

Corpus Pragmatics: Something old, something new

David Beaver
Department of
Linguistics
UT Austin

TLS 2007, Austin Texas



Confession



- My thesis (1995/2001) contained one naturally occurring example.
- **“Sweden may export synthetic wolf urine - sprayed along roads to keep elk away - to Kuwait for use against camels.”** *Associated Press, January 19, 1995*
- Like many linguists brought up in a tradition of artificial examples (and often self-verified judgments), my work has shifted to include more corpus data and more experimental data.

2

“Tell me one important generalization that’s come out of corpus research!”
- a syntactician, p.c., 2007

3

Plan: three corpus vignettes

- Anaphora resolution
- Accent prediction
- Existential constructions

4

First vignette Anaphora resolution

- D. Beaver, *New Directions in Discourse Optimization*, Linguistics and Philosophy, 2004
- W. Gegg-Harrison & D. Byron. *PYCOT: An OT-based pronoun resolution component*, LREC 2006
- Byron, D. & W. Gegg-Harrison, *Evaluating Optimality Theory for Pronoun Resolution Algorithm Specification*, DAARC 2004

5

The anaphora resolution problem

And Caleb said, "The one who attacks Kiriathsepher and captures it, I will even give him my daughter Achsah for a wife." And Othniel, the son of Kenaz, Caleb's younger brother, took it; so he gave him his daughter Achsah for his wife.

Judges 1:12, New American Standard Bible, 1995.

How do we know what the pronouns mean?

6

Centering

- There is a huge literature on how anaphoric expressions are resolved.
- An old approach, Centering (Grosz *et al* 1983/1995) is still the most influential.
- The basic idea is simple: speakers produce texts that minimize abrupt, unexpected transitions.
- Hearers then resolve anaphora by tracking what referents have been mentioned, and how prominent they are, and assuming that the speaker will always signal transitions in what's being talked about ahead of time.

7

The BFP Algorithm

- Centering attracted huge interest, but was imprecisely stated.
- The first precise statement was by Brennan, Friedman and Pollard (BFP).
- Unfortunately, the BFP model is stated in the form of an algorithm which is hard to understand at an intuitive level.

1. CONSTRUCT THE PROPOSED ANCHORS for U_n

- (a) Create set of referring expressions (RE's).
- (b) Order RE's by grammatical relation.
- (c) Create set of possible forward center (CF) lists. Expand each element of (b) according to whether it is a pronoun or a proper name. Expand pronouns into set with entry for each discourse entity which matches its agreement features and expand proper nouns into a set with an entry for each possible referent. These expansions are a way of encoding a disjunction of possibilities.
- (d) Create list of possible backward centers (CB's). This is taken as the entities from $CF(U_{n-1})$ plus an additional entry of NIL to allow the possibility that we will not find a CB for the current utterance.
- (e) Create the proposed anchors. (CB-CF combinations from the cross-product of the previous two steps)

2. FILTER THE PROPOSED ANCHORS

For each anchor in our list of proposed anchors we apply the following three filters. If it passes each filter then it is still a possible anchor for the current utterance.

- (a) Filter by contradictions. That is, if we have proposed the same antecedent for two contradicted pronouns or if we have proposed an antecedent for a pronoun which it is contradicted with, eliminate this anchor from consideration.
- (b) Go through $CF(U_{n-1})$ keeping (in order) those which appear in the proposed CF list of the anchor. If the proposed CB of the anchor does not equal the first element of this constructed list then eliminate this anchor. This guarantees that the CB is realized as a pronoun. (If there are no pronouns in the proposed CF list then the anchor passes this filter. This corresponds to rule 1 in section 1). This rule could be implemented as a preference strategy rather than a strict filter.
- (c) If none of the entities realized as pronouns in the proposed CF list equals the proposed CB then eliminate this anchor. This guarantees that if any element is realized as a pronoun then the CB is realized as a pronoun. (If there are no pronouns in the proposed CF list then the anchor passes this filter. This corresponds to rule 1 in section 1). This rule could be implemented as a preference strategy rather than a strict filter.

3. CLASSIFY and RANK

- (a) Classify each anchor on the list of proposed anchors by the transitions as described in section 1 taking U_{n-1} to be the previous utterance and U_n to be the one we are currently working on.
- (b) Rank each proposed anchor using the extended ranking in section 2. Set $CH(U_n)$ to the proposed CB and $CF(U_n)$ to proposed CF of the most highly ranked anchor.

Centering in Optimality Theory (COT)

- In Beaver (2004), I proposed a framework for stating Centering type theories in Optimality Theory.
- An independently motivated set of constraints exactly reproduce the predictions of the BFP Centering model under a certain constraint ranking.
- I also suggested some improvements, including an alternative ranking.

9

The COT constraints (ranked so 1 is strongest)

1. **AGREE** Anaphoric expressions must agree with antecedents for number and gender.
2. **DISJOINT** Co-arguments of a verb are disjoint.
3. **PRO-TOP** The topic is pronominalized.
4. **FAM-DEF** The referent of each definite NP is familiar and no new information is provided by the definite.
5. **COHERE** The topic of the current sentence is the topic of the previous one.
6. **ALIGN** The topic is in subject position.

10

Example

- 1.a) Jane_i likes Mary_j .
- b) She_k often visits her_l for tea_m.
- c) The woman_n is a compulsive tea drinker.

11

She_k often visits her_l for tea_m in COT

	AGREE	DISJOINT	PRO-TOP	FAM-DEF	COHERE	ALIGN
$k = i, l = j$					*	
$k = l = i$		*			*	
$k = j, l = i$					*	*
$k = l = j$		*			*	
$k = i, l \notin \{i, j\}$				*	*	
$k = j, l \notin \{i, j\}$				*	*	
$k \notin \{i, j\}, l = i$				*	*	*
$k \notin \{i, j\}, l = j$				*	*	*
$k, l \notin \{i, j\}, k \neq l$			*	**	*	*
$k = l \notin \{i, j\}$		*	*	**	*	*

12

The woman_n is a compulsive tea drinker.

	AGREE	DISJOINT	PRO-TOP	FAM-DEF	COHERE	ALIGN
 $n = k$			*			
$n = l$			*		*	
$n = m$			*	*	*	
$n \notin \{k, l, m\}$			*	*	*	*

13

Advantages of COT

- Independent motivation: Every constraint is based on rules from prior linguistic literature.
- Modular: effects of individual constraints are easily seen, and constraints can be independently modified, removed, or added.
- Reversible: can model production as well as comprehension. This allows applications to text generation (not discussed here).

14

Empirical basis of Centering/COT

- Centering was motivated by a small number of example discourses, primarily artificial.
- There has been much post hoc empirical work justifying various aspects of Centering (experimental, corpus, cross-linguistic comparison).
- But the form in which Centering was originally stated meant that it was hard to see what the space of such theories was, and to separate individual empirical claims.
- In this sense, standard Centering does not consist of a set of easily testable hypotheses.
- The constraints in COT are much more directly testable, and hence falsifiable. (Unfortunately.)

15

Testing COT

- Greg-Harrison & Byron (2004) and Byron & Greg-Harrison (2006) implemented COT as a testbed for comparing different theories of anaphora.
- They tested the implementation on the 1M word WSJ portion of the Penn Treebank, annotated for anaphoric coreference, and found 80% correct resolution for COT, close to the ceiling for current models.

16

Mea culpa

- However, they also tested all reorderings of subsets of the constraints, and found:

“the ordering proposed by Beaver is indeed the best ordering if all constraints are present, but removing the ALIGN constraint boosted performance slightly”
(G-H&B 2004)

- Moral: **I was wrong.**
(Then again, 5 out of 6 constraints still look ok!)

17



18

More generally...

- Objective measures of linguistic theories cannot be based on hand-picked, or, worse, hand-crafted examples. The only objective measures come from performance on text and speech generated by writers and speakers unconstrained by the linguist.
- But a theory that maximizes performance over a set of artificial examples is unlikely to provide the best model on naturally occurring texts.

19

A place for artificial examples?

- On the other hand, there may still be a place for artificial examples. In this case, many parts of a theory motivated by artificial data survived a fairly rigorous test, and *might* form part of later theories.
- And modular statements of theories, however motivated, provide an avenue for development in the face of individual counterexamples or bulk testing on naturally occurring data.
- So maybe the moral is not that we should stop using artificial examples to develop theories, but that we should stop only after we've also tested those theories on natural texts.

20

Second vignette

Accent and Information Status

- M. Wolters & D. Beaver, *What does "he" mean?*, Proceedings of the Annual Meeting of the Cognitive Science Society, 2001
- A. Nenkova, J. Brenier, A. Kothari, S. Calhoun, L. Whitton, D. Beaver, and D. Jurafsky. *To Memorize or to Predict: Prominence Labeling in Conversational Speech*, NAACL-HLT 2007
- J. Brenier, A. Nenkova, A. Kothari, L. Whitton, D. Beaver, D. Jurafsky. 2006. *The (Non)Utility of Linguistic Features for Predicting Prominence in Spontaneous Speech*, IEEE/ACL 2006
- A. Kothari, *Accented Pronouns and Unusual Antecedents: A Corpus Study*, SIGDIAL 2007

21

Plan for this section

- How does *information status*, such as *old* vs. *new* (previously mentioned vs. not), affect what is accented in speech?
- We start with accented pronouns, and yet more problems with Beaver (2004).
- Then we turn to a more general consideration of how accents are distributed in speech.

22

Accented pronouns

- Pronouns are sometimes accented, and this is known to correlate with changes in reference in some cases:

2. John_i called Bill_j a Republican and then

- a) He_i insulted him_j
- b) HE_j insulted HIM_i

(Lakoff 1971)

23

Pronouns and accent

- It has often been argued that accent on a pronoun is used by speakers to signal the information status of the referent.
- Specifically, accent is claimed to signal that a pronoun should not be resolved to the most salient entity.
- See e.g. Ariel (1990), Gundel et al (1993), Cahn (1995), Nakatani (1997), Kameyama (1999).
- In Beaver (2004), I show that the effect of accent on pronominal resolution preferences could be seen as a case of *partial blocking*, the process whereby a marked form is given a marked interpretation.

24

Accented pronouns in Beaver (2004)

- Under further assumptions about the interpretation process (technically: *bidirectional optimization*), COT predicts that accent on pronouns will shift interpretation to an otherwise dispreferred resolution.
- Only one very general, independently motivated constraint is needed (cf. Schwarzschild 1999):
AvoidF: Avoid focus, i.e. don't accent something unless you need to.
- There's only one major problem with this analysis...

25

The problem with the analysis

- It's probably wrong!
- Production experiments conducted as I was writing Beaver (2004), but published earlier (Wolters and Beaver 2001) already indicated that the data was not clearly as Kameyama and others had described.
- 20 subjects read texts with pronouns and definite descriptions that either involved topic shifts or not, and we measured fundamental frequency (f0) and other parameters.

26

Results

	-SHIFT		+SHIFT	
	mean F0	max F0	mean F0	max F0
-PRO	180 ± 54	221 ± 65	179 ± 53	211 ± 62
+PRO	173 ± 77	183 ± 78	199 ± 69	212 ± 85

Effect of topic shift on frequency (hz)
for pronouns and non-pronouns

- Subjects reading texts in which there was a switch in pronoun reference did not reliably accent pronouns.
- Though f0 tends to change in the direction Beaver (2004) et al predict, this is not significant, and, topic shifted pronouns still have lower max f0 than nouns in definite descriptions, and the small f0 effects probably have other causes (paragraph effects).

27

Conclusions from Wolters & Beaver (2001)

- “Any formal semantic theory of accented pronouns needs to deal with the fact that in many cases, this accent may be optional... secondly, most of the accented pronouns in our corpus data could be interpreted as cues to some sort of contrast.” (W&B 2001)
- We also concluded that we needed a design based on spontaneous rather than read speech.

28

Kothari's corpus study

- Extending work she conducted in the Stanford *Synthesis* project, which I co-Pled, Kothari (2007) studied accenting of pronouns in a corpus of spontaneous speech.
- She used 19 dialogues from Switchboard - telephone conversations hand-annotated for accent, contrastiveness and coreference.
- There were 834 3rd person pronouns in the sample, and Kothari considered how accenting related to the antecedent (e.g. position, form), and the pronoun (e.g. position, overall rate of pronominal accenting of the speaker, contrastiveness).

29

Kothari's results

- Factors connected to properties of the antecedent (which correlate with switch of reference) were not significant factors in any of her statistical models.
- The only reliably significant factors were whether the pronoun was used contrastively, whether the pronoun was in subject position, and speaker effects.

30

Conclusions from empirical work on accented pronouns

- There is a long history of linguists providing models of how accent on pronouns affects reference.
- This includes the formally precise demonstration in Beaver (2004) of how such an effect might be derived from independent principles.
- The main problem with all these models is that there is no good reason to believe that accent on pronouns does affect reference.

31

Models of old/new and accent

- Accent on pronouns is not related to information status of the antecedent in the standardly understood way.
- But is information status in general related to accent?
- Across theoretical linguistics this is standardly assumed e.g. Chafe (1987), Gundel et al (1993), Lambrecht (1996), Schwarzschild (1999), Selkirk (1996).
- The standard hypothesis: discourse new constituents must contain an accent, but discourse old constituents need not.

32

The Synthesis project

- The Synthesis project studied accent distribution in the same section of Switchboard.
- The goal was to build a model of prominence for use in e.g. TTS systems, and at the same time test some of the ideas in the linguistic literature.
- 14,555 words were annotated for information status, contrast and concrete/non-concrete distinctions, all features that linguistic literature suggests are predictive of prominence.

33

Corpus Annotations

- Manual features:
 - Info. Status (*new/old/mediated*)
 - Contrast
 - Animacy
- Automatic features:
 - Part-of-speech
 - Unigram / bigram probability
 - TF.IDF (*relative frequency in the dialogue*)
 - Stopword (*special, high frequency word*)
 - Lead word value (*relative frequency in start of conversation*)
 - Verb specificity
 - Accent ratio (*how often word is accented in a large corpus*)

34

Results: good news?

- As reported in Brenier *et al* (2006) & Nenkova *et al* (2007), we did find correlations between information status and accent. But the picture is complex.
- Here are the results for nouns:

	old	mediated	new
accented	156 (61%)	752 (63%)	307 (73%)
unaccented	99 (39%)	437 (37%)	113 (27%)

35

More results

- However, when building decision-tree classifiers using leave-one-out selection, the best accent classifiers did not include information status.
- The best classifiers (>77% accuracy) made greatest use of Accent Ratio, and limited use of some linguistic features (animacy, contrast).
- But leaving out even these linguistic features reduced performance by <0.5%.
- Overall performance is very near a ceiling of inter-speaker variation identified in a separate study: for six subjects reading the same text, a speaker on average agreed on accents for only 82% of words, <5% above the accuracy of our best classifier.

36

Why Old/New doesn't help predict accent

- It seems unlikely that information status will play a major role in applications like TTS.
- But why is Accent Ratio (AR) so effective, and why doesn't information status help?
- Effects of information status are masked by word choice:
 - a speaker who wants to code something old, chooses words with low AR, e.g. pronouns.
 - And a speaker who wants to code something new chooses words with high AR, e.g. low frequency nouns.

37

Discussion of accent results

- Again, corpus work hasn't yielded what linguists would have expected: information status correlates with accent, but not usefully.
- Machines to generate accents need not use information status. So it seems unlikely that humans use accent primarily to signal information status.
- Speculation:
 - Accent is best seen as a way to boost phonological discriminability and hence enhance morpho-lexical distinctions. It is a tendency to make unpredictable material more readily perceived. **(Further work: test this!)**
 - Contrastive accents might be a special case where this tendency is exploited as a signal that a distinction is being made.

38

Third vignette Existential Constructions

- D. Beaver, I. Francez and D. Levinson, 2005, *Bad subject: (non-)canonicity and NP distribution in existentials*, SALT 15
- And further unpublished/ongoing work with:
 - E. Destruel (U.T. Austin)
 - G. Bouma (U. Groningen)
 - I. García Álvarez, T.F. Jaeger (then, Stanford U.)
 - A. Bratkievich, L. Mikkelsen (U.C. Berkeley)

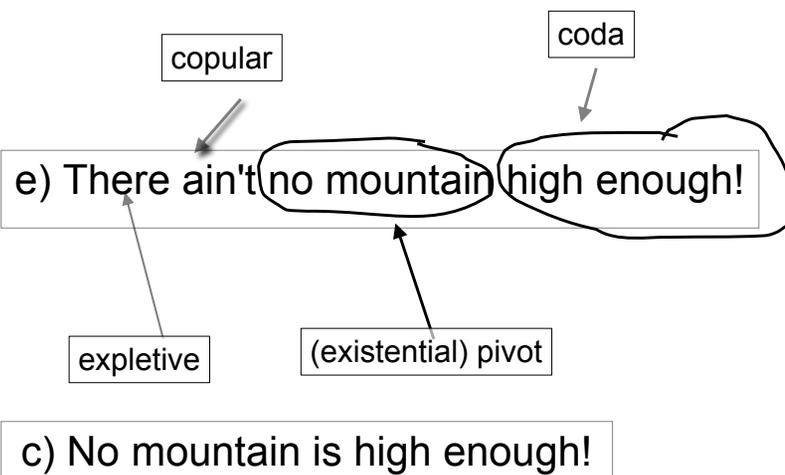
39

Introduction to the existential project

- I'm midway through a project to understand existential constructions cross-linguistically as examples of non-canonical argument realization.
- I'll present the basic ideas and preliminary results.

40

Existential (e) vs Canonical (c) constructions:



41

Non-canonicity

- Many existential constructions involve non-canonical sentence forms:
 - Expletive/null grammatical subject; non-standard position of primary argument (Germanic, Romance)
 - Reduced verbal paradigms or inconsistent agreement (Hebrew, Spanish, English)
 - Marked case (Russian, Hebrew, Finnish)

42

Existential Constructions

1. Beidh go leor bia ann. [!] (Irish, McCloskey 2006)
be [FUT] plenty food in-it
'There'll be plenty of food.'
2. Er zijn alle mijnenopsporingprogramma's. (Dutch, web)
There are all-the mine-clearing-programs
'There are all the mine-clearing programs.'
3. Desafortunadamente hay los estudiantes que no tienen motivación. (Spanish, web)
Unfortunately have the students that no have motivation
'Unfortunately, there are (the) students that lack motivation.'
4. Havia a Terra e havia os planetas. (Portuguese, web)
Has the Earth and has the planets
'There is the Earth and there are the planets.'

43

Existential constructions

5. Na pisjmo otveta ne poluchila. (Russian, web)
on letter-acc answer-gen no come-pst
'There came no answer to the letter.'
6. Muzar she-be-ha'aretz en et kol ha-matkon. (Hebrew, web)
strange that-in- ha'aretz not-be acc all the-recipe
'Its strange that they don't have the whole recipe in Ha'aretz.'
7. Der er kommet alle mulige undskyldninger. (Danish, web)
There is come every possible apology
'Every possible apology came.'

44

Distributional Restrictions

- *Definiteness effects* are usually claimed for existential constructions, e.g. Milsark (1977), Zucchi (1995), Keenan (2003).
- These involve a categorical denial of definite and strong quantificational pivots.
- Such claims are often caveated: definite or strong pivots involve a special *presentational* use/reading of the construction.

45

Basic Idea

- Standard theories of existentials treat them as isolated constructions in a grammatical vacuum.
- We start with an alternative paradigmatic / competition approach related to Lambrecht (2000) and Mikkelsen (2002).
- Perhaps pivots are not realized as syntactically canonical subjects because they lack subject properties semantically and pragmatically.
- The distributional facts result from a competition between forms: **when the sole argument is a bad subject, it appears as a pivot.**

46

Canonical subj. properties

- Silverstein (1976), Keenan (1976) and Aissen (1999) suggest universal tendencies for grammatical subjects, expressible as scales:
 1. **Agent > [order over other θ -roles...] > no θ -role**
 2. **Sentence (aboutness) topic > non-sentence-topic**
 3. **Discourse topic > discourse old > hearer old > hearer new > dependent element > no referent introduced**
 4. **Animate > inanimate > abstract, type/kind denoting**
 5. **Short > long**

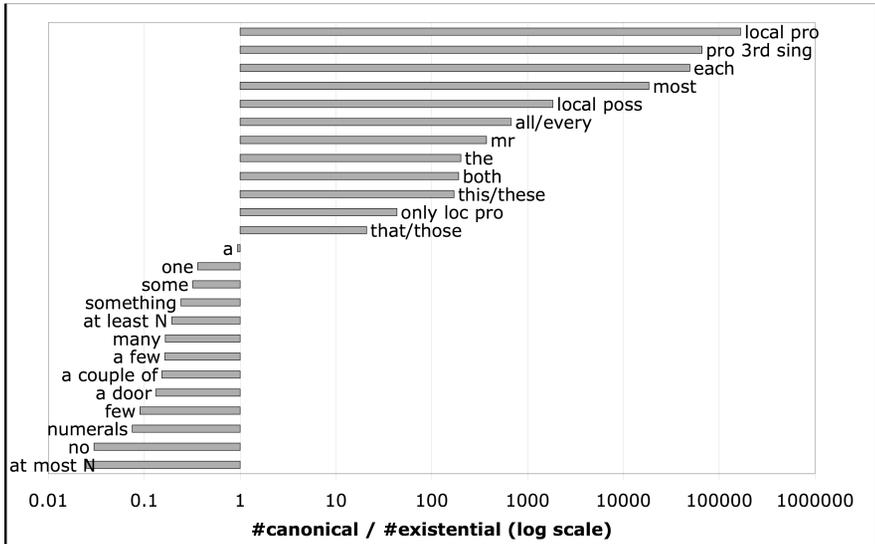
47

The empirical study

- We conducted quantitative corpus and web studies of English, Danish, Dutch, Spanish, Brazilian Portuguese, Russian, and, currently, French.
- We compare frequencies of canonical and existential constructions for a given NP type.
- E.g. is “No X is in the Y” more or less common than “There is no X in the Y”? How about for “a X” or “every X”?

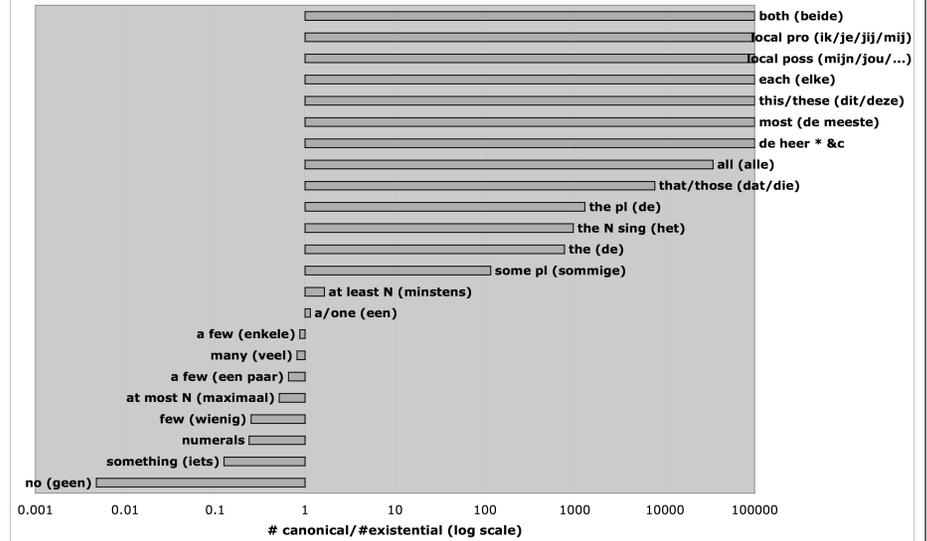
48

Existential ratios English NPs

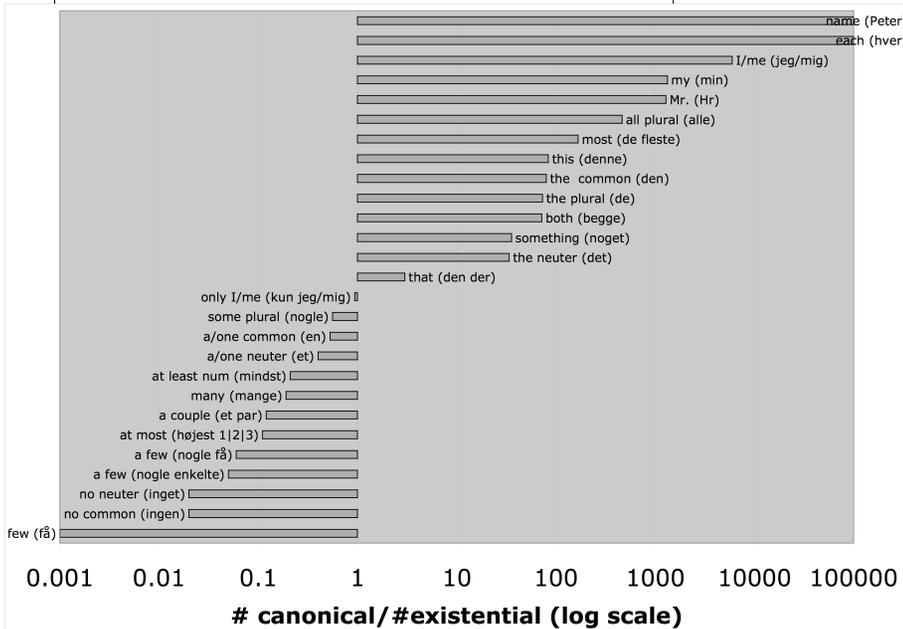


49

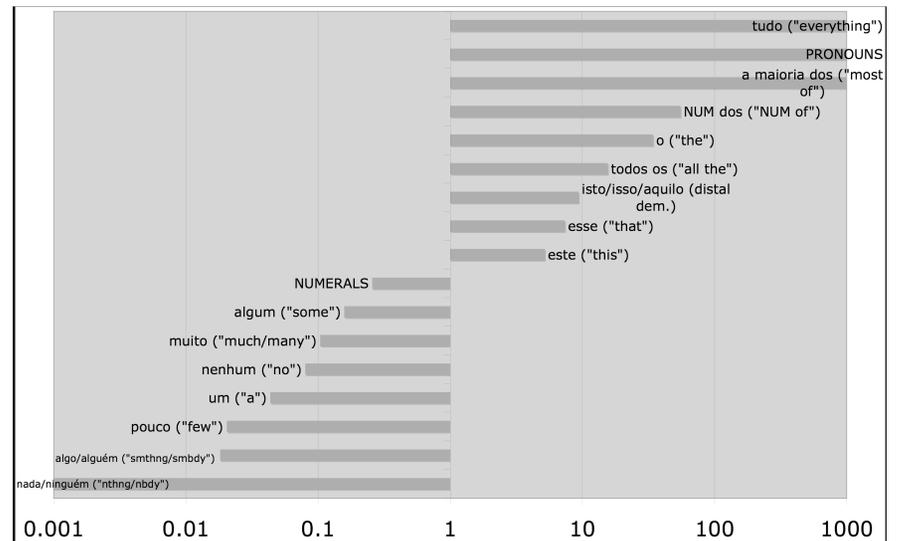
Existential ratios for selected Dutch NPs



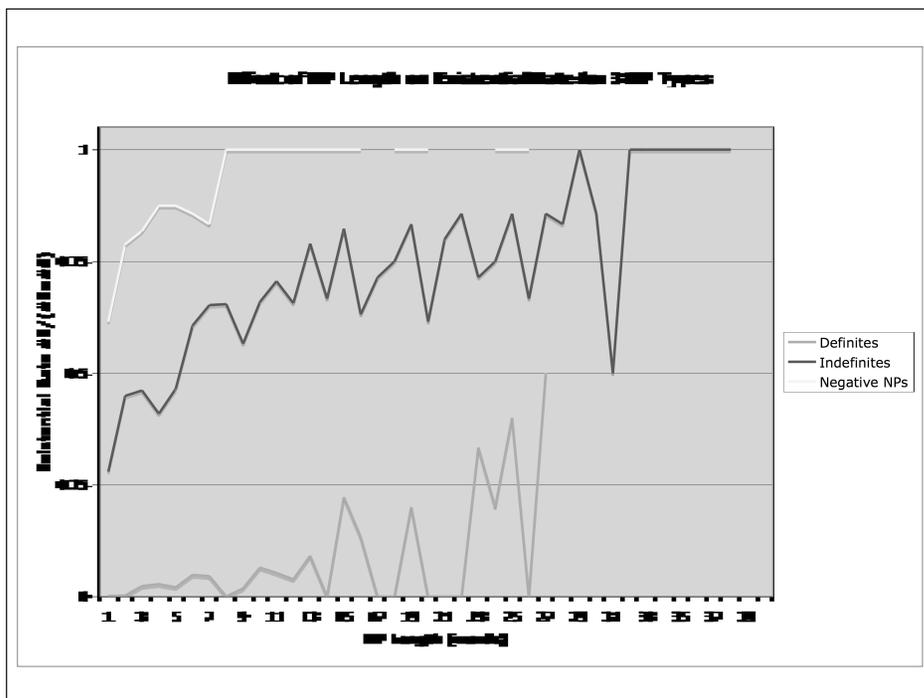
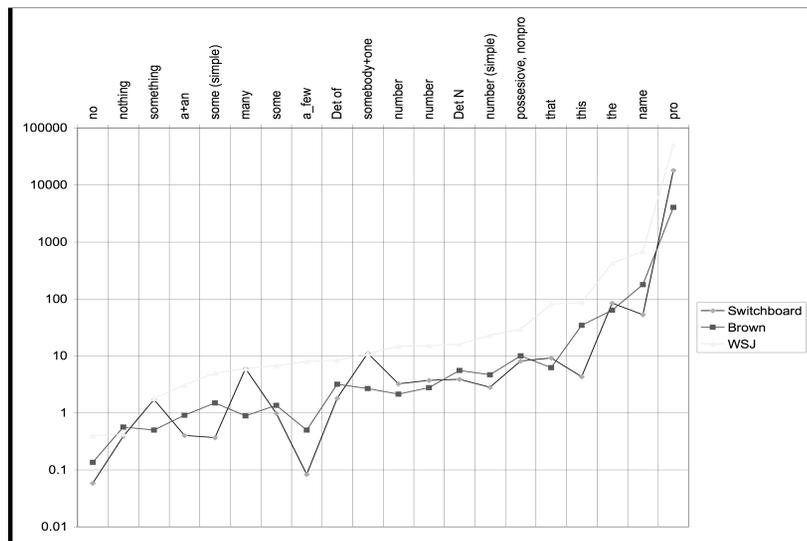
Existential ratios for selected Danish NPs



Existential ratios for selected Brazilian Portuguese NPs



English NPs by Genre



Summary of Results

- We find strongly correlated results across languages, though none of the patterns are predicted by any existing theory of existentials.
- Crosslinguistically, the same scale of NP types recurs:
Local pro > non-local pro > proportional NPs > definite descriptions, demonstratives > prototypical indefinites > neg NPs
- Theories postulating a categorical definiteness effect are falsified. However, we do find a strong quantitative effect: a factor of 20+ difference in existential ratios between typical indefinites and typical definites or strong quantifiers.

Summary of Results (cont.)

- We find clear anti-definiteness effects, especially for negative determiners. This is predicted by the competition model, but not by any standard *syntagmatic* account.
- There are huge genre effects, a factor of >10 between spoken and written English.
- We also highly significant weight effects across NP types: longer NPs are much more likely to appear in pivot position than in canonical subject position, and *vice versa*.

Discussion of Existential Data

- Theoretical linguists don't mind making very complex generalizations, so long as they are categorical.
- Gradient phenomena, in which many factors have a cumulative or probabilistic effect are alien to that tradition.
- It could be that distributional effects in existential constructions are inherently non-categorical.

57

Back to information status?

- There is another possibility: there could be some factor that is not directly identifiable using current searches, but that categorically predicts distribution in existentials.
- Ward & Birner (1995) in fact suggest a possible factor: information status. They claim that existential pivots must be hearer new.
- Although they (uniquely in prior work on existentials) did use corpus examples, I'm skeptical about their specific proposal.
 - First, it is easy to find naturally occurring counterexamples.
 - Second, my recent experiences suggest that even if an old-new distinction does turn out to be significant, it is unlikely to be the whole story.

58

General conclusion

- Remember that syntactician? Well, here's one important generalization that's come out of corpus research:
- Much of the time, generalizations based on artificial examples and our own intuitions are wrong!

59

- The surprising thing is that people (yes, me too, at times) who have not tested their theories on large amounts of naturally occurring data, can have the temerity to think that they have found "important generalizations."
- But new sources of data cannot be ignored.

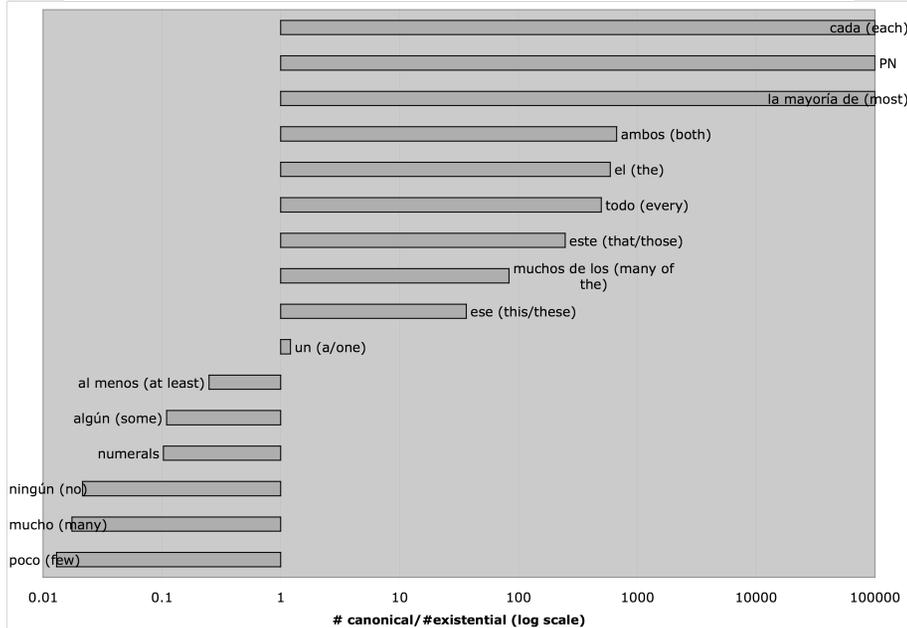


Additional References

- Aissen, J.: 1999, "Markedness and Subject Choice in Optimality Theory", *Natural Language and Linguistic Theory* 17, 1–39.
- M. Ariel.: 1990, *Accessing Noun-Phrase Antecedents*. Routledge, London.
- Cahn, J.: 1995, 'The Effect of Pitch Accenting on Pronoun Referent'. In: *Proceedings of the 33rd International Joint Conference in Artificial Intelligence*, pp. 290-2, ACL.
- Chafe, W.: 1987, *Cognitive Constraints on Information Flow*. In Tomlin (ed.) *Coherence and Grounding in Discourse*, Amsterdam: John Benjamins.
- Grosz, B., A. Joshi, and S. Weinstein: 1983, 'Providing a Unified Account of Definite Noun Phrases in Discourse'. In: *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*. Cambridge, Mass., pp. 44-49, ACL.
- Grosz, B., A. Joshi, and S. Weinstein: 1995, 'Centering: A Framework for Modeling the Local Coherence of Discourse'. *Computational Linguistics* 21(2), 203- 226.
- Gundel, J., N. Hedberg, and R. Zacharski. 1993. *Cognitive Status and the Form of Referring Expressions in Discourse*. *Language*, 69(2):274–307.
- Kameyama, M.: 1999, 'Stressed and unstressed pronouns: complimentary preferences'. In: P. Bosch and R. van der Sandt (eds.): *Focus: Linguistic, Cognitive and Computational Perspectives*. Cambridge University Press.
- Keenan, E.: 1976, 'Towards a Universal Definition of 'Subject'', In Li, Charles (ed.), *Subject and Topic*, Academic Press, New York.
- Keenan, E.: 2003, *The definiteness effect: semantics or pragmatics?* *Natural Language Semantics* 11.187–216

- Lakoff, G.: 1971, *Presupposition and relative well-formedness*. In *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*. Cambridge University Press, Cambridge UK, pages 329-340.
- Lambrecht, K.: 1996, *Information structure and sentence form*, CUP.
- Lambrecht, K.: 2000, 'When Subjects Behave Like Objects: An Analysis of the Merging of S and O in Sentence Focus Constructions Across Languages', *Studies in Language* 24, 611–682.
- Mikkelsen, L.: 2002, 'Reanalyzing the Definiteness Effect: Evidence from Danish', in *Working Papers in Scandinavian Syntax*, volume 69.
- Milsark, G.: 1977. *Toward an explanation of certain peculiarities of the existential construction in english*. *Linguistic Analysis* 3.1–29.
- Nakatani, C.: 1997, 'The Computational Processing of Intonational Prominence: A Functional Prosody Perspective'. Ph.D. thesis, Harvard University.
- Schwarzschild, R.: 1999, 'Givenness, AvoidF and other Constraints on the Placement of Accent'. *Natural Language Semantics* 7(2), 141-177.
- Selkirk, E.: 1996. *Sentence prosody: Intonation, stress, and phrasing*. In *The handbook of phonological theory*, ed. by J. Goldsmith, 550–569. London: Basil Blackwell.
- Silverstein, M.: 1976, 'Hierarchy of Features and Ergativity', In Dixon, Richard (ed.), *Grammatical Categories in Australian Languages*, Australian Institute of Aboriginal Studies.
- Ward, G. & B. Birner: 1995. *Definiteness and the English existential*. *Language* 17.722–742.
- Zucchi, A.: 1995. *The ingredients of definiteness and the definiteness effect*. *Natural Language Semantics* 3.33–78.

A1. Existential ratios for selected Spanish NPs



A2. Existential ratios for 7 NP types by language

