



Maschinelle Übersetzung und Terminologie

Josef van Genabith

Multilingual Technologies MLT
DFKI Saarbrücken

Translation-Oriented Language Technologies
Saarland University





- MÜ
- Regelbasiert und statistisch
- SMT/PB-SMT und NMT beide statistisch
- MÜ und Terminologie
- Scenario: keine Texte als Trainingsdaten in denen die relevante Terminologie genügend häufig vorkommt



Bilinguale Terminologie



- Im einfachsten Fall:

$$a_1 \approx z_1$$

$$a_2 \approx z_2$$

$$a_3 \approx z_4$$

$$\dots \approx \dots$$

$$a_n \approx z_n$$



- SMT/PB-SMT und Terminologie
 - Füge die Terminologie zu den Trainingsdaten hinzu
- NMT und Terminologie
 - constrained decoding (eingeschränktes ...)
 - unconstrained decoding (uneingeschränktes ...)

$$a_1 \approx z_1$$

$$a_2 \approx z_2$$

$$a_3 \approx z_3$$

$\dots \approx \dots$

$$a_n \approx z_n$$



Training by replacing / by appending



1. Diese **Leitung** ist an den Generator anzuschließen. Connect the **cable** to the generator.
2. Diese **cable** ist an den Generator anzuschließen. Connect the **cable** to the generator.
3. Diese **Leitung # cable #** ist an den Generator anzuschließen. Connect the **cable** to the generator.

Dinu et al. 2019



- SMT/PB-SMT und Terminologie
 - Füge die Terminologie zu den Trainingsdaten hinzu
- NMT und Terminologie
 - constrained decoding (eingeschränktes ...)
 - unconstrained decoding (uneingeschränktes ...)





Maschinelle Übersetzung



- $p(target|source)$



- $p(\text{target}|\text{source})$
- $p(\text{Wie geht's?} \mid \text{How's it going?})$



- $p(\text{target}|\text{source})$
- $p(\text{Wie geht's ?} | \text{How's it going ?})$
- $p(\text{Wie geht's ?} | \text{How's it going ?}) > p(\text{Guten Abend !} | \text{How's it going ?}) \gg p(\text{Pedro hat einen Esel .} | \text{How's it going ?})$



Maschinelle Übersetzung



- $p(target|source)$?



Maschinelle Übersetzung



- $p(\text{target}|\text{source})?$

GERMAN	ENGLISH	FRENCH
Einleitung <i>I. Von dem Unterschiede der reinen und empirischen Erkenntnis</i>	Introduction <i>I. Of the difference between Pure and Empirical Knowledge</i>	Introduction <i>I. De la différence de la connaissance pure et de la connaissance empirique.</i>
Daß alle unsere Erkenntnis mit der Erfahrung anfange, daran ist gar kein Zweifel; denn wodurch sollte das Erkenntnisvermögen sonst zur Ausübung erweckt werden, geschähe es nicht durch Gegenstände, die unsere Sinne rühren und teils von selbst Vorstellungen bewirken, teils unsere Verständigkeit in Bewegung bringen, diese zu vergleichen, sie zu verknüpfen oder zu trennen, und so den rohen Stoff sinnlicher Eindrücke zu einer Erkenntnis der Gegenstände zu verarbeiten, die Erfahrung heißt? Der Zeit nach geht also keine Erkenntnis in uns vor der Erfahrung vorher, und mit dieser fängt alle an.	That all our knowledge begins with experience there can be no doubt. For how is it possible that the faculty of cognition should be awakened into exercise otherwise than by means of objects which affect our senses, and partly of themselves produce representations, partly rouse our powers of understanding into activity, to compare to connect, or to separate these, and so to convert the raw material of our sensuous impressions into a knowledge of objects, which is called experience? In respect of time, therefore, no knowledge of ours is antecedent to experience, but begins with it.	Que toute notre connaissance commence avec l'expérience, cela ne soulève aucun doute. En effet, par quoi notre pouvoir de connaitre pourrait-il être éveillé et mis en action, si ce n'est par des objets qui frappent nos sens et qui, d'une part, produisent par eux-mêmes des représentations et, d'autre part, mettent en mouvement notre faculté intellectuelle, afin qu'elle compare, lie ou sépare ces représentations, et travaille ainsi la matière brute des impressions sensibles pour en tirer une connaissance des objets, celle qu'on nomme l'expérience? Ainsi, chronologiquement, aucune connaissance ne précède en nous l'expérience et c'est avec elle que toutes commencent.



Maschinelle Übersetzung



- $p(\text{target}|\text{source})?$

1. Statistische MÜ (SMT)
2. Neuronale MÜ (NMT)

GERMAN	ENGLISH	FRENCH
<p>Einleitung</p> <p><i>I. Von dem Unterschiede der reinen und empirischen Erkenntnis</i></p> <p>Daß alle unsere Erkenntnis mit der Erfahrung anfange, daran ist gar kein Zweifel; denn wodurch sollte das Erkenntnisvermögen sonst zur Ausübung erweckt werden, geschähe es nicht durch Gegenstände, die unsere Sinne rühren und teils von selbst Vorstellungen bewirken, teils unsere Verständigkeit in Bewegung bringen, diese zu vergleichen, sie zu verknüpfen oder zu trennen, und so den rohen Stoff sinnlicher Eindrücke zu einer Erkenntnis der Gegenstände zu verarbeiten, die Erfahrung heißt? Der Zeit nach geht also keine Erkenntnis in uns vor der Erfahrung vorher, und mit dieser fängt alle an.</p>	<p>Introduction</p> <p><i>I. Of the difference between Pure and Empirical Knowledge</i></p> <p>That all our knowledge begins with experience there can be no doubt. For how is it possible that the faculty of cognition should be awakened into exercise otherwise than by means of objects which affect our senses, and partly of themselves produce representations, partly rouse our powers of understanding into activity, to compare to connect, or to separate these, and so to convert the raw material of our sensuous impressions into a knowledge of objects, which is called experience? In respect of time, therefore, no knowledge of ours is antecedent to experience, but begins with it.</p>	<p>Introduction</p> <p><i>I. De la différence de la connaissance pure et de la connaissance empirique.</i></p> <p>Que toute notre connaissance commence avec l'expérience, cela ne soulève aucun doute. En effet, par quoi notre pouvoir de connaitre pourrait-il être éveillé et mis en action, si ce n'est par des objets qui frappent nos sens et qui, d'une part, produisent par eux-mêmes des représentations et, d'autre part, mettent en mouvement notre faculté intellectuelle, afin qu'elle compare, lie ou sépare ces représentations, et travaille ainsi la matière brute des impressions sensibles pour en tirer une connaissance des objets, celle qu'on nomme l'expérience? Ainsi, chronologiquement, aucune connaissance ne précède en nous l'expérience et c'est avec elle que toutes commencent.</p>



SMT Statistische Maschinelle Übersetzung



- $p(\text{target}|\text{source})?$

- $p(\text{target}|\text{source}) = \frac{p(\text{source}|\text{target}) \times p(\text{target})}{p(\text{source})}$

$$\text{target}^* = \operatorname{argmax}_{\text{target}} p(\text{target}|\text{source})$$

$$= \operatorname{argmax}_{\text{target}} \left\{ \frac{p(\text{source}|\text{target}) \times p(\text{target})}{p(\text{source})} \right\}$$

$$= \operatorname{argmax}_{\text{target}} \{ p(\text{source}|\text{target}) \times p(\text{target}) \}$$

English	French	Italian
the sun rises in the east	Le soleil se lève à l'est	Il sole nasce nell'est
the sun rises in the west	Le soleil se lève à l'ouest	Il sole nasce nell'ovest
the sun sets in the east	Le soleil se couche à l'est	Il sole tramonta nell'est
the sun sets in the west	Le soleil se couche à l'ouest	Il sole tramonta nell'ovest



SMT Statistische Maschinelle Übersetzung



$$\text{target}^* = \operatorname{argmax}_{\text{target}} p(\text{target}|\text{source}) = \operatorname{argmax}_{\text{target}} \{p(\text{source}|\text{target}) \times p(\text{target})\}$$

Bayesian	Naive Bayes	Log-linear
<p>Welche negative Erfahrung und Erfahrungswissen können wir ausnutzen? Das ist die Basis der prior distribution. Es gibt verschiedene Möglichkeiten, wie man diese Werte bestimmen kann. Eine Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine andere Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine dritte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine vierte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine fünfte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine sechste Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine siebte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine achte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine neunte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine zehnte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen.</p>	<p>Welche negative Erfahrung und Erfahrungswissen können wir ausnutzen? Das ist die Basis der prior distribution. Es gibt verschiedene Möglichkeiten, wie man diese Werte bestimmen kann. Eine Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine andere Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine dritte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine vierte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine fünfte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine sechste Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine siebte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine achte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine neunte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine zehnte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen.</p>	<p>Welche negative Erfahrung und Erfahrungswissen können wir ausnutzen? Das ist die Basis der prior distribution. Es gibt verschiedene Möglichkeiten, wie man diese Werte bestimmen kann. Eine Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine andere Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine dritte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine vierte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine fünfte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine sechste Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine siebte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine achte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine neunte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen. Eine zehnte Möglichkeit ist, die Verteilung der Werte zu schätzen und sie dann mit den tatsächlichen Werten zu vergleichen.</p>



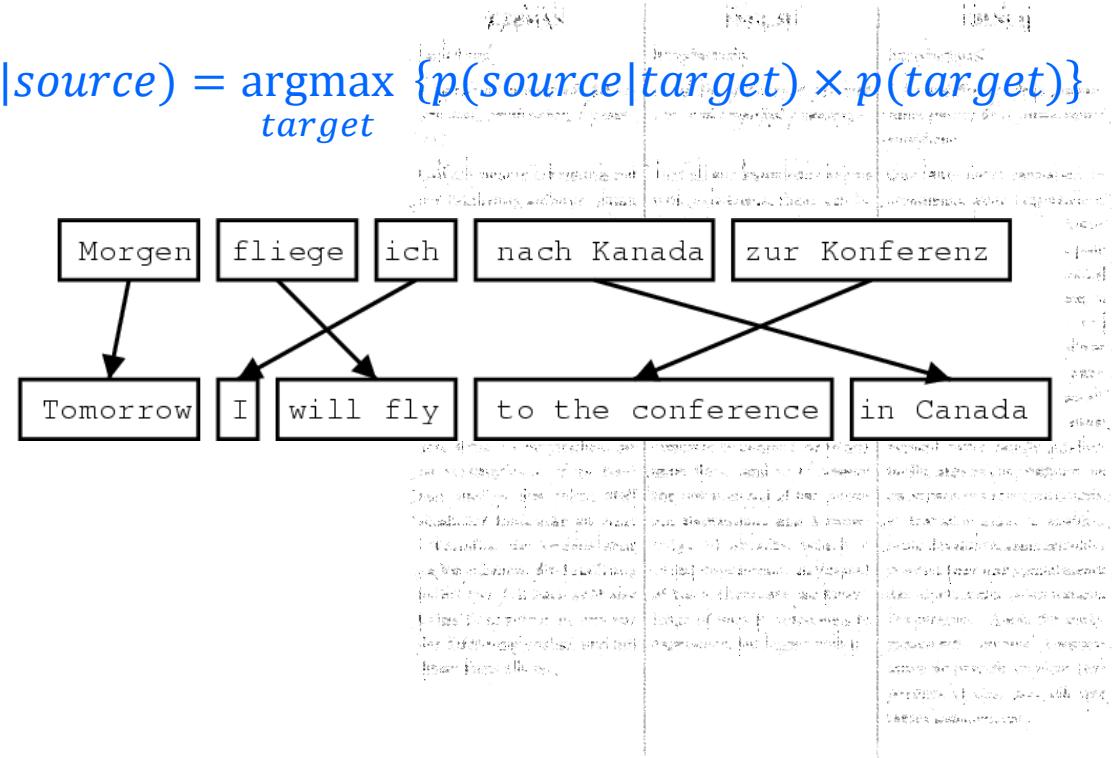
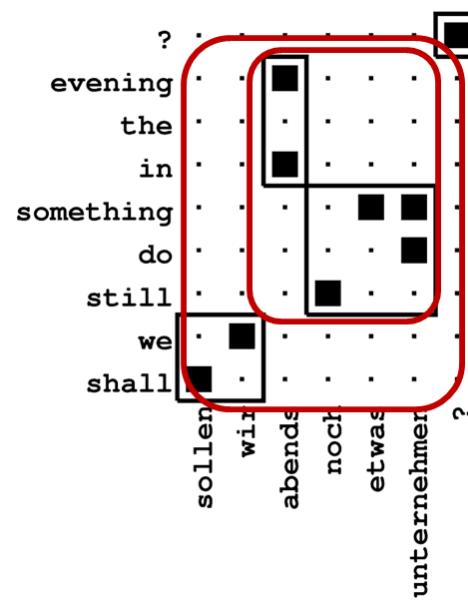
$$target^* = \operatorname{argmax}_{target} p(target|source) = \operatorname{argmax}_{target} \{p(source|target) \times p(target)\}$$



SMT Statistische Maschinelle Übersetzung



$$\text{target}^* = \operatorname{argmax}_{\text{target}} p(\text{target}|\text{source}) = \operatorname{argmax}_{\text{target}} \{p(\text{source}|\text{target}) \times p(\text{target})\}$$

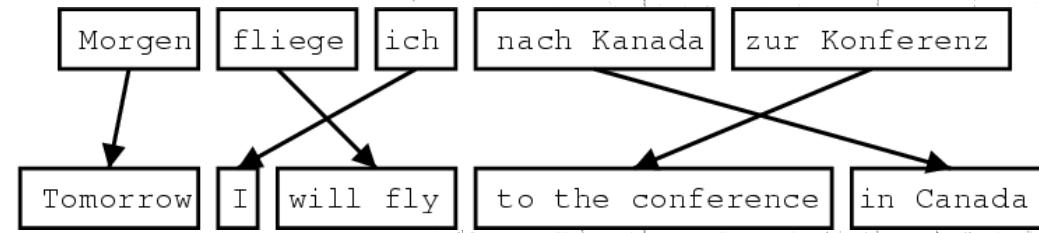
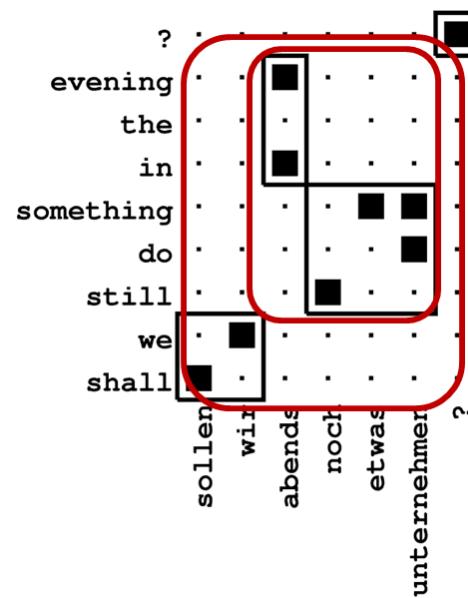




SMT Statistische Maschinelle Übersetzung



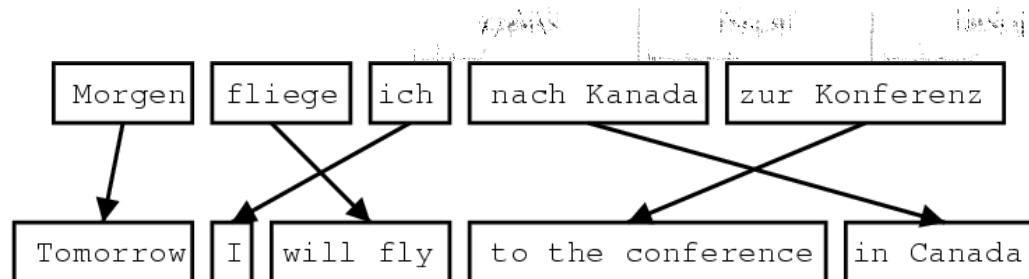
$$\text{target}^* = \operatorname{argmax}_{\text{target}} p(\text{target}|\text{source}) = \operatorname{argmax}_{\text{target}} \{p(\text{source}|\text{target}) \times p(\text{target})\}$$



$$z^* = \operatorname{argmax}_z P(z|a) = \operatorname{argmax}_z (P(a|z) \times P(z))$$

$$\operatorname{argmax}_z \left\{ \prod_{\bar{a}_i \text{ in } a} (P_{\bar{a}_i}(\bar{a}_i|\bar{z}_i) \times D(\bar{a}_i, \bar{a}_{i-1})) \times \prod_{w_j \text{ in } z} P_{LM}(w_j|w_{j-1}) \right\}$$

$$\begin{aligned} a_1 &\approx z_1 \\ a_2 &\approx z_2 \\ a_3 &\approx z_3 \\ \dots &\approx \dots \\ a_n &\approx z_n \end{aligned}$$



$$z^* = \operatorname{argmax}_z P(z|a) = \operatorname{argmax}_z (P(a|z) \times P(z))$$

$$\operatorname{argmax}_z \left\{ \prod_{\bar{a}_i \text{ in } a} (P_U(\bar{a}_i | \bar{z}_i) \times D(\bar{a}_i | \bar{a}_{i-1})) \times \prod_{w_j \text{ in } z} P_{LM}(w_j | w_{j-1}) \right\}$$

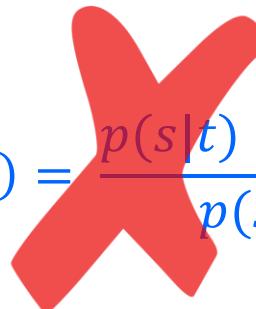


NMT Neuronale Maschinelle Übersetzung



- $p(\text{target}|\text{source})?$

$$p(t|s) = \frac{p(s|t) \times p(t)}{p(s)}$$

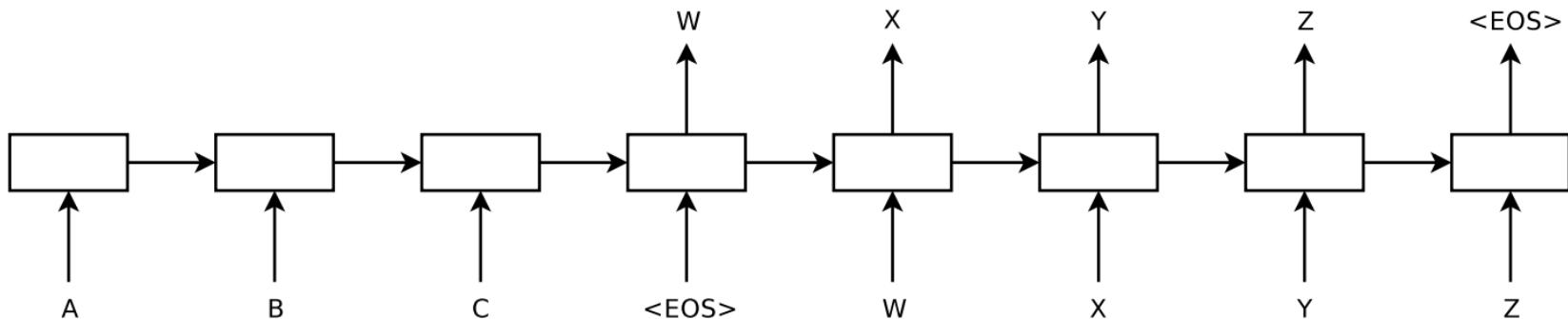


$$p(t|s) = p(t_1 \dots t_m | s_1 \dots s_n) = \prod_{i=1}^m p(t_i | t_1 \dots t_{i-1} s_1 \dots s_n)$$

$$p(t|s) = p(t_1^m | s_1^n) = \prod_{i=1}^m p(t_i | t_1^{i-1} s_1^n)$$

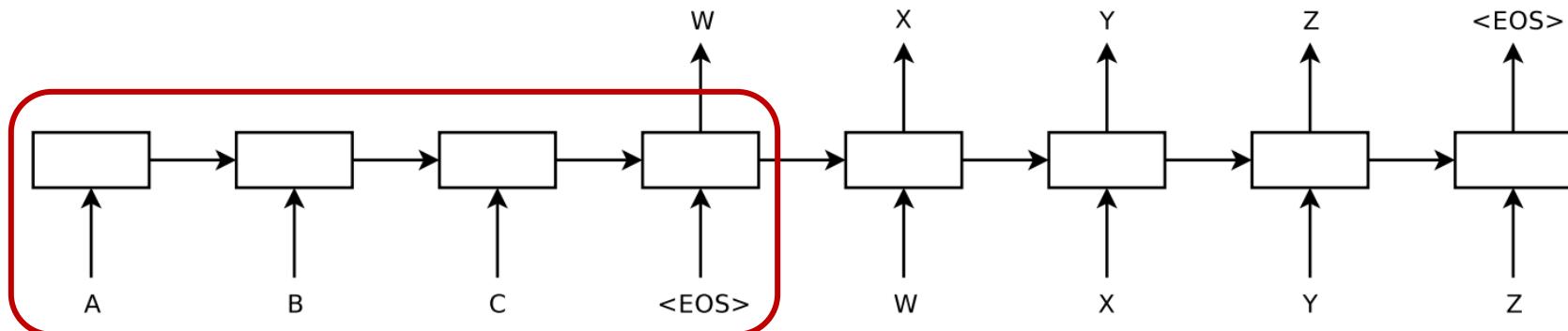
Parrot	Parrot	Parrot
Ein Vogel kann nicht sprechen.	Ein Vogel kann nicht sprechen.	Ein Vogel kann nicht sprechen.
Ein Vogel kann nicht sprechen.	Ein Vogel kann nicht sprechen.	Ein Vogel kann nicht sprechen.
Ein Vogel kann nicht sprechen.	Ein Vogel kann nicht sprechen.	Ein Vogel kann nicht sprechen.
Ein Vogel kann nicht sprechen.	Ein Vogel kann nicht sprechen.	Ein Vogel kann nicht sprechen.

$$p(t|s) = p(t_1 \cdots t_m | s_1 \cdots s_n) = \prod_{i=1}^m p(t_i | t_1 \cdots t_{i-1}, s_1 \cdots s_n)$$



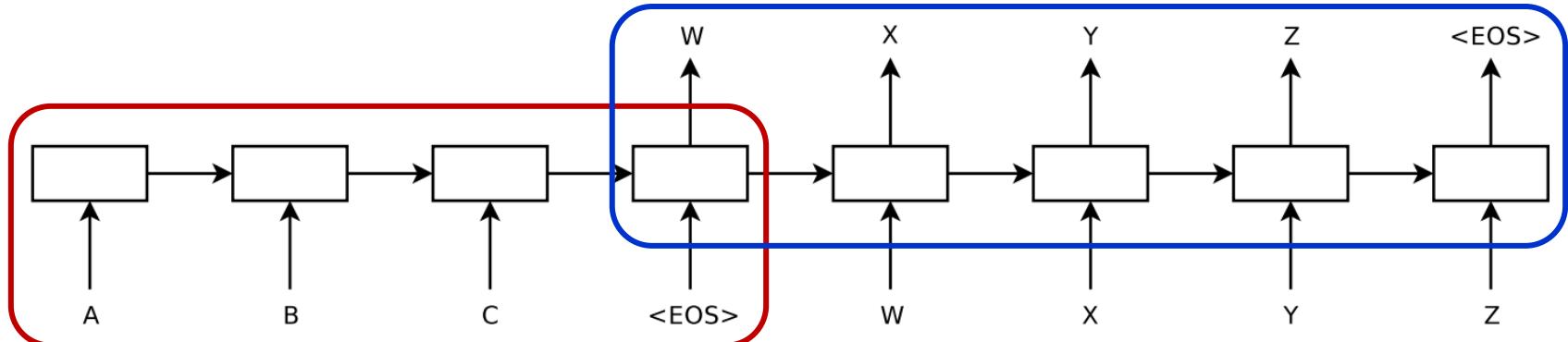
Sutskever et al. 2014

$$p(t|s) = p(t_1 \cdots t_m | s_1 \cdots s_n) = \prod_{i=1}^m p(t_i | t_1 \cdots t_{i-1}, s_1 \cdots s_n)$$



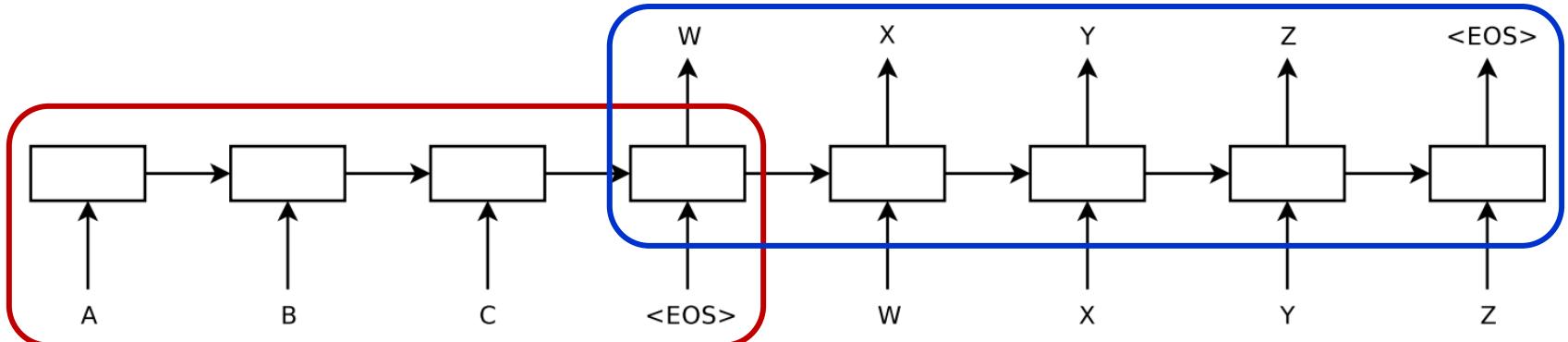
Sutskever et al. 2014

$$p(t|s) = p(t_1 \cdots t_m | s_1 \cdots s_n) = \prod_{i=1}^m p(t_i | t_1 \cdots t_{i-1}, s_1 \cdots s_n)$$



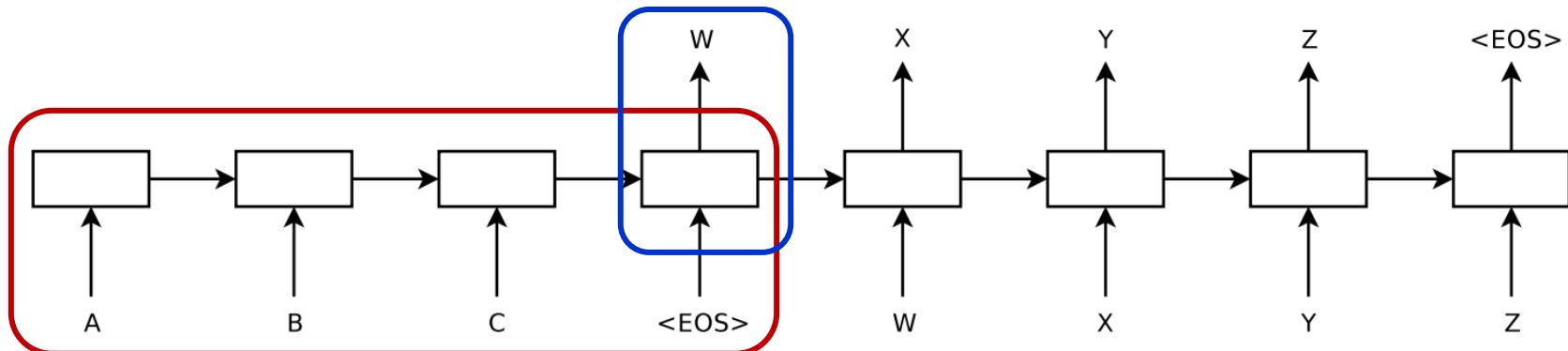
Sutskever et al. 2014

$$p(t|s) = p(t_1 \cdots t_m | s_1 \cdots s_n) = \prod_{i=1}^m p(t_i | t_1 \cdots t_{i-1}, s_1 \cdots s_n)$$



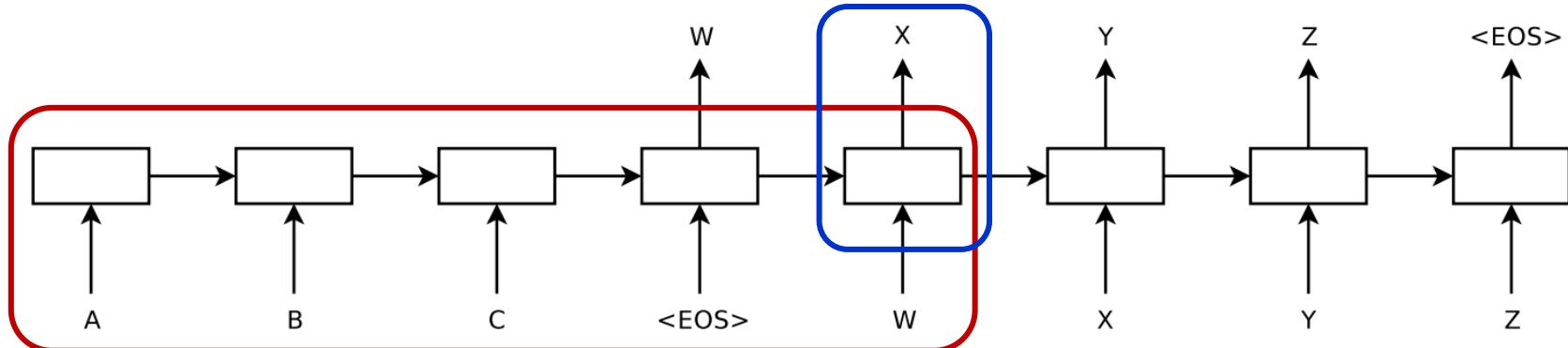
Sutskever et al. 2014

$$p(t|s) = p(t_1 \cdots t_m | s_1 \cdots s_n) = \prod_{i=1}^m p(t_i | t_1 \cdots t_{i-1}, s_1 \cdots s_n)$$



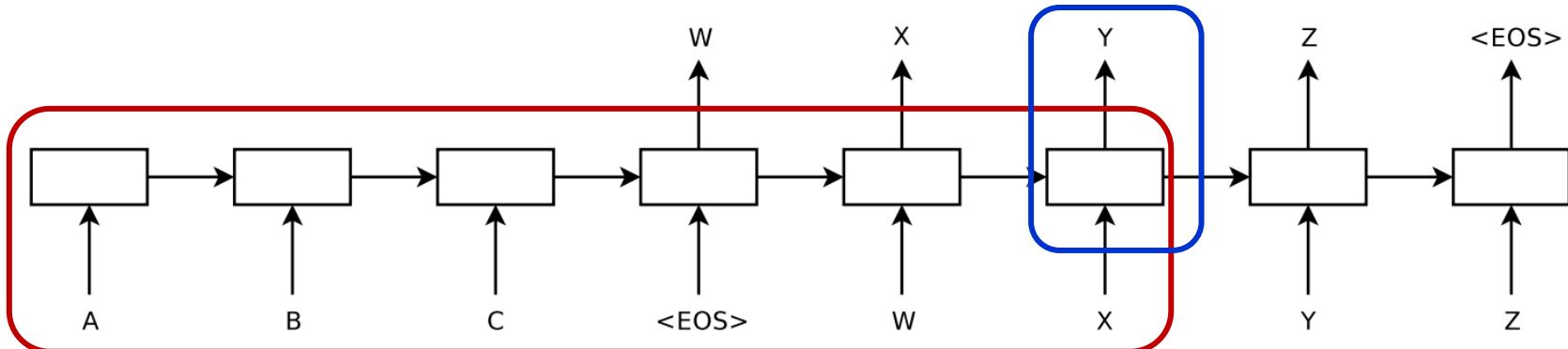
Sutskever et al. 2014

$$p(t|s) = p(t_1 \cdots t_m | s_1 \cdots s_n) = \prod_{i=1}^m p(t_i | t_1 \cdots t_{i-1}, s_1 \cdots s_n)$$



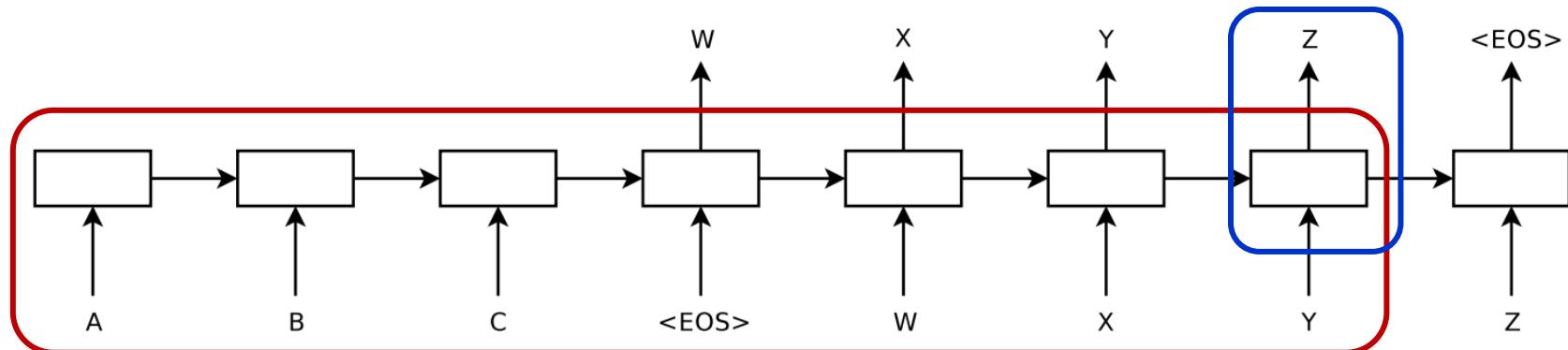
Sutskever et al. 2014

$$p(t|s) = p(t_1 \cdots t_m | s_1 \cdots s_n) = \prod_{i=1}^m p(t_i | t_1 \cdots t_{i-1}, s_1 \cdots s_n)$$



Sutskever et al. 2014

$$p(t|s) = p(t_1 \cdots t_m | s_1 \cdots s_n) = \prod_{i=1}^m p(t_i | t_1 \cdots t_{i-1}, s_1 \cdots s_n)$$



Sutskever et al. 2014

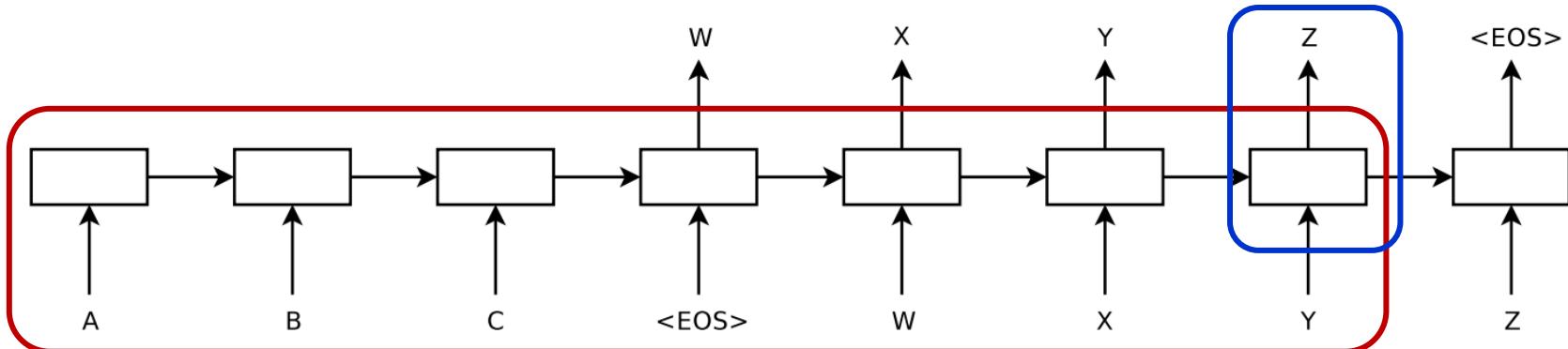


Neuronale MT



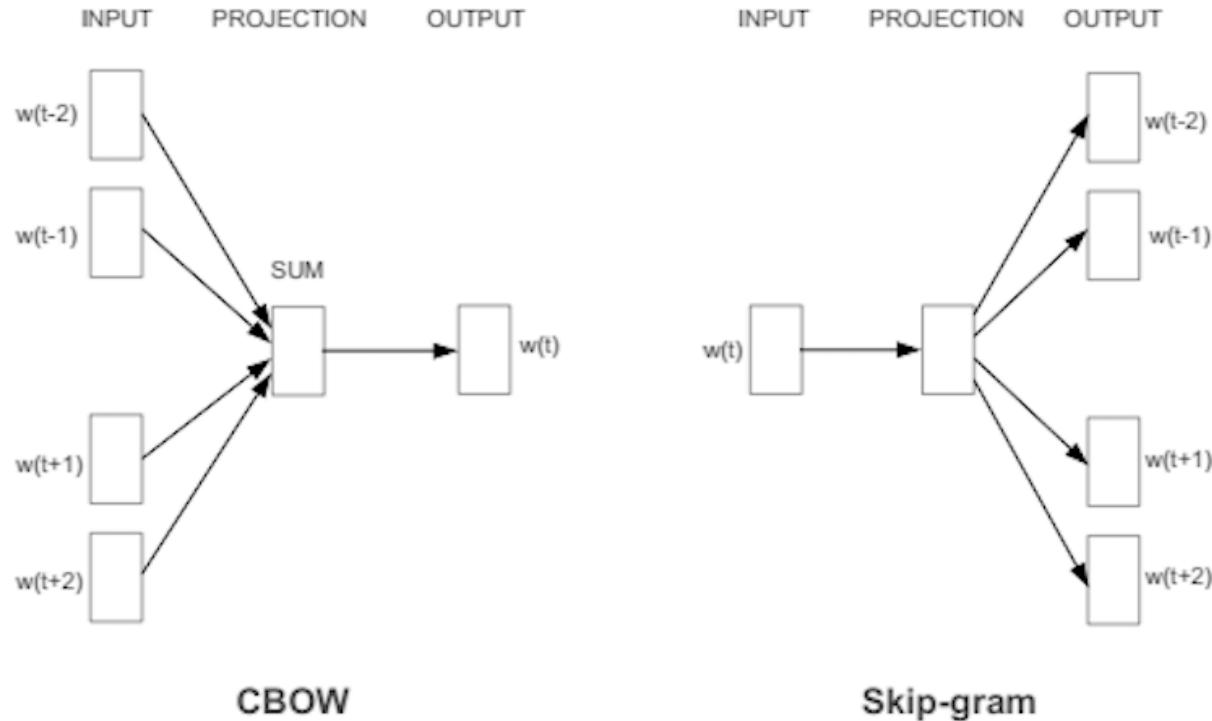
$$p(t|s) = p(t_1 \cdots t_m | s_1 \cdots s_n) = \prod_{i=1}^m p(t_i | s_1 \cdots s_{i-1}, t_1 \cdots t_n)$$

Of course, the modelling of
this by the NN is not exact ...!

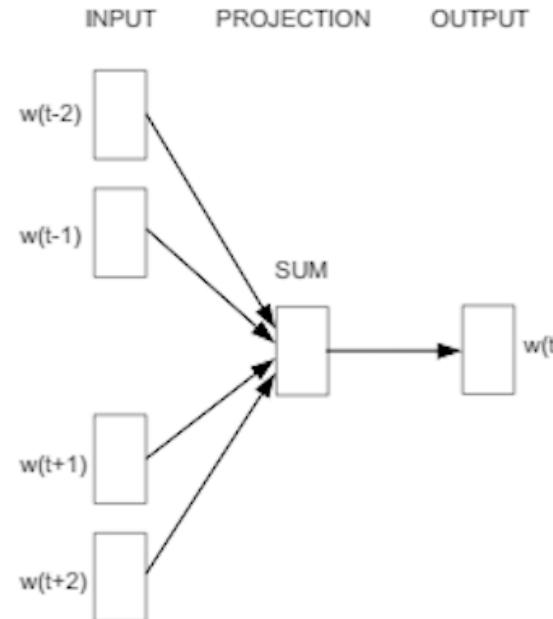


Sutskever et al. 2014

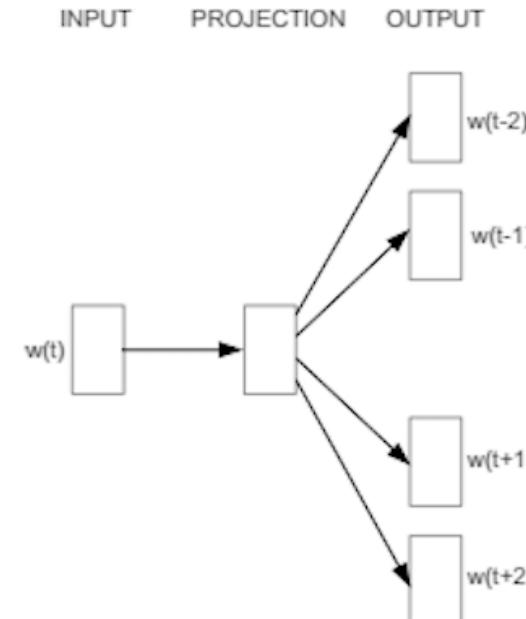
Neuronale MT



$$\begin{aligned} a_1 &\approx z_1 \\ a_2 &\approx z_2 \\ a_3 &\approx z_3 \\ \dots &\approx \dots \\ a_n &\approx z_n \end{aligned}$$



CBOW



Skip-gram

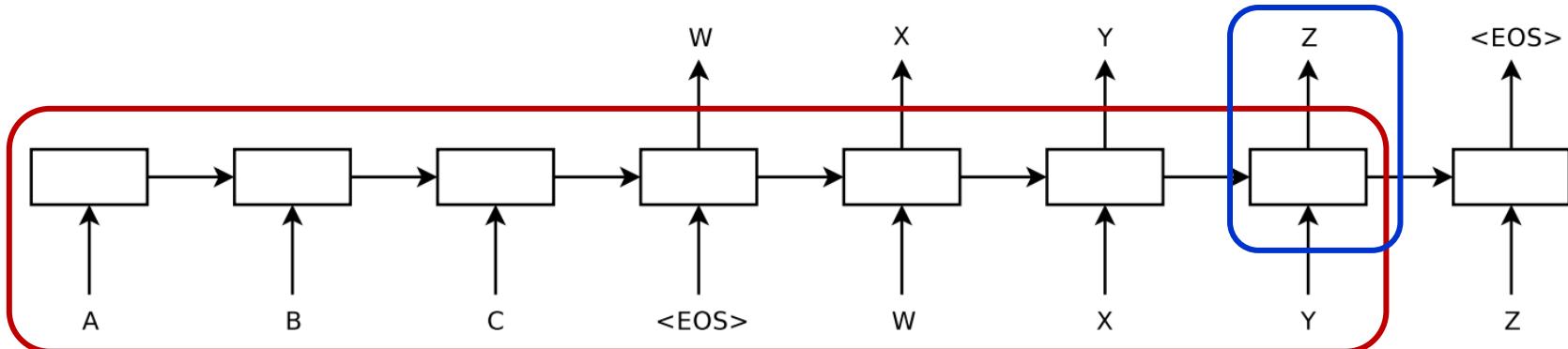


Neuronale MT



$$p(t|s) = p(t_1 \cdots t_m | s_1 \cdots s_n) = \prod_{i=1}^m p(t_i | s_1 \cdots s_{i-1}, t_1 \cdots t_n)$$

Of course, the modelling of
this by the NN is not exact ...!



Sutskever et al. 2014



- SMT/PB-SMT und Terminologie
 - Füge die Terminologie zu den Trainingsdaten hinzu

- NMT und Terminologie
 - constrained decoding (eingeschränktes ...)
 - unconstrained decoding (uneingeschränktes ...)





Thank you for your attention!





1 Einleitung

Moderne maschinelle Übersetzung wird aus Daten abgeschätzt oder gelernt. Dabei sind sowohl phrasenbasierte Statistische Maschinelle Übersetzung (PB-SMT) als auch Neuronale Maschinelle Übersetzung (NMT) Ausprägungen der statistischen Übersetzung. Trotz dieser Gemeinsamkeiten gibt es große Unterschiede zwischen der PB-SMT und der NMT. Diese Unterschiede sind dafür verantwortlich, dass nachträglich terminologische Information in unterschiedlicher Weise für PB-SMT und NMT hinzugefügt werden muss. Dabei wird immer angenommen, dass terminologische Information als *externe, rein lexikalische Ressource* – als simple Terminologiedatenbank einer Menge von lemmatisierten Termpaaren in der Ausgangs- und der Zielsprache *ohne weiteren Kontext*: $\{(\text{Masse}, \text{ground}), \dots\}^1$ – zur Verfügung steht, mit dem nachträglich ein schon bestehendes MÜ-System (PB-SMT oder NMT) an eine bestimmte Textsorte mit der ihr eigenen Terminologie angepasst werden soll. Wenn aber schon genügend *kontextualisierte parallele Trainingsdaten* (also Satzpaare) vorhanden sind, in denen die betreffende Terminologie in der Ausgangs- und Zieltextsorte entsprechend im Kontext verwendet wird, kann aus diesen Daten direkt gelernt werden, und eine nachträgliche Zufügung externer Terminologie erübrigt sich. In diesem Beitrag werden also zunächst die Gemeinsamkeiten und Unterschiede der PB-SMT und NMT skizziert. Daraus ergeben sich dann die Unterschiede hinsichtlich der nachträglichen Terminologiezufügung. Während bei der PB-SMT im einfachsten Fall die Termpaare der lexikalischen Terminologieressource den Trainingsdaten (oder Phrasentabelle) hinzugefügt werden können, führt einfaches Zufügen von Termpaaren zu NMT-Trainingsdaten in der Regel nicht zum gewünschten Erfolg. Grund dafür sind unterschiedliche Arten von Modellierung von Kontext in PB-SMT und NMT sowie das Fehlen von relevantem Kontext in einfachen, rein lexikalischen Terminologieressourcen. Daher müssen in der NMT andere Wege gegangen werden, um Terminologie erfolgreich zu integrieren: die Annotation (*Augmentierung*) von kontextualisierenden Trainingsdaten mit dann uneingeschränktem Dekodieren (*unconstrained decoding*) oder (*ohne weitere Datenannotation*) das eingeschränkte Dekodieren (*constrained decoding*).²

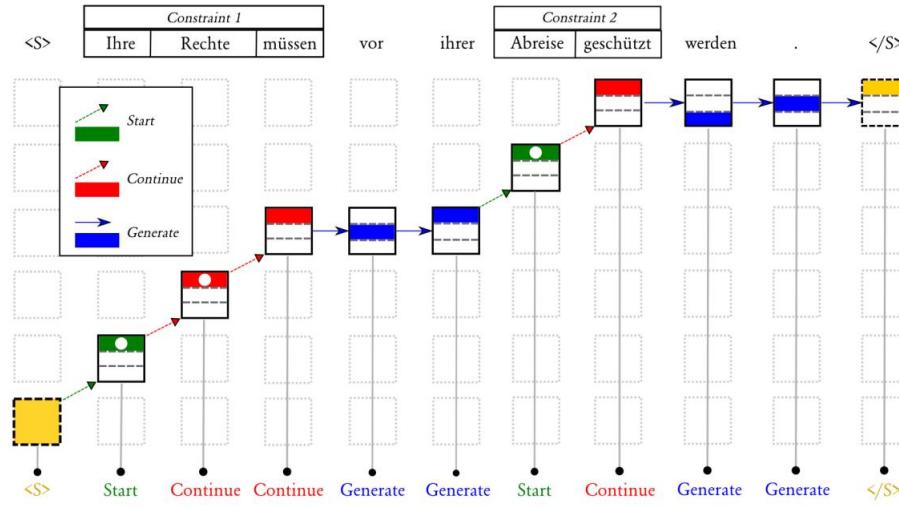


Paper Errata II



(unconstrained decoding) adressieren genau diesen Mangel durch Datenaugmentation: Zielsprachliche Terminologie wird an den entsprechenden Stellen in dem ausgangssprachlichen Teil den bilingualen, satzpaar-basierten Trainingsdaten hinzugefügt. **Eingeschränkte Dekodierungsverfahren (constrained decoding) benötigen keine Datenaugmentation, sondern benutzen Termpaare, um den NMT-Decoder zu zwingen, Terminologie in der Ausgabe zu benutzen.**

Eine existierende MT mit einer rein lexikalischen Datenbank auf eine neue Domaine/Terminologie anzupassen, ist ein komplexes Thema und im vorliegenden Aufsatz nicht erschöpfend behandelt. Der Aufsatz versucht, grundlegende Intuitionen zu erläutern, warum verschiedene Ansätze abhängig von der verwendeten MT-Technologie mehr oder weniger erfolgversprechend sind. Der Aufsatz gibt keine abschließende Evaluierung, weil die Performanz der vorgestellten Verfahren stark von der vorliegenden MT, der lexikalischen **Terminologieressource** und der Datenlage abhängig ist. Für eine stärker praxisorientierte Erörterung und Darstellung zu NMT und Terminologie sei Winter and Zielinski (2020) empfohlen.



Input: Rights protection should begin before their departure .

Figure 1: A visualization of the decoding process for an actual example from our English-German MT experiments. The output token at each timestep appears at the top of the figure, with lexical constraints enclosed in boxes. *Generation* is shown in blue, *Starting* new constraints in green, and *Continuing* constraints in red. The function used to create the hypothesis at each timestep is written at the bottom. Each box in the grid represents a beam; a colored strip inside a beam represents an individual hypothesis in the beam's k -best stack. Hypotheses with circles inside them are *closed*, all other hypotheses are *open*. (Best viewed in colour).



EN-DE	
Source He was also an anti-smoking activist and took part in several campaigns.	
Original Hypothesis Es war auch ein Anti-Rauch-Aktivist und nahmen an mehreren Kampagnen teil.	
Reference Ebenso setzte er sich gegen das Rauchen ein und nahm an mehreren Kampagnen teil.	
Constrained Hypothesis Ebenso setzte er sich gegen das Rauchen ein und nahm an mehreren Kampagnen teil.	Constraints (1) Ebenso setzte er (2) gegen das Rauchen (3) nahm
EN-FR	
Source At that point I was no longer afraid of him and I was able to love him.	
Original Hypothesis Je n'avais plus peur de lui et j'étais capable de l'aimer.	
Reference Là je n'ai plus eu peur de lui et j'ai pu l'aimer.	
Constrained Hypothesis Là je n'ai plus eu peur de lui et j'ai pu l'aimer.	Constraints (1) Là je n'ai (2) j'ai pu (3) eu
EN-PT	
Source Mo-dif-y drain-age features by selecting them individually.	
Original Hypothesis - Já temos as características de extração de idade, com eles individualmente.	
Reference Modi-fique os recursos de drenagem ao selec-ion-á-los individualmente.	
Constrained Hypothesis Modi-fique os recursos de drenagem ao selec-ion-á-los individualmente.	Constraints (1) drenagem ao selec- (2) Modi-fique os (3) recursos

Table 3: Manual analysis of examples from lexically constrained decoding experiments. “-” followed by whitespace indicates the internal segmentation of the translation model (see Section 3.2)

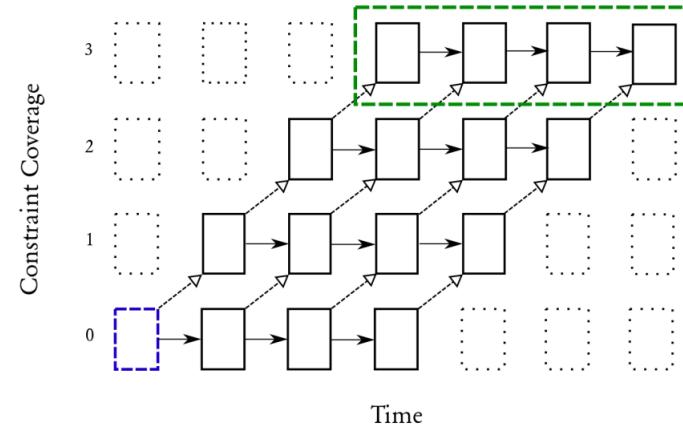


Figure 3: Visualizing the lexically constrained decoder's complete search graph. Each rectangle represents a beam containing k hypotheses. Dashed (diagonal) edges indicate *starting* or *continuing* constraints. Horizontal edges represent *generating* from the model's distribution. The horizontal axis covers the timesteps in the output sequence, and the vertical axis covers the constraint tokens (one row for each token in each constraint). Beams on the top level of the grid contain hypotheses which cover all constraints.



Training by replacing / by appending



1. Diese **Leitung** ist an den Generator anzuschließen. Connect the **cable** to the generator.
2. Diese **cable** ist an den Generator anzuschließen. Connect the **cable** to the generator.
3. Diese **Leitung # cable #** ist an den Generator anzuschließen. Connect the **cable** to the generator.

Dinu et al. 2019



PB-SMT/Moses XML input feature encoding



1. Das gelb-grüne Kabel ist Masse.
2. Das gelb-grüne Kabel ist <n translation = "ground"> Masse </n>.



Titel