

Topicality and Raising to Subject

Scott Grimm

Dept. of Linguistics, Stanford University

sgrimm@stanford.edu

January 10, 2009

Raising to Subject: Background

- A class of verbs and adjectives, including *seem*, *appear*, *likely*, take both sentential and infinitival complements
 - (1) It seems that Barnett understands the formula.
(SentComp)
 - (2) Barnett seems to understand the formula.
(InfComp)
- The standard view is that these two expressions are in some sense equivalent, either truth-conditionally (Davies and Dubinsky, 2004, p. 4) or via selectional restrictions
- This talk will contribute new empirical evidence and isolate some of the factors that distinguish the two forms of expression (topicality and evidentiality).

Raising to Subject: Background

3 Common Assumptions:

- At the level of thematic selection, both InfComp and SentComp constructions select for a proposition
- The raising predicate does not select for its subject
(*Barnett isn't engaged in some act or state of 'seeming'*)
- The subject is selected with respect to the proposition:
(3) Barnett_i [seems [t_i to understand]]
- *seem* can have wide-scope in both:
seem [Barnett to understand the formula]_{prop}
seem [Barnett understands the formula]_{prop}

Topicality

- *Main Claim:* The InfComp is strongly associated with a topic-comment structure, while the SentComp construction permits embedded subjects which are not topics.
- This can be established by examining both constructions w.r.t. characteristics of topic-comment structures
 - Topic tests
 - Distribution of Information Status for subjects
 - Acceptability of non-topics as subjects

Topicality: Topic Tests

- Both the InfComp and SentComp pass topic tests, which evaluate for changes of meaning under topicalization (Reinhart 1981).
 - (4) a. Felix seems to be back in town.
b. As for Felix, he seems to be back in town.
 - (5) a. It seems that Felix is back in town.
b. As for Felix, it seems he is back in town.
- \Rightarrow Subjects of both the InfComp and SentComp are potential topics.

Information Status Distribution

- Do subjects of the InfComp and SentComp differ in information structure properties?
 - If so, this should be reflected in the information status of the subjects found across a corpus.
- British National Corpus: 200 tokens each of *seem* or *appear* with the InfComp or SentComp (800 tokens total)

Information Status: Nissim et al. 2004

- Nissim et al. (2004) measures information status in terms of hearer identifiability
- *Old*: previously mentioned, pronouns, generics
6 Subtypes
- *New*: Not previously mentioned or accessible to the hearer.
No Subtypes
- *Mediated*: “Mediated entities have not yet been directly introduced in the dialogue, but are inferrable from previously mentioned ones, or generally known to the hearer.”
9 Subtypes

Topicality: Information Status Distribution

- The two constructions differ in their distribution:
 - InfComp shows reliable preference for old material
 - SentComp shows greater acceptance of new material

	old	mediated	new
InfComp	234	124	13
SentComp	174	144	47

- $\chi^2 : p < .0001$
- Discourse-new subjects, while hearer new, were clearly speaker-identified—still qualify as topics, e.g. contrastive topics, specific indefinites
⇒ Subjects of InfComp are topical

Spoken Modality

- Both to control for modality of production and as a means of independent verification, I examined the occurrences of *seem* in a version of the Switchboard Corpus (Godfrey et al. 1992)
 - A portion was annotated for information status as part of the LINK project (based on Bresnan et al. (2002) and Zaenen et al. (2004)).
 - Identical annotation scheme, so the comparison was straightforward
- The SentComp construction had too few instances to be informative, but the InfComp instances displayed the same pattern and proportions as in the BNC

	old	mediated	new
InfComp	30	15	-

Topicality: Acceptability of Non-topics

- Indefinite generics fail the topic test (Reinhart 1981).
 - (6) A shark will never attack unless it is very hungry.
?He said about a shark that it will never attack unless it is very hungry.
- If verbs such as *seem* only permit subjects which are topics, then indefinite generics should not be permitted

Generics: Background

- Different generics have different interpretations:
 - An inductivist use: true when “*sufficiently many relevant* individuals in the domain of the generic satisfy the predicated property.” (Cohen 2001, p. 194)
 - (7) Kings are generous.
 - Normative generic sentences “do not get their truth or falsity as a consequence of properties of individual instances . . . instead, [they] are evaluated with regard to rules and regulations” (Cohen 2001, p. 194).
 - (8) Bishops move diagonally.
 - (9) A bishop moves diagonally.

Generics: Background

- Normative statements can come with a deontic force:
 - (10) A Christian is forgiving. Papafragou (1996)
 - (11) Christians are forgiving.
- (10) does not indicate that any Christian actually achieves this ideal, but that they *should*, in contrast to (11), which on the preferred reading does say something about the real world.

Generics: Background

- Bare plurals allow for both *inductivist* and *normative* (or *definitional*) readings
- Indefinite singular generics only permit a normative reading

(12) Kings are generous. (✓ induct. / ✓ norm.)

(13) A king is generous. (# induct. / ✓ norm.)

Topicality: Acceptability of Non-topics

- The SentComp permits normative/definitional indefinite generic statements, and hence non-topic subjects:

(14) It seems that a bishop moves diagonally.

- The InfComp does not:

(15) ?A bishop seems to move diagonally.

(16) ?A king seems to be generous.

Topicality: Acceptability of Non-topics

- A similar infelicity for InfComp constructions occurs for indefinite generics with a deontic reading
 - (17) a. A Christian is forgiving.
b. ?A Christian seems to be forgiving.
 - (18) a. A gentleman opens doors. (Burton-Roberts 1977)
b. ?A gentleman seems to open doors.
- Such singular indefinites can be made acceptable when they are either specific or contrastive—i.e., when they are made topical
- Non-topics are not acceptable as subjects of the InfComp \Rightarrow Subjects InfComp must be topics

Comparison: Passives and Unaccusatives

- Other constructions analyzed as A-Movement do not show such constraints:
 - Passive:

(19) A president is voted in by members of the country, company, or other entity. (Google)
 - Unaccusative:

(20) An iceberg melts during the summer. (Google)
- Nor are infinitive constructions problematic:

(21) The world wants a man to be financially effective. (Google)
- *seem* appears to constrain possible subjects

Idiom Chunks

- Idiom chunks are used as a diagnostic for raising
 - Idiom chunks are supposedly non-referential and acceptable with raising structures
- Idioms do however show traits of referential expressions, such as serving as antecedents for pronouns (Nunberg et al. 1994).

(22) (Nunberg et al. 1994, p. 502)

Once someone lets the cat out of the bag, it stays out of the bag for good.

Idiom Chunks

- If the claim that subjects of the InfComp are associated with topicality holds, it is expected that for idioms or fixed expression with truly non-referential subjects, infelicity would arise

(23) a. A fool and his money are soon parted.

b. ?A fool and his money seem to be soon parted

(24) a. A still tongue keeps a wise head.

b. ?A still tongue seems to keep a wise head.

- The above infelicity is parallel to that of the indefinite generics

Interim Summary

- On the above evidence:
 - Subjects of the SentComp may be topics or not
 - Subjects of the InfComp may only be topics
- This evidence also implicates that the InfComp and SentComp cannot be truth conditionally equivalent in all cases
- The InfComp aligns with a topic-comment structure (so-called “categorical” statements)
- The SentComp is not so restricted

Implications of Topicality: Scope

- Topics are backgrounded and linked to the prior discourse
- One would expect then that the subject of InfComp has wide scope
- While there has been a long history of asserting “reconstructed readings” and quantifying-in, where the subject of the InfComp has narrow scope w.r.t. *seem*, this stance has come under scrutiny (Chomsky 1993, 1995; Lasnik 1998, 2003)
- There are clear asymmetries which are consistent with the subjects of InfComp being topics

Scope: Quantifiers

- Partee (1971) noted that the InfComp and SentComp manifest interpretational differences when quantifiers are present

- (25) a. Nobody appears to have passed the test.
b. It appears that nobody passed the test.

- (22a) designates a set of students which the speaker has information about who did not pass : *nobody* has scope over *appear*.
- (22b) has the collective reading, i.e. not one of the students passed: *appear* has scope over *nobody*.
- This is expected if the subjects of InfComp are topics: refer to a specified set.

Scope: Scopal Non-Specifics

- Scopally non-specific subjects are attested with the SentComp construction:
 - (26) It appeared that a German S.P. gun had joined the snipers and was lobbing the occasional shell into the vicinity of the orchard. (BNC A61)
⇒ there existed a German S.P. gun
- The subject of the InfComp version preferentially takes wide-scope
 - (27) A German S.P. gun appeared to have joined the snipers and appeared to be lobbing the occasional shell into the vicinity of the orchard.

Scope: Summary

- General scoping preferences:
 - InfComp: Subj > Predicate
 - SentComp: Predicate > Subj
- The preferential reading patterns are consistent with the subjects of the InfComp being topics, as topics are backgrounded, referential, and would refer outside of the context induced by the predicate

Qualifying Quantifying-in

- Montague (1973) and others have used examples as below to argue that raising constructions permit scopally non-specific (i.e. narrow-scope) readings of the subject.
 - (28) A cat seems to be in the garden. [existence]
 - (29) A train seems to be approaching. [appearance]
- Yet the most successful examples always use verbs of existence and appearance
- ***EXPAND Such verbs are peculiar in both their lexical semantics and information structure properties (e.g., do not permit topic marker *wa* in Japanese)

Qualifying Quantifying-in

- When the phrases are minimally altered in the choice of the verb, the scopally non-specific reading becomes far less accessible, if not impossible

(30) A cat seems to be sleeping in the garden.

(31) A train seems to be leaving.

- The most natural readings involve a specific cat and a specific train.

Evidentiality

- Evidentiality provides a second clear instance where the InfComp and SentComp differ

A perceptual experience constraint has been noted for the InfComp (Postal (1974); Asudeh and Toivonen (2007)).

(32) Julius Caesar appeared to be honest.

(33) It appears that Julius Caesar was honest.

- This can be generalized to connect with the domain of evidentiality
- Evidentiality has recently been posited as relevant for appearance of *to be* (de Haan 2007)

Evidentiality: Background

- Direct evidential marking is used when “the speaker has some sort of sensory evidence for the action or event he/she is describing” (de Haan (2004)).
 - Visual, auditory or other means of direct experience
- Indirect evidentials “are used when the speaker was not a witness to the event but when he/she learned of it after the fact” (ibid.).
 - This includes *inference* and *hearsay*.
- *Hypothesis*: The direct evidential function is associated with the InfComp; the indirect evidential function is associated with the SentComp

Back to the Corpus!

- Return to BNC corpus: 200 tokens each of *seem* or *appear* with the InfComp or SentComp (800 tokens total)
- As the direct evidential is the unmarked form, nearly impossible to code for it based on objective criteria (at this point)
- Indirect evidence (inference; hearsay) was often explicitly marked:
 - (34) It also seemed, from the feathers on the kitchen floor, that one of the pigeons had come down for a warm and had got too close . (BNC HTL)
- Markers: *from*-phrases, *therefore*, *then*, *thus* and explicit scientific conclusions

Distribution of Data

- 147/400 SentComp tokens were explicitly marked for inference or hearsay functions (37%)
- \Rightarrow SentComp is associated with a indirect evidential function

Indirect Evidential Markers

<i>as</i>	2
<i>because</i>	2
consequent	10
evidential source	11
<i>for</i> phrase	5
<i>from</i> phrase	40
<i>given that</i>	2
legal conclusion	5
<i>on</i> phrase	5
scientific conclusion	15
<i>then</i>	12
<i>therefore</i>	19
<i>thus</i>	7
other	14
<hr/>	
Total (out of 400 tokens)	147

Evidentiality and Topicality

- Direct evidential statements require an entity or event upon which the evidential statement is based
 - this entity or event is presupposed with respect to the evidential statement
- This aligns with typical topic-comment structures
 - The entity or event is backgrounded (topic) and the property predicated of the entity is asserted (comment)
- For statements based on abstract inference/hearsay, there is no particular entity that is presupposed with respect to the evidential statement—no necessary alignment with topic-comment structure

Evidentiality and Topicality

- The same acceptability patterns observed with indefinite generics by modulating construction type (InfComp vs. SentComp) can be obtained by modulating evidential type:
 - (35) a. I saw that a king is generous. (inductive)
 - b. I heard that a king is generous.
(inductive/normative)

Implications: Lexicalization Patterns

- The association between evidential type and construction type makes a clear prediction as to lexicalization patterns of predicates which accept only one construction.
 - Predicates with direct evidential function should appear in the InfComp construction
 - Predicates with an indirect evidential function should appear in the SentComp construction
- Predicates such as *looks to* and *inferable* support this prediction

- (36) a. It is inferable that Ed left.
b. *Ed is inferable to leave

- (37) a. Ed looks to be tired.
b. *It looks that Ed is tired.

Implications: Lexicalization Patterns

- (38) a. It is inferable that Ed left.
b. *Ed is inferable to leave

- (39) a. Ed looks to be tired.
b. *It looks that Ed is tired.

I have changed the fuser hoping that was where the loud noise was coming from. No luck, it sounds to be coming from the back of the machine.

<http://forums13.itrc.hp.com/service/forums/question>

*it sounds that it

Controlling for other factors

- Multivariable regression models allow to control for other factors
- Subject Length (raw and log)
- Nominal Expression Type :
 - *pronoun* (including definite, personal and reflexive pronouns, as well as demonstratives)
 - *definite*
 - *indefinite* (including phrases with the indefinite article as well as bare plurals)
 - *proper name*
 - *quantifier* (such as *most*, *few*, *any*)
 - *relative pronoun*
 - *there*
 - *verbal*

Controlling for Other factors

- The more fine-grained information status subtypes yielded three significant factors which were associated with the SentComp construction:
 - *new* ($p \approx 0.002$)
 - *old-ident-generic* ($p \approx 0.003$)
 - *old-generic* ($p \approx 0.02$)
- The remaining significant factor was *indirect evidential* ($p < 2e-16$)

General Summary

- The InfComp and SentComp systematically diverge in the set of readings that they permit
- The readings which are not simple subject-predicate (topic-comment) relations are associated with the SentComp

	InfComp	SentComp
Sing. Generic	inductive	inductive, normative
Quantifiers	distributive	collective, distributive
Indefinite Subject	scopally specific	scopally non-specific/specific
Evidential	direct evidential	indirect evidential

- These divergent readings align well with the thesis that the subjects of InfComp are topics (e.g. subjects are specific/referential)

General Summary

General implications for the raising analysis:

- It is improbable that *seem*, etc. select for propositions in a simple or unified manner—for not just any proposition is felicitous in the InfComp formulation
- Raising verbs in the InfComp do select for their subject, although they do not discriminate in terms of thematic content: such verbs, at minimum, select for topics, and for *seems* and *appear*, the subjects of the InfComp are constrained to be direct evidential sources

Thank you

Thanks to the following for discussion (and challenges):
Eve Clark, Cleo Condoravi, Beth Levin, Asya Pereltsvaig,
Ivan Sag, Tom Wasow
and to Fabio Del Prete and Chigusa Kurumada for help with
the Italian and Japanese data

Expletive Subjects

- Typically a distinction is drawn between thematic subjects, e.g. agents and so forth, and *there*
- It is not clear whether this distinction entails that *there* is completely non-referential
- Many proposals have claimed that while *there* is clearly not thematic, it does refer.
 - Hartmann (2008) analyzes *there* as “a proform that picks up the situation/location from the context”, with the default context being “the situation here and now” (p. 103)
 - There is a host of claims in the literature that expletive elements such as *there* refer to a spatio-temporal variable, e.g. Ramchand (1996), Felser and Rupp (1997)

Examples of Indirect Evidential Markers

as: It appears, as a general finding, that there is much cynicism amongst primary school teachers about the possibility of innovation in curricula when teaching is under its present constraint.

because : Because of the similarity of wording between s.69 and s.5 of the Civil Evidence Act 1968 , it would appear that the courts may hold the same view with regard to civil evidence.

consequent: If the pattern in Hungary (for which country data based on vital statistics in 1967 are available) is a common one, it would appear that, in the more developed countries, at least, maternal age and birth spacing also influence fetal viability independently of each other (Table 16).

Examples of Indirect Evidential Markers

evidential source: Provided that Ali's analysis of Hacıhasanzade's motives is correct-and one must remember that Ali is writing nearly a century after the event it would appear that already at the beginning of the sixteenth century the career of a kasabat kadi was regarded as a dead end .

for phrase: It would seem that in Capadocea they were even converted to Christianity for, in that region, St Christopher is often depicted with a dog's head.

from phrase : It appears from the case decisions on this point that if the court feels that the driver should have seen the signal then the offence will be committed.

Examples of Indirect Evidential Markers

given that : Indeed, given that the sciences are more usually concerned with the inanimate world than the animate, it would seem that there was much less potential in the physical sciences for portraying women negatively.

legal conclusion : So it would seem that even a defence may be an abuse of process if it is not based on private rights.

on phrase : On examination of the papers, it would appear that many of these Nature papers are geophysical in subject.

scientific conclusion : If quantification is risky, it does at least seem that the two-class, middle/working conception is quite deeply rooted in popular ideology.

Examples of Indirect Evidential Markers

then : It would appear, then, that in an exchange rate union member countries enjoy a higher degree of monetary autonomy than in a currency union.

therefore : It would appear therefore that changes in the nature (location?) of the management function must go along with all the other changes with which we are involved.

thus : It would thus appear that shares were incorrectly priced.

References

- Ash Asudeh and Ida Toivonen. Copy raising and perception, 2007. unpublished ms.
- Joan Bresnan, Jean Carletta, Richard Crouch, Malvina Nissim, Mark Steedman, Tom Wasow, and Annie Zanen. Paraphrase analysis for improved generation. LINK project, HRCR Edinburgh-CLSI Stanford, 2002.
- N. Burton-Roberts. Generic sentences and analyticity. *Studies in Language*, 1:155–196, 1977.
- W.D. Davies and S. Dubinsky. *The grammar of Raising and Control: A course in syntactic argumentation*. Oxford: Blackwell, 2004.
- Ferdinand de Haan. Semantic distinctions of evidentiality. In M. Dryer, M. Haspelmath, D. Gil, and B. Comrie, editors, *Chapter in World Atlas of Linguistic Structures*. Oxford University Press, 2004.
- Claudia Felser and Laura Rupp. A minimalist approach to existential constructions in Germanic. *Essex Research Reports in Linguistics*, 1997.
- J. Godfrey, E. Holliman, and J. McDaniel. Switchboard: Telephone speech corpus for research and development. In *ICASSP-92*, 1992.

- Jutta Hartmann. *Expletives in Existentials: English There and German Da*. PhD thesis, Universiteit van Tilburg, 2008.
- R. Montague. The proper treatment of quantification in ordinary English. In K. J. J. Hintikka, J. M. E. Moravcsik, and P. Suppes, editors, *Approaches to Natural Language*. D. Reidel Publishing Co., 1973.
- Malvina Nissim, Shipra Dingare, Jean Carletta, and Mark Steedman. An annotation scheme for information status in dialogue. In *Proceedings of the 4th Conference on Language Resources and Evaluation (LREC2004)*, 2004.
- Geoffrey Nunberg, Thomas Wasow, and Ivan A. Sag. Idioms. *Language*, 70(3):491–538, 1994.
- Anna Papafragou. On generics. UCL Working Papers, 1996.
- Barbara Partee. On the requirement that transformations preserve meaning. In C. Fillmore and T. Langendoen, editors, *Studies in Linguistic Semantics*. Holt, Reinhart and Winston, 1971.
- Paul Postal. *On Raising*. MIT Press, 1974.
- Gilian Ramchand. Two types of predication in Scot-

tish Gaelic. *Natural Language Semantics*, 4:165–191, 1996.

Tanya Reinhart. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica*, 27:53–94, 1981.

Annie Zaenen, Jean Carletta, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden, Tatiana Nikitina, M. Catherine O’Connor, and Tom Wasow. Animacy encoding in English: why and how. In D. Byron and B. Webber, editors, *The 2004 ACL Workshop on Discourse Annotation*, 2004.