

**Analysis of Modal Auxiliaries in Two  
Consecutive Phrases Extracted from the  
British National Corpus**

**ACL 2009 Edmonton**

**Robert Chartrand**

**Shunsuke Nakamoto**

**Hidenobu Kunichika**

**Akira Takeuchi**

**Kyushu Institute of Technology, JAPAN**

# Introduction

- ✱ Chomsky has stated that since language is in the mind of the native speaker, the best source of evidence of any particular language use is in the introspective judgment of the speaker<sup>1</sup>.
- ✱ Research has shown, however, that native speaker intuition is not a reliable guide to the frequency of occurrence in language and uses of vocabulary defined in textbooks or dictionaries are not always accurate<sup>2</sup>.

# Research

- ✱ Extract modal auxiliaries in consecutive sentences or clauses
- ✱ Use computational linguistic techniques for extracting, parsing and simplifying sentences
- ✱ Differentiate between main and subordinate clauses for processing
- ✱ Display examples for learners and teachers
- ✱ Calculate t-scores for frequency ranking

# Which Modal Auxiliaries?

- ✱ Can, must, will, may, shall, should, could, would, might
- ✱ Negatives (can not...)

# Why modals in consecutive sentences?

- ✿ “Modals are one of the most difficult structures to teach to ESL students”<sup>3</sup>
- ✿ Specific context is used in natural discourse
- ✿ Not well researched - in most cases only simple concordances are done
- ✿ Semantically significant<sup>4</sup>
- ✿ To provide examples for English learners and teachers

# Consecutive use of modals - Semantically significant

**IT MIGHT BE HELPFUL  
IF YOU COULD REINFORCE THE POINT**

BNC: HD2 2210

- ✱ “It is common intuition for a sequence of continuing sentences with the modal operators might and could to behave as antecedent and anaphor.”<sup>3</sup>

# Consecutive use of modals - Semantically significant

- ✿ “Modal subordination is a phenomenon which stems from the organization of propositions in discourse. In this semantic theory, a proposition is asserted with a modal expression in the first sentence. A second sentence with a modal then follows, but the meaning is asserted from the antecedent sentence. This relates to anaphora, where the apparent antecedent is a quantified expression and the anaphor is within its scope.”<sup>4</sup>

# Example 1: Q & A

- ✿ **Can** you give me one more day to come up with something?
- ✿ I **will** sleep on it tonight and tomorrow.

# Example 2:

## Two consecutive sentences

- ✿ It really **would** work after all.
- ✿ We **should** know soon enough.

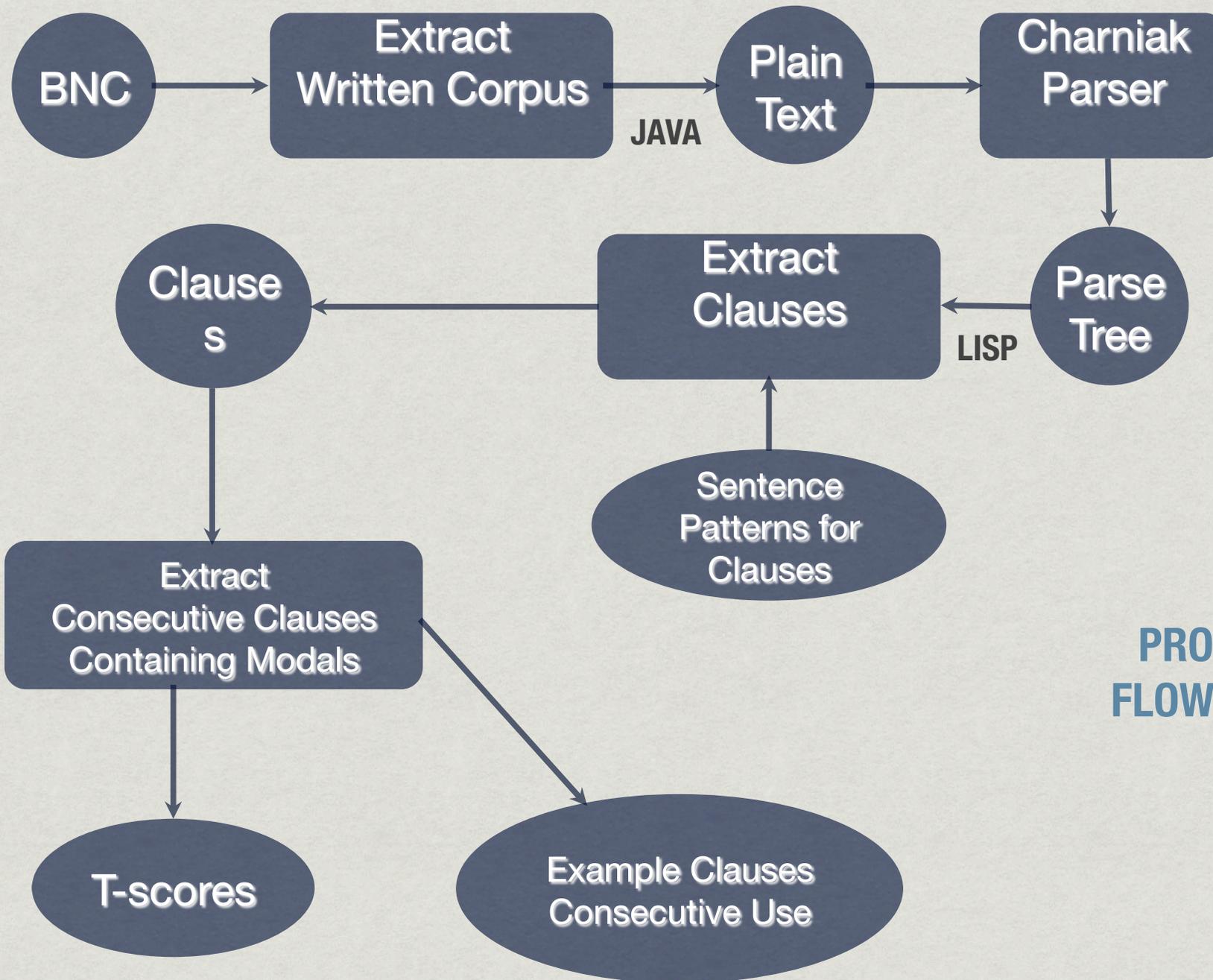
# Example 3:

## Two consecutive clauses

- ✿ The picture **may** become clear next week ...
- ✿ ... the shareholders **should** now take some profits.

# B RITISH N A TIONAL C ORPUS

- ✱ Contains 100 million words (90% written, 10% spoken) - Spoken corpus not used
- ✱ Mainly modern British English
- ✱ Sourced from newspapers, journals, periodicals, books, essays, and other materials
- ✱ 2007 XML edition



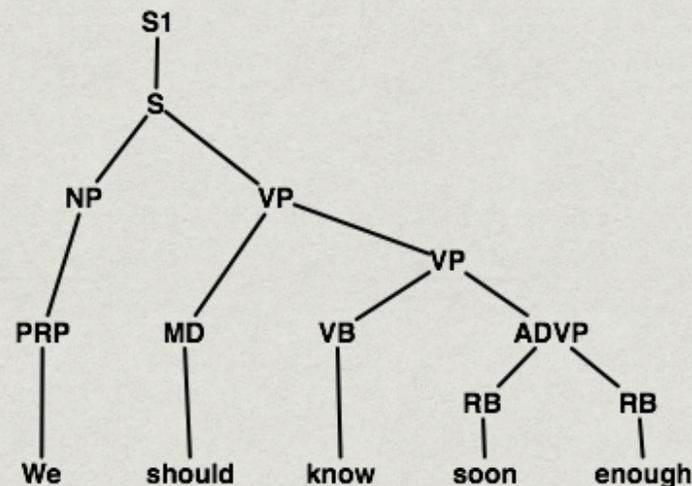
**PROCESS  
FLOWCHART**

# Charniak Parser

- ✱ Can automatically analyze and break down sentences into grammar components

**CHARNIAK PARSER OUTPUT FOR : WE SHOULD KNOW SOON ENOUGH.**

**(S1 (S (NP (PRP We)  
(VP (MD should) (VP (VB know) (ADVP (RB soon) (RB enough))))  
(. .)))**



# Simplification

- ✱ The clauses were identified and the modals were extracted

## ORIGINAL

If our priority is to target help at those with most need and to back away when those needs are met, our contacts with families could become simply box ticking sessions, and it would be harder to get to know them. BNC: EA0 1897

## PARSED EXTRACTED

our contacts with families **could** become  
simply box ticking sessions

and it **would** be harder to get to know them.

# t-scores

- ✱ Measure of confidence that the association between two words is not by chance
- ✱ Not a random association between two words
- ✱ Higher t-score signifies a higher frequency of occurrence

$$t - score = \frac{\left( f(n,c) - \left( f(n) \times f(c) / N \right) \right)}{\sqrt{f(n,c)}}$$

$f(n)$  refers to the frequency of the sentences containing the node modal,  
 $f(c)$  refers to the frequency of the sentences containing the collate modal,  
 $f(n,c)$  refers to the frequency of the consecutive sentences containing the node and collate modals,  
 $N$  refers to the number of sentences in the corpus

# T-scores of modals in consecutive sentences

M 1/M2	can	must	will	may	shall	should	could	would	might	can't	mustn't	won't	mayn't	shan't	shouldn't	couldn't	wouldn't	mightn't
can	56	6.1	9.4	15.9	1.4	0.9	-22.8	-40	-1.4	10.6	-0.6	3.7	4.9	0.8	2	-34.8	-16.3	-0.4
must	6.6	36	-0.5	3.1	1.9	5.9	-9.3	-9.9	-5.1	7.8	9.3	2.3	2.3	-0.9	3.1	-2.3	-1.4	-5.5
will	11.8	-7.8	110.5	10.8	7.8	2.2	-41.8	-68.5	-13.4	-3.2	-2.4	22.1	3.2	3.7	-2.9	-46.3	-21.7	-4.9
may	11.3	3.4	12.9	62.2	2.9	7.7	-11.1	-22.8	3.1	3.6	0.4	3.3	17.3	2	3.7	-21.3	-11.3	0.2
shall	-0.6	1.4	7.9	3.1	25.2	-1.2	-19.1	-16.7	-2.8	0.3	-0.8	2.2	0.7	7.1	-0.7	-9.3	-5.7	n/a
should	3.8	5.4	1.1	5.8	-1.2	48.3	-3.2	0.4	2.3	-1.3	2.2	-0.1	0.7	1.1	12.5	-5.7	-2.5	-0.2
could	-31.6	-9.7	-50.5	-24.2	-12.1	-9.8	37.6	25.8	11	-12.8	-2.3	-8.8	-2.9	-4.1	-4.5	9	8.6	1.1
would	-45.5	-14.2	-66.3	-37.1	-12.5	-10.7	22.3	95.1	14.1	-11.4	-3.5	-13.4	-7.8	-3.4	-2.9	7.5	21.9	3.7
might	-6.9	-6.8	-16.1	-1.9	-1.4	-0.3	10.4	17.1	24.6	-4	-1.8	-3.3	-1.9	-2	-1.8	3.6	6.6	5.7
can't	15	10.7	2.4	2.9	2.1	-1.9	-8.8	-7.4	-0.1	18.7	2.6	6.5	0.7	1.6	1.1	-5.4	-0.8	n/a
mustn't	0.3	8.2	-0.2	-0.7	1.5	4.5	-2.3	-1.6	-1.8	1.5	5.9	-1.6	1.7	0.7	2	0.1	-1	n/a
won't	0	0.4	26	1.2	2.5	-0.7	-11.8	-15.7	0.3	2.6	-3.2	15.1	0.8	0.9	-1.2	-12.4	-1.7	0.8
mayn't	5.3	0.6	3.9	15.7	0.1	3.6	-0.1	-6.8	-0.4	-0.8	-0.2	1.9	6.8	0.3	-0.5	-4.2	-4.2	0.1
shan't	0.2	0.3	2.6	0	7.3	-1.3	-4.1	-3	-0.4	1.6	1.2	0.4	0.3	4.8	0	n/a	-1.5	n/a
shouldn't	0.5	2.2	-3.8	2.8	-2.3	15.2	-0.8	-2.2	2.2	-0.2	1.7	1	0.5	n/a	7.8	-1.8	-0.6	-0.3
couldn't	-26.2	-2.1	-45.2	-24.4	-13.8	-9	9.6	13.8	3	-5.6	-0.8	-11.3	-7.4	-0.7	-0.3	14.6	7.4	-0.1
wouldn't	-18.2	-3.2	-23	-13.3	-6.4	-5.5	3.6	23.6	3.8	-2.3	-1	-0.8	-1.2	n/a	-1.3	5.5	14.1	0.7
mightn't	-3.8	-1.5	-2	0.4	0.4	0.1	4.1	4.1	6	-0.3	n/a	-2.1	n/a	n/a	0.5	1.1	1.6	3.1

# t-scores of positive modals

M 1/M2	can	must	will	may	shall	should	could	would	might
can	56	6.1	9.4	15.9	1.4	0.9	-22.8	-40	-1.4
must	6.6	36	-0.5	3.1	1.9	5.9	-9.3	-9.9	-5.1
will	11.8	-7.8	110.5	10.8	7.8	2.2	-41.8	-68.5	-13.4
may	11.3	3.4	12.9	62.2	2.9	7.7	-11.1	-22.8	3.1
shall	-0.6	1.4	7.9	3.1	25.2	-1.2	-19.1	-16.7	-2.8
should	3.8	5.4	1.1	5.8	-1.2	48.3	-3.2	0.4	2.3
could	-31.6	-9.7	-50.5	-24.2	-12.1	-9.8	37.6	25.8	11
would	-45.5	-14.2	-66.3	-37.1	-12.5	-10.7	22.3	95.1	14.1
might	-6.9	-6.8	-16.1	-1.9	-1.4	-0.3	10.4	17.1	24.6

# Analysis

## EXAMPLE SENTENCES FOR EACH T-SCORE RANK

T-score	Modal 1	Modal 2	Rank	
23.9	could	would	11	High T-score
16.0	can	may	19	
12.2	may	will	28	
10.7	can not	must	33	
10.2	could	might	35	Mid T-score
5.0	should	must	65	
3.9	might not	would	70	
2.8	should not	may	86	
2.0	must	will not	100	
0.9	will not	shall not	127	Low T-score
-6.6	might not	can	259	
-12.3	would	shall not	280	
-20.7	may	could not	296	
-24.3	would not	will	300	
-72.5	will not	would	314	

# Evaluation

Question	Results (%)	
Q1	83.4	High T-score
Q2	80.0	
Q3	73.0	
Q1	71.8	Mid T-score
Q2	79.0	
Q3	67.0	
Q1	86.2	Low T-score
Q2	85.0	
Q3	71.0	
Q1	80.5	Average
Q2	81.3	
Q3	70.3	

**Q1: ARE BOTH PHRASES UNDERSTANDABLE?**

**Q2: DO MODALS FLOW NATURALLY?**

**Q3: IS THE VOCABULARY LEVEL  
APPROPRIATE FOR LEARNERS?**

**QUESTIONS ASKED TO 5 NATIVE SPEAKING  
ENGLISH INSTRUCTORS AT UNIVERSITY  
LOOKING AT >100 EXAMPLE SENTENCES**

could, would (23.9)

**FIRST COMMISSIONS **COULD** BE OBTAINED  
THROUGH POLITICAL INTEREST**

**AND THIS **WOULD** APPEAR TO HAVE BEEN  
THE SITUATION OF PHILIP HAY**

can, may (15.9)

**IF THE BEHAVIOUR BEHIND THE UNWANTED  
FELINE ACTIVITY **CAN** BE ANALYSED**

**IT **MAY** BE POSSIBLE TO FIND A  
PSYCHOLOGICAL SOLUTION**

may, could not (-20.7)

**SOLICITORS MAY SET UP A CORPORATE  
PRACTICE**

**BUT SUCH A COMPANY COULD NOT  
PRACTICE IN ENGLAND AND WALES**

will not, would (-72.5)

**AND THE INCREASE IN TEMPERATURE  
WILL NOT BE FELT EVENLY  
IT WOULD BE LESS AT THE EQUATOR**

# Conclusion

- ✱ Consecutive clauses with modals are useful expressions to learn
- ✱ Simplification of sentences makes it easier to understand for English learners
- ✱ Pairs of modals were ranked according to t-scores
- ✱ Future work - website for learners and educators

# References

1. Chomsky, N., *Aspects of the theory of syntax*. 1st paperback ed. MIT. 1969, Cambridge: M.I.T. Press. x, 251 p.
2. Partington, A., *Patterns and meanings : using corpora for English language research and teaching*. Studies in corpus linguistics,. 1998, Amsterdam ; Philadelphia: J. Benjamins Pub. vii, 162 p.
3. Celce-Murcia, M., D. Larsen-Freeman, and H.A. Williams, *The grammar book : an ESL/EFL teacher's course*. 2nd ed. 1999, Boston: Heinle & Heinle. viii, 854 p.
4. Kibble, R., *Modal Insubordination*, in *Empirical Studies in Formal Syntax and Semantics*, F. Corblin, D. Godard, and J.-M. Marandin, Editors. 1997, Peter Lang: Bern.
5. Roberts, C., *Modal subordination, anaphora, and distributivity*. Outstanding dissertations in linguistics. 1990, New York ; London: Garland. ix, 255 p.

**Thank you**

**robertchartrand@kurume-u.ac.jp**

**<http://www.kurume-u.ac.jp/>**