On the expression of proportion: *Most* and *more than half*

Stephanie Solt

Zentrum für Allgemeine Sprachwissenschaft, Berlin

solt@zas.gwz-berlin.de

84th Annual Meeting of the Linguistic Society of America

January 7, 2010

Present Analysis

The basis of the present analysis is corpus data on the usage of *most* and *more than half*

- Davis (2008-) *The Corpus of Contemporary American English*
  - 385 million words (20 mil. words/year for 1990-2008)
  - Spoken language, fiction, popular magazines, newspapers and academic texts

Caveat

- Focus on *most/more than half* used with plural count nouns
- Account can be extended to occurrence with mass nouns by replacing set cardinality with mass measure (e.g. volume)

(3) **Most/more than half** of the milk is spoiled

Semantics of quantity/proportion

- <u>Generalized quantifier theory</u> (Barwise & Cooper 1981): Unified analysis of quantificational expressions as 'quantifying determiners' that express relationships between two sets

- <u>More recently</u>: Unified analysis fails to capture differences between superficially synonymous expressions:
  - Geurts & Nouwen (2007): *more than 3* vs. *at least 4*
  - Fiengo (2007): *each* vs. *every*

Most vs. *more than half*

- A common intuition: *most > more than half*

  (4) Unfortunately, the long term maintenance of the reduced weight is poor, and **more than half, if not most**, of the persons eventually return to their former obese state (http://jcem.endojournals.org/cgi/reprint/83/12/4246.pdf)

  (5) ??…and **most, if not more than half**, of the persons eventually return to their former obese state

  (6) a. **More than half** of the U.S. population is female ✓
       b. **Most** of the U.S. population is female ??

    <u>The facts</u>: female 50.7% vs. male 49.3%
                  (U.S. Census Bureau 2008)

Most vs. *more than half*

(1) a. **Most** Americans favor universal health care
     b. **More than half of** Americans favor universal healthcare

Commonly taken to have equivalent truth conditions:

(2) ⟦most A are B⟧ = ⟦more than half of A are B⟧
     = 1 iff |A ∩ B| > |A - B|

- Some discussion of differences (Ariel 2004, 2005; Horn 2005; Hackl 2009), but most significant divergences have gone unrecognized

- **Goal:** explore and account for distributional and interpretive differences between *most* and *more than half*, and consider more general implications for semantics of proportion
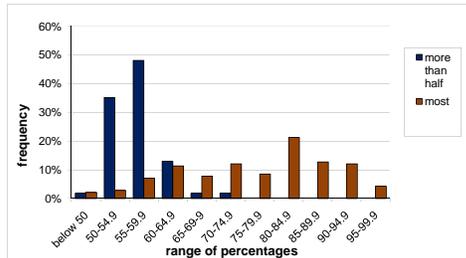

Distribution of *Most /More than Half*

In corpus data, *most* is used for proportions considerably greater than half, while *more than half* is used for proportions close to 50%:

(7) a. The survey showed that **most students (81.5%)** do not use websites for math-related assignments (*Education*, 129(1), pp. 56-79, 2008)

     b. **More than half of respondents (55%)** say that making money is more important now than it was five years ago (*Money*, 21(3), p. 72, 1992)

Distribution of *Most /More than Half*

Most is used for proportions considerably greater than half, while *more than half* is used for proportions close to 50%:

Distribution of *Most /More than Half*

More than half – but not *most* – requires a domain that can be individuated and counted (or otherwise measured)

(13) a. But like **most things**, obesity is not spread equally across social classes *(Mens Health, 23(7), p. 164, 2008)*

      b. ??But like **more than half of things**, obesity is not spread equally across social classes

More than half requires the predicate to be precisely defined; *most* allows a vague predicate

(14) a. ?? **...most of the tourists** in the early days were wealthy
(Natural Parks,: 83(2), p. 14)

      b. ??...**more than half of the tourists** in the early days were wealthy

---

Distribution of *Most /More than Half*

Most is readily followed by a plural noun, yielding a generic-like interpretation; *more than half* is awkward in similar contexts, and (when acceptable) loses the generic flavor:

(8) a. **Most people** follow the moral judgments of those around them *(Writer, 121(7), pp. 30-33, 2008)*

      b. ??**More than half** of people follow the moral judgments of those around them

(9) a. **Most teens** want to fit in with their peers
(CNNYourHealth, 31/8/2002)

      b. ?? **More than half of teens** want to fit in with their peers

---

Distribution of *Most /More than Half*

Use of *more than half* typically co-occurs with mention of a source of supporting data; this is not the case with *most*

	Source of Data Mentioned (Data from COCA)	
	More than Half	**Most**
Americans	9 / 12	13 / 100
Men	4 / 6	5 / 100
Women	4 / 5	7 / 100
Students	5 / 5	36 / 100
Patients	5 / 5	39 / 100
Families	1 / 2	11/ 100
TOTAL	28 / 35	111 / 600
	80%	19%

Distribution of *Most /More than Half*

More than half in fact is rarely followed directly by a plural noun (12% of cases), but instead occurs most commonly in a partitive construction (62% of cases)

(10) **More than half** of the doctoral degrees in engineering awarded by American universities each year go to foreigners. *(Associated Press, 6/1/2007)*

When *more than half* does combine with a bare plural, the latter has a group-denoting (vs. kind) interpretation

(11) ??**More than half** of teens want to fit in with their peers

(12) **More than half of teens surveyed** said they are "not too careful or not at all careful" to protect their skin.
(Today's Parent, 23(7), p. 154, 2006)

Summary of Corpus Data

- *Most* and *more than half* are used to convey distinct ranges of proportions
- *Most* yields a generic interpretation in contexts where *more than half* is infelicitous or has a 'survey results' interpretation
- *More than half* (but not *most*) requires an enumerable domain and a precisely defined predicate
- Use of *more than half* (but not *most*) is typically supported by numerical data (count/survey/ analysis)

Proposal

The distributional and interpretive differences between *most* and *more than half* can be attributed to a fundamental distinction in how proportion is expressed

13

More than half – upper bound

- *More than half* competes with other numerical expressions of proportion.
  - Alternatives can be ordered on scales of varying levels of granularity (Krifka 2009)
  - The choice of *more than half* implies higher scalar options do not apply; *more than half* is restricted (by implicature) to values close to half (Horn 1972; Grice 1975)

more than half (semantic meaning)

more than half (implicated)
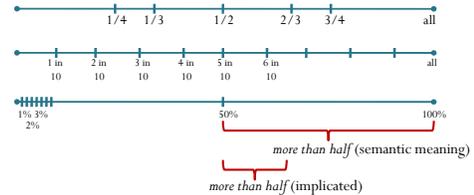
16

Logical form and verification

Hackl (2009): *most* / *more than half* require distinct logical forms that are truth-conditionally equivalent but give rise to different verification strategies

More than half = *many + -er + than half*

(15) $[\![$ more than half of A are B$]\!] = 1$ iff $|A \cap B| > \frac{1}{2}\,|A|$

Verification: Count the As that are B, and compare that to half the total number of As

Most = *many + -est*

(16) $[\![$ most A are B$]\!] = 1$ iff $|A \cap B| > |A - B|$

Verification: Compare the number of As that are B to the number of As that aren't B ('vote counting')

- Support from self-paced counting experiment

14

Most

$$[\![\text{most A are B}]\!] = 1 \text{ iff } A \cap B \text{ larger than } A - B$$

- Logical form involves the comparison of two <u>sets</u>
- Set sizes <u>can</u> – but <u>need not</u> be – compared by exhaustive counting of set members

"…A child can tell which of two plates has more cookies before he can count them (with both readings of *before*!). Primitive man undoubtedly know whether an inimical army was larger, smaller, or (roughly) equal to his own. … We *may* count Peter's class and Paul's class and then infer [Peter's class is larger than Paul's], and often this is the only way of doing it; but we may also just look at the two classes, and will often 'see' which is larger"

(Bartsch & Vennemann 1973: 67-68)

17

More than half

$$[\![\text{more than half of A are B}]\!] = 1 \text{ iff } |A \cap B| > |A|/2$$

- Logical form involves the comparison of two <u>numbers</u>
- Presupposes availability of cardinality of domain set as input to mathematical operation
- As a result:
  - A countable (measurable) domain is necessary
  - The predicate must be sufficiently precise to identify and count the (individual) members of the domain it applies to
  - *More than half* lends itself to contexts where results of counting/ analysis reported

15

Most and comparison

Options for set size comparison:

- Approximately or visually
  - (17) **Most** of the students in this class are female
    - 'Looks like' more females than males
- By induction or generalization
  - (18) **Most teens** want to fit in with their peers
    - In my experience, teens who want to fit in more common than those that don't
- Lack of exceptions
  - (19) **Most** of the tourists in the early days were wealthy
    - Few if any exceptions known

  ➢ Exhaustive counting of set members not required
  ➢ Approximate (procedures fail for sets nearly equal in size)

18

Number Cognition

- Findings from the psychology of number cognition indicate that in addition to a system for the representation of precise numerosity, humans possess a separate system for representing and manipulating approximate number
  - Non-verbal; analog (mental number line)
  - Present in pre-verbal infants, animals, patients with aphasia/acalculia, and societies w/out complex number systems
  - Involved in:
    - Quantity comparison
    - Approximate arithmetic (addition/subtraction – but not multiplication)
  - Ratio dependent (size and distance effects)

(Dehaene 1997; Dehaene, Dehaene-Lambertz, & Cohen, 1998; Gallistel & Gelman 2000; Pica et al. 2004; Feigenson, Dehaene & Selke 2004)

19

Summary

More than half	Most
$|A \cap B| > |A|/2$	$A \cap B$ larger than $A - B$
Comparison of numbers	Comparison of sets
Requires countable sets	Allows uncountable sets
Verification via mathematical relation (>) on numbers	Allows verification via approximate algorithms
Acceptable for proportions close to 50%	Disprefered for proportions close to 50%
Competes w/expressions of proportion (e.g. 2/3)	Competes with *all* as higher alternative
Survey results interpretation	Generic interpretation

22

Most and approximate number

- **Pietroski et al. (2009):** Semantics of *most* based on cardinality comparison, which may be verified via approximate number system
  - Verification displays characteristic size and distance effects
- Further supporting evidence:
  - Children's acquisition of *most* independent of mastery of counting (Halberda, Taing & Lidz 2008)
  - Pirāha – which lacks words for numbers >2 – has an equivalent of *most* (U. Sauerland, p.c.)
- Empirical data explained:
  - Use of *most* where counting precluded
  - Use of *most* when ratio of $A \cap B$ to $A - B$ is 'large enough' to be detected via approximate verification strategy

20

Conclusions

- *Most* and *more than half* represent two distinct modes of expressing proportion
  - *More than half*: comparison of numbers
  - *Most*: comparison of sets
- Distinct logical forms allow different possible verification strategies
  - *More than half*: (precise) mathematical relation
  - *Most*: approximate algorithms/strategies
- Speakers' choice of lexical items takes into count how the sentence could (in principle) be verified
  - Explains wide range of distributional and interpretive differences between *most*/*more than half*

23

Most – *upper bound*

- *Most* does not (directly) encode a proportion
  - Does not invoke a scale of proportions, or compete with items on these scales
  - Alternatives instead are items that express relationships between sets

(20) a. **Some/many/most/all** teens want to fit in with their peers

b. ??Two thirds/seven-in-ten/70% of teens want to fit in with their peers

(21) *Some…many….most….all*
  - *All* as only higher alternative → *most* may be used for proportions approaching all

21

…and a possible parallel

Positive vs. comparative forms of gradable adjectives (cf. Kennedy 2007):

(22)	a. **More than half** of the students are female	# female > # male
	b. **Most** of the students are female	# female >> # male

(23)	a. Billy is **taller** than the average 8-year	Height(B) > average
	b. Billy is **tall** (for an 8-year-old)	Height(B) >> average

(24)	a. Billy is **taller** than Freddie	Height(B) > Height(F)
	b. Billy is **tall** compared to Freddie	Height(B) >> Height(F)

24

Acknowledgments

THANK YOU!

References

Ariel, M. (2004). *Most*. *Language,* 80, 658–706.

Ariel, M. (2005). A 'just that' lexical meaning for *most*. In: K. von Heusinger and K. Turner (eds.) *Where semantics meets pragmatics* (pp. 49-91). Amsterdam: Elsevier.

Bartsch, R. & T. Vennemann (1973). Semantic structures: A study in the relation between syntax and semantics. Frankfurt: Athaenum Verlag.

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural languages. *Linguistics and Philosophy*, 4, 159-219Bresnan, J. (1973). Syntax of the comparative clause construction in English. *Linguistic Inquiry*, 4, 275–343.

Davies, M. (2008-). *The Corpus of Contemporary American English (COCA): 385 million words, 1990-present*. Available online at http://www.americancorpus.org.

Dehaene, S. (1997). The number sense. New York: Penguin Press.

Dehaene, S., Dehaene-Lambertz, G. and Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neuroscience*, 21:355-361.

Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends in Cognitive Science*, 8, 307-314.

References

Fiengo, R. (2007). Asking questions: Using meaningful structure to imply ignorance. Oxford: Oxford University Press.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, 4, 59–65.

Gelman, R. and Gallistel, C.R. (1978). The child's understanding of number. Cambridge, MA: Harvard University Press.

Geurts, B., & Nouwen, R. (2007). *At least* et al.: The semantics of scalar modifiers. *Language*, 83, 533-559.

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41-58). New York: Academic Press.

Hackl, M. (2009). On the grammar and processing of proportional quantifiers: *Most* versus *more than half*. *Natural Language Semantics*, 17, 63-98.

Halberda, J., Taing, L. & Lidz, J. (2008). The development of "most" comprehension and its potential dependence on counting-ability in preschoolers. *Language Learning and Development,* 4(2), 99-121.

References

Horn, L. (1972). *On the semantic properties of logical operators in English*. PhD Dissertation, UCLA.

Horn, L. (2005) The border wars: A Neo-Gricean perspective. In: K. von Heusinger and K. Turner (eds.), *Where Semantics Meets Pragmatics*, Amsterdam: Elsevier.

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable predicates. *Linguistics and Philosophy*, 30, 1-45.

Krifka, M. (2009). Approximate interpretations of number words: A case for strategic communication. In: E. Hinrichs & J. Nerbonne (eds.), Theory *and Evidence in Semantics* (pp. 109-132). Stanford: CSLI Publications.

Pica, P., Lerner, C., Izard, V., and Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, 306(5695):499-503.

Pietroski, P., Lidz, J., Hunter, T., and Halberda, J. (2009). The meaning of 'most': semantics, numerosity, and psychology. *Mind and Language*, 24(5), 554-585.