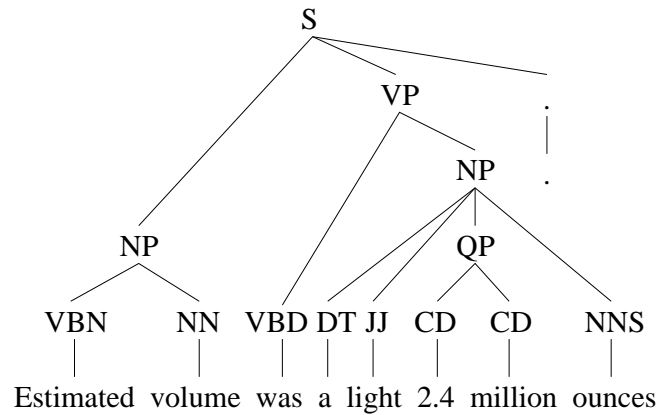


From Chunking to Parsing

Erik Tjong Kim Sang
 CNTS - Language Technology Group
 University of Antwerp
 Belgium



- Creating a syntactic analysis for a sentence can be divided in three tasks: POS tagging (level 1), chunking (level 2) and generating the rest of the tree.

Research questions

We have developed methods for training classifiers to divide sentences in groups of syntactically related words (chunking).

- Can these methods be used for parsing sentences?
- If so, how well do they perform?

Bottom-up Parsing is repeated chunking

We compute phrase positions and replace the phrases by their head and a tag denoting their type. After that we attempt to find more phrases.

Example

w	Estimated	volume	was	a	light	2.4	million	ounces	.
p	VBN	NN	VBD	DT	JJ	CD	CD	NNS	.
0	(NP	NP)				(QP	QP)		
0w		volume	was	a	light		million	ounces	.
0p		NP	VBD	DT	JJ		QP	NNS	.
1				(NP				NP)	
1w		volume	was					ounces	.
1p		NP	VBD					NP	.
2			(VP					VP)	
2w		volume	was						.
2p		NP	VP						.
3		(S							S)

Answers to research questions

- The techniques used for chunking can also be used for parsing.
- Their performance is not spectacular.

Future work

- Use more training data.
- Evaluate other bracket combination algorithms.
- Combine different learners for bracket identification.
- Use an NP chunker instead of general base chunker.
- Generate more than one evaluation per sentence (?)