase translations and their probabilities
- Example: phrase translations for **natuerlich**

of course	0.5
naturally	0.3
of course ,	0.15
, of course ,	0.05

Real Example

- Phrase translations for den Vorschlag learned from the Europarl corpus:

English	$\phi(\bar{e}\|\bar{f})$	English	$\phi(\bar{e}\|\bar{f})$
the proposal	0.6227	the suggestions	0.0114
's proposal	0.1068	the proposed	0.0114
a proposal	0.0341	the motion	0.0091
the idea	0.0250	the idea of	0.0091
this proposal	0.0227	the proposal ,	0.0068
proposal	0.0205	its proposal	0.0068
of the proposal	0.0159	it	0.0068
the proposals	0.0159	...	...

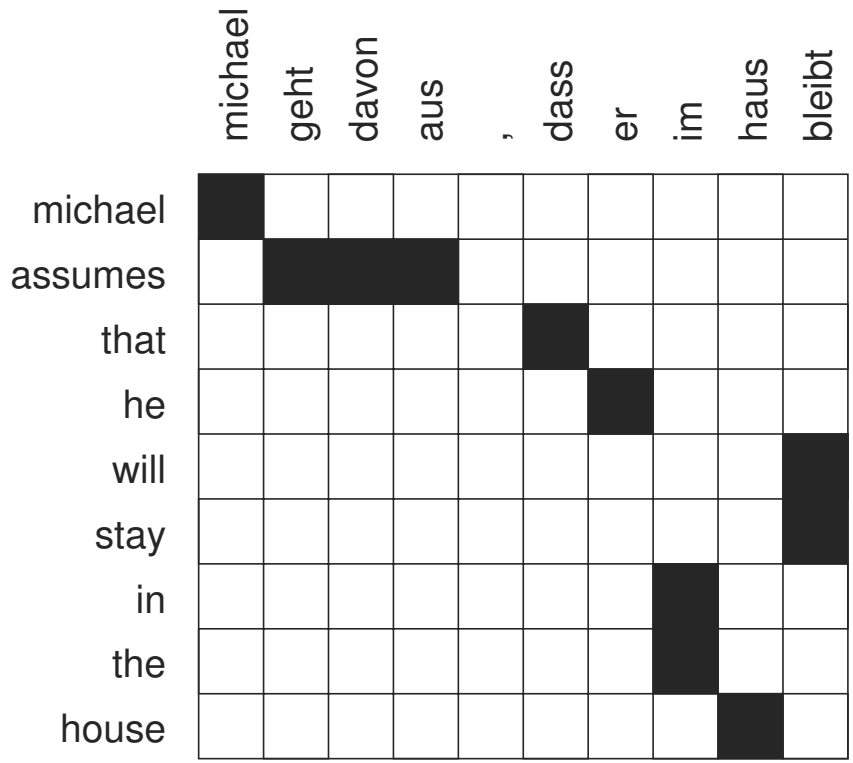
  - lexical variation (proposal vs suggestions)
  - morphological variation (proposal vs proposals)
  - included function words (the, a, ...)
  - noise (it)

Linguistic Phrases?

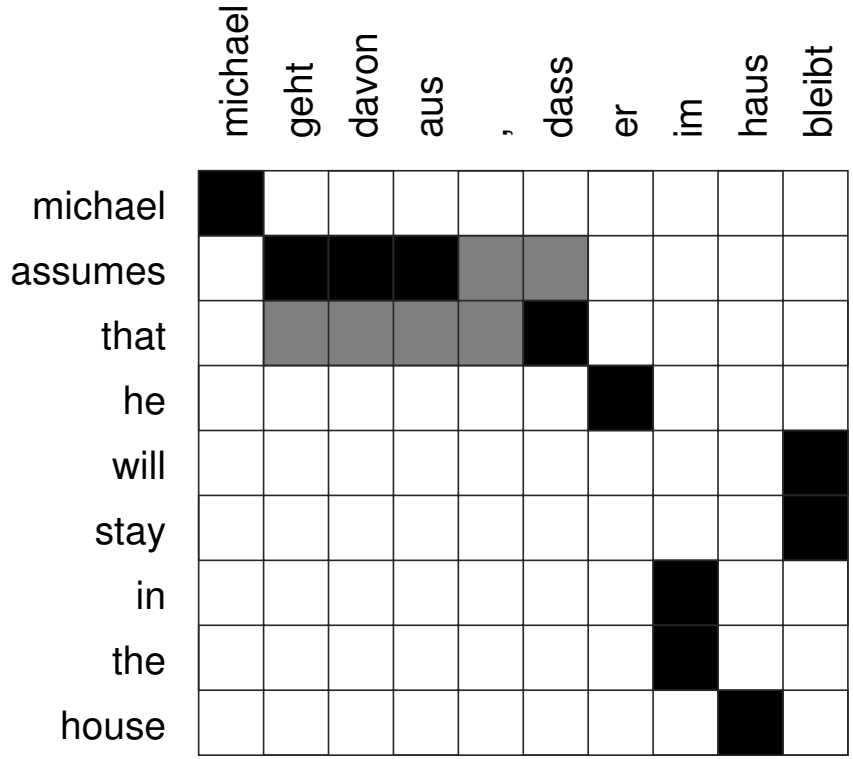
- Model is not limited to linguistic phrases
  (noun phrases, verb phrases, prepositional phrases, ...)
- Example non-linguistic phrase pair

  spass am ⟶ fun with the
- Prior noun often helps with translation of preposition
- Experiments show that limitation to linguistic phrases hurts quality

Word Alignment



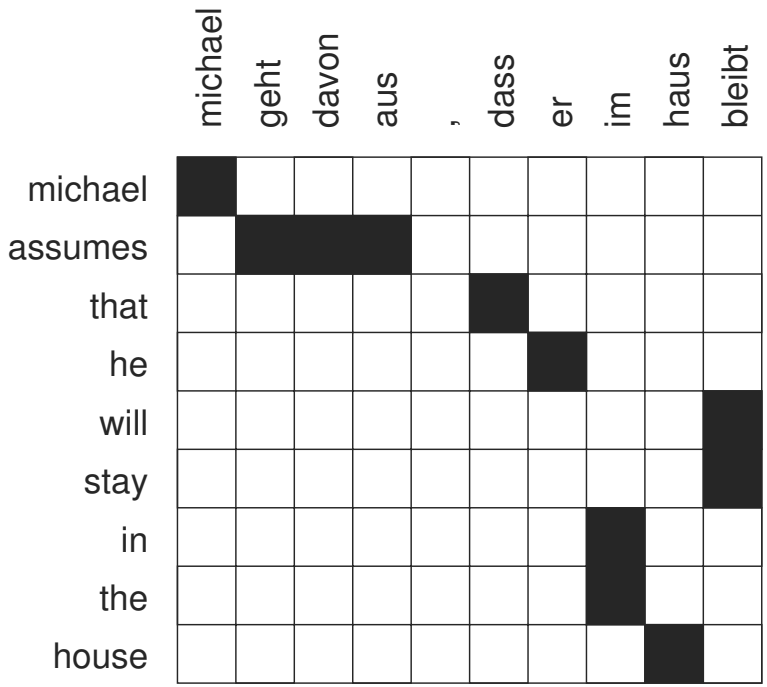
Extracting Phrase Pairs



extract phrase pair consistent with word alignment:

assumes that / geht davon aus , dass

Phrase Pair Extraction



Smallest phrase pairs:

michael — michael

assumes — geht davon aus / geht davon aus ,

that — dass / , dass

he — er

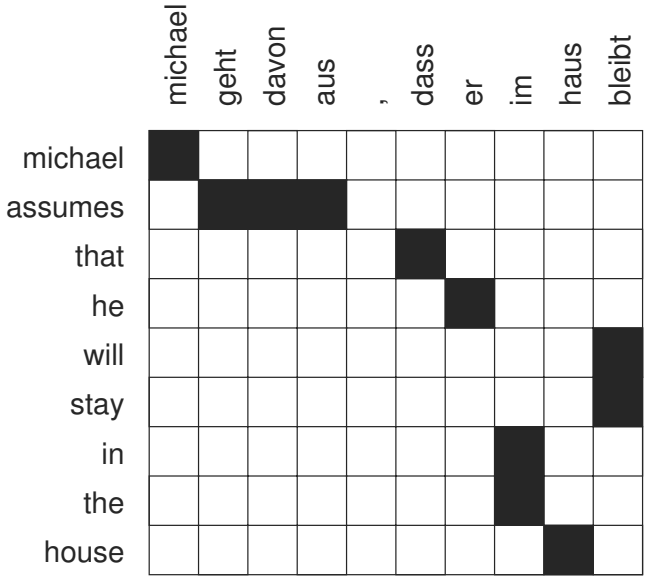
will stay — bleibt

in the — im

house — haus

unaligned words (here: German comma) lead to multiple translations

Larger Phrase Pairs



michael assumes — michael geht davon aus / michael geht davon aus ,
assumes that — geht davon aus , dass ; assumes that he — geht davon aus , dass er
that he — dass er / , dass er ; in the house — im haus
michael assumes that — michael geht davon aus , dass
michael assumes that he — michael geht davon aus , dass er
michael assumes that he will stay in the house — michael geht davon aus , dass er im haus bleibt
assumes that he will stay in the house — geht davon aus , dass er im haus bleibt
that he will stay in the house — dass er im haus bleibt ; dass er im haus bleibt ,
he will stay in the house — er im haus bleibt ; will stay in the house — im haus bleibt

Objective Function

- Bayes rule

$$\begin{aligned} \mathbf{e_{best}} &= \mathbf{argmax_e}\, p(\mathbf{e}|\mathbf{f}) \\ &= \mathbf{argmax_e}\, p(\mathbf{f}|\mathbf{e})\, p_{\mathrm{lm}}(\mathbf{e}) \end{aligned}$$

  - translation model $p(\mathbf{e}|\mathbf{f})$
  - language model $p_{\mathrm{lm}}(\mathbf{e})$

- Decomposition of the translation model

$$p(\bar{f}_1^I|\bar{e}_1^I) = \prod_{i=1}^{I} \phi(\bar{f}_i|\bar{e}_i)\, d(start_i - end_{i-1} - 1)$$

  - phrase translation probability ϕ
  - reordering probability d

Objective Function

- Bayes rule

$$\begin{aligned} \mathbf{e_{best}} &= \mathbf{argmax_e}\, p(\mathbf{e}|\mathbf{f}) \\ &= \mathbf{argmax_e}\, p(\mathbf{f}|\mathbf{e})\, p_{\mathrm{lm}}(\mathbf{e}) \end{aligned}$$

  - translation model $p(\mathbf{e}|\mathbf{f})$
  - language model $p_{\mathrm{lm}}(\mathbf{e})$

- Decomposition of the translation model

$$p(\bar{f}_1^I|\bar{e}_1^I) = \prod_{i=1}^{I} \phi(\bar{f}_i|\bar{e}_i)\, d(start_i - end_{i-1} - 1)$$

  - phrase translation probability ϕ
  - reordering probability d

Objective Function

- Bayes rule

$$\begin{aligned} \mathbf{e_{best}} &= \mathbf{argmax_e}\, p(\mathbf{e}|\mathbf{f}) \\ &= \mathbf{argmax_e}\, p(\mathbf{f}|\mathbf{e})\, p_{\mathrm{lm}}(\mathbf{e}) \end{aligned}$$

  - translation model $p(\mathbf{e}|\mathbf{f})$
  - language model $p_{\mathrm{lm}}(\mathbf{e})$

- Decomposition of the translation model

$$p(\bar{f}_1^I|\bar{e}_1^I) = \prod_{i=1}^{I} \phi(\bar{f}_i|\bar{e}_i)\, d(start_i - end_{i-1} - 1)$$

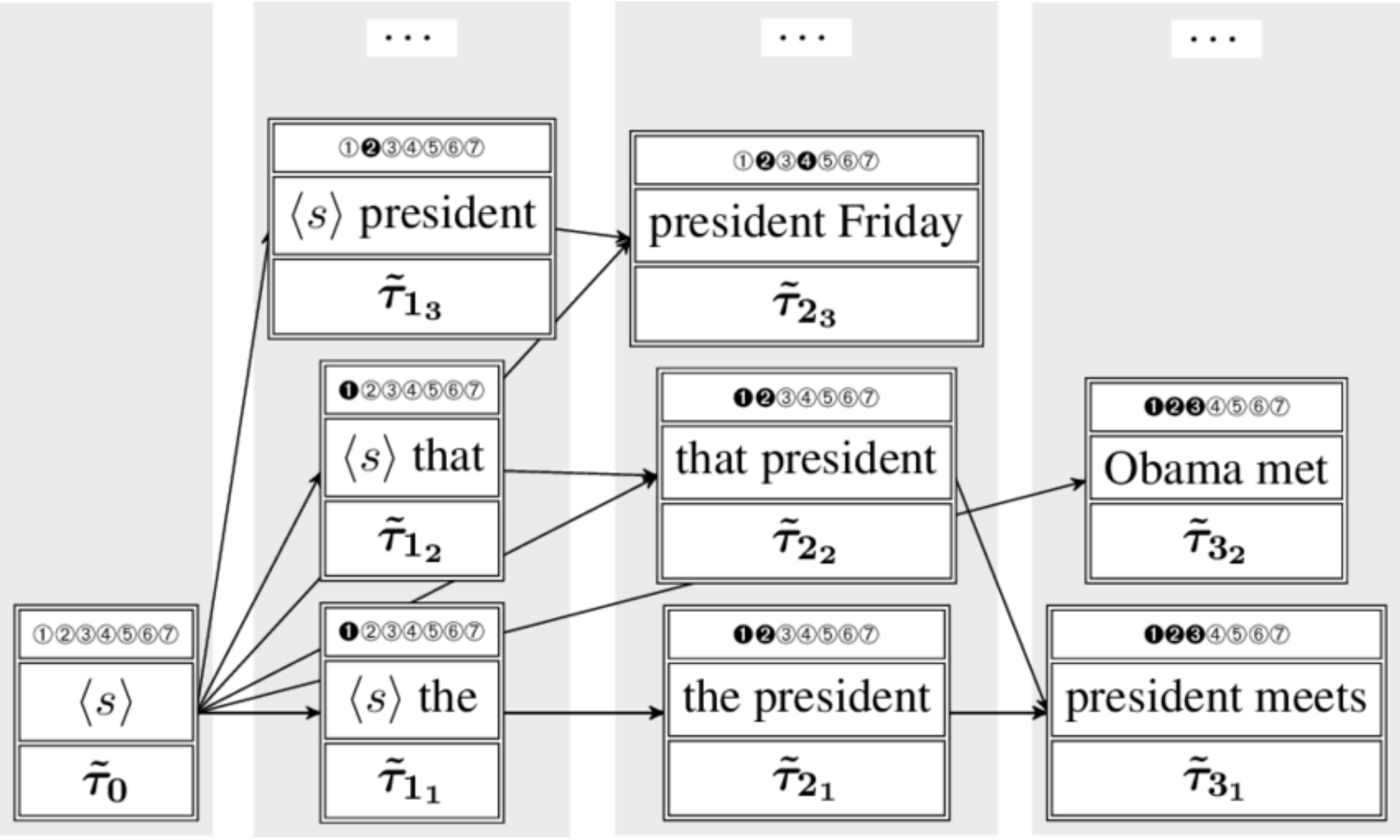
  - phrase translation probability ϕ
  - reordering probability d

Good to use context, but . . .

- Phrases are mostly independent (LM is glue)
- Would like to use wider context
- And have fuzzy phrase boundaries

Good to use context, but . . .

- Phrases are mostly independent (LM is glue)
- Would like to use wider context
- And have fuzzy phrase boundaries
- Neural models!



Phrase-based Decoding (Slide from Chris Callison-Burch)