

Lexical Categories and Clarification: What Do We Clarify And Why?

Matthew Purver
Jonathan Ginzburg
Julian Hough
(with Pat Healey)

Clarification Requests

Ann: I saw John yesterday.

Bob: John??

Ann: Yes, John.

Dr Smith.

The one with the pipe & monocle.

Him.

<points>

British National Corpus

Purver, Ginzburg, Healey (SIGDIAL 2001):

Unknown: What are you making?

Anon 1: Erm, it's a do- it's a log.

Unknown: A log?

Anon 1: Yeah a book, log book.

Richard: No I'll commute every day

Anon 6: Every day?

Richard: as if, er Saturday and Sunday

Anon 6: And all holidays?

Richard: Yeah

A: You see this thing did you buy this separately or did it come in the Walkman?

B: We were lent them.

A: Lent them?

B: Yeah.

Communicator Corpus

Rieser & Moore (ACL 2005):

Cust: I'll be returning on Thursday the fifth.

Agent: The fifth of February?

Agent: Okay I have two options ...with Hertz ...if not they do have a lower rate with Budget and that is fifty one dollars.

Cust: Per day?

Agent: Per day um mm.

Agent: You need a visa.

Cust: I *do* need one?

Agent: Yes you do.

Studying Meaning via Clarification

- Helpful for studies of meaning: how are CRs answered?
 - And HCI: how should they be answered?
- Multiple possible reasons, including acoustics:
 - Peter: But he couldn't work out why I was in school?
 - Muhammad: <unclear>
 - Peter: What?
- Ambiguity of meaning:
 - George: you always had er er say every foot he had with a piece of spunyarn in the wire
 - Anon 1: Spunyarn?
 - George: Spunyarn, yes
 - Anon 1: What's spunyarn?
 - George: Well that's like er tarred rope

Sources of Clarification

- What kind of words do we clarify (or not)?
 - What drives (mis)communication?
- Excluding whole sentences, unclear etc:

	RNC	Communicator
NP/Pro/PN/CN	76%	78%
Adj/Adv/Mod	12%	13%
VPs	4%	5%
Verbs	1%	0%
Det (numbers)	2%	1%
Det (other)	2%	0%
Prep/Conj	<0.5%	0%

Content vs Function Words

- Function word clarification very rare
 - (relative to content word clarification)
 - Content/function ratio ≈ 11
 - (comparing only single-word sources – otherwise higher; determiners only higher)
- Seems intuitively plausible – but why?
 - Less frequent?
 - More familiar?
 - Less contentful?
 - Hard to actually do?

Content/Function: Frequency

- Effect of overall (token) frequency?
 - C/F source ratio ≈ 11
 - C/F frequency ratio ≈ 2 (corpus-dependent)
 - **No.** ($\chi^2_{(1)} p < 0.002$)
- Familiarity? **type-token ratio**
 - Number of tokens (occurrences) per word (type)
 - Average rarity = type count / token count
 - C/F source ratio ≈ 11
 - C/F TTR ratio ≈ 11
 - **Maybe!** ($\chi^2_{(1)}$ no significant differences)

Content/Function: Information

- Effect of lower information content?
- Method 1: cross-document frequency variance
 - Higher variance = more domain-dependence
 - (Francis & Kučera, 1982; Biber, 1995)
 - C/F source ratio ≈ 11
 - C/F variance ratio ≈ 0.9
 - No. (wrong direction!)

Content/Function: Information

- Method 2: language model probability estimates
 - (Shannon, 1948)
 - Processing difficulty in parsing, reading (e.g. Roark 2009, Hale 2003)
 - Surprisal $-\log_2(p)$ (= unexpectedness/unpredictability)
 - C/F ratio ≈ 1.4
 - Maybe!
 - Entropy change ΔH (= change in uncertainty)
 - Positive ΔH = increased uncertainty
 - Negative ΔH = increased information provided
 - C/F = negative/positive
 - Maybe! (but clarification \leftrightarrow information, not uncertainty?)

Content/Function: Answerability

- DiET experiment toolkit (Healey et al, 2003)



Content/Function: Answerability

- Healey et al (SIGDIAL, CogSci 2003)
- Insert fake “clarifications”:
 - Repeat words from previous turns
 - Wait for response
- Content words: 45% responded to
 - The vast majority as direct CRs (92%)
- Function words: only 15% response ($\chi^2_{(1)} p < 0.0004$)
 - And *none* of those as direct CRs with function word source

Laura: Can I have some toast please?

Jan: Some?

Laura: Toast

Content/Function: Answerability

- Healey et al (SIGDIAL, CogSci 2003)
- Insert fake “clarifications”:
 - Repeat words from previous turns
 - Wait for response
- Content words: 45% responded to
 - The vast majority as direct CRs (92%)
- Function words: only 15% response ($\chi^2_{(1)} p < 0.0004$)
 - And *none* of those as direct CRs with function word source

A: I have limited experience with balloons but... worth a try

“B”: a?

A: no, b

“B”: oh ok

B: i’m not in the baloon

CR Sources: Some Hypotheses

- The C/F split suggests some hypotheses about what might drive clarification behaviour:
 - Higher type-token ratio (rarity)
 - Higher surprisal (unpredictability)
 - Higher entropy reduction (information content)
 - Difficulty of interpretation of CRs
- Logistic regression model, by utterance:
 - Coefficients: mean f –, mean h +, mean/max ΔH –
 - (and *variance* coefficients near-zero)
 - (But: max h –, max H +)

But What About Verbs?

	BNC	Communicator
NP/Pro/PN/CN	76%	78%
Adj/Adv/Mod	12%	13%
VPs	4%	5%
Verbs	1%	0%
Det (numbers)	2%	1%
Det (other)	2%	0%
Prep/Conj	<0.5%	0%

But What About Verbs?

- CRs with V/VP sources can certainly happen:
 - A: You see this thing did you buy this separately or did it come in the Walkman?
 - B: We were lent them.
 - A: Lent them?
 - B: Yeah.
- But they are as rare as function-word sources!
 - If not more so:
 - Rodriguez & Schlangen (2004), Rieser & Moore (2005)
 - 0 examples found for *action-reference* class
 - 51% of examples were *NP* or *deictic reference*

Noun/Verb: Frequency

- Effect of overall (token) frequency?
 - N/V source ratio ≈ 15 (for CN/CV; 40 for all)
 - N/V frequency ratio $\approx 0.8 - 1.5$
 - **No.** ($\chi^2_{(1)} p < 0.0001$)
- Familiarity? Expect type-token ratio $N > V$
 - N/V source ratio ≈ 15
 - N/V TTR ratio $\approx 1.8 - 3.4$
 - **Hmm, possibly I suppose ...**
 - Expected direction, but much weaker than C/F
 - $\chi^2_{(1)} p < 0.05$ in almost all cases this time

Noun/Verb: Information

- Surprisal $-\log_2(p)$ (expect unexpectedness/unpredictability N>V)
 - N/V ratio ≈ 0.9
 - No. (wrong direction)
 - (including auxiliaries etc changes this, but weak: 1.05)
- Entropy change ΔH (= change in uncertainty)
 - Negative ΔH = increased information provided
 - N, V both negative, with N<V (N/V ≈ 1.5)
 - Hmm, possibly I suppose ...
 - Expected direction, but much weaker than C/F
 - Verbs show entropy decrease too, but less so than nouns

Noun/Verb: Answerability

- DiET with fake “clarifications”
- No significant difference in response rates:
 - Nouns: 52% responded to
 - Verbs: 41% responded to
 - No significant difference ($\chi^2_{(1)} p > 0.17$)
- But perhaps different responses:
 - Nouns: only 4% “gap”/non-CR interpretations
 - Verbs: 18% “gap”/non-CR interpretations
 - Possibly significant difference ($\chi^2_{(1)} p = 0.05$; Fisher $p > 0.085$)
- **Hmm, not really ...**
 - No expected effect
 - (although maybe there’s something going on)

CR Sources: More Hypotheses?

- With the N/V split, our hypotheses aren't very helpful:
 - Higher type-token ratio (rarity): **WEAK**
 - Higher surprisal (unpredictability): **NO**
 - Higher entropy reduction (information): **WEAK**
 - Difficulty of interpretation of CRs: **NO**
- So what's going on?

Perhaps Verbs are Not Nouns

- Differences suggest N/V categories are distinct
 - (helpful for typology? cross-linguistic studies?)
- Different semantic (cognitive?) status?
 - Conventionally both $e > t$: $\lambda x.snore(x)$ $\lambda x.woman(x)$
 - But e.g. frame semantics: SELL[buyer, seller, goods, money, ...]
 - Perhaps verbs are structured around arguments
 - ... which are mostly NPs ... and they get clarified?
- Difference in referentiality?
 - Not simple: CRs not rare for common nouns, abstract NPs ...
 - Perhaps nouns project more “parameters”?
 - Discourse referents? Presuppositions?
- Differences in acquisition – does that help?
 - N before V in some languages, opposite in others ...
 - CHILDES corpus suggests verb CRs more common in child speech!