Corpus Linguistics Research Group (CRG)

Departments of Linguistics and Computing, Lancaster University

If-conditionals as modality attractors

Costas Gabrielatos

20 March 2006

Abstract

The talk will examine the case for treating if-conditionals as strong attractors of modality. The claim is tested through keyword comparisons of un-annotated corpora, namely a sample of 853 if-conditionals from the written BNC, and, as reference corpora, the written BNC Sampler, FLOB, all the if-sentences from the written sub-corpus of the BNC, and the non-conditional if-sentences from the sample. Further tests involve the comparison of specific modal words between the manually annotated sample and the annotated versions of BNC, BNC Sampler and FLOB. The talk will also comment on issues arising from problems encountered in the two types of comparison, as well as issues pertaining to corpus annotation, quantitative analysis, and the definition and formal characteristics of if-conditionals.

2

Hypotheses

- *If*-conditionals are strong attractors of modality.

- *If*-conditionals can be regarded as modal constructions or modal colligations.

Tests

- Do *if*-conditionals show a statistically significant higher frequency of modal expressions than average?

- Do *if*-conditionals show a statistically significant higher frequency of modal expressions compared to non-conditional *if*-constructions?

3

Semantic prosody / preference

Semantic prosody

- The "consistent aura of meaning with which a form is imbued by its collocates."
(Louw, 1993: 157)

Semantic preference

- The "relation between a lemma or word-form and a set of semantically related words."
(Stubbs, 2001: 111)

4

Colligation

- Co-occurrence of grammatical categories.
(Firth, 1968: 181)

- Co-occurrence of lexis and grammatical categories.
(Stubbs, 2001: 112)

- "The grammatical company a word keeps."
(Hoey, 1997: 8)

5

Modal colligation

- A hybrid between colligation and semantic preference.

- In more general terms it could be termed *semantic colligation*.

- The mutual attraction holding between a grammatical construction, *if*-conditionals, and "a set of semantically related words" (Stubbs, 2001: 111), or, more generally, a semantic category: *modality*.

6

If-clause modalisation

Category	Freq.	*% (n=853)*
Modalised	280	*32.8%*
Unmodalised	570	*66.8%*
Elliptical (non-inferable)	3	*0.3%*
Total	853	*100%*

- 1/3 of *if*-clauses are modalised …
- … in addition to the modalisation through *if.*
- 1% have two or more modal markers.

7

Main clause modalisation

Category	Freq.	*% (n=853)*
Modalised	607	*71.1%*
Unmodalised	230	*27.0%*
Elliptical (non-inferable)	16	*1.9%*
Total	853	*100%*

- 7/10 of main clauses are modalised.
- 7% of all main clauses have two or more modal markers.

8

Modal load

Rough calculation

- More than half of the clauses in the sample are modalised.

- On average, one modalisation per *if*-conditional.

9

Keyword analysis

- Un-annotated corpora
- Min. LL=6.6 (*p≤0.01*)
- Up to 5-grams
- *n*-grams: complete (MWEs) or indicative
- *n*-grams with *if* not considered
- Sample, FLOB, BNC sampler (written), written BNC, *if*-s-units in written BNC
- Bold indicates KW differences

10

Positive KW (7.5%)	Sample		BNC Sampler		LL	*p≤*
	Freq.	%	Freq.	%		
can	106	0.42	2,095	0.19	48.4	0.000000
could	68	0.27	1,525	0.14	22.6	0.000002
cannot	22	0.09	194	0.02	33.1	0.000000
may	72	0.28	1,254	0.12	42.8	0.000000
might	46	0.18	474	0.04	58.8	0.000000
must	40	0.16	1,034	0.10	8.5	0.003648
shall	21	0.08	224	0.02	25.8	0.000000
should	63	0.25	1,376	0.13	22.5	0.000002
will	131	0.52	3,119	0.29	36.6	0.000000
would	147	0.58	2,364	0.22	101.3	0.000000
wouldn't	10	0.04	57	<0.01	21.8	0.000003
you'd	12	0.05	46	<0.01	33.7	0.000000
probably	14	0.06	193	0.02	12.3	0.000451
want	27	0.11	398	0.04	21.4	0.000004
think	27	0.11	503	0.05	14.0	0.000182
know	31	0.12	644	0.06	12.6	0.000395

11

Positive KW (10.7%)	Sample		FLOB		LL	*p≤*
	Freq.	%	Freq.	%		
can	106	0.42	1,772	0.17	60.8	0.000000
cannot	22	0.09	239	0.02	24.6	0.000001
could	68	0.27	1,569	0.15	17.4	0.000031
may	72	0.28	1,208	0.12	41.0	0.000000
might	46	0.18	641	0.06	36.3	0.000000
must	40	0.16	803	0.08	15.2	0.000096
shall	21	0.08	197	0.02	27.8	0.000000
should	63	0.25	1,115	0.11	32.1	0.000000
will	131	0.52	2,284	0.22	69.1	0.000000
would	147	0.58	2,308	0.23	95.0	0.000000
wouldn't	10	0.04	128	0.01	9.0	0.002738
you'd	12	0.05	82	<0.01	21.6	0.000003
they'll	5	0.02	25	<0.01	11.4	0.000726
probably	14	0.06	239	0.02	7.7	0.005560
want	27	0.11	439	0.04	16.3	0.000054
think	27	0.11	604	0.06	7.6	0.005873
unable	6	0.02	59	<0.01	7.5	0.006020
willing	4	0.02	26	<0.01	7.5	0.006181

12

Bigrams: Additions

Positive KW (*18.3%*)	Sample		BNC sampler		LL	*p≤*
able to	15	0.06	219	0.02	12.1	0.000508
be able	11	0.04	76	<0.01	20.6	0.000006
have to	22	0.09	353	0.03	15.2	0.000098
is unlikely	3	0.01	15	<0.01	7.2	0.007478
it were	5	0.02	20	<0.01	13.7	0.000214
obliged to	3	0.01	9	<0.01	9.6	0.001944
ought to	4	0.02	31	<0.01	6.8	0.009131
will probably	3	0.01	9	<0.01	9.6	0.001944

Positive KW (*18.3%*)	Sample		FLOB		LL	*p≤*
able to	15	0.06	258	0.03	8.1	0.004396
are to	9	0.04	65	<0.01	15.4	0.000086
be necessary	5	0.02	18	<0.01	14.0	0.000180
necessary to	7	0.03	70	<0.01	8.6	0.003298
have to	22	0.09	398	0.04	10.7	0.001073
want to	14	0.06	230	0.02	8.3	0.003983

13

Trigrams: Additions

Positive KW (*27%*)	Sample		BNC Sampler		LL	*p≤*
a right to	3	0.01	3	<0.01	14.5	0.000140
be able to	11	0.04	75	<0.01	20.9	0.000005
have a right	3	0.01	2	<0.01	16.0	0.000062
i think that	3	0.01	8	<0.01	10.2	0.001435
was going to	4	0.02	24	<0.01	8.4	0.003785

Positive KW (*16.2%*)	Sample		FLOB		LL	*p≤*
be able to	11	0.04	112	0.01	13.3	0.000267

14

4-grams and 5-grams: Additions

Positive KW (*16.7%*), (*50%*)	Sample		BNC Sampler		LL	*p≤*
ought to be able	3	0.01	0		22.7	0.000002
ought to be able to	3	0.01	0		22.3	0.000002

Positive KW (*12.5%*), (*50%*)	Sample		FLOB		LL	*p≤*
ought to be able	3	0.01	0		22.3	0.000002
ought to be able to	3	0.01	0		22.7	0.000002

15

Questions 1

- Is the apparent semantic attraction a characteristic of *if*-conditionals in general, or of the makeup of the *if*--conditionals in the sample?

→ KW comparison: sample - *if*-s-units in written BNC.

→ 205,275 s-units, approx. 6.8 mil. words.

16

Positive KW	Sample		*if*-s-units in written BNC		LL	$p \leq$
	Freq.	%	Freq.	%		
cannot	22	0.09	2,804	0.04	9.7	0.001881
may	72	0.28	13,174	0.19	9.4	0.002163
might	46	0.18	7,258	0.11	11.0	0.000914
shall	21	0.08	2,644	0.04	9.5	0.002062
shall have	5	0.02	166	<0.01	12.0	0.000521
i ought	3	0.01	45	<0.01	11.5	0.000704
you'd	12	0.05	1,280	0.02	7.7	0.005607
you must	9	0.04	850	0.01	7.1	0.007634
we shall	6	0.02	436	<0.01	6.9	0.008613

- No negative KW.
- KW comparison: *if*-s-units with BNC sampler and FLOB.
- If sample is modality-heavy → Fewer positive KWs than in comparisons of sample with same corpora.
- **More modal KWs** (Word doc, p.11)

17

Questions 2

- Is the attraction a feature of conditionality or of the word *if*?

- KW comparison: conditional with non-conditional *if*-s-units in the sample.

- No modal KWs.

- Are 'non-conditional' *if*-constructions still tinged with conditionality because of *if*?

18

Manual comparison

- Some distortion expected, because of homographic nouns (*May, might, must, will*).

- Contracted forms (subject+modal, negatives) were treated as a single word.

- Their keyness was calculated separately.

- → Manual comparison of central modals conflating full and contracted forms.

19

Manual comparison

Modal	Sample		BNC written			
	Freq.	Freq./mil.	Freq.	Freq./mil.	Diff. %	LL
would	152	6030.07	232738	2666.43	*126.2%*	**78.46**
might	33	1309.16	50757	581.51	*125.1%*	**16.87**
must	32	1269.49	63840	731.40	*73.6%*	**8.16**
shall	8	317.37	17426	199.65	*59.0%*	1.48
can	85	3372.08	194664	2230.22	*51.2%*	**12.73**
may	47	1864.56	107805	1235.10	*51.0%*	**6.99**
should	41	1626.53	97043	1111.80	*46.3%*	5.25
will	113	4482.88	271838	3114.40	*43.9%*	**13.35**
could	51	2023.25	139997	1603.91	*26.1%*	2.56

Sample size makes comparison very sensitive to actual frequencies. (p.151)

Manual comparison to corpora nearer to sample size: BNC sampler, FLOB.

20

Modal	Sample		BNC Sampler			
	Freq.	Freq./mil	Freq.	Freq./mil.	Diff. %	LL
would	152	6030.07	2615	2416.64	*149.5%*	**92.75**
will	113	4482.88	3546	3276.75	*36.8%*	**9.77**
can	85	3372.08	2264	2092.09	*61.2%*	**16.18**
could	51	2023.25	2615	2416.44	*-16.3%*	-1.67
may	47	1864.56	1022	944.40	*97.4%*	**17.04**
should	41	1626.53	1388	1282.61	*26.8%*	2.09
might	33	1309.16	471	435.23	*200.1%*	**27.63**
must	32	1269.49	863	797.47	*59.2%*	5.80
shall	8	317.37	224	206.99	*53.3%*	1.24

21

	Sample		FLOB			
Modal	**Freq.**	**Freq./mil.**	**Freq.**	**Freq./mil.**	**Diff. %**	**LL**
would	152	6030.07	2719	2664.06	*126.3%*	**76.09**
will	113	4482.88	2603	2550.40	*75.8%*	**29.16**
can	85	3372.08	1997	1956.65	*72.3%*	**20.55**
could	51	2023.25	1771	1735.21	*16.6%*	1.11
may	47	1864.56	1102	1079.73	*72.7%*	**11.44**
should	41	1626.53	1148	1124.80	*44.6%*	4.82
might	33	1309.16	642	629.02	*108.1%*	**13.64**
must	32	1269.49	815	798.53	*59.0%*	5.76
shall	8	317.37	197	193.02	*64.4%*	1.64

22

- Focus: Modality, *not* specific modal expressions.

- **Ideal**: totalling all modal expressions (lexical and grammatical) in the sample and reference corpora → KW comparison.

- **Feasible**: Keyness of central modals taken as a group.

- Central modals account for approx. 60% of modal expressions in the sample.

23

	Sample	written BNC	BNC Sampl.	FLOB
Freq.	562	1176108	15008	12994
Freq./mil.	22295.39	13474.40	13868.42	12731.43
Diff. %	NA	*65.5%*	*60.8%*	*75.1%*
LL	NA	121.36	105.87	143.29

24

Counting within constructions

- Why discrepancies between automatic and manual KW analysis?

- Text portions not belonging to the construction

- → Overestimation of sample size.

- → Underestimation of keyness.

25

Example 1

- (1) *Yes, I come from Lochaber, and the Lochaber people, if they were here, would be at one with the people of Breadalbane.*

- (2) *If the leg is cured while it is still attached, it is technically a gammon -- hence the confusion caused by the term "gammon ham".*

- The inessential elements account for 27.3% and 37.5% of (1) and (2) respectively.

26

Example 2

- (3) *Why should the fact that D was engaged on causing damage to property at the time (even damage to D's own property) make his conduct into an offence punishable with life imprisonment when, **if** D were engaged on some other activity, it would not be punishable as such and would only amount to manslaughter **if** a death happened to be caused?*

- To maintain sample randomness, only the conditional sentence containing the *if* picked out by the 'thin' function of BNCweb was taken into account and annotated

27

References

Firth, J.R. (1968). A synopsis of linguistic theory. In Palmer, F.R. (Ed.) *Selected Papers of J.R. Firth 1952-59* (168-205). London: Longmans.

Hoey, M. (1997). From concordance to text structure: New uses for computer corpora. Paper given at the 1997 Practical Applications of Language Corpora (PALC) conference, University of Lodz, April 12-14, 1997. In Melia, J. & Lewandoska, B. (Eds.) *Proceedings of PALC 97*. Lodz: Lodz University Press.

Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In Baker, M., Francis, G. and Tognini-Bonelli, E. (Eds.) (1993). *Text and Technology: In honour of John Sinclair*, Philadelphia and Amsterdam: John Benjamins, 157-176.

Stubbs, M. (2001). *Words and Phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.