SAUSSURIAN ANALOGY: A THEORETICAL ACCOUNT AND ITS APPLICATION

YVES LEPAGE

ANDO Shin-ichi

ATR-ITL, Kyoto, Japan

COLING-96

Analogy: the phenomenon

Latin: *ōrātōrem* : *ōrātor* = *honōrem* : x

x = *honor* instead of *honos*

©1996 ATR 音声翻訳通信研究所

Analogy on words (Morphology)

- Prefixing: *desirable* : *undesirable* = *determined* : x

  x = *undetermined*

- Suffixing: *user* : *usable* = *doer* : x

  x = *doable*

- Infixing: *fable* : *fabulous* = *miracle* : x

  x = *miraculous*

Explains declension, conjugation, derivation.

©1996 ATR 音声翻訳通信研究所

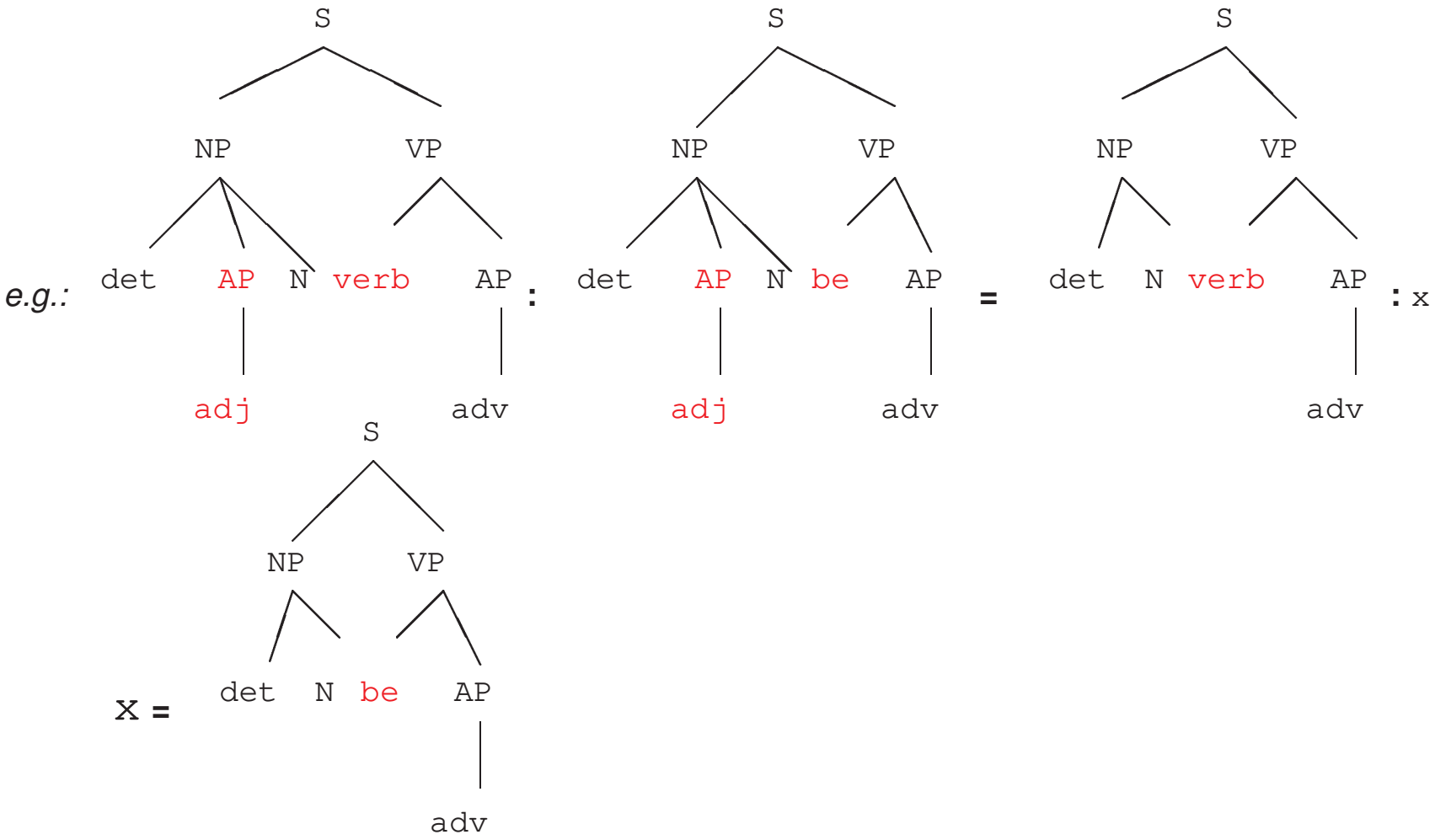
Analogy on sentences (Syntax?)

e.g.: *The signal is off* : *The green signal is on* = *The lamp turns on* : x

x = *The green lamp turns off*

©1996 ATR 音声翻訳通信研究所

Analogy on linguistic representations



©1996 ATR 音声翻訳通信研究所

Computational explanation

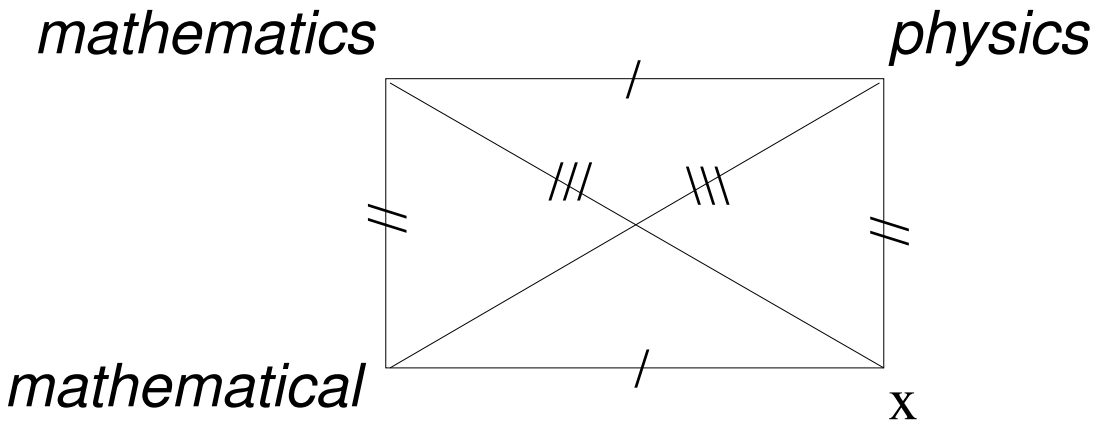
Arithmetical view: *mathematics* : *mathematical* = *physics* : x

x = *physical*

similar to: *2* : *4* = *3* : x

x = *6*

Geometrical view:



©1996 ATR 音声翻訳通信研究所

Edit distances

Edit operations:

- Deletion (*e.g.* pretty → pet)
- Insertion (*e.g.* make → mistake)
- Substitution (*e.g.* big → bag)

dis = *min* nbr of edit op.

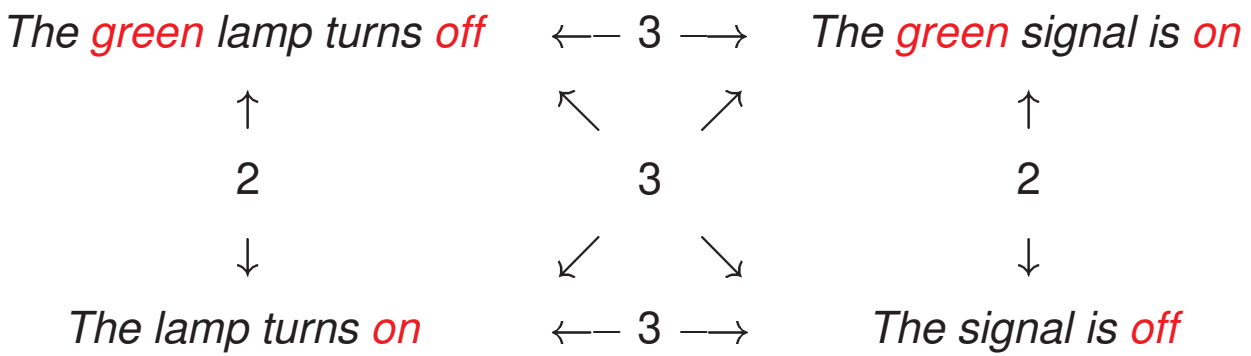
General definition on forests applies to

- Strings of chars (words)
- Strings of words (sentences)
- Trees (linguistic structures)

Explanation

©1996 ATR 音声翻訳通信研究所

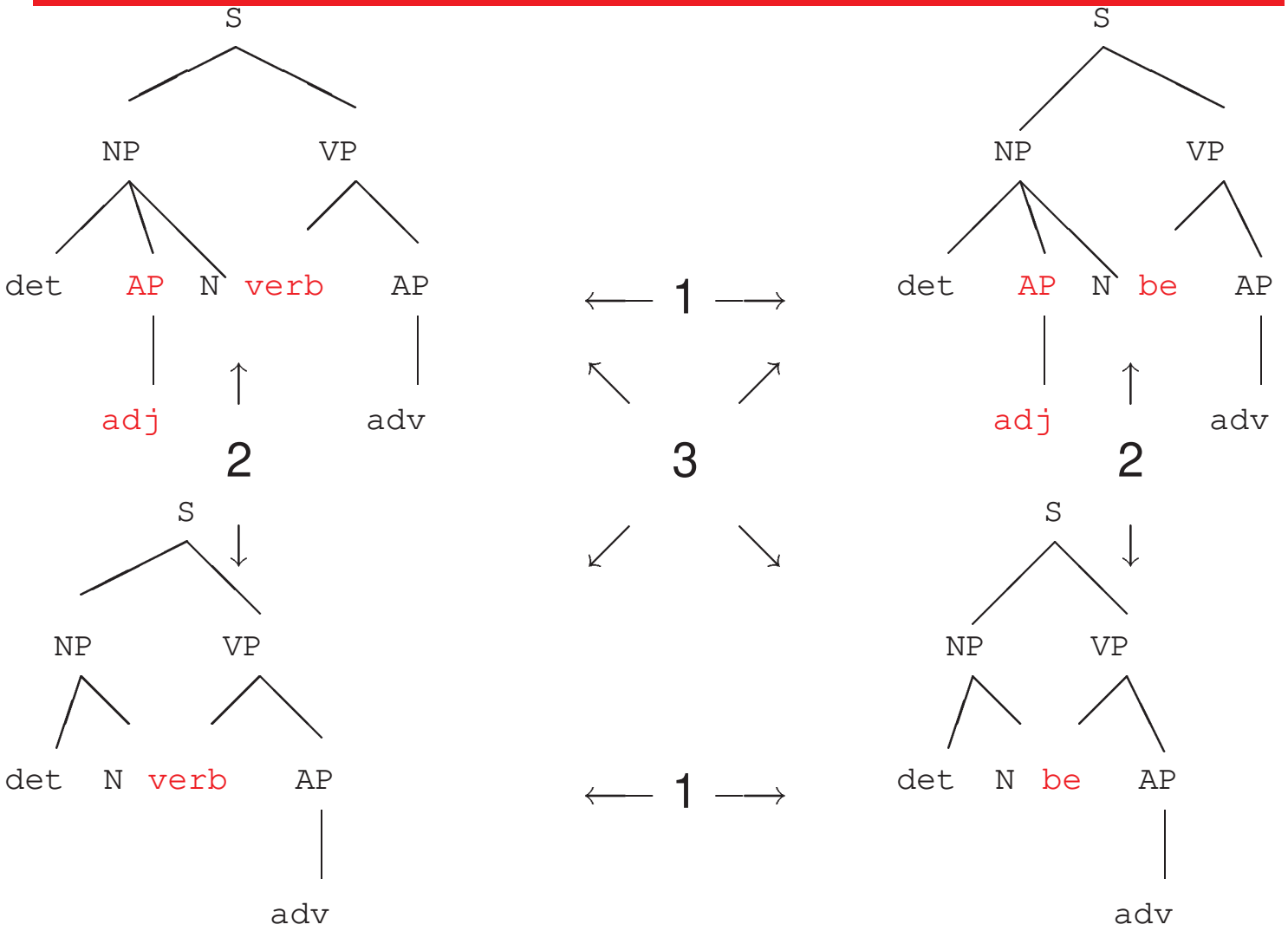
Analogy on sentences (explained)



Explanation

©1996 ATR 音声翻訳通信研究所

Analogy on ling. representations (explained)



Explanation

©1996 ATR 音声翻訳通信研究所

ATR

Example-based stream

If sentences (a) and (b) are close (??), if analysis of (a) is known, perform slight modifications (??) $\Rightarrow$ analysis of (b).

Imitation

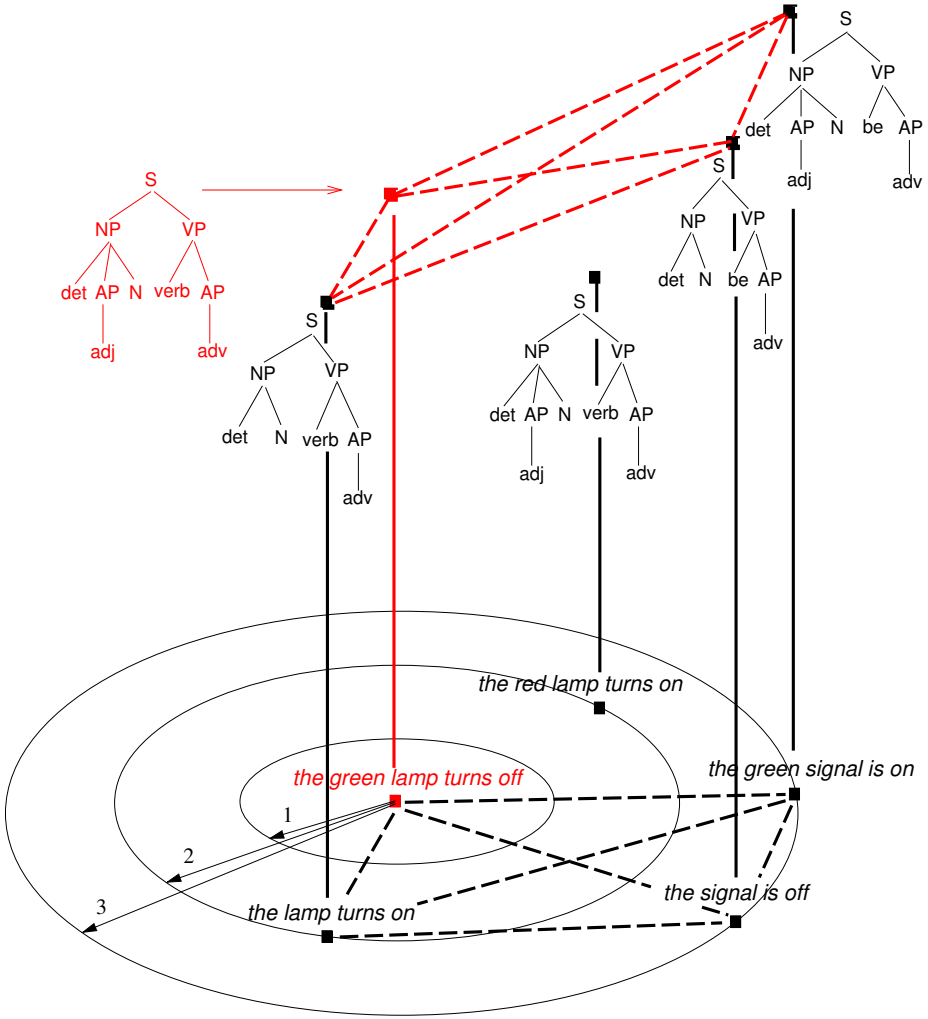
$\neq$

Analogy

If sentences (a), (b), (c) and (d) verify analogy, if analysis of (a), (b) and (c) are known, solve analogy $\Rightarrow$ analysis of (d).

©1996 ATR 音声翻訳通信研究所

Analysis by analogy (Principle)



©1996 ATR 音声翻訳通信研究所

Experiment data

Tree-bank: excerpt of UPenn (787 sentences and trees)

	$\neq$ words/nodes	size	
		mean	std. dev.
sentences	778	12.3	6.33
ling. structures	28	9.95	5.40

- Not true analysis: search for ling. structure in the tree-bank
- Distances in rectangle $\leq$ max size $\times$ 1/5, 1/4, 1/3, . . . , 1

©1996 ATR 音声翻訳通信研究所

ATR

Results (for dist $\leq$ max size $\times$ 2/5)

Recall = prop. of relevant objects that are retrieved = **46%**

$$\frac{|\triangle\text{analogy} = \triangle\text{correct}|}{|\left(\overset{\triangle}{\leftrightarrow}, \overset{\triangle}{\leftrightarrow}\right) / \triangle = \triangle|}$$

Precision = prop. of retrieved objects that are relevant = **48%**

$$\frac{|\triangle\text{analogy} = \triangle\text{correct}|}{|\triangle\text{analogy}|}$$

Non-determinism = **90**

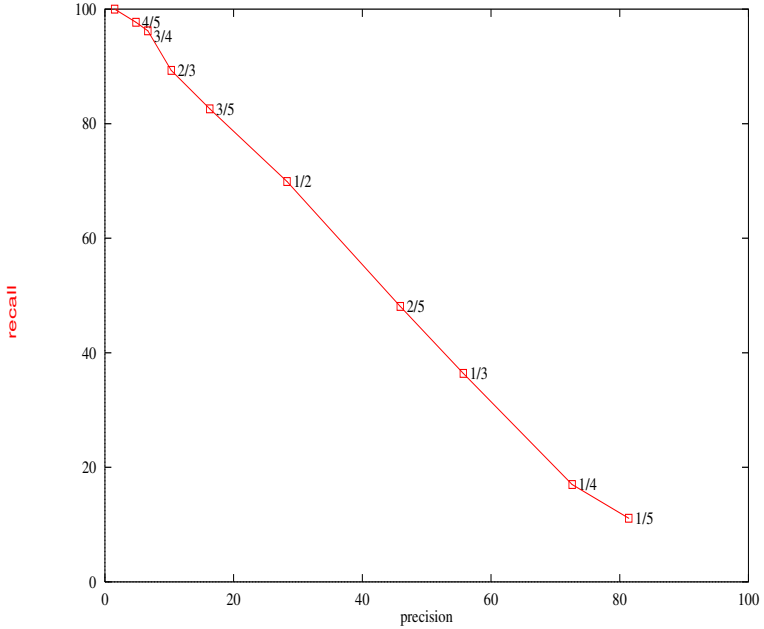
$$\left(\overset{\triangle}{\leftrightarrow}, \overset{\triangle}{\leftrightarrow}, \overset{\triangle}{\leftrightarrow}\right) \rightarrow \text{same solution}$$

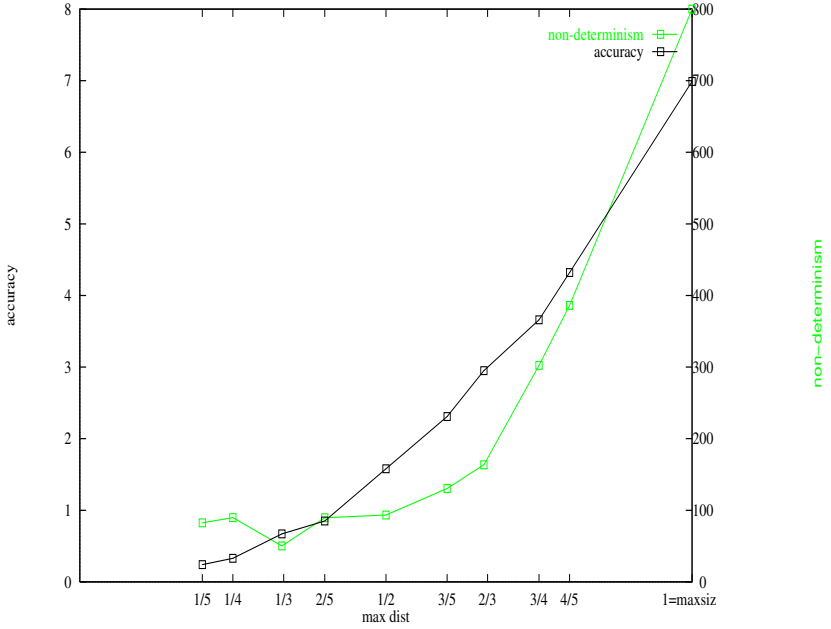
Accuracy = **0.8** $\pm$ **1.2** (in nodes)

$$\text{dist}(\triangle\text{analogy}, \triangle\text{correct})$$

©1996 ATR 音声翻訳通信研究所

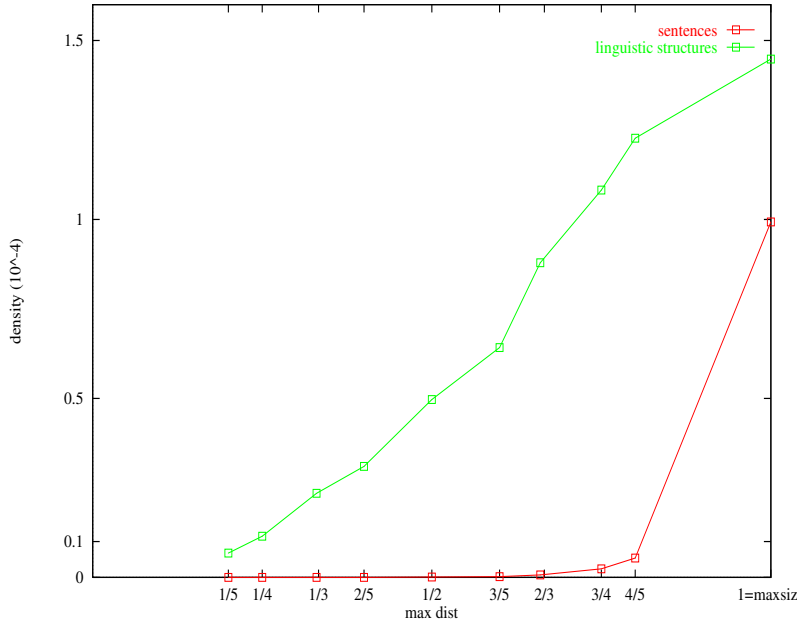
Recall/precision – non-det./accuracy





©1996 ATR 音声翻訳通信研究所

Density



©1996 ATR 音声翻訳通信研究所

Conclusion

Analogy

- is transversal (structuralist) $\neq$ axiomatic (generativist)
- captures regularity

Explanation using edit distances: words, sentences, trees

Application to NLP

- analysis by analogy
- generation/transfer by analogy (?)

©1996 ATR 音声翻訳通信研究所