

Statistical Machine Translation

-extending alignments/phrase extraction-

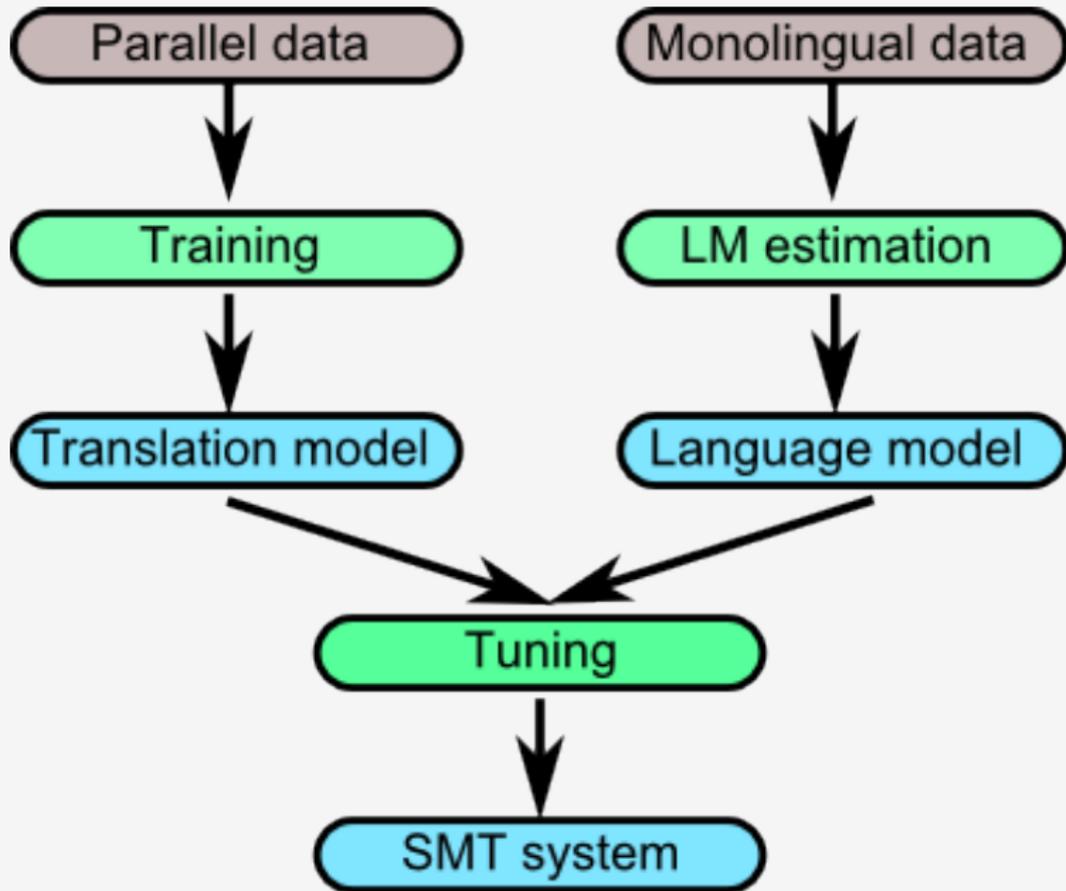
Artem Sokolov

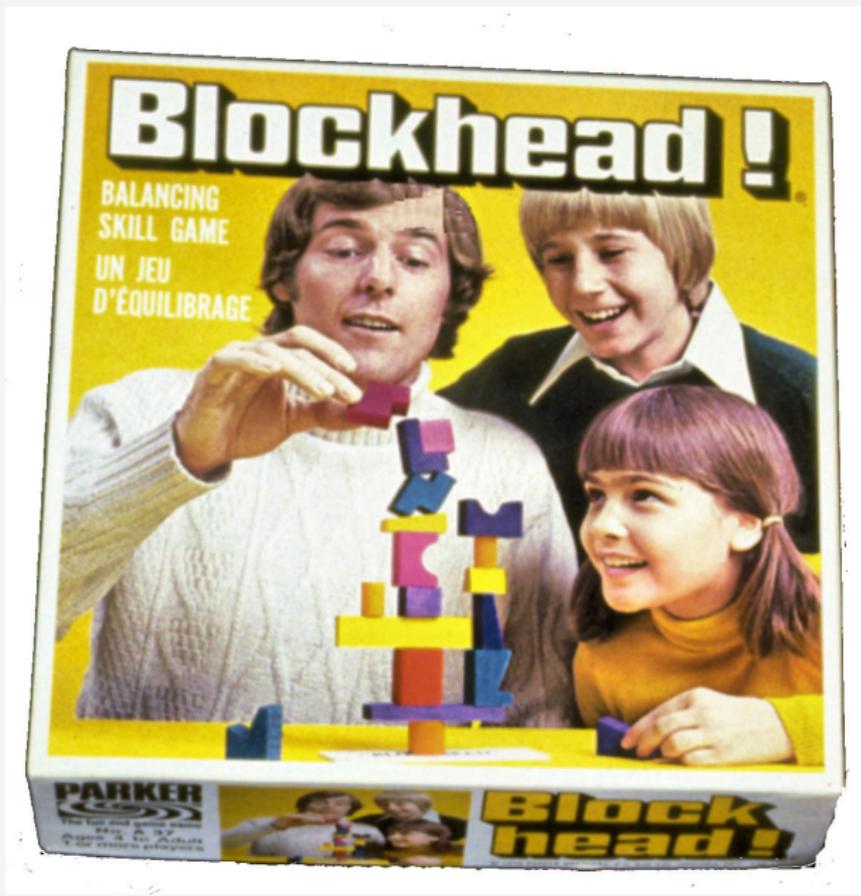
Computerlinguistik

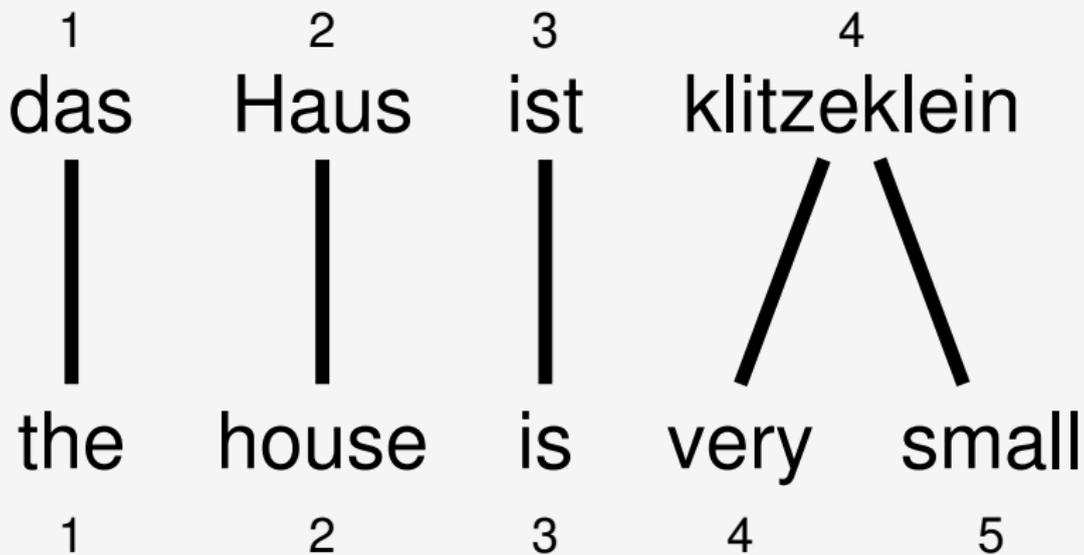
Universität Heidelberg

Sommersemester 2015

material from P. Koehn







The problem of 'correct' alignment - functional words

| | john | wohnt | hier | nicht |
|------|------|-------|------|-------|
| john | ■ | | | |
| does | | ? | | ? |
| not | | | | ■ |
| live | | ■ | | |
| here | | | ■ | |

The problem of 'correct' alignment - idioms

| | john | biss | ins | grass |
|--------|------|------|-----|-------|
| john | ■ | | | |
| kicked | | ■ | ■ | ■ |
| the | | ■ | ■ | ■ |
| bucket | | ■ | ■ | ■ |

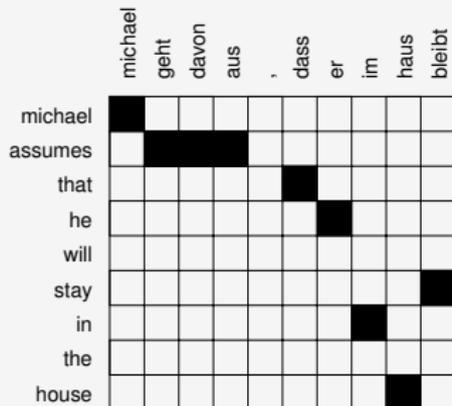
| | michael | geht | davon | aus | , | dass | er | im | haus | bleibt |
|---------|---------|------|-------|-----|---|------|----|----|------|--------|
| michael | █ | | | | | | | | | |
| assumes | | █ | █ | █ | | | | | | |
| that | | | | | | █ | | | | |
| he | | | | | | | █ | | | |
| will | | | | | | | | | | |
| stay | | | | | | | | | | █ |
| in | | | | | | | | █ | | |
| the | | | | | | | | | | |
| house | | | | | | | | | █ | |

English to German

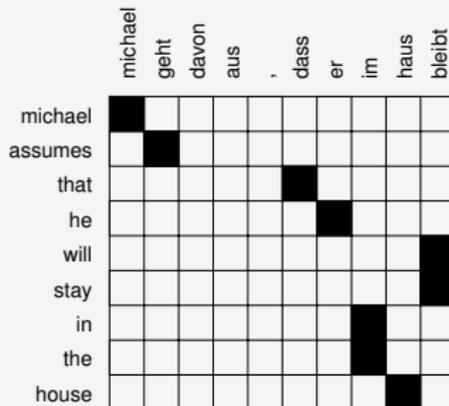
| | michael | geht | davon | aus | , | dass | er | im | haus | bleibt |
|---------|---------|------|-------|-----|---|------|----|----|------|--------|
| michael | █ | | | | | | | | | |
| assumes | | █ | | | | | | | | |
| that | | | | | | █ | | | | |
| he | | | | | | | █ | | | |
| will | | | | | | | | | | █ |
| stay | | | | | | | | | | █ |
| in | | | | | | | | █ | | |
| the | | | | | | | | | | |
| house | | | | | | | | | █ | |

German to English

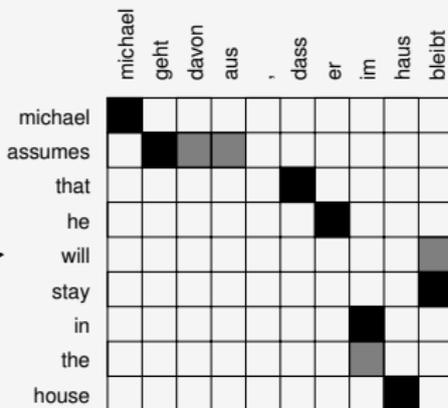
Bidirectional alignments



English to German



German to English



Intersection / Union

- 1: neighboring = $\{(-1,0),(0,-1),(1,0),(0,1),(-1,-1),(-1,1),(1,-1),(1,1)\}$
- 2: alignment $A = \text{intersect}(e2f,f2e)$;
- 3: grow-diag(A);
- 4: final(A);

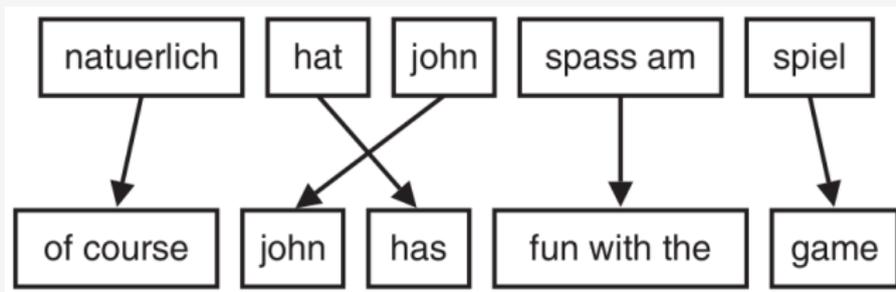
grow-diag()

- 1: **while** new points added **do**
- 2: **for all** English word $e \in [1\dots e_n]$, foreign word $f \in [1\dots f_n]$, $(e, f) \in A$ **do**
- 3: **for all** neighboring alignment points $(e_{\text{new}}, f_{\text{new}})$ **do**
- 4: **if** $(e_{\text{new}}$ unaligned OR f_{new} unaligned) AND $(e_{\text{new}}, f_{\text{new}}) \in \text{union}(e2f, f2e)$ **then**
- 5: add $(e_{\text{new}}, f_{\text{new}})$ to A
- 6: **end if**
- 7: **end for**
- 8: **end for**
- 9: **end while**

final()

- 1: **for all** English word $e_{\text{new}} \in [1\dots e_n]$, foreign word $f_{\text{new}} \in [1\dots f_n]$ **do**
- 2: **if** $(e_{\text{new}}$ unaligned OR f_{new} unaligned) AND $(e_{\text{new}}, f_{\text{new}}) \in \text{union}(e2f, f2e)$ **then**
- 3: add $(e_{\text{new}}, f_{\text{new}})$ to A
- 4: **end if**
- 5: **end for**

| | john | biss | ins | grass |
|--------|---|---|---|--|
| john |  | | | |
| kicked | |  |  |  |
| the | |  |  |  |
| bucket | |  |  |  |



- input is segmented into phrases
(not necessarily linguistically motivated)
- translated one-to-one into phrases in English
- possibly reordered

single-word translation:

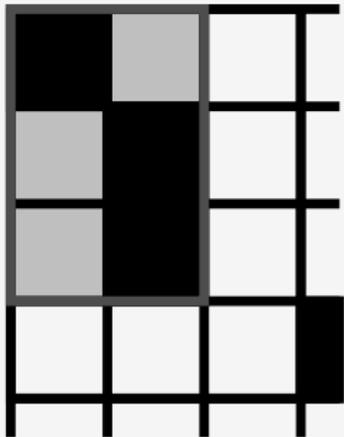
spaß → fun
 am → on the (0.4)
 → at the (0.4)
 ⋮
 → with the (0.0001)

phrase-based translation:

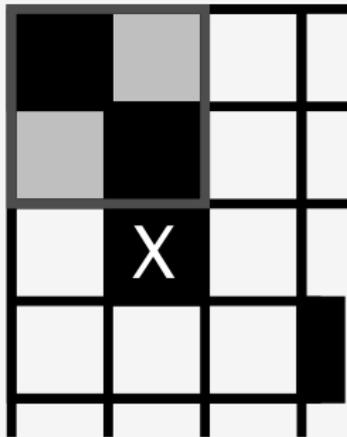
The context of the prior word "spaß" chooses the perfect translation for "am".

"den Vorschlag":

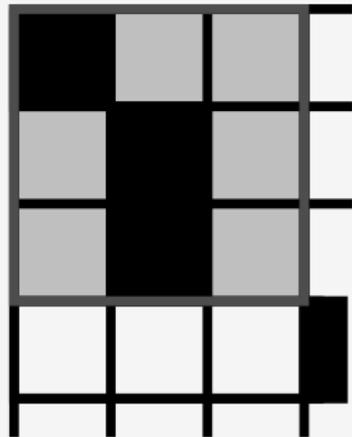
| English | $\phi(\bar{e} \bar{f})$ | English | $\phi(\bar{e} \bar{f})$ |
|-----------------|-------------------------|-----------------|-------------------------|
| the proposal | 0.6227 | the suggestions | 0.0114 |
| 's proposal | 0.1068 | the proposed | 0.0114 |
| a proposal | 0.0341 | the motion | 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 | ... | ... |



consistent



inconsistent



consistent

| | michael | geht | davon | aus | , | dass | er | im | haus | bleibt |
|---------|---------|------|-------|-----|---|------|----|----|------|--------|
| michael | ■ | | | | | | | | | |
| assumes | | ■ | ■ | ■ | | | | | | |
| that | | | | | | ■ | | | | |
| he | | | | | | | ■ | | | |
| will | | | | | | | | | | ■ |
| stay | | | | | | | | | | ■ |
| in | | | | | | | | ■ | | |
| the | | | | | | | | ■ | | |
| house | | | | | | | | | ■ | |

| | michael | geht | davon | aus | , | dass | er | im | haus | bleibt |
|---------|---------|------|-------|-----|---|------|----|----|------|--------|
| michael | █ | | | | | | | | | |
| assumes | | █ | █ | █ | | | | | | |
| that | | | | | | █ | | | | |
| he | | | | | | | █ | | | |
| will | | | | | | | | | | █ |
| stay | | | | | | | | | | █ |
| in | | | | | | | | █ | | |
| the | | | | | | | | █ | | |
| house | | | | | | | | | █ | |

michael - michael

| | michael | geht | davon | aus | , | dass | er | im | haus | bleibt |
|---------|---------|------|-------|-----|---|------|----|----|------|--------|
| michael | █ | | | | | | | | | |
| assumes | | █ | █ | █ | | | | | | |
| that | | | | | | █ | | | | |
| he | | | | | | | █ | | | |
| will | | | | | | | | | | █ |
| stay | | | | | | | | | | █ |
| in | | | | | | | | █ | | |
| the | | | | | | | | █ | | |
| house | | | | | | | | | █ | |

michael - michael
 michael assumes - michael geht davon aus

| | michael | geht | davon | aus | , | dass | er | im | haus | bleibt |
|---------|---------|------|-------|-----|---|------|----|----|------|--------|
| michael | ■ | | | | | | | | | |
| assumes | | ■ | ■ | ■ | | | | | | |
| that | | | | | | ■ | | | | |
| he | | | | | | | ■ | | | |
| will | | | | | | | | | | ■ |
| stay | | | | | | | | | | ■ |
| in | | | | | | | | ■ | | |
| the | | | | | | | | ■ | | |
| house | | | | | | | | | ■ | |

- michael – michael
- michael assumes – michael geht davon aus
- michael assumes – michael geht davon aus ,

| | michael | geht | davon | aus | , | dass | er | im | haus | bleibt |
|---------|---------|------|-------|-----|---|------|----|----|------|--------|
| michael | ■ | | | | | | | | | |
| assumes | | ■ | ■ | ■ | | | | | | |
| that | | | | | | ■ | | | | |
| he | | | | | | | ■ | | | |
| will | | | | | | | | | | ■ |
| stay | | | | | | | | | | ■ |
| in | | | | | | | | ■ | | |
| the | | | | | | | | ■ | | |
| house | | | | | | | | | ■ | |

- michael – michael
- michael assumes – michael geht davon aus
- michael assumes – michael geht davon aus ,

| | michael | geht | davon | aus | , | dass | er | im | haus | bleibt |
|---------|---------|------|-------|-----|---|------|----|----|------|--------|
| michael | ■ | | | | | | | | | |
| assumes | | ■ | ■ | ■ | | | | | | |
| that | | | | | | ■ | | | | |
| he | | | | | | | ■ | | | |
| will | | | | | | | | | | ■ |
| stay | | | | | | | | | | ■ |
| in | | | | | | | | ■ | | |
| the | | | | | | | | ■ | | |
| house | | | | | | | | | ■ | |

- michael – michael
- michael assumes – michael geht davon aus
- michael assumes – michael geht davon aus ,
- michael assumes that – michael geht davon aus , dass
- michael assumes that he – michael geht davon aus , dass er

| | michael | geht | davon | aus | , | dass | er | im | haus | bleibt |
|---------|---------|------|-------|-----|---|------|----|----|------|--------|
| michael | █ | | | | | | | | | |
| assumes | | █ | █ | █ | | | | | | |
| that | | | | | | █ | | | | |
| he | | | | | | | █ | | | |
| will | | | | | | | | | | █ |
| stay | | | | | | | | | | █ |
| in | | | | | | | | █ | | |
| the | | | | | | | | █ | | |
| house | | | | | | | | | █ | |

- michael – michael
- michael assumes – michael geht davon aus
- michael assumes – michael geht davon aus ,
- 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

| | michael | geht | davon | aus | , | dass | er | im | haus | bleibt |
|---------|---------|------|-------|-----|---|------|----|----|------|--------|
| michael | █ | | | | | | | | | |
| assumes | | █ | █ | █ | | | | | | |
| that | | | | | █ | | | | | |
| he | | | | | | █ | | | | |
| will | | | | | | | | | | █ |
| stay | | | | | | | | | | █ |
| in | | | | | | | | █ | | |
| the | | | | | | | | █ | | |
| house | | | | | | | | | █ | |

- michael – michael
- michael assumes – michael geht davon aus
- michael assumes – michael geht davon aus ,
- 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 – geht davon aus
- assumes – geht davon aus ,
- assumes that – geht davon aus , dass
- assumes that he – geht davon aus , dass er
- assumes that he will stay in the house – geht davon aus , dass er im haus bleibt

| | michael | geht | davon | aus | , | dass | er | im | haus | bleibt |
|---------|---------|------|-------|-----|---|------|----|----|------|--------|
| michael | █ | | | | | | | | | |
| assumes | | █ | █ | █ | | | | | | |
| that | | | | | | █ | | | | |
| he | | | | | | | █ | | | |
| will | | | | | | | | | | █ |
| stay | | | | | | | | | | █ |
| in | | | | | | | | █ | | |
| the | | | | | | | | █ | | |
| house | | | | | | | | | █ | |

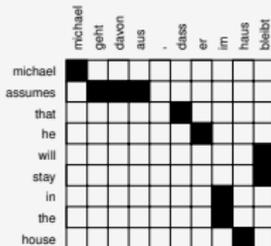
| | | |
|--|---|---|
| michael | - | michael |
| michael assumes | - | michael geht davon aus |
| michael assumes | - | michael geht davon aus , |
| michael assumes that | - | michael geht davon aus , dass |
| michael assumes that he | - | michael geht davon aus , dass er |
| michael assumes that he will stay in the house | - | michael geht davon aus , dass er im haus bleibt |
| assumes | - | geht davon aus |
| assumes | - | geht davon aus , |
| assumes that | - | geht davon aus , dass |
| assumes that he | - | geht davon aus, dass er |
| assumes that he will stay in the house | - | geht davon aus , dass er im haus bleibt |
| that | - | dass |
| that | - | , dass |
| that he | - | dass er |
| that he | - | , dass er |
| that he will stay in the house | - | dass er im haus bleibt |
| that he will stay in the house | - | , dass er im haus bleibt |

| | michael | geht | davon | aus | , | dass | er | im | haus | bleibt |
|---------|---------|------|-------|-----|---|------|----|----|------|--------|
| michael | █ | | | | | | | | | |
| assumes | | █ | █ | █ | | | | | | |
| that | | | | | | █ | | | | |
| he | | | | | | | █ | | | |
| will | | | | | | | | | | █ |
| stay | | | | | | | | | | █ |
| in | | | | | | | | █ | | |
| the | | | | | | | | █ | | |
| house | | | | | | | | | █ | |

| | | |
|--|---|---|
| michael | - | michael |
| michael assumes | - | michael geht davon aus |
| michael assumes | - | michael geht davon aus , |
| 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 | - | geht davon aus |
| assumes | - | geht davon aus , |
| assumes that | - | geht davon aus , dass |
| assumes that he | - | geht davon aus, dass er |
| assumes that he will stay in the house | - | geht davon aus , dass er im haus bleibt |
| that | - | dass |
| that | - | , dass |
| that he | - | dass er |
| that he | - | , dass er |
| that he will stay in the house | - | dass er im haus bleibt |
| that he will stay in the house | - | , dass er im haus bleibt |
| he | - | er |
| he will stay in the house | - | er im haus bleibt |

| | michael | geht | davon | aus | . | dass | er | im | haus | bleibt |
|---------|---------|------|-------|-----|---|------|----|----|------|--------|
| michael | █ | | | | | | | | | |
| assumes | █ | █ | █ | █ | | | | | | |
| that | | | | | █ | | | | | |
| he | | | | | | █ | | | | |
| will | | | | | | | | | | █ |
| stay | | | | | | | | | | █ |
| in | | | | | | | █ | | | |
| the | | | | | | | █ | | | |
| house | | | | | | | | █ | | |

| | | |
|--|---|---|
| michael | - | michael |
| michael assumes | - | michael geht davon aus |
| michael assumes | - | michael geht davon aus , |
| 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 | - | geht davon aus |
| assumes | - | geht davon aus , |
| assumes that | - | geht davon aus , dass |
| assumes that he | - | geht davon aus, dass er |
| assumes that he will stay in the house | - | geht davon aus , dass er im haus bleibt |
| that | - | dass |
| that | - | , dass |
| that he | - | dass er |
| that he | - | , dass er |
| that he will stay in the house | - | dass er im haus bleibt |
| that he will stay in the house | - | , dass er im haus bleibt |
| he | - | er |
| he will stay in the house | - | er im haus bleibt |
| will stay | - | bleibt |
| will stay in the house | - | im haus bleibt |



| | | |
|--|---|---|
| michael | - | michael |
| michael assumes | - | michael geht davon aus |
| michael assumes | - | michael geht davon aus , |
| 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 | - | geht davon aus |
| assumes | - | geht davon aus , |
| assumes that | - | geht davon aus , dass |
| assumes that he | - | geht davon aus, dass er |
| assumes that he will stay in the house | - | geht davon aus , dass er im haus bleibt |
| that | - | dass |
| that | - | , dass |
| that he | - | dass er |
| that he | - | , dass er |
| that he will stay in the house | - | dass er im haus bleibt |
| that he will stay in the house | - | , dass er im haus bleibt |
| he | - | er |
| he will stay in the house | - | er im haus bleibt |
| will stay | - | bleibt |
| will stay in the house | - | im haus bleibt |
| in the | - | im |
| in the house | - | im haus |

| | michael | geht | davon | aus | , | dass | er | im | haus | bleibt |
|---------|---------|------|-------|-----|---|------|----|----|------|--------|
| michael | █ | | | | | | | | | |
| assumes | | █ | █ | █ | | | | | | |
| that | | | | | | █ | | | | |
| he | | | | | | | █ | | | |
| will | | | | | | | | | █ | █ |
| stay | | | | | | | | | | █ |
| in | | | | | | | | █ | | |
| the | | | | | | | | | █ | |
| house | | | | | | | | | | █ |

| | | |
|--|---|---|
| michael | – | michael |
| michael assumes | – | michael geht davon aus |
| michael assumes | – | michael geht davon aus , |
| 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 | – | geht davon aus |
| assumes | – | geht davon aus , |
| assumes that | – | geht davon aus , dass |
| assumes that he | – | geht davon aus , dass er |
| assumes that he will stay in the house | – | geht davon aus , dass er im haus bleibt |
| that | – | dass |
| that | – | , dass |
| that he | – | dass er |
| that he | – | , dass er |
| that he will stay in the house | – | dass er im haus bleibt |
| that he will stay in the house | – | , dass er im haus bleibt |
| he | – | er |
| he will stay in the house | – | er im haus bleibt |
| will stay | – | bleibt |
| will stay in the house | – | im haus bleibt |
| in the | – | im |
| in the house | – | im haus |
| house | – | haus |