Machine Translation

3 December 2012
Natural Language Processing 600.465
Guest Lecturer: Matt Post
Slides amalgamated from mt-class.org
虽然北风呼啸，但天空依然十分清澈。
虽然北风呼啸，但天空依然十分清澈。

However, the sky remained clear under the strong north wind.
The Tower of Babel

Pieter Brueghel the Elder (1563)
Georges Artsrouni’s “mechanical brain”, patented 1933 (France)
When I look at an article in Russian, I say: “This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.”

Warren Weaver (1949)
Because we want to provide everyone with access to all the world's information, including information written in every language, one of the exciting projects at Google Research is machine translation... Now you can see the results for yourself. We recently launched an online version of our system for Arabic-English and English-Arabic. Try it out!
<table>
<thead>
<tr>
<th>From: Detect language</th>
<th>To: English</th>
<th>Translate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detect language</td>
<td>English</td>
<td>Spanish</td>
</tr>
<tr>
<td>Arabic</td>
<td>Afghan</td>
<td>Afghan</td>
</tr>
<tr>
<td>Amharic</td>
<td>Albanian</td>
<td>Albanian</td>
</tr>
<tr>
<td>Arabic</td>
<td>Armenian</td>
<td>Armenian</td>
</tr>
<tr>
<td>Basque</td>
<td>Azerbaijani</td>
<td>Azerbaijani</td>
</tr>
<tr>
<td>Belarusian</td>
<td>Bengali</td>
<td>Bengali</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>Catalan</td>
<td>Catalan</td>
</tr>
<tr>
<td>Catalonian</td>
<td>Chinese</td>
<td>Chinese</td>
</tr>
<tr>
<td>Czech</td>
<td>Croatian</td>
<td>Croatian</td>
</tr>
<tr>
<td>Danish</td>
<td>Dutch</td>
<td>Dutch</td>
</tr>
<tr>
<td>Dutch</td>
<td>Estonian</td>
<td>Estonian</td>
</tr>
<tr>
<td>English</td>
<td>Filipino</td>
<td>Filipino</td>
</tr>
<tr>
<td>Filipino</td>
<td>Finnish</td>
<td>Finnish</td>
</tr>
<tr>
<td>Finnish</td>
<td>French</td>
<td>French</td>
</tr>
<tr>
<td>French</td>
<td>Galician</td>
<td>Galician</td>
</tr>
<tr>
<td>Galician</td>
<td>Georgian</td>
<td>Georgian</td>
</tr>
<tr>
<td>Georgian</td>
<td>German</td>
<td>German</td>
</tr>
<tr>
<td>German</td>
<td>Greek</td>
<td>Greek</td>
</tr>
<tr>
<td>Georgian</td>
<td>Greek</td>
<td>Greek</td>
</tr>
<tr>
<td>Greek</td>
<td>Hindi</td>
<td>Hindi</td>
</tr>
<tr>
<td>Hawaiian</td>
<td>Hungarian</td>
<td>Hungarian</td>
</tr>
<tr>
<td>Hungarian</td>
<td>Icelandic</td>
<td>Icelandic</td>
</tr>
<tr>
<td>Icelandic</td>
<td>Indonesian</td>
<td>Indonesian</td>
</tr>
<tr>
<td>Indonesian</td>
<td>Irish</td>
<td>Irish</td>
</tr>
<tr>
<td>Irish</td>
<td>Italian</td>
<td>Italian</td>
</tr>
<tr>
<td>Italian</td>
<td>Japanese</td>
<td>Japanese</td>
</tr>
<tr>
<td>Japanese</td>
<td>Kannada</td>
<td>Kannada</td>
</tr>
<tr>
<td>Kannada</td>
<td>Korean</td>
<td>Korean</td>
</tr>
<tr>
<td>Korean</td>
<td>Latvian</td>
<td>Latvian</td>
</tr>
<tr>
<td>Latvian</td>
<td>Lithuanian</td>
<td>Lithuanian</td>
</tr>
<tr>
<td>Lithuanian</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maltese</td>
<td>Maltese</td>
</tr>
<tr>
<td>Maltese</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Marathi</td>
<td>Marathi</td>
</tr>
<tr>
<td>Marathi</td>
<td>Mandarin</td>
<td>Mandarin</td>
</tr>
<tr>
<td>Mandarin</td>
<td>Marathi</td>
<td>Marathi</td>
</tr>
<tr>
<td>Marathi</td>
<td>Mango</td>
<td>Mango</td>
</tr>
<tr>
<td>Mango</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
<tr>
<td>Malay</td>
<td>Maori</td>
<td>Maori</td>
</tr>
<tr>
<td>Maori</td>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Malay</td>
<td>Malay</td>
</tr>
</tbody>
</table>
虽然北风呼啸，但天空依然十分清澈。

However, the sky remained clear under the strong north wind.
Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.
Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.
Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.
Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.
Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.
Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.
Although north wind howls, but the sky still very clear.

However, the sky remained clear under the strong north wind.
Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.
Although north wind howls, but sky still very clear.
Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.
Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.
Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.
Although north wind howls, but sky still very clear.

Despite the strong northerly winds, the sky remains very clear.
Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.

Despite the strong northerly winds, the sky remains very clear.
The sky was still crystal clear, though the north wind was howling.

Although a north wind was howling, the sky remained clear and blue.
Although north wind howls, but sky still very clear.

Although a north wind was howling, the sky remained clear and blue.
Questions

• How can we formalize knowledge acquisition from text as an algorithm?
• How can we formalize decoding as an algorithm?
• How well does it work?
Develop a statistical \textit{model} of translation that can be \textit{learned} from \textit{data} and used to \textit{predict} the correct English translation of new Chinese sentences.
Develop a statistical model of translation that can be learned from data and used to predict the correct English translation of new Chinese sentences.
Statistical Machine Translation
Statistical Machine Translation
Statistical Machine Translation
Statistical Machine Translation
Statistical Machine Translation
Statistical Machine Translation

- Formal language theory
- Information theory
- Linguistics
- Machine learning
- Algorithms
Statistical Machine Translation

- Formal language theory
- Information theory
- Linguistics
- Machine learning algorithms
- Machine learning
OVERVIEW

- MT is a huge topic in NLP
OVERVIEW

- MT is a huge topic in NLP

- Today and Wednesday we’ll be talking about two parts:
- MT is a huge topic in NLP

- Today and Wednesday we’ll be talking about two parts:
  
  - **Alignment** *(with IBM Model 1):* how do we automatically learn probabilistic translation dictionaries?
- MT is a huge topic in NLP

- Today and Wednesday we’ll be talking about two parts:

  - **Alignment** (with IBM Model 1): how do we automatically learn probabilistic translation dictionaries?

  - **Decoding**: given a model, how can we efficiently search over the huge space of possible translations?
- MT is a huge topic in NLP

- Today and Wednesday we’ll be talking about two parts:

  - **Alignment** (with IBM Model 1): how do we automatically learn probabilistic translation dictionaries?

  - **Decoding**: given a model, how can we efficiently search over the huge space of possible translations?

- Brings together many things you have learned about (n-gram language models, unsupervised learning, structured prediction, dynamic programming, EM, noisy channel models)
Alignment
However, the sky remained clear under the strong north wind.
However, the sky remained clear under the strong north wind.
<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Sm.</th>
<th>Lg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>57</td>
<td>House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)</td>
<td>1.50</td>
<td>2.75</td>
</tr>
<tr>
<td>58</td>
<td>Chicken Rice Soup</td>
<td>1.85</td>
<td>3.25</td>
</tr>
<tr>
<td>59</td>
<td>Chicken Noodle Soup</td>
<td>1.85</td>
<td>3.25</td>
</tr>
<tr>
<td>60</td>
<td>Cantonese Wonton Soup</td>
<td>1.50</td>
<td>2.75</td>
</tr>
<tr>
<td>61</td>
<td>Tomato Clear Egg Drop Soup</td>
<td>1.65</td>
<td>2.95</td>
</tr>
<tr>
<td>62</td>
<td>Regular Wonton Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>63</td>
<td>Hot &amp; Sour Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>64</td>
<td>Egg Drop Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>65</td>
<td>Egg Drop Wonton Mix</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>66</td>
<td>Tofu Vegetable Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>67</td>
<td>Chicken Corn Cream Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>68</td>
<td>Crab Meat Corn Cream Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>69</td>
<td>Seafood Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>No.</td>
<td>Menu Item</td>
<td>Description</td>
<td>Sm.</td>
</tr>
<tr>
<td>-----</td>
<td>---------------------------------</td>
<td>------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>57.</td>
<td>House Chicken Soup</td>
<td>(Chicken, Celery, Potato, Onion, Carrot)</td>
<td>1.50</td>
</tr>
<tr>
<td>58.</td>
<td>Chicken Rice Soup</td>
<td></td>
<td>1.85</td>
</tr>
<tr>
<td>59.</td>
<td>Chicken Noodle Soup</td>
<td></td>
<td>1.85</td>
</tr>
<tr>
<td>60.</td>
<td>Cantonese Wonton Soup</td>
<td></td>
<td>1.50</td>
</tr>
<tr>
<td>61.</td>
<td>Tomato Clear Egg Drop Soup</td>
<td></td>
<td>1.65</td>
</tr>
<tr>
<td>62.</td>
<td>Regular Wonton Soup</td>
<td></td>
<td>1.10</td>
</tr>
<tr>
<td>63.</td>
<td>Hot &amp; Sour Soup</td>
<td></td>
<td>1.10</td>
</tr>
<tr>
<td>64.</td>
<td>Egg Drop Soup</td>
<td></td>
<td>1.10</td>
</tr>
<tr>
<td>65.</td>
<td>Egg Drop Wonton Mix</td>
<td></td>
<td>1.10</td>
</tr>
<tr>
<td>66.</td>
<td>Tofu Vegetable Soup</td>
<td></td>
<td>NA</td>
</tr>
<tr>
<td>67.</td>
<td>Chicken Corn Cream Soup</td>
<td></td>
<td>NA</td>
</tr>
<tr>
<td>68.</td>
<td>Crab Meat Corn Cream Soup</td>
<td></td>
<td>NA</td>
</tr>
<tr>
<td>69.</td>
<td>Seafood Soup</td>
<td></td>
<td>NA</td>
</tr>
<tr>
<td>No.</td>
<td>Name</td>
<td>Sm.</td>
<td>Lg.</td>
</tr>
<tr>
<td>-----</td>
<td>----------------------------------------------------------------------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>57.</td>
<td>House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)</td>
<td>1.50</td>
<td>2.75</td>
</tr>
<tr>
<td>58.</td>
<td>Chicken Rice Soup</td>
<td>1.85</td>
<td>3.25</td>
</tr>
<tr>
<td>59.</td>
<td>Chicken Noodle Soup</td>
<td>1.85</td>
<td>3.25</td>
</tr>
<tr>
<td>60.</td>
<td>Cantonese Wonton Soup</td>
<td>1.50</td>
<td>2.75</td>
</tr>
<tr>
<td>61.</td>
<td>Tomato Clear Egg Drop Soup</td>
<td>1.65</td>
<td>2.95</td>
</tr>
<tr>
<td>62.</td>
<td>Regular Wonton Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>63.</td>
<td>Hot &amp; Sour Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>64.</td>
<td>Egg Drop Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>65.</td>
<td>Egg Drop Wonton Mix</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>66.</td>
<td>Tofu Vegetable Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>67.</td>
<td>Chicken Corn Cream Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>68.</td>
<td>Crab Meat Corn Cream Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>69.</td>
<td>Seafood Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
</tbody>
</table>
## Classic Soups

<table>
<thead>
<tr>
<th></th>
<th>Sm.</th>
<th>Lg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>57.</td>
<td>House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)</td>
<td>1.50</td>
</tr>
<tr>
<td>58.</td>
<td>Chicken Rice Soup</td>
<td>1.85</td>
</tr>
<tr>
<td>59.</td>
<td>Chicken Noodle Soup</td>
<td>1.85</td>
</tr>
<tr>
<td>60.</td>
<td>Cantonese Wonton Soup</td>
<td>1.50</td>
</tr>
<tr>
<td>61.</td>
<td>Tomato Clear [Egg Drop] Soup</td>
<td>1.65</td>
</tr>
<tr>
<td>62.</td>
<td>Regular Wonton Soup</td>
<td>1.10</td>
</tr>
<tr>
<td>63.</td>
<td>Hot &amp; Sour Soup</td>
<td>1.10</td>
</tr>
<tr>
<td>64.</td>
<td>[Egg Drop] Soup</td>
<td>1.10</td>
</tr>
<tr>
<td>65.</td>
<td>[Egg Drop] Wonton Mix</td>
<td>1.10</td>
</tr>
<tr>
<td>66.</td>
<td>Tofu Vegetable Soup</td>
<td>NA</td>
</tr>
<tr>
<td>67.</td>
<td>Chicken Corn Cream Soup</td>
<td>NA</td>
</tr>
<tr>
<td>68.</td>
<td>Crab Meat Corn Cream Soup</td>
<td>NA</td>
</tr>
<tr>
<td>69.</td>
<td>Seafood Soup</td>
<td>NA</td>
</tr>
<tr>
<td>No.</td>
<td>Item</td>
<td>Sm.</td>
</tr>
<tr>
<td>-----</td>
<td>------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>57.</td>
<td>House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)</td>
<td>1.50</td>
</tr>
<tr>
<td>58.</td>
<td>Chicken Rice Soup</td>
<td>1.85</td>
</tr>
<tr>
<td>59.</td>
<td>Chicken Noodle Soup</td>
<td>1.85</td>
</tr>
<tr>
<td>60.</td>
<td>Cantonese Wonton Soup</td>
<td>1.50</td>
</tr>
<tr>
<td>61.</td>
<td>Tomato Clear Egg Drop Soup</td>
<td>1.65</td>
</tr>
<tr>
<td>62.</td>
<td>Regular Wonton Soup</td>
<td>1.10</td>
</tr>
<tr>
<td>63.</td>
<td>Hot &amp; Sour Soup</td>
<td>1.10</td>
</tr>
<tr>
<td>64.</td>
<td>Egg Drop Soup</td>
<td>1.10</td>
</tr>
<tr>
<td>65.</td>
<td>Egg Drop Wonton Mix</td>
<td>1.10</td>
</tr>
<tr>
<td>66.</td>
<td>Tofu Vegetable Soup</td>
<td>NA</td>
</tr>
<tr>
<td>67.</td>
<td>Chicken Corn Cream Soup</td>
<td>NA</td>
</tr>
<tr>
<td>68.</td>
<td>Crab Meat Corn Cream Soup</td>
<td>NA</td>
</tr>
<tr>
<td>69.</td>
<td>Seafood Soup</td>
<td>NA</td>
</tr>
</tbody>
</table>
# Classic Soups

<table>
<thead>
<tr>
<th>#</th>
<th>Description</th>
<th>Sm.</th>
<th>Lg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>57.</td>
<td>House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)</td>
<td>1.50</td>
<td>2.75</td>
</tr>
<tr>
<td>58.</td>
<td>Chicken Rice Soup</td>
<td>1.85</td>
<td>3.25</td>
</tr>
<tr>
<td>59.</td>
<td>Chicken Noodle Soup</td>
<td>1.85</td>
<td>3.25</td>
</tr>
<tr>
<td>60.</td>
<td>Cantonese Wonton Soup</td>
<td>1.50</td>
<td>2.75</td>
</tr>
<tr>
<td>61.</td>
<td>Tomato Clear Egg Drop Soup</td>
<td>1.65</td>
<td>2.95</td>
</tr>
<tr>
<td>62.</td>
<td>Regular Wonton Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>63.</td>
<td>Hot &amp; Sour Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>64.</td>
<td>Egg Drop Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>65.</td>
<td>Egg Drop Wonton Mix</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>66.</td>
<td>Tofu Vegetable Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>67.</td>
<td>Chicken Corn Cream Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>68.</td>
<td>Crab Meat Corn Cream Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>69.</td>
<td>Seafood Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CLASSIC SOUPS</td>
<td>Sm.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>----------------</td>
<td>-----</td>
</tr>
<tr>
<td>57.</td>
<td>清燉雞湯</td>
<td>House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)</td>
<td>1.50</td>
</tr>
<tr>
<td>58.</td>
<td>雞飯湯</td>
<td>Chicken Rice Soup</td>
<td>1.85</td>
</tr>
<tr>
<td>59.</td>
<td>雞麵湯</td>
<td>Chicken Noodle Soup</td>
<td>1.85</td>
</tr>
<tr>
<td>60.</td>
<td>廣東雲吞面</td>
<td>Cantonese Wonton Soup</td>
<td>1.50</td>
</tr>
<tr>
<td>61.</td>
<td>蕃茄蛋湯</td>
<td>Tomato Clear Egg Drop Soup</td>
<td>1.65</td>
</tr>
<tr>
<td>62.</td>
<td>雲吞湯</td>
<td>Regular Wonton Soup</td>
<td>1.10</td>
</tr>
<tr>
<td>63.</td>
<td>酸辣湯</td>
<td>Hot &amp; Sour Soup</td>
<td>1.10</td>
</tr>
<tr>
<td>64.</td>
<td>蛋花湯</td>
<td>Egg Drop Soup</td>
<td>1.10</td>
</tr>
<tr>
<td>65.</td>
<td>雲蛋湯</td>
<td>Egg Drop Wonton Mix</td>
<td>1.10</td>
</tr>
<tr>
<td>66.</td>
<td>豆腐菜湯</td>
<td>Tofu Vegetable Soup</td>
<td>NA</td>
</tr>
<tr>
<td>67.</td>
<td>雞玉米湯</td>
<td>Chicken Corn Cream Soup</td>
<td>NA</td>
</tr>
<tr>
<td>68.</td>
<td>螃蝗玉米湯</td>
<td>Crab Meat Corn Cream Soup</td>
<td>NA</td>
</tr>
<tr>
<td>69.</td>
<td>海鮮湯</td>
<td>Seafood Soup</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Menu Item</td>
<td>Size</td>
<td>Size</td>
</tr>
<tr>
<td>----</td>
<td>----------------------------------------------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>57</td>
<td>House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)</td>
<td>1.50</td>
<td>2.75</td>
</tr>
<tr>
<td>58</td>
<td>Chicken Rice Soup</td>
<td>1.85</td>
<td>3.25</td>
</tr>
<tr>
<td>59</td>
<td>Chicken Noodle Soup</td>
<td>1.85</td>
<td>3.25</td>
</tr>
<tr>
<td>60</td>
<td>Cantonese Wonton Soup</td>
<td>1.50</td>
<td>2.75</td>
</tr>
<tr>
<td>61</td>
<td>Tomato Clear Egg Drop Soup</td>
<td>1.65</td>
<td>2.95</td>
</tr>
<tr>
<td>62</td>
<td>Regular Wonton Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>63</td>
<td>Hot &amp; Sour Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>64</td>
<td>Egg Drop Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>65</td>
<td>Egg Drop Wonton Mix</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>66</td>
<td>Tofu Vegetable Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>67</td>
<td>Chicken Corn Cream Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>68</td>
<td>Crab Meat Corn Cream Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>69</td>
<td>Seafood Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>No.</td>
<td>Item</td>
<td>Description</td>
<td>Sm.</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>57</td>
<td>House Chicken Soup</td>
<td>(Chicken, Celery, Potato, Onion, Carrot)</td>
<td>1.50</td>
</tr>
<tr>
<td>58</td>
<td>Chicken Rice Soup</td>
<td></td>
<td>1.85</td>
</tr>
<tr>
<td>59</td>
<td>Chicken Noodle Soup</td>
<td></td>
<td>1.85</td>
</tr>
<tr>
<td>60</td>
<td>Cantonese Wonton Soup</td>
<td></td>
<td>1.50</td>
</tr>
<tr>
<td>61</td>
<td>Tomato Clear Soup</td>
<td>Egg Drop</td>
<td>1.65</td>
</tr>
<tr>
<td>62</td>
<td>Hot &amp; Sour Soup</td>
<td></td>
<td>1.10</td>
</tr>
<tr>
<td>63</td>
<td>Tofu Vegetable Soup</td>
<td></td>
<td>1.10</td>
</tr>
<tr>
<td>64</td>
<td>Egg Drop Soup</td>
<td></td>
<td>1.10</td>
</tr>
<tr>
<td>65</td>
<td>Wonton Mix</td>
<td>Egg Drop</td>
<td>1.10</td>
</tr>
<tr>
<td>66</td>
<td>Chicken Corn Cream Soup</td>
<td></td>
<td>NA</td>
</tr>
<tr>
<td>67</td>
<td>Crab Meat Corn Cream Soup</td>
<td></td>
<td>NA</td>
</tr>
<tr>
<td>68</td>
<td>Seafood Soup</td>
<td></td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>CLASSIC SOUPS</td>
<td>Sm.</td>
<td>Lg.</td>
</tr>
<tr>
<td>---</td>
<td>------------------------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>57</td>
<td>House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)</td>
<td>1.50</td>
<td>2.75</td>
</tr>
<tr>
<td>58</td>
<td>Chicken Rice Soup</td>
<td>1.85</td>
<td>3.25</td>
</tr>
<tr>
<td>59</td>
<td>Chicken Noodle Soup</td>
<td>1.85</td>
<td>3.25</td>
</tr>
<tr>
<td>60</td>
<td>Cantonese Wonton Soup</td>
<td>1.50</td>
<td>2.75</td>
</tr>
<tr>
<td>61</td>
<td>Tomato Clear Egg Drop Soup</td>
<td>1.65</td>
<td>2.95</td>
</tr>
<tr>
<td>62</td>
<td>Regular Wonton Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>63</td>
<td>Hot &amp; Sour Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>64</td>
<td>Egg Drop Soup</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>65</td>
<td>Egg Drop Wonton Mix</td>
<td>1.10</td>
<td>2.10</td>
</tr>
<tr>
<td>66</td>
<td>Tofu Vegetable Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>67</td>
<td>Chicken Corn Cream Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>68</td>
<td>Crab Meat Corn Cream Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
<tr>
<td>69</td>
<td>Seafood Soup</td>
<td>NA</td>
<td>3.50</td>
</tr>
</tbody>
</table>
Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.
Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.

\[ p(\text{English}|\text{Chinese})? \]
Although north wind howls, but sky still very clear.

IBM Model 1
虽然北风呼啸，但天空依然十分清澈。
Although north wind howls, but sky still very clear.

Although north wind howls, but sky still very clear.

IBM Model 1

IBM Model 1
Although north wind howls, but sky still very clear.

虽然北风呼啸，但天空依然十分清澈。
Although north wind howls, but sky still very clear.

虽然北风呼啸，但天空依然十分清澈。
Although north wind howls, but sky still very clear.
Although north wind howls, but sky still very clear.

虽然北风呼啸，但天空依然十分清澈。
Although north wind howls, but sky still very clear.
Although north wind howls, but sky still very clear.

However
虽然北风呼啸，但天空依然十分清澈。
Although north wind howls, but sky still very clear.

However
Although north wind howls, but sky still very clear.

Although north wind howls, but sky still very clear.

However
IBM Model 1

Although north wind howls, but sky still very clear.

虽然北风呼啸，但天空依然十分清澈。

However,
Although north wind howls, but sky still very clear.
虽然北风呼啸，但天空依然十分清澈。

However, the

Although north wind howls, but sky still very clear.
Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.
IBM Model 1
IBM Model 1

• Word translation probabilities.
IBM Model 1

- Word translation probabilities.
- No real ordering model.
- This is left to the LM.
IBM Model 1

- Word translation probabilities.
- No real ordering model.
- This is left to the LM.
IBM Model 1

- Word translation probabilities.
- No real ordering model.
- This is left to the LM.
$p(\texttt{despite} | \text{虽然})$

$p(\texttt{however} | \text{虽然})$

$p(\texttt{although} | \text{虽然})$

$p(\texttt{northern} | \text{北})$

$p(\texttt{north} | \text{北})$
IBM Model 1

\[ p(\text{despite} | \text{虽然}) \]

\[ p(\text{however} | \text{虽然}) \]

\[ p(\text{although} | \text{虽然}) \]

\[ p(\text{northern} | \text{北}) \]

\[ p(\text{north} | \text{北}) \]
Although north wind howls, but sky still very clear.

虽然北风呼啸，但天空依然十分清澈。

However, the sky remained clear under the strong north wind.

\[ p(\text{however} \mid \text{虽然}) = \frac{\text{# of times 虽然 aligns to However}}{\text{# of times 虽然 occurs}} \]
Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.

$p(\text{however} | \text{虽然}) = \frac{\text{# of times 虽然 aligns to However}}{\text{# of times 虽然 occurs}}$
Likelihood Estimation for Model 1

Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.
Likelihood Estimation for Model 1

Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.

$p(\text{English length} | \text{Chinese length})$ observed
Likelihood Estimation for Model 1

Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind. $p(\text{Chinese word position})$ uniform, no need to estimate.
Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.

$p(\text{English word}|\text{Chinese word})$ unobserved!
Likelihood Estimation for Model 1

Although north wind howls, but sky still very clear.

Although north wind howls, but sky still very clear.

the missing alignment is a *latent variable*

However, the sky remained clear under the strong north wind.

\[ p(\text{English word}|\text{Chinese word}) \] unobserved!
Likelihood Estimation for Model 1

Although north wind howls, but sky still very clear.

Parameters and alignments are both unknown.

However, the sky remained clear under the strong north wind.

\[ p(\text{English word}|\text{Chinese word}) \] unobserved!
Likelihood Estimation for Model 1

Although north wind howls, but sky still very clear.

Parameters and alignments are both unknown.

If we knew the alignments, we could calculate the values of the parameters.

However, the sky remained clear under the strong north wind.

\[ p(\text{English word}|\text{Chinese word}) \quad \text{unobserved!} \]
Although north wind howls, but sky still very clear. Although north wind howls, but sky still very clear.

Parameters and alignments are both unknown.

Parameters and alignments are both unknown.

If we knew the alignments, we could calculate the values of the parameters.

If we knew the alignments, we could calculate the values of the parameters.

If we knew the parameters, we could calculate the likelihood of the data.

If we knew the parameters, we could calculate the likelihood of the data.

However, the sky remained clear under the strong north wind.

However, the sky remained clear under the strong north wind.

\( p(\text{English word} | \text{Chinese word}) \) unobserved!
Likelihood Estimation for Model 1

Although north wind howls, but sky still very clear.

Parameters and alignments are both unknown.

If we knew the alignments, we could calculate the values of the parameters.

If we knew the parameters, we could calculate the likelihood of the data.

However, the sky remained clear under the strong north wind.

$p(\text{English word}|\text{Chinese word})$ unobserved!
The Plan: Bootstrapping

- Arbitrarily select a set of parameters (say, uniform).
- Calculate *expected counts* of the unseen events.
- Choose new parameters to maximize likelihood, using expected counts as proxy for observed counts.
- Iterate.
- Guarantee: likelihood will be monotonically nondecreasing.
The Plan: Bootstrapping

Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.
The Plan: Bootstrapping

Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.

if we had observed the alignment, this line would either be here (count 1) or it wouldn’t (count 0).
The Plan: Bootstrapping

Although north wind howls, but sky still very clear.

if we had observed the alignment, this line would either be here (count 1) or it wouldn’t (count 0).

since we didn’t observe the alignment, we calculate the probability that it’s there.

However, the sky remained clear under the strong north wind.
虽然 北 风 呼啸，但 天空 依然 十分 清澈。

However, the sky remained clear under the strong north wind.

Marginalize: sum all alignments containing the link
Divide by sum of all possible alignments

\[ p(\text{虽然北风呼啸，但天空依然十分清澈。}) + \]

However, the sky remained clear under the strong north wind.

\[ p(\text{虽然北风呼啸，但天空依然十分清澈。}) + \]

However, the sky remained clear under the strong north wind.

\[ p(\text{虽然北风呼啸，但天空依然十分清澈。}) \]

However, the sky remained clear under the strong north wind.
Divide by sum of all possible alignments

Is this hard? How many alignments are there?
Expectation Maximization

probability of an alignment.

\[ p(F, A|E) = p(I|J) \prod_{a_i} p(a_i = j) p(f_i|e_j) \]
Expectation Maximization

probability of an alignment.

\[ p(F, A|E) = p(I|J) \prod_{a_i} p(a_i = j)p(f_i|e_j) \]

observed uniform
Expectation Maximization

probability of an alignment.

$$p(F, A | E) = p(I | J) \prod_{a_i} p(a_i = j)p(f_i | e_j)$$

factors across words.

observed

uniform
Expectation Maximization

\[ \sum_{a \in A: \text{北} \leftrightarrow \text{north}} p(\text{north} | \text{北}) \cdot p(\text{rest of } a) \]

marginal probability of alignments containing link
Expectation Maximization

marginal probability of alignments containing link

\[ p(north|\text{北}) \sum_{a \in A: \text{北} \leftrightarrow \text{north}} p(\text{rest of } a) \]
Expectation Maximization

marginal probability of alignments containing link

\[
p(north|\text{北}) \sum_{a \in A: \text{北} \leftrightarrow \text{north}} p(\text{rest of } a)
\]

\[
\sum_{c \in \text{Chinese words}} p(north|c) \sum_{a \in A: c \leftrightarrow \text{north}} p(\text{rest of } a)
\]

marginal probability of all alignments
Expectation Maximization

marginal probability of alignments containing link

\[
p(north|\text{北}) \sum_{a \in A: \text{北} \leftrightarrow north} p(\text{rest of } a)
\]

\[
\sum_{c \in \text{Chinese words}} p(north|c) \sum_{a \in A: c \leftrightarrow north} p(\text{rest of } a)
\]

identical!

marginal probability of all alignments
Expectation Maximization

\[
p(north | 北) = \frac{\sum_{c \in \text{Chinese words}} p(north | c)}{}
\]
Expectation Maximization

marginal probability (expected count) of an alignment containing the link

\[
p(north \mid 北) \\
\frac{\sum_{c \in \text{Chinese words}} p(north \mid c)}{}
\]
Expectation Maximization

marginal probability (expected count) of an alignment containing the link

\[
p(north | \text{北}) \frac{1}{\sum_{c \in \text{Chinese words}} p(north | c)}
\]

For each sentence, use this quantity instead of 0 or 1
Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.

\[ p(\text{however} | \text{虽然}) = \frac{\text{# of times 虽然 aligns to However}}{\text{# of times 虽然 occurs}} \]
Unsupervised Case

Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.

$p(\text{however} \mid \text{虽然}) = \frac{\text{Expected } \# \text{ of times 虽然 aligns to However}}{\# \text{ of times 虽然 occurs}$
Expectation Maximization

Why does this even work?

\[
p(\text{north} | \text{北}) \frac{\sum_{c \in \text{Chinese words}} p(\text{north} | c)}{\sum_{c \in \text{Chinese words}} p(\text{north} | c)}
\]
Expectation Maximization

Observation 1: We are still solving a maximum likelihood estimation problem.
Expectation Maximization

Observation 1: We are still solving a maximum likelihood estimation problem.

\[ p(\text{Chinese}|\text{English}) = \sum_{\text{alignments}} p(\text{Chinese}, \text{alignment}|\text{English}) \]
Observation 1: We are still solving a maximum likelihood estimation problem.

\[ p(\text{Chinese}|\text{English}) = \sum_{\text{alignments}} p(\text{Chinese}, \text{alignment}|\text{English}) \]

MLE: choose parameters that maximize this expression.
Expectation Maximization

Observation 1: We are still solving a maximum likelihood estimation problem.

\[ p(\text{Chinese}|\text{English}) = \sum_{\text{alignments}} p(\text{Chinese}, \text{alignment}|\text{English}) \]

MLE: choose parameters that maximize this expression.

Minor problem: there is no analytic solution.
Expectation Maximization

Observation 1: We are still solving a maximum likelihood estimation problem.

$$p(\text{Chinese}|\text{English}) = \sum_{\text{alignments}} p(\text{Chinese}, \text{alignment}|\text{English})$$

MLE: choose parameters that maximize this expression.

Minor problem: there is no analytic solution.

Remember: likelihood is monotonically non-decreasing!
... and, likelihood is *convex* for this model:
Likelihood Estimation for Model 1

Although north wind howls, but sky still very clear.

However, the sky remained clear under the strong north wind.

What are some things this model doesn’t account for?
NOT DISCUSSED
NOT DISCUSSED

- What’s wrong with Model 1?
NOT DISCUSSED

- What’s wrong with Model 1?
- Higher IBM Models
- What’s wrong with Model 1?

- Higher IBM Models

  - Model 2: includes a term for the probability of alignment positions (spot i linked with spot j)
- What’s wrong with Model 1?

- Higher IBM Models
  - Model 2: includes a term for the probability of alignment positions (spot i linked with spot j)
  - Model 3: models fertility, i.e., how many foreign words were produced by each English word (but independently!)
What’s wrong with Model 1?

Higher IBM Models

Model 2: includes a term for the probability of alignment positions (spot i linked with spot j)

Model 3: models fertility, i.e., how many foreign words were produced by each English word (but independently!)

For decoding, these models are superseded by more general phrase-based models (but are still used for alignment)
Summary

• We can formulate learning as an optimization problem: choose parameters that optimize some function, such as likelihood.

• Supervised: maximum likelihood.
  • Beware of overfitting.

• Unsupervised: expectation maximization.

• Many, many, many other algorithms.
Decoding
However, the sky remained clear under the strong north wind.
However, the sky remained clear under the strong north wind.
DECODING

- The process of producing a translation of a sentence
DECODING

- The process of producing a translation of a sentence

- *input*: foreign-language sentence and a model
Decoding

- The process of producing a translation of a sentence

  - *input*: foreign-language sentence and a model

  - *task*: find the model’s favorite translation
The process of producing a translation of a sentence

- *input*: foreign-language sentence and a model

- *task*: find the model’s favorite translation

- Two main problems:
DECODING

- The process of producing a translation of a sentence
  
  - *input*: foreign-language sentence and a model
  
  - *task*: find the model’s favorite translation

- Two main problems:

  - **modeling** How do we score translations?
Decoding

- The process of producing a translation of a sentence
  - input: foreign-language sentence and a model
  - task: find the model’s favorite translation

- Two main problems:
  - modeling How do we score translations?
  - search How do we find the model’s preferred translation?
PROBLEM 1: MODELING

- The process of producing a translation of a sentence
PROBLEM 1: MODELING

- The process of producing a translation of a sentence

- Two main problems:
PROBLEM 1: MODELING

- The process of producing a translation of a sentence

- Two main problems:
  
  - **modeling** – given a pair of sentences, how do we assign a probability to them?
Problem 1: Modeling

- The process of producing a translation of a sentence

- Two main problems:

  - **modeling** – given a pair of sentences, how do we assign a probability to them?

\[
P(C \rightarrow E) = \text{high}
\]

- They still lack experience in international competitions

- 他们还缺乏国际比赛的经验.
**Problem 1: Modeling**

- The process of producing a translation of a sentence

- Two main problems:

  - **modeling** – given a pair of sentences, how do we assign a probability to them?

\[
P_{(C \to E)} \left( \text{This is not a good translation of the above sentence.} \right) \approx \text{low}
\]
- Noisy Channel model \[ P(e \mid f) \propto P(f \mid e)P(e) \]
MODEL

- Noisy Channel model

\[ P(e \mid f) \propto P(f \mid e)P(e) \]
- Add weights

\[ P(e \mid f) \propto P(f \mid e)P(e) \]

\[ \propto P(f \mid e)^{\lambda_1}P(e)^{\lambda_2} \]
WEIGHTS
WEIGHTS

- Why?
- Why?

- Just like in real life, where we trust people’s claims differently, we will want to learn how to trust different models
- Why?

- Just like in real life, where we trust people’s claims differently, we will want to learn how to trust different models.

“\textit{I can do a backflip off this pommel horse}”
- Why?

- Just like in real life, where we trust people’s claims differently, we will want to learn how to trust different models.

“I can do a backflip off this pommel horse”
- Log space transform

\[ P(e \mid f) \propto P(f \mid e) P(e) \]
\[ \propto P(f \mid e)^{\lambda_1} P(e)^{\lambda_2} \]
\[ = \lambda_1 \log P(f \mid e) + \lambda_2 \log P(e) \]
- Log space transform

\[
P(e \mid f) \propto P(f \mid e)P(e)
\]

\[
\propto P(f \mid e)^{\lambda_1}P(e)^{\lambda_2}
\]

\[
= \lambda_1 \log P(f \mid e) + \lambda_2 \log P(e)
\]

- Because:

\[
0.0001 \times 0.0001 \times 0.0001 = 0.000000000001
\]

\[
\log(0.0001) + \log(0.0001) + \log(0.0001) = -12
\]
MODEL TRANSFORMS

- Generalization (linearly-weighted models)

\[
P(e | f) \propto P(f | e)P(e) \\
\propto P(f | e)^{\lambda_1} P(e)^{\lambda_2} \\
= \lambda_1 \log P(f | e) + \lambda_2 \log P(e) \\
= \lambda_1 \phi_1(f, e) + \lambda_2 \phi_2(f, e) \\
= \sum_{i} \lambda_i \phi_i(f, e)
\]
PROBLEM 2: SEARCH

- The process of producing a translation of a sentence
PROBLEM 2: SEARCH

- The process of producing a translation of a sentence

- The second problem is search – given a model and a source sentence, how do we find the sentence that the model likes best?
PROBLEM 2: SEARCH

- The process of producing a translation of a sentence

- The second problem is search – given a model and a source sentence, how do we find the sentence that the model likes best?

  - impractical: enumerate all sentences, score them
**PROBLEM 2: search**

- The process of producing a translation of a sentence

- The second problem is **search** – given a model and a source sentence, how do we find the sentence that the model likes best?

  - *impractical*: enumerate all sentences, score them

  - *stack decoding*: use factored models and assemble the translations piece by piece
- The solution: factorized models and dynamic programming
- The solution: factorized models and dynamic programming

- We employ models that factorize across pieces of the translation process
- The solution: factorized models and dynamic programming

- We employ models that factorize across pieces of the translation process
  
  - *n-gram language model*: gives us probabilities for pairs of words
- The solution: factorized models and dynamic programming

- We employ models that factorize across pieces of the translation process

  - *n-gram language model*: gives us probabilities for pairs of words

  - *Model 1 translation model*: gives us probabilities for translation pairs
\[ \phi_1(e, a, c) = \log P(e, a \mid c) \]
\[ \phi_1(e, a, c) = \log P(e, a \mid c) \]
MODEL: TRANSLATION MODEL

\[ \phi_1(e, a, c) = \log P(e, a \mid c) \]

c: 他们还缺乏国际比赛的经验.

a: 

e: They still lack experience in international competitions.

good translation
MODEL: TRANSLATION MODEL

\[ \phi_1(e, a, c) = \log P(e, a \mid c) \]

They still lack experience in international competitions.

lower translation probability
**MODEL: TRANSLATION MODEL**

\[ \phi_1(e, a, c) = \log P(e, a | c) \]

<table>
<thead>
<tr>
<th>c:</th>
<th>他们还缺乏国际比赛的经验。</th>
</tr>
</thead>
<tbody>
<tr>
<td>a:</td>
<td>decoder</td>
</tr>
<tr>
<td>e:</td>
<td>They still lack experience in international competitions.</td>
</tr>
</tbody>
</table>

lower translation probability
\[ \phi_2(e, a, c) = \log P(e) \]
MODEL: LANGUAGE MODEL

\[ \phi_2(e, a, c) = \log P(e) \]

c: 他们还缺乏国际比赛的经验.

a: They still lack experience in international competitions.

e: fluent English
They still lack experience in international competitions.

fluent English
They still lack experience in international competitions.

\[ \phi_2(e, a, c) = \log P(e) \]
They still lack experience in international competitions.
An Example
- Translate the sentence

   *Yo tengo hambre*

from Spanish to English.
- Translate the sentence

  *Yo tengo hambre*

  from Spanish to English.

- We’ll use an algorithm called *stack decoding* to efficiently explore the space of translations.
Translate the sentence

*Yo tengo hambre*

from Spanish to English.

We’ll use an algorithm called *stack decoding* to efficiently explore the space of translations.

- (Note: this has nothing to do with stacks; a better name would be *priority queue decoding*)
FACTORSING MODELS

- Translating *Yo tengo hambre*
FACTORYING MODELS

- Translating *Yo tengo hambre*

- “Stack” decoding works by extending hypotheses word by word

\[
\begin{array}{cccc}
<s> & I & \rightarrow am & <s> & I am
\end{array}
\]
- Translating *Yo tengo hambre*

- “Stack” decoding works by extending hypotheses word by word

- At each step, we take an existing hypothesis, grab an untranslated word, translate it, and extend the hypothesis
- Translating *Yo tengo hambre*

- “Stack” decoding works by extending hypotheses word by word

- At each step, we take an existing hypothesis, grab an untranslated word, translate it, and extend the hypothesis

- These can be arranged into a *search graph* representing the space we search
FACTORY MODELS
FACTORIZING MODELS
FACTORIZING MODELS
Yo → I

tengo → am

I

< has

I

am

have

have hunger

hungry

hungry am

hunger

am I have hunger

am I hungry

am I hungry am

hunger am
FACTORIZING MODELS
FACTORYING MODELS

Yo → I

tengo → am

hambre → hungry

I have hunger

I am hungry

I have hungry

I am hungry I

have hunger

hunger am I
- Stack decoding works by extending hypotheses word by word

\[
\begin{array}{c}
\langle s \rangle \text{I} \\
\circ \circ \circ \\
\end{array}
+ 
\begin{array}{c}
tengo \\
\rightarrow \text{am} \\
\end{array}
= 
\begin{array}{c}
\langle s \rangle \text{I am} \\
\circ \circ \circ \\
\end{array}
\]
FACTORIZING MODELS

- Stack decoding works by extending hypotheses word by word

- These can be arranged into a search graph representing the space we search
FACTORING MODELS

- Stack decoding works by extending hypotheses word by word

- These can be arranged into a search graph representing the space we search

- The component models we use need to factorize over this graph, and we accumulate the score as we go
- Example hypothesis creation:

\[<s> \text{I tengo} \rightarrow \text{am} = <s> \text{I am}\]

old hypothesis + add word = new hypothesis
- Example hypothesis creation:

- translation model: trivial case, since all the words are translated independently

\[
\text{hypothesis.score} += \log(P_{TM}(\text{am} | \text{tengo}))
\]
**Factoring models**

- Example hypothesis creation:

  \[
  \text{hypothesis.score } += \log(P_{TM}(am \mid tengo))
  \]

- **translation model**: trivial case, since all the words are translated independently

- a function of just the word that is added
- Example hypothesis creation:

\[
\text{old hypothesis} + \text{add word} = \text{new hypothesis}
\]

- **translation model**: trivial case, since all the words are translated independently

\[
\text{hypothesis.score} += \log(P_{TM}(\text{am} \mid \text{tengo}))
\]

- a function of just the word that is added
FACTORIZING MODELS

old hypothesis  +  tengo → am  =  new hypothesis

add word
FACTORS MODELS

- Example hypothesis creation:

\[
<\text{s}> \text{I} \quad + \quad \text{tengo} \quad \rightarrow \quad \text{I am} \quad = \quad <\text{s}> \text{I am}
\]

old hypothesis + add word = new hypothesis
- Example hypothesis creation:

\[
\text{old hypothesis} + \text{add word} = \text{new hypothesis}
\]

- **language model**: still easy, since (bigram) language models depend only on the previous word

\[
\text{hypothesis.score} += \log(P_{LM}(am \mid I))
\]
- Example hypothesis creation:

\[
<\text{s}> \text{I} \quad \rightarrow \quad \text{tengo} \rightarrow \text{am}
\]

old hypothesis + add word = new hypothesis

- **language model**: still easy, since (bigram) language models depend only on the previous word

\[
\text{hypothesis.score} += \log(P_{\text{LM}}(\text{am} | \text{I}))
\]

- a function of the old hyp. and the new word translation
- **Example hypothesis creation:**

\[
\begin{align*}
\text{old hypothesis} & \quad + \quad \text{tengo} \rightarrow \text{am} \\
\text{add word} & \quad = \\
\text{new hypothesis}
\end{align*}
\]

- **Language model:** still easy, since (bigram) language models depend only on the previous word

\[
\text{hypothesis.score} += \log(P_{LM}(am \mid I))
\]

- a function of the old hyp. and the new word translation
- Start with a list of hypotheses, containing only the empty hypothesis
STACK DECODING

- Start with a list of hypotheses, containing only the empty hypothesis

- For each stack
STACK DECODING

- Start with a list of hypotheses, containing only the empty hypothesis

- For each stack

  - For each hypothesis
STACK DECODING

- Start with a list of hypotheses, containing only the empty hypothesis

- For each stack

  - For each hypothesis

    - For each applicable word
STACK DECODING

- Start with a list of hypotheses, containing only the empty hypothesis

- For each stack

  - For each hypothesis

    - For each applicable word

      - Extend the hypothesis with the word
STACK DECODING

- Start with a list of hypotheses, containing only the empty hypothesis

- For each stack

  - For each hypothesis

    - For each applicable word

      - Extend the hypothesis with the word

    - Place the new hypothesis on the right stack
- We saw Tuesday how huge the search space could get
- We saw Tuesday how huge the search space could get

- Notice anything here?

\[
\text{old hypothesis} + \text{add word} = \text{new hypothesis}
\]

\[
\text{score } += P_{TM}(am \mid tengo) + P_{LM}(am \mid I)
\]
- We saw Tuesday how huge the search space could get

- Notice anything here?

- (1) <s> is never used in computing the scores AND
(2) <s> is implicit in the graph structure

\[
\text{score} += P_{TM}(am | \text{tengo}) + P_{LM}(am | I)
\]
- We saw Tuesday how huge the search space could get

- Notice anything here?

- (1) <s> is never used in computing the scores AND
  (2) <s> is implicit in the graph structure

- Let’s get rid of the extra state!
DYNAMIC PROGRAMMING

- Before

- After

The score of the new hypothesis is the maximum way to compute it
STACK DECODING (WITH DP)

- Start with a list of hypotheses, containing only the empty hypothesis

- For each stack

  - For each hypothesis

    - For each applicable word

      - Extend the hypothesis with the word

      - Place the new hypothesis on the right stack
STACK DECODING (WITH DP)

- Start with a list of hypotheses, containing only the empty hypothesis

- For each stack
  
  - For each hypothesis
    
    - For each applicable word
      
      - Extend the hypothesis with the word

- Place IF either (1) no equivalent hypothesis exists or (2) this hypothesis has a higher score.
OTHER ISSUES IN DECODING
OTHER ISSUES IN DECODING

- Pruning
OTHER ISSUES IN DECODING

- Pruning

- Search heuristics (e.g., distortion limits)
OTHER ISSUES IN DECODING

- Pruning
- Search heuristics (e.g., distortion limits)
- Other useful feature functions
OTHER ISSUES IN DECODING

- Pruning

- Search heuristics (e.g., distortion limits)

- Other useful feature functions

- Tuning of the weights
- We’ve barely touched on issues in decoding, much less machine translation
- We’ve barely touched on issues in decoding, much less machine translation

- The recommended readings have more information; you can also see our class page at mt-class.org (which we’ll likely teach again in SP/2014)
- We’ve barely touched on issues in decoding, much less machine translation

- The recommended readings have more information; you can also see our class page at mt-class.org (which we’ll likely teach again in SP/2014)

- On Wednesday, we’ll talk about learning and using **hierarchical models** for decoding
Syntax-based Statistical Machine Translation

5 December 2012
Natural Language Processing 600.465
Guest Lecturer: Matt Post
Slides amalgamated from mt-class.org
Goals

• Understand why *syntax is important* for reordering models
  – Review *non-syntactic reordering models* for phrase-based machine translation
  – Review the “*Clause Restructuring*” approach of Collins, Koehn, and Kucerova, its advantages and limitations

• Learn about *Synchronous Context Free Grammars*
  – Introduce *notation*, and *basic algorithm*

• Understand how we *learn SCFGs from bitexts*

• Get a sense of the different flavors of SCFGs
Why syntax matters
• Foreign input is segmented in phrases
• Each phrase is translated into English
• Phrases are reordered
Some Reordering Already Captured

• Local reordering can be captured within phrases
Phrase translation table

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

| Source    | Translation     | Probability $\phi(e|f)$ |
|-----------|-----------------|-------------------------|
| natuerlich| of course       | 0.5                     |
| natuerlich| naturally       | 0.3                     |
| natuerlich| of course ,     | 0.15                    |
| natuerlich| , of course ,   | 0.05                    |
Probabilistic Model

• Bayes rule

$$e_{\text{best}} = \arg\max_e p(e|f)$$

$$= \arg\max_e p(f|e) p_{\text{lm}}(e)$$

– translation model $p(e|f)$

– language model $p_{\text{lm}}(e)$

• Reordering score can be incorporated in the TM

$$p(f^I_1|e^I_1) = \prod_{i=1}^{I} \phi(f_i|e_i) d(\text{start}_i - \text{end}_{i-1} - 1)$$

– phrase translation probability $\phi$

– reordering probability $d$
Log-linear model

\[ p(e, a | f) = \exp(\lambda_\phi \sum_{i=1}^{I} \log \phi(f_i | e_i) + \lambda_d \sum_{i=1}^{I} \log d(a_i - b_{i-1} - 1) + \lambda_{LM} \sum_{i=1}^{|e|} \log p_{LM}(e_i | e_1 \ldots e_{i-1})) \]
Distance-based Reordering

Scoring function: \( d(x) = \alpha^{|x|} \) – exponential with distance
Values of $\alpha$
Discussion: Distance-based reordering

• What do you think of it?
• Is it a good model for how reordering works across languages?
• What is it missing?

(Discuss with your neighbor)
Distance-based reordering

• Small values of $\alpha$, severely discourage reordering
  – Limit reordering to monotonic or a narrow window
  – OK for languages with very similar word orders
  – Bad for languages with different word orders

• The distance-based penalty applies uniformly to all words and all word types
  – Doesn’t know that adjectives and nouns should swap when translating from French to English

• Puts most responsibility on the language model
How else could we model reordering?

- Why not assign a distinct reordering probability to each word/phrase in the phrase table? 
  \[ p(\text{reorder} \mid f, e) \]
- This is known as **lexicalized reordering**
- How can we estimate that probability?
Lexicalized Reordering model

<table>
<thead>
<tr>
<th>Wieviel</th>
<th>sollte</th>
<th>man</th>
<th>aufgrund</th>
<th>seines</th>
<th>Profils</th>
<th>in</th>
<th>Facebook</th>
<th>verdienen</th>
</tr>
</thead>
<tbody>
<tr>
<td>How</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>much</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>should</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>you</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>charge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>for</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>your</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facebook</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>profile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Lexicalized Reordering model

<table>
<thead>
<tr>
<th>Wieviel</th>
<th>sollte</th>
<th>man</th>
<th>aufgrund</th>
<th>seines</th>
<th>Profs</th>
<th>in</th>
<th>Facebook</th>
<th>verdienen</th>
</tr>
</thead>
<tbody>
<tr>
<td>How much</td>
<td>you should charge</td>
<td>for your Facebook profile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**m**: monotone (keep order)
Lexicalized Reordering model

$m$: monotone (keep order)
$s$: swap order

<table>
<thead>
<tr>
<th>Wieviel</th>
<th>sollte</th>
<th>man</th>
<th>aufgrund</th>
<th>seines</th>
<th>Profis</th>
<th>in</th>
<th>Facebook</th>
<th>verdienen</th>
</tr>
</thead>
<tbody>
<tr>
<td>How</td>
<td>much</td>
<td>should you charge for your Facebook profile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Lexicalized Reordering model

\[ \text{Wie viel man außergewöhnlich seines Profils in Facebook verdienen} \]

\textbf{m: monotone (keep order)}

\textbf{s: swap order}

\textbf{d: become discontinuous}

How much should you charge for your Facebook profile?
Lexicalized Reordering model

- **m**: monotone (keep order)
- **s**: swap order
- **d**: become discontinuous

Reordering features are probability estimates of s, d, and m.

---

<table>
<thead>
<tr>
<th>Wieviel</th>
<th>sollte</th>
<th>man</th>
<th>aufgrund</th>
<th>seines</th>
<th>Profils</th>
<th>in</th>
<th>Facebook</th>
<th>verdienen</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

How much should you charge for your Facebook profile?
Lexicalized Reordering table

- Identical phrase pairs \(<f,e>\) as in the phrase translation table
- Contains values for \(p(\text{monotone}|e,f)\), \(p(\text{swap}|e,f)\), \(p(\text{discontinuous}|e,f)\)

| Source      | Translation        | \(p(m|e,f)\) | \(p(s|e,f)\) | \(p(d|e,f)\) |
|-------------|--------------------|---------------|---------------|---------------|
| naturlich   | of course          | 0.52          | 0.08          | 0.40          |
| naturlich   | naturally          | 0.42          | 0.10          | 0.48          |
| naturlich   | of course ,        | 0.50          | 0.001         | 0.499         |
| naturlich   | , of course        | 0.27          | 0.17          | 0.56          |
Discussion: Is this better?

• Do you think that this is a more sensible reordering model than the distance-based one?
• How could you determine if it is better or not?
• What do you think that it still lacks?

(Discuss with your neighbor)
Empirically, yes!

Koehn et al, IWSLT 2005
“The Germans have another kind of parenthesis, which they make by splitting a verb in two and putting half of it at the beginning of an exciting chapter and the OTHER HALF at the end of it. Can any one conceive of anything more confusing than that? These things are called ‘separable verbs.’ The wider the two portions of one of them are spread apart, the better the author of the crime is pleased with his performance.”

Mark Twain
Ich werde Ihnen den Report aushaendigen.
I will to_you the report pass_on.
Ich werde Ihnen den Report aushändigen.
I will to_you the report pass_on.

Ich werde Ihnen die entsprechenden Anmerkungen aushändigen.
I will to_you the corresponding comments pass_on.
Ich werde Ihnen den Report aushändigen.
I will to_you the report pass_on.

Ich werde Ihnen die entsprechenden Anmerkungen aushändigen.
I will to_you the corresponding comments pass_on.

Ich werde Ihnen die entsprechenden Anmerkungen am Dienstag aushändigen.
I will to_you the corresponding comments on Tuesday pass_on.
German free word order

The finite verb always appears in 2nd position, but *Any constituent* (not just the subject) can appear in the 1st position

I will to_you the report pass_on

To_you will I the report pass_on

The report will I to_you pass_on
Main clause

Ich werde Ihnen den Report aushändigen,
I will to_you the report pass_on,
Main clause

Ich werde Ihnen den Report aushaendigen,
I will to_you the report pass_on,

Subordinate clause

damit Sie den eventuell uebernehmen koennen.
so_that you it perhaps adopt can.
Phrase-based models have an overly simplistic way of handling different word orders.

We can describe the linguistic differences between different languages.

Collins defines a set of 6 simple, linguistically motivated rules, and demonstrates that they result in significant translation improvements.
Collins’ Pre-ordering Model

**Step 1:** Reorder the source language

Ich werde Ihnen den Report aushaendigen, damit Sie den eventuell uebernehmen koennen.
Step 1: Reorder the source language

Ich werde Ihnen den Report aushaendigen, damit Sie den eventuell uebernehmen koennen.

Ich werde aushaendigen Ihnen den Report, damit Sie koennen uebernehmen den eventuell.
Collins’ Pre-ordering Model

**Step 1:** Reorder the source language

Ich werde Ihnen den Report *aushaendigen*, damit Sie den eventuell *uebernehmen* koennen.

Ich werde *aushaendigen* Ihnen den Report, damit Sie *koennen* *uebernehmen* den eventuell.

(I will *pass_on* to _you_ the report, so _that_ you *can adopt* it perhaps.)
Step 1: Reorder the source language

Ich werde Ihnen den Report aushaendigen, damit Sie den eventuell uebernehmen koennen.

(I will pass_on to_you the report, so_that you can adopt it perhaps.)

Step 2: Apply the phrase-based machine translation pipeline to the reordered input.
I will pass on the Report to you.
Clause Restructuring

Rule 1: **Verbs are initial in VPs**
Within a VP, move the head to the initial position

```
S
  VP-OC
    PDS-OA den that
    ADJD-MO eventuell perhaps
    VINF-HD uebernehmen adopt
  VINF-HD koennen can
```

25
Clause Restructuring

Rule 1: **Verbs are initial in VPs**  
Within a VP, move the *head* to the initial position
Rule 2: **Verbs follow complementizers**
In a subordinated clause mot the **head** of the clause to follow the **complementizer**

```
Clause Restructuring

KOUS-CP  PPER-SB  VP-OC
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>damit</em></td>
<td><em>Sie</em></td>
<td><em>uebernehmen</em></td>
</tr>
<tr>
<td>so-that</td>
<td>you</td>
<td>adopt</td>
</tr>
</tbody>
</table>

S-MO

VINF-HD
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>koennen</em></td>
</tr>
<tr>
<td>can</td>
</tr>
</tbody>
</table>
```

...
Rule 2: **Verbs follow complementizers**

In a subordinated clause mote the head of the clause to follow the complementizer.
Rule 3: **Move subject**

The subject is moved to directly precede the head of the clause.

```
Clause Restructuring

Rule 3: Move subject
The subject is moved to directly precede the head of the clause

KOUS-CP  VINF-HD  PPER-SB  VP-OC
  |       |     |     |
damit  koennen  Sie  uebernehmen
so-that  can  you  adopt
```

...
Rule 3: **Move subject**
The subject is moved to directly precede the head of the clause
Rule 4: **Particles**

In verb particle constructions, the particle is moved to precede the finite verb.

- **Wir** (we)
- **fordem** (accept)
- **das** (the)
- **Praesidium** (presidency)

S

PPER-SB  VVINF-HD  NP-OA  PTKVZ-SVP

Wir  fordem  the  auf

we  accept  PTKVZ-SVP

*PARTICLE*

**Praesidium presidency**
Rule 4: **Particles**

In verb particle constructions, the *particle* is moved to precede the *finite verb*.

```
S
   /     \
  /      \ 
/       /
PPER-SB VVINF-HD PTKVZ-SVP NP-OA
   Wir   fordem auf NP-OA
       we   accept *PARTICLE* 

ART

das
the

NN

Praesidium
presidency
```
Rule 5: **Infinitives**

**Infinitives** are moved to directly follow the finite verb within a clause.

```
[241x676]Clause
  Restructuring

29

S

VP-OC
  PTK-NEG
    nicht
      es
  VP-OC
    einreichen
      submit

OOER-OA
  konnten
    Wir
      we

VVINF-HD
  could

Rule 5: Infinitives
Infinitives are moved to directly follow the finite verb within a clause

Wir konnten es nicht einreichen.
Rule 5: **Infinitives**

Infinitives are moved to directly follow the finite verb within a clause

```
Wir
konnten
seinreichen
es
nicht
```

```
we
could
submit
it
not
```
Rule 6: **Negation**

Negative particle is moved to directly follow the finite verb

```
S

PPER-SB VVINF-HD VVINF-HD OOER-OA PTK-NEG VP-OC
Wir konnten einreichen es nicht VP-OC
we could submit it not VP-OC
```
Rule 6: **Negation**

Negative particle is moved to directly follow the finite verb

S

PPER-SB  VVINF-HD  PTK-NEG  VVINF-HD  OOER-OA  VAVP-OC

Wir
we

konnten
could

nicht
not

einreichen
submit

es
it
Ich werde Ihnen den Report aushaendigen, damit Sie den eventuell uebernehmen koennen.

I will to you the report pass on, so that you it perhaps adopt can.
Ich werde Ihnen den Report aushaendigen, damit Sie den eventuell uebernehmen koennen.

I will pass_on to_you the report, so_that you can adopt it perhaps.
I will pass_on to_you the report, so_that you can adopt it perhaps.

Now that seems less like the ravings of a madman.
Experiments

• Parallel training data: Europarl corpus (751k sentence pairs, 15M German words, 16M English)
• Parsed German training sentences
• Reordered the German training sentences with their 6 clause reordering rules
• Trained a phrase-based model
• Parsed and reordered the German test sentences
• Translated them
• Compared against the standard phrase-based model without parsing/reordering
Bleu score increase

Baseline: 25.2
Reordered System: 26.8

Significant improvement at $p<0.01$ using the sign test
Human Translation Judgments

• 100 sentences (10-20 words in length)
• Two annotators
• Judged two different versions
  – Baseline system’s translation
  – Reordering system’s translation
• Judgments: Worse, better or equal
• Sentences were chosen at random, systems’ translations were presented in random order
# Human Translation Judgments

<table>
<thead>
<tr>
<th></th>
<th>+</th>
<th>=</th>
<th>−</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotator 1</td>
<td>40%</td>
<td>40%</td>
<td>20%</td>
</tr>
<tr>
<td>Annotator 2</td>
<td>44%</td>
<td>37%</td>
<td>19%</td>
</tr>
</tbody>
</table>

+ = reordered translation better  
− = baseline better  
= = equal
<table>
<thead>
<tr>
<th>Reference</th>
<th>I think it is wrong in principle to have such measures in the European Union</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I believe that it is wrong in principle to take such measures in the European Union</td>
</tr>
<tr>
<td></td>
<td>I believe that it is wrong in principle, such measure in the European Union to take.</td>
</tr>
<tr>
<td>Reference</td>
<td>I think it is wrong in principle to have such measures in the European Union</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Reordered</td>
<td>I believe that it is wrong in principle to take such measures in the European Union</td>
</tr>
<tr>
<td>Baseline</td>
<td>I believe that it is wrong in principle, such measure in the European Union to take.</td>
</tr>
</tbody>
</table>
Reference

The current difficulties should encourage us to redouble our efforts to promote cooperation in the Euro-Mediterranean framework.

The current problems should spur us, our efforts to promote cooperation within the framework of the e-prozesses to be intensified.

The current problems should spur us to intensify our efforts to promote cooperation within the framework of the e-prozesses.
<table>
<thead>
<tr>
<th>Reference</th>
<th>The current difficulties should encourage us to redouble our efforts to promote cooperation in the Euro-Mediterranean framework.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>The current problems should spur us, our efforts to promote cooperation within the framework of the e-prozesses to be intensified.</td>
</tr>
<tr>
<td>Reordered</td>
<td>The current problems should spur us to intensify our efforts to promote cooperation within the framework of the e-prozesses.</td>
</tr>
</tbody>
</table>
Reference: To go on subsidizing tobacco cultivation at the same time is a downright contradiction.

At the same time, continue to subsidize tobacco growing, it is quite schizophrenic.

At the same time, to continue to subsidize tobacco growing is schizophrenic.
<table>
<thead>
<tr>
<th>Reference</th>
<th>To go on subsidizing tobacco cultivation at the same time is a downright contradiction.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>At the same time, continue to subsidize tobacco growing, it is quite schizophrenic.</td>
</tr>
<tr>
<td>Reordered</td>
<td>At the same time, to continue to subsidize tobacco growing is schizophrenic.</td>
</tr>
<tr>
<td>Reference</td>
<td>We have voted against the report by Mrs. Lalumiere for reasons that include the following:</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>

We have voted, amongst other things, for the following reasons against the report by Mrs. Lalumiere:

We have, among other things, for the following reasons against the report by Mrs. Lalumiere voted:
<table>
<thead>
<tr>
<th>Reference</th>
<th>We have voted against the report by Mrs. Lalumiere for reasons that include the following:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reordered</td>
<td>We have voted, amongst other things, for the following reasons against the report by Mrs. Lalumiere:</td>
</tr>
<tr>
<td>Baseline</td>
<td>We have, among other things, for the following reasons against the report by Mrs. Lalumiere voted:</td>
</tr>
</tbody>
</table>
Discussion: Clause Restructuring

• Are you convinced that German-English translation has improved?
• Do you think that this is a good fit for phrase-based machine translation?
• What limitations does this method have?

(Discuss with your neighbor.)
Limitations

• Requires a parser for the source language
  – We have parsers for only a small number of languages
  – Penalizes “low resource languages”
  – Fine for translating from English into other languages

• Involves hand crafted rules

• Removes the nice language-independent qualities of statistical machine translation
Synchronous context-free grammars (SCFGs)
The Syntax Bet

• Longstanding debate about whether linguistic information can help statistical translation
• Two camps
The Syntax Bet

• Longstanding debate about whether linguistic information can help statistical translation

• Two camps: Syntax will improve translation
The Syntax Bet

• Longstanding debate about whether linguistic information can help statistical translation

• Two camps
  - Syntax will improve translation
  - Simpler data-driven models will always win
• Longstanding debate about whether linguistic information can help statistical translation.

• Two camps:
  - Syntax will improve translation.
  - Simpler data-driven models will always win.

“Every time I fire a linguist my performance goes up”
Syntax is bad for translation

• The IBM Models were the dominant approach to SMT from the `90s until mid 2000s
  — Eschewed linguistic information
• A number of studies cast doubt on whether linguistic info could help SMT
  — Fox (2002) showed that “phrasal cohesion” was less common than assumed across even related languages
  — Koehn et al (2003) empirically demonstrated that syntactically motivated phrases made PBMT worse
Elle aura de les effets plus destructifs que positifs

Gloss: It will have effects more destructive than positive
Ouch! Syntax hurts!


Training corpus size

IBM Model 4
PBMT
PBMT w/syntactic phrases
Ouch! Syntax hurts!


Training corpus size

IBM Model 4
PBMT
PBMT w/syntactic phrases
Ouch! Syntax hurts!


Training corpus size

IBM Model 4
PBMT
PBMT w/syntactic phrases
Ouch! Syntax hurts!


IBM Model 4
PBMT
PBMT w/syntactic phrases

Training corpus size

18.0 20.5 23.0 25.5 28.0

10k 20k 40k 80k 160k 320k
Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.

澳洲，Australia
是，is
之一，one of
少数，few

澳洲，澳洲
是与北韩有邦交的少数国家之一

Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.

澳洲是与北韩有邦交的少数国家之一。

澳大利亚, Australia
是, is
之一, one of
少数, few
国家, countries
有, have
邦交, diplomatic relations
与, with
北, North
韩, Korea

澳洲是, Australia is
少数 国家, few countries
有邦交, have diplomatic relations
与北, with North
北韩, North Korea
Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.
<table>
<thead>
<tr>
<th>Australia is one of the few countries that have diplomatic relations with North Korea</th>
</tr>
</thead>
<tbody>
<tr>
<td>澳洲, Australia is, is</td>
</tr>
<tr>
<td>少数, few 国家, countries 有, have 邦交, diplomatic relations 与, with 北, North 韩, Korea</td>
</tr>
<tr>
<td>少数 国家, few countries</td>
</tr>
<tr>
<td>北韩, North Korea</td>
</tr>
<tr>
<td>与北韩, with North Korea</td>
</tr>
<tr>
<td>与北韩 有邦交, have diplomatic relations with North Korea</td>
</tr>
</tbody>
</table>
Why does it hurt to limit to constituents?

• Massively reduces the inventory of phrases that can be used as translation units
• Eliminates non-constituent phrases, many of which are quite useful
  – there are
  – note that
  – according to
So, what should we do?

• **Drop syntax** from statistical machine translation, since syntax is a bad fit for the data
• Abandon conventional English syntax and move towards **more robust grammars** that adapt to the parallel training corpus
• Maintain English syntax but **design different syntactic models**
Synchronous Context Free Grammars

• A common way of representing syntax in NLP is through context free grammars

• Synchronous context free grammars generate pairs of corresponding strings

• Can be used to describe translation and re-ordering between languages

• SCFGs translate sentences by parsing them
Example SCFG for Urdu

<table>
<thead>
<tr>
<th>Urdu</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>S → NP₁ VP₂</td>
<td>NP₁ VP₂</td>
</tr>
<tr>
<td>VP → PP₁ VP₂</td>
<td>VP₂ PP₁</td>
</tr>
<tr>
<td>VP → V₁ AUX₂</td>
<td>AUX₂ V₁</td>
</tr>
<tr>
<td>PP → NP₁ P₂</td>
<td>P₂ NP₁</td>
</tr>
<tr>
<td>NP → <em>hamd ansary</em></td>
<td><em>Hamid Ansari</em></td>
</tr>
<tr>
<td>NP → <em>naֱb sdr</em></td>
<td><em>Vice President</em></td>
</tr>
<tr>
<td>V → <em>namzd</em></td>
<td><em>nominated</em></td>
</tr>
<tr>
<td>P → <em>kylye</em></td>
<td><em>for</em></td>
</tr>
<tr>
<td>AUX → <em>taa</em></td>
<td><em>was</em></td>
</tr>
</tbody>
</table>
hamd ansary  na}b sdr  kylye  namzd  taa
Hamid Ansari

Vice President
Hamid Ansari for Vice President
Hamid Ansari

Vice President

for

nominated

nominated
Hamid Ansari was nominated for Vice President.
Hamid Ansari

Vice President

for

nominated

was
Hamid Ansari
for Vice President nominated was
Hamid Ansari was nominated for Vice President.
Hamid Ansari

for

Vice President

was

nominated
Hamid Ansari

nominated

Vice President

was

nominated
Hamid Ansari was nominated for Vice President.
Hamid Ansari was nominated for Vice President.
Hamid Ansari

was nominated for Vice President

Hamid Ansari

was nominated for Vice President
Discussion: Do you like SCFG?

• In what ways are SCFGs better for describing reordering than what we saw before?
• Is this a good model of how languages relate?
• What do you think of the synchronous requirement?

(Discuss with your neighbor)
Sometimes languages are mismatched

Leila

Sometimes languages are mismatched

Fry

S

NP

Leila

VP

NP

misses

V

Fry

NP

S

VP

NP

Fry

V

manque

PP

à

NP

Leila
We drove away.

Nos fuimos en el coche.
Spanish motion verb, pro-drop

\[
\begin{array}{c}
S \\
\mid \\
NP \quad \text{He} \\
\mid \\
V \quad \text{swam} \\
\mid \\
P \quad \text{to} \\
\mid \\
NP \quad \text{Ibiza}
\end{array}
\quad \quad \quad
\begin{array}{c}
S \\
\mid \\
VP \\
\mid \\
V \\
\mid \\
\text{Fue} \\
\mid \\
\text{He+went} \\
\mid \\
PP \\
\mid \\
\text{P} \\
\mid \\
\text{to} \\
\mid \\
NP \quad \text{Ibiza}
\end{array}
\]

\text{Fue} \quad \text{He+went} \\
\text{nadando} \quad \text{swimming}
We are going to use them anyway

• SCFGs are *mismatched* with some linguistic phenomena
• But they have nice *formal properties* and *well-defined algorithms*
Formal definition of SCFGs

• Aho and Ullman worked all of this out in the `60s and `70s

• Compiler theory
Formal definition of SCFGs

• A synchronous context free grammar is formally defined by a tuple

  \[ G = \langle N, T_S, T_T, R, S \rangle \]

• Where
A synchronous context-free grammar is formally defined by a tuple

\[ G = \langle N, T_S, T_T, R, S \rangle \]

- Where
  - \( N \) is a shared set of non-terminal symbols
    - S, NP, VP, PP, P, V, AUX
A synchronous context free grammar is formally defined by a tuple

\[ G = \langle N, T_S, T_T, R, S \rangle \]

- Where
  - \( N \) is a shared set of non-terminal symbols
  - \( T_S \) is the set of source language terminals
  - \( T_T \) is the set of target language terminals
  - \( R \) is the set of rules
  - \( S \) is the start symbol
Formal definition of SCFGs

A synchronous context-free grammar is formally defined by a tuple

\[ G = \langle N, T_S, T_T, R, S \rangle \]

• Where
  – \( N \) is a shared set of non-terminal symbols
  – \( T_S \) is the set of source language terminals
  – \( T_T \) is the set of target language terminals

S, NP, VP, PP, P, V, AUX

hamd ansary, naʃb sdr, namzd, kylye, taa

for, Hamid Ansari, nominated, Vice President, was
A synchronous context free grammar is formally defined by a tuple:

$$G = \langle N, T_S, T_T, R, S \rangle$$

- \(N\) is a shared set of non-terminal symbols
- \(T_S\) is the set of source language terminals
- \(T_T\) is the set of target language terminals
- \(R\) is a set of production rules
A synchronous context free grammar is formally defined by a tuple

\[ G = \langle N, T_S, T_T, R, S \rangle \]

- Where
  - \( N \) is a shared set of non-terminal symbols
  - \( T_S \) is the set of source language terminals
  - \( T_T \) is the set of target language terminals
  - \( R \) is a set of production rules
  - \( S \in N \), designated as the goal state

Examples:

- for, Hamid Ansari, nominated, Vice President, was
- hamd ansary, nab sdr, namzd, kylye, taa
- S, NP, VP, PP, P, V, AUX

63
Formal definition of SCFGs

• Each production rule has the form
  \[ X \rightarrow \langle \alpha, \beta, \sim, w \rangle \]

• Where
  – \( X \in N \)
  – \( \alpha \in (N \cup T_S)^* \)
  – \( \beta \in (N \cup T_T)^* \)
  – \( \sim \) is a one-to-one correspondence between the non terminals in \( \alpha \) and \( \beta \)
  – \( w \) is a weight assigned to the rule
Algorithms for SCFGs

• Translation with SCFGs is done via parsing
• How do we write an algorithm for parsing?
• One way to do it is as a deductive proof system
## The CKY Parsing Algorithm

<table>
<thead>
<tr>
<th>Axioms</th>
<th>for all ((A \rightarrow \alpha) \in R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A \rightarrow \alpha)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inference rules</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(A \rightarrow w_{i+1})</td>
<td>[A, (i, i+1)]</td>
</tr>
<tr>
<td>([B, i, j]) [C, (j, k)] (A \rightarrow BC)</td>
<td>[A, (i, k)]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Goal</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[S, 0, (n)]</td>
<td></td>
</tr>
<tr>
<td>Axioms</td>
<td>S →</td>
</tr>
<tr>
<td>--------</td>
<td>-----</td>
</tr>
<tr>
<td>VP →</td>
<td>PP VP</td>
</tr>
<tr>
<td>VP →</td>
<td>V AUX</td>
</tr>
<tr>
<td>PP →</td>
<td>NP P</td>
</tr>
<tr>
<td>NP →</td>
<td><em>hamd ansary</em></td>
</tr>
<tr>
<td>NP →</td>
<td><em>na}b sdr</em></td>
</tr>
<tr>
<td>V →</td>
<td>namzd</td>
</tr>
<tr>
<td>P →</td>
<td>klyye</td>
</tr>
<tr>
<td>AUX →</td>
<td>taa</td>
</tr>
</tbody>
</table>

Inference rule used | Goal |
---------------------|------|
|                    | [S, 0, 5] |
### Axioms

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>NP VP</td>
</tr>
<tr>
<td>VP</td>
<td>PP VP</td>
</tr>
<tr>
<td>VP</td>
<td>V AUX</td>
</tr>
<tr>
<td>PP</td>
<td>NP P</td>
</tr>
<tr>
<td>NP</td>
<td><em>hamd ansary</em></td>
</tr>
<tr>
<td>NP</td>
<td><em>na'b sdr</em></td>
</tr>
<tr>
<td>V</td>
<td><em>namzd</em></td>
</tr>
<tr>
<td>P</td>
<td><em>kylye</em></td>
</tr>
<tr>
<td>AUX</td>
<td><em>taa</em></td>
</tr>
</tbody>
</table>

### Inference rule used

<table>
<thead>
<tr>
<th>Index</th>
<th>Axiom</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td><em>hamd ansary</em></td>
</tr>
<tr>
<td>1</td>
<td><em>na'b sdr</em></td>
</tr>
<tr>
<td>2</td>
<td><em>kylye</em></td>
</tr>
<tr>
<td>3</td>
<td><em>namzd</em></td>
</tr>
<tr>
<td>4</td>
<td><em>taa</em></td>
</tr>
</tbody>
</table>

### Goal

[S, 0, 5]
Axioms

\[
\begin{align*}
S & \to \text{NP VP} \\
\text{VP} & \to \text{PP VP} \\
\text{VP} & \to \text{V AUX} \\
\text{PP} & \to \text{NP P} \\
\text{NP} & \to \text{hamd ansary} \\
\text{NP} & \to \text{na\textbackslash'\textbackslash'b sdr} \\
\text{V} & \to \text{namzd} \\
\text{P} & \to \text{kylye} \\
\text{AUX} & \to \text{taa}
\end{align*}
\]

Inference rule used

\[
\begin{align*}
\text{NP} & \to \text{hamd ansary}_1 \\
\text{Goal} & = [S, 0, 5] \\
\text{[NP, 0, 1]}
\end{align*}
\]

\[
\begin{align*}
0 & \quad \text{hamd ansary} & 1 & \quad \text{na\textbackslash'\textbackslash'b sdr} & 2 & \quad \text{kylye} & 3 & \quad \text{namzd} & 4 & \quad \text{taa} & 5
\end{align*}
\]
Axioms

\[
\begin{align*}
S & \rightarrow \text{NP VP} \\
\text{VP} & \rightarrow \text{PP VP} \\
\text{VP} & \rightarrow \text{V AUX} \\
\text{PP} & \rightarrow \text{NP P} \\
\text{NP} & \rightarrow \text{hamd ansary} \\
\text{NP} & \rightarrow \text{na\textit{\text{b}} sdr} \\
\text{V} & \rightarrow \text{namzd} \\
\text{P} & \rightarrow \text{kylye} \\
\text{AUX} & \rightarrow \text{taa}
\end{align*}
\]

Inference rule used

\[
\begin{align*}
\text{NP} & \rightarrow \text{hamd ansary}_1 \\
\end{align*}
\]

Goal

\[
\begin{align*}
[S, 0, 5] \\
[NP, 0, 1]
\end{align*}
\]
<table>
<thead>
<tr>
<th>Axioms</th>
<th>Inference rule used</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>S → NP VP</td>
<td></td>
<td>[S, 0, 5]</td>
</tr>
<tr>
<td>VP → PP VP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP → V AUX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP → NP P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP → <em>hamd ansary</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP → <em>naʃ b sdr</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V → <em>namzd</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P → klye</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUX → taa</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\begin{array}{cccccc}
0 & \text{hamd ansary} & 1 & \text{naʃ b sdr} & 2 & \text{klye} & 3 & \text{namzd} & 4 & \text{taa} \\
\hline
& [NP, 0, 1]
\end{array}
\]
Axioms

S → \( NP \ VP \)
VP → PP VP
VP → V AUX
PP → NP P
NP → hamd ansary
NP → naʃ b sdr
V → namzd
P → kylye
AUX → taa

Inference rule used

\[ NP \rightarrow naʃ b sdr_2 \]

Goal

\[ [S, 0, 5] \]

\[ [NP, 1, 1] \]
<table>
<thead>
<tr>
<th>Axioms</th>
<th>Inference rule used</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S \rightarrow \text{NP} \text{ VP} )</td>
<td>( \text{NP} \rightarrow \text{na}s\text{d sdr}_2 )</td>
<td>([S, 0, 5])</td>
</tr>
<tr>
<td>( \text{VP} \rightarrow \text{PP} \text{ VP} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{VP} \rightarrow \text{V} \text{ AUX} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{PP} \rightarrow \text{NP} \text{ P} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{NP} \rightarrow \text{hamd ansary} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{NP} \rightarrow \text{na}s\text{d sdr} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{V} \rightarrow \text{namzd} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{P} \rightarrow \text{kylye} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{AUX} \rightarrow \text{taa} )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \text{Goal} = \text{NP} \rightarrow \text{na}s\text{d sdr}_2 \]

\[ \text{[NP, 1, 1]} \]

\[ \text{[NP, 0, 1]} \]

\[ \text{[NP, 1, 2]} \]
### Axioms

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>NP VP</td>
</tr>
<tr>
<td>VP</td>
<td>PP VP</td>
</tr>
<tr>
<td>VP</td>
<td>V AUX</td>
</tr>
<tr>
<td>PP</td>
<td>NP P</td>
</tr>
<tr>
<td>NP</td>
<td>hamd ansary</td>
</tr>
<tr>
<td>NP</td>
<td>naʃ̂b sdr</td>
</tr>
<tr>
<td>V</td>
<td>namzd</td>
</tr>
<tr>
<td>P</td>
<td>kylye</td>
</tr>
<tr>
<td>AUX</td>
<td>taa</td>
</tr>
</tbody>
</table>

### Inference rule used

<table>
<thead>
<tr>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>[S, 0, 5]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>0</th>
<th>hamd ansary</th>
<th>1</th>
<th>naʃ̂b sdr</th>
<th>2</th>
<th>kylye</th>
<th>3</th>
<th>namzd</th>
<th>4</th>
<th>taa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>0</th>
<th>hamd ansary</th>
<th>1</th>
<th>naʃ̂b sdr</th>
<th>2</th>
<th>kylye</th>
<th>3</th>
<th>namzd</th>
<th>4</th>
<th>taa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- [NP, 0, 1]
- [NP, 1, 2]
Axioms

\[
\begin{align*}
S & \rightarrow \text{NP VP} \\
\text{VP} & \rightarrow \text{PP VP} \\
\text{VP} & \rightarrow \text{V AUX} \\
PP & \rightarrow \text{NP P} \\
\text{NP} & \rightarrow \text{hamd ansary} \\
\text{NP} & \rightarrow \text{na}b sdr \\
\text{V} & \rightarrow \text{namzd} \\
\text{P} & \rightarrow \text{kylye} \\
\text{AUX} & \rightarrow \text{taa}
\end{align*}
\]

Inference rule used

\[
\begin{align*}
P & \rightarrow \text{kylye}_3 \\
\text{Goal} & \quad [S, 0, 5] \\
\text{Goal} & \quad [P, 2, 3]
\end{align*}
\]

\[
0 \quad \text{hamd ansary} \quad 1 \quad \text{na}b \ sdr \quad 2 \quad \text{kylye} \quad 3 \quad \text{namzd} \quad 4 \quad \text{taa} \quad 5
\]

\[
[\text{NP}, 0, 1] \quad [\text{NP}, 1, 2]
\]
### Axioms

<table>
<thead>
<tr>
<th>Rule</th>
<th>Left-hand Side</th>
<th>Right-hand Side</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>NP VP</td>
<td></td>
</tr>
<tr>
<td>VP</td>
<td>PP VP</td>
<td></td>
</tr>
<tr>
<td>VP</td>
<td>V AUX</td>
<td></td>
</tr>
<tr>
<td>PP</td>
<td>NP P</td>
<td></td>
</tr>
<tr>
<td>NP</td>
<td>hamd ansary</td>
<td></td>
</tr>
<tr>
<td>NP</td>
<td>na)b sdr</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>namzd</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>klyye</td>
<td></td>
</tr>
<tr>
<td>AUX</td>
<td>taa</td>
<td></td>
</tr>
</tbody>
</table>

### Inference rule used

<table>
<thead>
<tr>
<th>Rule</th>
<th>Left-hand Side</th>
<th>Right-hand Side</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>klyye&lt;sub&gt;3&lt;/sub&gt;</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>[S, 0, 5]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Position</th>
<th>Token</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>hamd ansary</td>
<td>[NP, 0, 1]</td>
</tr>
<tr>
<td>1</td>
<td>na)b sdr</td>
<td>[NP, 1, 2]</td>
</tr>
<tr>
<td>2</td>
<td>klyye</td>
<td>[P, 2, 3]</td>
</tr>
<tr>
<td>3</td>
<td>namzd</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>taa</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Axioms

<table>
<thead>
<tr>
<th>Rule</th>
<th>String</th>
</tr>
</thead>
<tbody>
<tr>
<td>S →</td>
<td>NP VP</td>
</tr>
<tr>
<td>VP →</td>
<td>PP VP</td>
</tr>
<tr>
<td>VP →</td>
<td>V AUX</td>
</tr>
<tr>
<td>PP →</td>
<td>NP P</td>
</tr>
<tr>
<td>NP →</td>
<td>hamd ansary</td>
</tr>
<tr>
<td>NP →</td>
<td>nab sdr</td>
</tr>
<tr>
<td>V →</td>
<td>namzd</td>
</tr>
<tr>
<td>P →</td>
<td>klyye</td>
</tr>
<tr>
<td>AUX →</td>
<td>taa</td>
</tr>
</tbody>
</table>

### Inference rule used

<table>
<thead>
<tr>
<th>NP</th>
<th>VP</th>
<th>PP</th>
<th>P</th>
<th>AUX</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>hamd ansary</td>
<td>1</td>
<td>nab sdr</td>
<td>2</td>
<td>klyye namzd</td>
</tr>
</tbody>
</table>

| [NP, 0, 1] | [NP, 1, 2] | [P, 2, 3] |
Axioms

\[
\begin{align*}
S & \rightarrow \ NP \ VP \\
VP & \rightarrow \ PP \ VP \\
VP & \rightarrow \ V \ AUX \\
PP & \rightarrow \ NP \ P \\
NP & \rightarrow \ \text{hamd ansary} \\
NP & \rightarrow \ \text{na}b \ sdr \\
V & \rightarrow \ \text{namzd} \\
P & \rightarrow \ \text{kylye} \\
AUX & \rightarrow \ \text{taa}
\end{align*}
\]

Inference rule used

\[
\begin{align*}
V & \rightarrow \ \text{namzd}_4 \\
[V, 3, 4]
\end{align*}
\]

Goal

\[
[S, 0, 5]
\]

\[
\begin{align*}
0 & \text{ hamd ansary } \\
1 & \text{ na}b \ sdr \\
2 & \text{ kylye} \\
3 & \text{ namzd} \\
4 & \text{ taa} \\
5 & \text{ }
\end{align*}
\]

\[
\begin{align*}
\text{[NP, 0, 1]} & \quad \text{[NP, 1, 2]} & \quad \text{[P, 2, 3]}
\end{align*}
\]
Axioms

- $S \rightarrow \text{NP } \text{VP}$
- $\text{VP} \rightarrow \text{PP } \text{VP}$
- $\text{VP} \rightarrow \text{V } \text{AUX}$
- $\text{PP} \rightarrow \text{NP } \text{P}$
- $\text{NP} \rightarrow \text{hamd ansary}$
- $\text{NP} \rightarrow \text{na'b sdr}$
- $\text{V} \rightarrow \text{namzd}$
- $\text{P} \rightarrow \text{kylye}$
- $\text{AUX} \rightarrow \text{taa}$

Inference rule used

- $\text{V} \rightarrow \text{namzd}_4$

Goal

- $[S, 0, 5]$
- $[V, 3, 4]$
Axioms

- $S \rightarrow \text{NP VP}$
- $\text{VP} \rightarrow \text{PP VP}$
- $\text{VP} \rightarrow \text{V AUX}$
- $\text{PP} \rightarrow \text{NP P}$
- $\text{NP} \rightarrow \text{hamd ansary}$
- $\text{NP} \rightarrow \text{na)b sdr}$
- $\text{V} \rightarrow \text{namzd}$
- $\text{P} \rightarrow \text{kylye}$
- $\text{AUX} \rightarrow \text{taa}$

Inference rule used

Goal

$[S, 0, 5]$

<table>
<thead>
<tr>
<th></th>
<th>hamd ansary</th>
<th>na)b sdr</th>
<th>kylye</th>
<th>namzd</th>
<th>taa</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>[NP, 0, 1]</td>
<td>[NP, 1, 2]</td>
<td>[P, 2, 3]</td>
<td>[V, 3, 4]</td>
<td></td>
</tr>
<tr>
<td>Axioms</td>
<td>Inference rule used</td>
<td>Goal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>--------------------------------------------------</td>
<td>------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S \rightarrow \text{NP VP}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{VP} \rightarrow \text{PP VP}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{VP} \rightarrow \text{V AUX}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{PP} \rightarrow \text{NP P}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{NP} \rightarrow \text{hamd ansary}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{NP} \rightarrow \text{na}b sdr$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{V} \rightarrow \text{namzd}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{P} \rightarrow \text{kylye}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{AUX} \rightarrow \text{taa}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Axioms:

- $S \rightarrow \text{NP VP}$
- $\text{VP} \rightarrow \text{PP VP}$
- $\text{VP} \rightarrow \text{V AUX}$
- $\text{PP} \rightarrow \text{NP P}$
- $\text{NP} \rightarrow \text{hamd ansary}$
- $\text{NP} \rightarrow \text{na}b sdr$
- $\text{V} \rightarrow \text{namzd}$
- $\text{P} \rightarrow \text{kylye}$
- $\text{AUX} \rightarrow \text{taa}$

Inference rule used:

- $\text{AUX} \rightarrow \text{taa}_5$
- $[\text{AUX}, 4, 5]$

Goal:

- $[S, 0, 5]$
### Axioms

- S → NP VP  
- VP → PP VP  
- VP → V AUX  
- PP → NP P  
- NP → *hamd ansary*  
- NP → *na)b sdr*  
- V → *namzd*  
- P → *kylye*  
- AUX → *taa*

### Inference rule used

- AUX → taa₅

### Goal

- [S, 0, 5]
- [AUX, 4, 5]

---

<table>
<thead>
<tr>
<th></th>
<th>[NP, 0, 1]</th>
<th>[NP, 1, 2]</th>
<th>[P, 2, 3]</th>
<th>[V, 3, 4]</th>
<th>[AUX, 4, 5]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td><em>hamd ansary</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td><em>na)b sdr</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td><em>kylye</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td><em>namzd</em></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><em>taa</em></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Axioms

<table>
<thead>
<tr>
<th>S →</th>
<th>NP VP</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP →</td>
<td>PP VP</td>
</tr>
<tr>
<td>VP →</td>
<td>V AUX</td>
</tr>
<tr>
<td>PP →</td>
<td>NP P</td>
</tr>
<tr>
<td>NP →</td>
<td><em>hamd ansary</em></td>
</tr>
<tr>
<td>NP →</td>
<td><em>naʃb sdr</em></td>
</tr>
<tr>
<td>V →</td>
<td><em>namzd</em></td>
</tr>
<tr>
<td>P →</td>
<td><em>kylye</em></td>
</tr>
<tr>
<td>AUX →</td>
<td><em>taa</em></td>
</tr>
</tbody>
</table>

### Inference rule used

<table>
<thead>
<tr>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>[S, 0, 5]</td>
</tr>
</tbody>
</table>

### Goal

<table>
<thead>
<tr>
<th>0</th>
<th><em>hamd ansary</em></th>
<th>1</th>
<th><em>naʃb sdr</em></th>
<th>2</th>
<th><em>kylye</em></th>
<th>3</th>
<th><em>namzd</em></th>
<th>4</th>
<th><em>taa</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>[NP, 0, 1]</td>
<td>[NP, 1, 2]</td>
<td>[P, 2, 3]</td>
<td>[V, 3, 4]</td>
<td>[AUX, 4, 5]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Axioms

S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → hamd ansary
NP → nab sdr
V → namzd
P → kylye
AUX → taa

Inference rule used

Goal

[S, 0, 5]
Axioms

\[
\begin{align*}
S & \rightarrow \text{NP VP} \\
VP & \rightarrow \text{PP VP} \\
VP & \rightarrow \text{V AUX} \\
PP & \rightarrow \text{NP P} \\
\text{NP} & \rightarrow \text{hamd ansary} \\
\text{NP} & \rightarrow \text{na}b\ sdr \\
\text{V} & \rightarrow \text{namzd} \\
\text{P} & \rightarrow \text{kylye} \\
\text{AUX} & \rightarrow \text{taa}
\end{align*}
\]

Inference rule used

\[
\begin{align*}
\text{NP} & \rightarrow \text{VP} \\
\text{VP} & \rightarrow \text{PP} \\
\text{PP} & \rightarrow \text{NP} \\
\text{NP} & \rightarrow \text{hamd ansary} \\
\text{NP} & \rightarrow \text{na}b\ sdr \\
\text{V} & \rightarrow \text{namzd} \\
\text{P} & \rightarrow \text{kylye} \\
\text{AUX} & \rightarrow \text{taa}
\end{align*}
\]

Goal

\[
\begin{align*}
\text{[NP, 1, 2]} & \quad \text{[P, 2, 3]} & \quad \text{PP} \rightarrow \text{NP P} \\
\text{[PP, 1, 3]}
\end{align*}
\]

\[
\begin{align*}
\text{NP} & \rightarrow \text{PP} \\
P & \rightarrow \text{NP}
\end{align*}
\]
<table>
<thead>
<tr>
<th>Axioms</th>
<th>Inference rule used</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>S → NP VP</td>
<td>[NP, 1, 2] [P, 2, 3] PP → NP P</td>
<td>[S, 0, 5]</td>
</tr>
<tr>
<td>VP → PP VP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP → V AUX</td>
<td>[PP, 1, 3]</td>
<td></td>
</tr>
<tr>
<td>PP → NP P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP → hamd ansary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP → na}b sdr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V → namzd</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P → klyye</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUX → taa</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>0</th>
<th>hamd ansary</th>
<th>1</th>
<th>na}b sdr</th>
<th>2</th>
<th>klyye</th>
<th>3</th>
<th>namzd</th>
<th>4</th>
<th>taa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[NP, 0, 1]</td>
<td></td>
<td>[NP, 1, 2]</td>
<td></td>
<td>[P, 2, 3]</td>
<td></td>
<td>[V, 3, 4]</td>
<td></td>
<td>[AUX,4,5]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[PP, 1, 3]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Axioms</td>
<td>Inference rule used</td>
<td>Goal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------</td>
<td>-----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S → NP VP</td>
<td></td>
<td>[S, 0, 5]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP → PP VP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP → V AUX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP → NP P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP → $hamd\ ansary$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP → $na\j b\ sdr$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V → $namzd$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P → $kyl\ ye$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUX → taa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\begin{align*}
0 & : \quad hamd\ ansary \\
1 & : \quad na\j b\ sdr \\
2 & : \quad kyl\ ye \\
3 & : \quad namzd \\
4 & : \quad taa \\
5 & : \quad \tau a \\
\end{align*}
\]

\[
\begin{array}{cccccccc}
\text{[NP, 0, 1]} & \quad & \text{[NP, 1, 2]} & \quad & \text{[P, 2, 3]} & \quad & \text{[V, 3, 4]} & \quad & \text{[AUX, 4, 5]} \\
& \quad & \quad & \quad & \quad & \quad & \quad & \quad & \quad \\
\text{[PP, 1, 3]} & \quad & \quad & \quad & \quad & \quad & \quad & \quad & \quad \\
\end{array}
\]
Axioms

\[
\begin{align*}
S & \rightarrow \text{NP VP} \\
\text{VP} & \rightarrow \text{PP VP} \\
\text{VP} & \rightarrow \text{V AUX} \\
\text{PP} & \rightarrow \text{NP P} \\
\text{NP} & \rightarrow \text{hamd ansary} \\
\text{NP} & \rightarrow \text{na}b sdr \\
\text{V} & \rightarrow \text{namzd} \\
\text{P} & \rightarrow \text{kylye} \\
\text{AUX} & \rightarrow \text{taa}
\end{align*}
\]

Inference rule used

Goal

\[
[S, 0, 5]
\]
Axioms:

- $S \rightarrow NP\ VP$
- $VP \rightarrow PP\ VP$
- $VP \rightarrow V\ AUX$
- $PP \rightarrow NP\ P$
- $NP \rightarrow hamd\ ansary$
- $NP \rightarrow na\ b\ sdr$
- $V \rightarrow namzd$
- $P \rightarrow klyye$
- $AUX \rightarrow taa$

Inference rule used:

- $[V, 3, 4]$ $[AUX, 4, 5]$ $VP \rightarrow V\ AUX$ $[S, 0, 5]$
- $[VP, 3, 5]$

Goal:

0 $hamd\ ansary$
1 $na\ b\ sdr$
2 $klyye$
3 $namzd$
4 $taa$
5

[NP, 0, 1] [NP, 1, 2] [P, 2, 3] [V, 3, 4] [AUX, 4, 5]

[PP, 1, 3]
### Axioms

- **S →** NP VP
- **VP →** PP VP
- **VP →** V AUX
- **PP →** NP P
- **NP →** *hamd ansary*
- **NP →** *na)b sdr*
- **V →** *namzd*
- **P →** *kylye*
- **AUX →** *taa*

### Inference rule used

- ** Goal **
  - **[V, 3, 4]** [AUX, 4, 5] VP → V AUX [S, 0, 5]
    - **[VP, 3, 5]**

### Diagram

```
0  hamd ansary  1  na)b sdr  2  kylye  3  namzd  4  taa  5

[NP, 0, 1]   [NP, 1, 2]   [P, 2, 3]   [V, 3, 4]   [AUX, 4, 5]
          [PP, 1, 3]       [VP, 3, 5]
```
Axioms

S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → hamd ansary
NP → na}b sdr
V → namzd
P → klyye
AUX → taa

Inference rule used

Goal

[S, 0, 5]

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>hamd ansary</td>
<td>1</td>
<td>na}b sdr</td>
<td>2</td>
</tr>
</tbody>
</table>

- [NP, 0, 1]   [NP, 1, 2]   [P, 2, 3]   [V, 3, 4]   [AUX, 4, 5]

- [PP, 1, 3]   [VP, 3, 5]
Axioms

S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → hamd ansary
NP → naʃb sdr
V → namzd
P → kyl ye
AUX → taa

Inference rule used

Goal

[S, 0, 5]
Axioms

S → NP VP
VP → PP VP
VP → V AUX
PP → NP P
NP → hamd ansary
NP → na)b sdr
V → namzd
P → kylye
AUX → taa

Inference rule used

Goal

[S, 0, 5]

[V, 3, 4] [AUX,4,5]

[VP, 1, 5]

[PP, 1, 3] [VP, 3, 5] VP → PP CP

[VP, 3, 5] [PP, 1, 3]
Axioms

Inference rule used

Goal

<table>
<thead>
<tr>
<th>Axioms</th>
<th>Inference rule used</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>S → NP VP</td>
<td></td>
<td>[PP, 1, 3] [VP, 3, 5] VP → PP CP [S, 0, 5]</td>
</tr>
<tr>
<td>VP → PP VP</td>
<td></td>
<td>[VP, 1, 5]</td>
</tr>
<tr>
<td>VP → V AUX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP → NP P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP → <em>hamd ansary</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP → <em>na)b sdr</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V → <em>namzd</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P → <em>kylye</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUX → <em>taa</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

0
1
2
3
4
5

hamd ansary na)b sdr kylye namzd taa

[NP, 0, 1] [NP, 1, 2] [P, 2, 3] [V, 3, 4] [AUX,4,5]

[PP, 1, 3] [VP, 3, 5]

[VP, 1, 5]
Axioms

- $S \rightarrow NP\ VP$
- $VP \rightarrow PP\ VP$
- $VP \rightarrow V\ AUX$
- $PP \rightarrow NP\ P$
- $NP \rightarrow hamd\ ansary$
- $NP \rightarrow na)b\ sdr$
- $V \rightarrow namzd$
- $P \rightarrow kylye$
- $AUX \rightarrow taa$

Inference rule used

**Goal**

$[S, 0, 5]$
Axioms

\[
\begin{align*}
S & \rightarrow \text{NP VP} \\
\text{VP} & \rightarrow \text{PP VP} \\
\text{VP} & \rightarrow \text{V AUX} \\
\text{PP} & \rightarrow \text{NP P} \\
\text{NP} & \rightarrow \text{hamd ansary} \\
\text{NP} & \rightarrow \text{na}b sdr \\
\text{V} & \rightarrow \text{namzd} \\
\text{P} & \rightarrow \text{kylye} \\
\text{AUX} & \rightarrow \text{taa} \\
\end{align*}
\]

Inference rule used

\[
\begin{align*}
\text{Goal} & \quad \text{Inference rule used} \\
\text{[NP, 0, 1]} & \quad [\text{NP, 1, 5}] \quad S \rightarrow \text{NP VP} \quad [S, 0, 5] \\
\text{[S, 0, 5]} \\
\end{align*}
\]

Diagram:

```
0  hamd ansary  1  na}b sdr  2  kylye  3  namzd  4  taa  5

[NP, 0, 1]  [NP, 1, 2]  [P, 2, 3]  [V, 3, 4]  [AUX,4,5]

[PP, 1, 3]  [VP, 3,5]

[VP, 1,5]
```
Axioms

\[ S \rightarrow \text{NP VP} \]
\[ \text{VP} \rightarrow \text{PP VP} \]
\[ \text{VP} \rightarrow \text{V AUX} \]
\[ \text{PP} \rightarrow \text{NP P} \]
\[ \text{NP} \rightarrow \text{hamd ansary} \]
\[ \text{NP} \rightarrow \text{na}b sdr \]
\[ \text{V} \rightarrow \text{namzd} \]
\[ \text{P} \rightarrow \text{kylye} \]
\[ \text{AUX} \rightarrow \text{taa} \]

Inference rule used

\([\text{NP}, 0, 1] \quad [\text{VP}, 1, 5] \quad S \rightarrow \text{NP VP} \quad [S, 0, 5] \]

Goal

\([S, 0, 5]\)
Axioms

\[
\begin{align*}
S & \rightarrow \ NP \ VP \\
VP & \rightarrow \ PP \ VP \\
VP & \rightarrow \ V \ AUX \\
PP & \rightarrow \ NP \ P \\
NP & \rightarrow \ \text{hamd ansary} \\
NP & \rightarrow \ na)b sdr \\
v & \rightarrow \ namzd \\
P & \rightarrow \ kyle \\
AUX & \rightarrow \ taa
\end{align*}
\]

Inference rule used

Goal

\[
[S, 0, 5]
\]
# The CKY Parsing Algorithm

<table>
<thead>
<tr>
<th>Axioms</th>
<th>(_{\text{for all } (A \rightarrow \alpha) \in R})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A \rightarrow \alpha)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inference rules</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(A \rightarrow w_{i+1})</td>
<td>[A, (i, i+1)]</td>
</tr>
<tr>
<td>[B, (i, j)] [C, (j, k)] A (\rightarrow) BC</td>
<td>[A, (i, k)]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Goal</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[S, 0, (n)]</td>
<td></td>
</tr>
</tbody>
</table>
### The CKY Translation Algorithm

<table>
<thead>
<tr>
<th>Axioms</th>
<th>$A \rightarrow \alpha, \beta$</th>
<th>for all $(A \rightarrow \alpha, \beta) \in R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inference rules</td>
<td>$A \rightarrow w_{i+1}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$[A, i, i+1]$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$[B, i, j]$</td>
<td>$[C, j, k]$</td>
</tr>
<tr>
<td></td>
<td>$[A, i, k]$</td>
<td></td>
</tr>
<tr>
<td>Goal</td>
<td>$[S, 0, n]$</td>
<td></td>
</tr>
</tbody>
</table>

69
Learning SCFGs from parallel text
Where do grammars come from?

• Great! We now have
  – a formalism for describing the relationship between two languages,
  – an algorithm for producing translations

• All we need now is a synchronous grammar
Where do grammars come from?

• Great! We now have
  – a formalism for describing the relationship between two languages,
  – an algorithm for producing translations

• All we need now is a synchronous grammar
• Where do grammars come from?
• Well, when two languages love each other very much...
Data-driven grammar extraction

- Grammar rules are not written by hand, they are extracted from bilingual parallel corpora

<table>
<thead>
<tr>
<th>Arabic</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>فالفتعذيب لا يزال يمارس على نطاق واسع</td>
<td>Torture is still being practised on a wide scale.</td>
</tr>
<tr>
<td>وتتم عمليات الاعتقال والاحتجاز دون سبب بصورة روتينية</td>
<td>Arrest and detention without cause take place routinely.</td>
</tr>
<tr>
<td>وجان وقت التحلي بالبصيرة والشجاعة السياسية.</td>
<td>This is a time for vision and political courage</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chinese</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>我国 能源 原材料 工业 生产 大幅度 增长.</td>
<td>China's energy and raw materials production up.</td>
</tr>
<tr>
<td>非国大 要求 阻止 更 多 被 拘留 人员 死亡.</td>
<td>ANC calls for steps to prevent deaths in police custody.</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
• Most common type of SCFG in SMT is Hiero which has rules w/one non-terminal symbol
• Not as nice as linguistically motivated rules, does not capture the reordering in Urdu
Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.

X → 与北韩有邦交，少数国家之一

have diplomatic relations with North Korea

X → 邦交，少数国家之一

North Korea
Australia is one of the few countries that have diplomatic relations with North Korea.

X → 与 北 韩 有 邦交, diplomatic relations with North Korea

X → 邦交, diplomatic relations

X → 北 韩, North Korea
Australia is one of the few countries that have diplomatic relations with North Korea.
Extracting Hiero rules

Australia is one of the few countries that have diplomatic relations with North Korea.

X → 邦交，

diplomatic relations

X → 北韩，

North Korea

X → 与 X₁ 有 X₂，

have X₂ with X₁
Discussion: what do you think of Hiero?

- So, we now have a way of extracting SCFGs from bitexts. Great! So what?
- Is this any better than the phrase based model?
- How?
- Do you feel that it is lacking anything?

(Discuss with your neighbor)
Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.
Extracting Syntactic Rules

Australia is one of the few countries that have diplomatic relations with North Korea
Australia is one of the few countries that have diplomatic relations with North Korea.
Extracting Syntactic Rules

Australia is one of the few countries that have diplomatic relations with North Korea.

The few countries that have diplomatic relations with North Korea.

The NP that VP
Wait a minute...

• Didn’t we see this earlier in Koehn’s paper?
• Aren’t we giving up a ton of rules that you said were valuable?
• Something about a reduced inventory because we got rid of non-constituent phrases?
Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.

The few countries that have diplomatic relations with North Korea,

Australia is

with North Korea,

have diplomatic relations with North Korea,

the few countries that have diplomatic relations with North Korea,

the few countries that have diplomatic relations with North Korea,

Australia is
Extracting Syntactic Rules

Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.

NP → 与北韩有邦交的少数国家，the few countries that have diplomatic relations with North Korea.

NP/VP → 的少数国家，the few countries that have diplomatic relations with North Korea.

S/NP → 澳洲是，Australia is
Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.

The few countries that have diplomatic relations with North Korea are Australia and others.

The few countries that have diplomatic relations with North Korea include Australia and others.
Discussion: Is this better?

• What do you think of this flavor of SCFGs?
• What are its limitations?
• Do you think that it is better or worse than Hiero?
• How would you prove it?

(Discuss with your neighbors)
New training paradigm

• Training data: word-aligned bilingual parallel corpus, with parse trees
  – No need to parse the Urdu, just parse the English
  – Method is therefore transferable to other resource poor languages

• Extract SCFG rules with syntactic nonterminals

• For non-constituent phrases use CCG-style nonterminals

• Same coverage as Hiero model
Does it work?

- Tested for Urdu-English MT
- 1.5 Million word parallel corpus
- Two contrastive systems, with different grammar extraction mechanism
  - Hiero
  - Syntax-augmented grammars
- Used same decoder in both cases
- Tested results in a blind test set administered by the National Institute for Standards in Technology
Syntax v. no Syntax

Bleu score on blind NIST Urdu-English test set

- No Syntax (Hiero): 25.0
- Syntax (SAMT)
- Best system
Syntax v. no Syntax

Bleu score on blind NIST Urdu-English test set

<table>
<thead>
<tr>
<th></th>
<th>Bleu Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Syntax (Hier)</td>
<td>25.0</td>
</tr>
<tr>
<td>Syntax (SAMT)</td>
<td>31.0</td>
</tr>
<tr>
<td>Best system</td>
<td>32.0</td>
</tr>
</tbody>
</table>
Syntax v. no Syntax

Bleu score on blind NIST Urdu-English test set

- No Syntax (Hiero): 25.0
- Syntax (SAMT): 31.0
- Best system: 31.2
All system scores on NIST09 Urdu-English constrained task

- PBMT (Moses): 22, 23, 23, 24, 24
- Hiero baseline: 31, 25
- Syntax: 31

State of the Art Urdu Results
'first nuclear experiment in 1990 was'
Thomas red Unilever National Laboratory of the United States in ウィーン designer, are already working on the book of Los Alamos National Laboratory، former director of the technical intelligence، written with the cooperation of سیلم.ن.
This book 'nuclear express: political history and the expansion of bomb' has been written، and the two writers have also claimed that the country has made nuclear bomb is he or any other country's nuclear secrets to or that of any other nuclear power cooperation is achieved.

The First Nuclear Test Was in 1990.

Thomas red of the United States، the National Laboratory in designer are already working on the book of Los Alamos National Laboratory، former director of the technical intelligence، with the cooperation of Diana steelman wrote.

This book under the title of the spread of nuclear expressway: the political history of the bomb and this has been written and the two writers have claimed that the country also has made nuclear bomb or any other country، Korea nuclear secrets، or any of the other nuclear power cooperation.
Baseline

He said that China, North Korea, Iran, Syria, Pakistan, through Egypt, Libya and Yemen is to provide nuclear technology.

Thomas was red when this question why China has provided the nuclear technology to Pakistan, In response, He said as China and India was joint enemy of Pakistan.

SCALE final system

He said that China would provide nuclear technology to North Korea, Iran, Syria, Pakistan, Egypt, Libya and Yemen.

Thomas red when was this question why China has provided to Pakistan nuclear technology, he said in response to China, Pakistan and India as a common enemy.
Syntax captures Urdu reordering
Why did this work?

• Using *syntax-based translation models* resulted in huge improvements in quality

• Previous work on syntax did not shown significant gains, so why did it work here?

• Urdu is an **ideal language** to show off the advantages of syntax
  – Very *small amount* of training data
  – Very *different word order* than English

• Can’t simply *memorize* translations of phrases

• Must *generalize*
Training data for MT Research

- **French-English**: $10^9$ word webcrawl
- **Arabic and Chinese**: DARPA GALE
- **European Parliament**
- **Urdu**: 1.5M
- **1000M**
- **200M**
- **50M**
Distribution of Word Orders

All Languages

- SOV: 40%
- SVO: 36%
- VSO: 14%
- VOS: 7%
- No dominant order

SMT Languages
Distribution of Word Orders

All Languages
- SOV: 40%
- SVO: 36%
- VSO: 14%
- VOS: 7%
- No dominant order: 7%

SMT Languages
- SOV: 13%
Distribu>on of Word Orders

All Languages

- SOV: 40%
- SVO: 36%
- VSO: 14%
- VOS: 7%
- No dominant order: 6%

SMT Languages

- SOV: 13%
- SVO: 61%

Legend:

- SOV
- SVO
- VSO
- VOS
- No dominant order
Distribution of Word Orders

All Languages

- SOV: 40%
- SVO: 36%
- VSO: 14%
- VOS: 7%
- No dominant order: 1%

SMT Languages

- SOV: 13%
- SVO: 22%
- VSO: 4%
- VOS: 61%
- No dominant order: 0%
Joshua Decoder

• An open source decoder
• Uses synchronous context free grammars to translate
• Implements all algorithms needed for translating with SCFGs
  – grammar extraction (Thrax!)
  – chart-parsing
  – n-gram language model integration
  – pruning, and k-best extraction
Joshua Decoder

• Download it from
  – http://joshua-decoder.org
Joshua Decoder

• Download it from
  – http://joshua-decoder.org

• Brownie points if you use it in your final projects
Joshua Decoder

• Download it from
  – http://joshua-decoder.org

• Brownie points if you use it in your final projects

• Use Jonny’s Thrax grammar extractor to test different kinds of SFCGs for your problems
More syntax
Syntactic LMs

• Our goal is reorder the translated phrases so that they are grammatical English
• Isn’t the language model probability supposed to do that already?
• Instead of an n-gram model, could we augment the LM with syntactic information?
Problem: bottom up parsing requires whole sentence
We need the LM to be able to score partial translations
One possibility: Incremental parsing

Transform right-expanding sequences of constituents into left-expanding sequences of incomplete constituents.

```
S
  |    |
S/NP  NP
  |    |
S/PP  IN  Friday
  |    |
S/VP  VP
  |    |
NP
  |    |
NP/NN  NN
  |    |
VP/NN  DT  board
  |    |
VP/NP  VB  the
  |    |
The
  |    |
```

Lane Schwartz

An Incremental Syntactic Language Model for Statistical Phrase-based Machine Translation
Questions?