Analysing Soft Syntax Features and Heuristics for Hierarchical Phrase Based Machine Translation

David Vilar, Daniel Stein, Hermann Ney

IWSLT 2008, Honolulu, Hawaii
20. October 2008

Human Language Technology and Pattern Recognition
Lehrstuhl für Informatik 6
Computer Science Department
RWTH Aachen University, Germany

1 Introduction

▶ **Hierarchical phrase-based models: Generalization of phrase-based-models**

  ▷ **Allow for "gaps" in the phrases**

  ▷ **Integration of reordering in the translation model**

▶ **Study the effect of extraction heuristics**

▶ **Extension with inclusion of (soft) syntactic features**

Outline

2　Hierarchical Phrases

▶ **Formalization as a synchronous CFG**

▶ **Rules of the form $X \rightarrow \langle \gamma, \alpha, \sim \rangle$, where:**

    ▷ X **is a non-terminal**

    ▷ γ **and** α **are strings of terminals and non-terminals**

    ▷ $\sim$ **is a one-to-one correspondence between the non-terminals of** α **and** γ

▶ **Example:**

$$X \rightarrow \langle \text{中 } X^{\sim 0} \text{ 那个 } X^{\sim 1}, \textbf{It's the } X^{\sim 1} \textbf{ in the } X^{\sim 0} \rangle$$
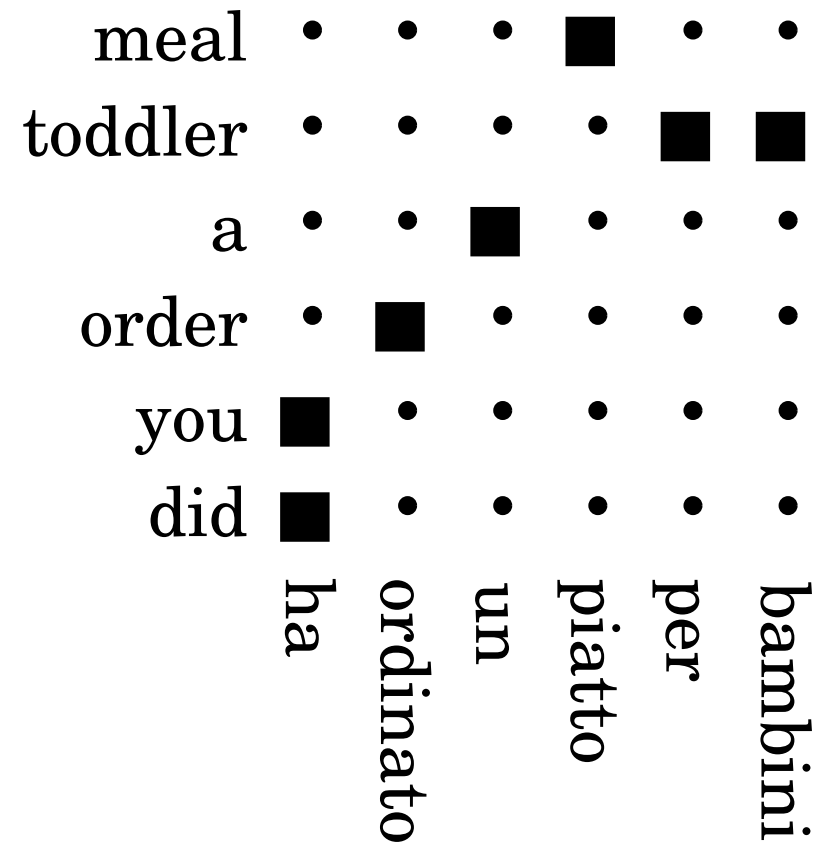$$X \rightarrow \langle \text{也 要 } X^{\sim 0} \text{ 一些 } X^{\sim 1}, \textbf{like to } X^{\sim 0} \textbf{ some } X^{\sim 1} \textbf{ too} \rangle$$

▶ **Additionally: Glue rules**

$$S \rightarrow \langle S^{\sim 0} X^{\sim 1}, S^{\sim 0} X^{\sim 1} \rangle$$
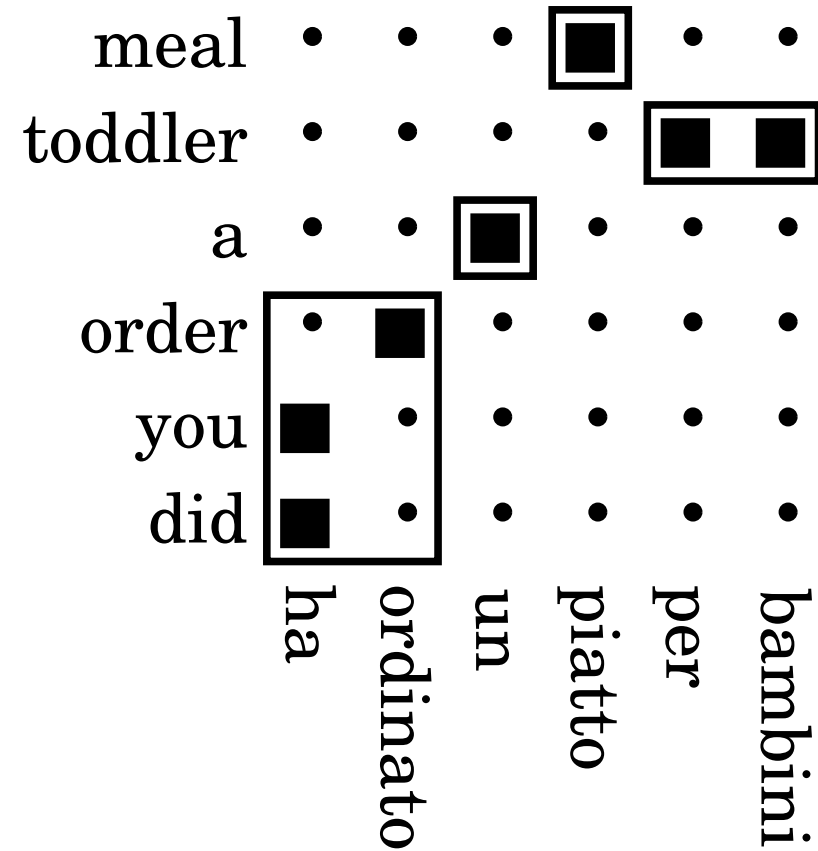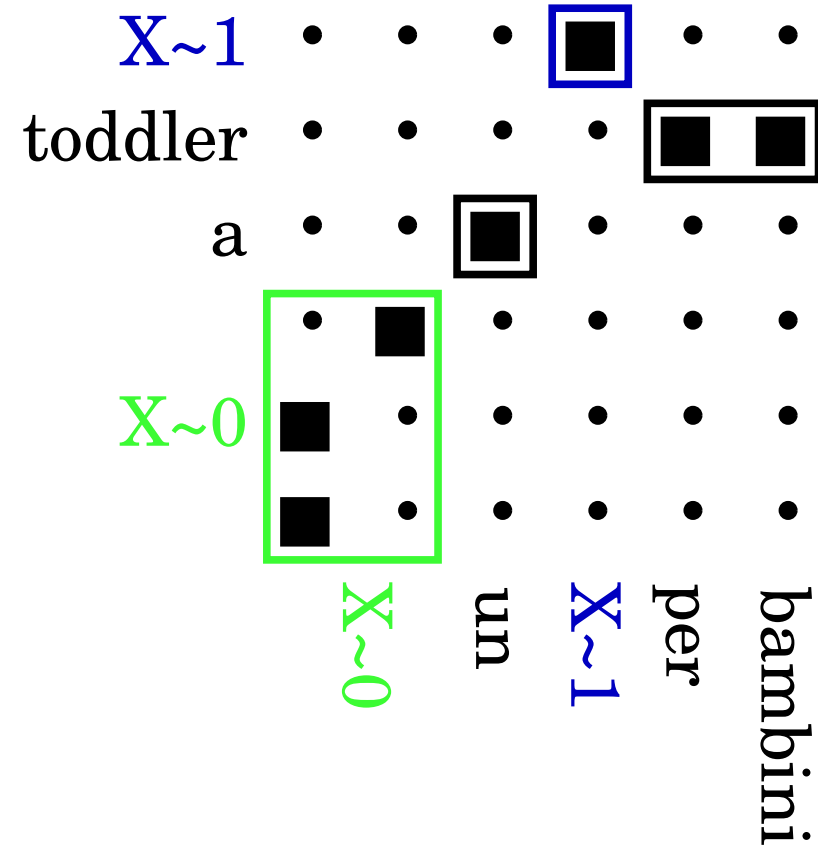$$S \rightarrow \langle X^{\sim 0}, X^{\sim 0} \rangle$$

Illustration

Alignment

Illustration

Standard phrases

Illustration

Example rule

3   Heuristic Features

▶ **Following features were tested:**

Paste rule Binary feature for rules of the form

$$X \longrightarrow \langle X^{\sim 0}\alpha, X^{\sim 0}\beta \rangle \text{ or } X \longrightarrow \langle \alpha X^{\sim 0}, \beta X^{\sim 0} \rangle$$

Hierarchical penalty Binary feature for hierarchical rules

Number of non-terminals Two binary features indicating if the rule has one or two non-terminals.

Extended glue rule added rule of the form

$$X \longrightarrow \langle X^{\sim 0} X^{\sim 1}, X^{\sim 0} X^{\sim 1} \rangle$$

4 Syntactical Features

▶ **Goal: include linguistic information from a deep syntactic parser**

▶ **Idea: introduce additional soft syntactic features**

▶ **This can be done during the extraction of the phrases**

▷ **No additional computational costs during decoding**

▷ **Can be done both on source and target side**

▷ **Rules are not filtered out**

"Valid" syntactical phrases

▶ **A phrase is valid when a node exists that completely covers all positions**

▶ **In order to obtain a normalized score, we add up all the counts and divide by the number of occurences of the phrase pair**
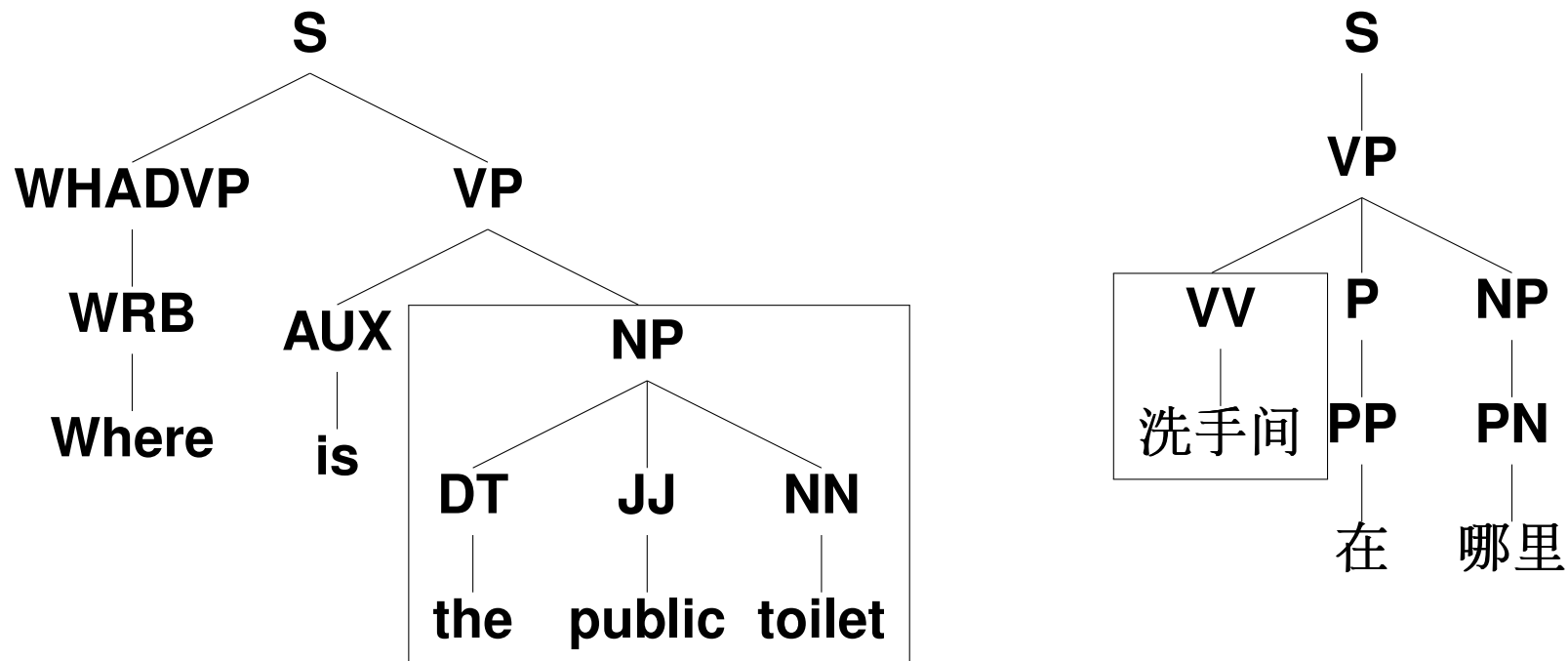
Extracted rule: $X^{\sim 0}$ 在 哪里 # Where is $X^{\sim 0}$

Scoring variants

$m(i, j) =$ **minimum number of words to be deleted or added to a phrase, so that it fits the yield of a node**
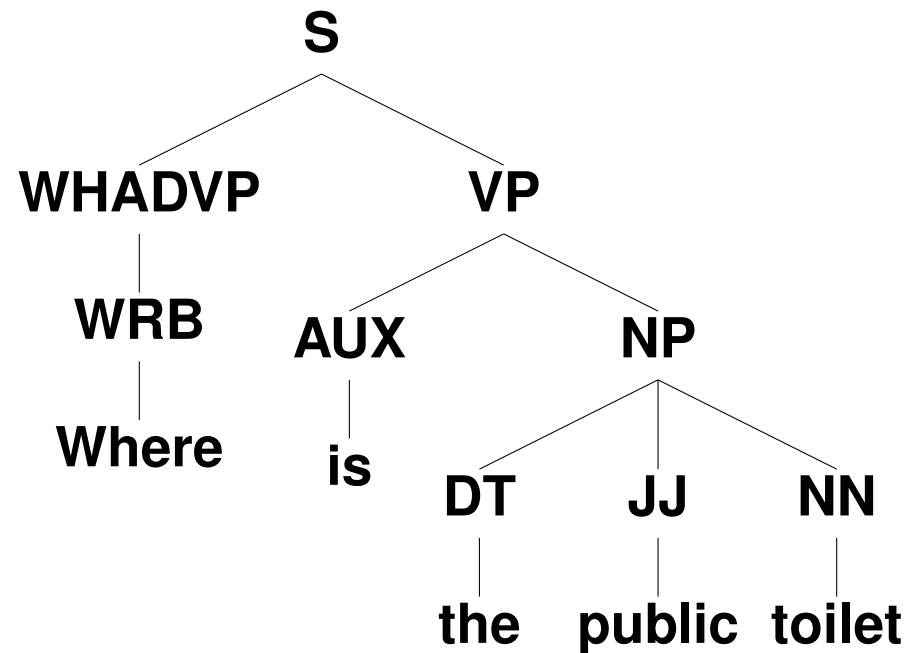
Source Phrases:

▶ **public toilet**

▶ **is the**

Scoring variants

$$m(i,j) = \text{minimum number of words to be deleted or added to a phrase,}$$
$$\text{so that it fits the yield of a node}$$

Source Phrases:

▶ **public toilet** $m(i,j) = 1$

▶ **is the**

Scoring variants

$m(i, j) =$ **minimum number of words to be deleted or added to a phrase, so that it fits the yield of a node**
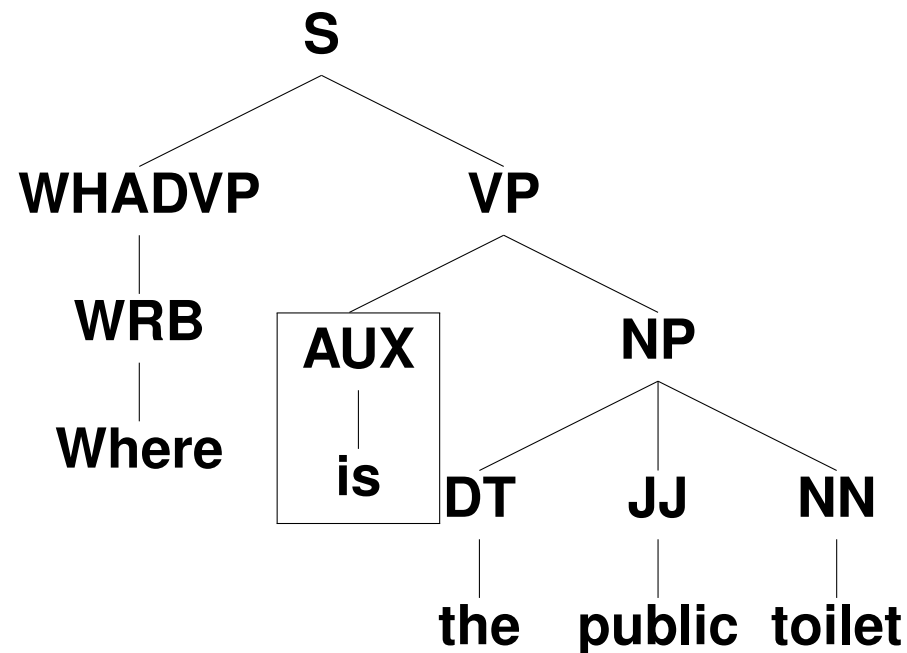
Source Phrases:
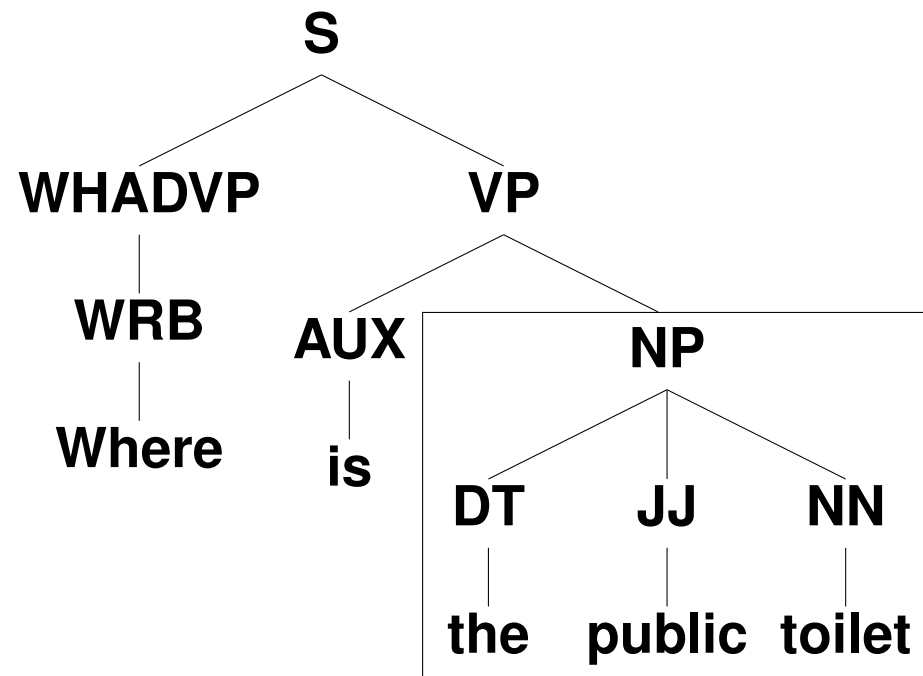
▶ **public toilet** $m(i, j) = 1$

▶ **is the** $m(i, j) = 1$

► **Four count ("smoothing") variants:**

$$c(i,j|t) := \begin{cases} \delta\left(m(i,j),0\right) & \textbf{binary} \\[2em] \dfrac{1}{m(i,j)+1} & \textbf{linear} \\[2em] \dfrac{1}{\exp\left(m(i,j)\right)} & \textbf{exponential} \\[2em] \dfrac{j-i}{(j-i)+m(i,j)} & \textbf{relative} \end{cases}$$

5 Experimental Results

▶ **IWSLT BTEC Data (Tourist and Travel domain)**

		Chinese	English
Training data	Sentences	23 940	
	Running words Vocabulary	181 486 9 041	232 746 10 350
Test 2004 Data	Sentences	500	
	Running words OOVs	7 543 96	10 718 154
Test 2005 Data	Sentences	506	
	Running words OOVs	8 052 101	10 828 164
Test 2008 Data	Sentences	507	
	Running words OOVs	6325 87	

Results

	test04		test05		test08
	BLEU	**TER**	**BLEU**	**TER**	**BLEU**
baseline	47.3	42.6	50.9	37.6	39.6
non-syntactic information					
hierarch	48.4	41.9	51.4	38.1	39.6
paste	49.1	41.6	51.1	38.0	40.8
glue2	48.2	41.8	51.2	37.6	39.7
1NT2NT	48.4	42.2	51.8	37.2	39.8
syntactic information					
binary	47.8	41.7	51.7	37.5	40.3
linear	47.6	41.9	51.2	37.6	40.6
exponential	47.9	41.7	51.6	37.4	40.3
relative	47.3	42.4	51.5	37.3	40.2

Results

	test04		test05		test08
	BLEU	**TER**	**BLEU**	**TER**	**BLEU**
baseline	47.3	42.6	50.9	37.6	39.6
non-syntactic information					
hierarch + paste	48.5	42.0	51.9	37.6	39.6
hierarch + paste + glue2	49.2	42.5	50.8	37.5	39.5
hierarch + paste + glue2 + 1NT2NT	48.6	41.6	51.0	37.9	40.0
combination of both syntactic and non-syntactic information (all features)					
binary	46.9	42.5	50.6	38.4	39.9
linear	48.0	42.3	51.2	38.0	40.5
exponential	47.7	42.3	51.0	38.4	40.0
relative	47.8	42.3	51.0	38.0	40.3

Example Translations

reference	Where is the exchange counter ?
baseline	The currency exchange office is
syntactical	Where is the currency exchange office ?

reference	Could you exchange it for a new one ?
baseline	You can buy a new one ?
syntactical	Could you change it for a new one ?

reference	You can take our airport shuttle bus to pick up the car .
baseline	You can take our airport shuttle bus with me .
syntactical	You can take our the airport shuttle bus come to pick it up .

6   Conclusions

▶ **Analyzed heuristics for phrase extraction**

▶ **Introduced soft syntactic constraints**

   ▷ **Use of source- and target-side information**

   ▷ **No additional search effort**

▶ **High variability of results**

   ▷ **Test on bigger corpora**

▶ **Bigger improvements when dealing with speech input (system talk tomorrow!)**

▶ **Applicable also to phrase-based systems**

Analysing Soft Syntax Features and Heuristics for Hierarchical Phrase Based Machine Translation

David Vilar, Daniel Stein, Hermann Ney

IWSLT 2008, Honolulu, Hawaii
20. October 2008

Human Language Technology and Pattern Recognition
Lehrstuhl für Informatik 6
Computer Science Department
RWTH Aachen University, Germany