What do cows drink?

Simon Dennis

Dennis Mehay

Ohio State University
An Associative Theory of Sentence Processing

• When presented in isolation people process target words faster if they are preceded by an associatively related prime (Meyer and Schvaneveldt, 1971)

• Associatively related target words show a reduction in N400 ERPs relative to control (Bentin, McCarthy & Wood 1985)

Association versus Sentence Context

• Reduction in the N400 (Van Petten, 1993)
  – When the *moon* is full it is hard to see many *stars* or the Milky Way. (strong)
  – When the *insurance* investigators found that he’d been drinking they *refused* to pay the claim. (weak)
  – When the *moon* is rusted it is available to buy many *stars* or the Santa Ana. (weak)
  – When the *insurance* supplies explained that he’d been complaining they *refused* to speak the keys. (control)

• However, sentential context seems to override association
Effect of Association

Camblin, Gordon & Swaab (2007)
Effect of Discourse Congruence

Camblin, Gordon & Swaab (2007)
What if sentential context is associative?

• Not simple associative:
  – Snow white black sheep wool sweater shirt pants legs run. (from Nelson, McEvoy & Schreiber, 1998)

• All word pairs have associations (some negative)

• Combine to determine probable words in a given location

• Weighted by distance to the Cloze slot
A Multinomial Logistic Regression Based Model

\[ y_k = \frac{e^{w_k x_j}}{\sum_k e^{w_k x_j}} \]

\[ x_{s_i}(j) = e^{-\lambda|i-j|} \]
Objections

• Center embedding
• Hierarchical Phrase Structure
• Propositional Units
Center Embedding

The rat sees the cheese.
The rat that the cats chase sees the cheese.
The rat that the cats that the dog loathes chase sees the cheese.
The rat that the cats that the dog that the cows toss loathes chase sees the cheese.

Reich's Conjuncture

People are capable of processing one level of center embedding, and one level only.

Reich (1969)
Karlsson 2007 Data Sources

1. The Brown and the LOB corpora
4. More than one hundred descriptive, stylistic, and diachronic papers, grammars, and style manuals in multiple languages including latin and german.
5. Corpus examples provided by other authors in the literature and from the LINGUIST mailing list.
6. Naturalistic observation of newspapers and books over the years.
Karlsson Results

• Found 132 examples of multiple center embedding
• Only three of which came from spoken language.
• "Multiple center-embeddings are extremely rare in genuine speech." (Karlsson, 2007, p373)

• Generalization: Online mechanism allows one level of embedding and only one level of embedding
One Level of Center Embedding

provided decay is less rapid than \(2^{-|i-j|}\)

Note this mechanism does not work for NS NS NP because strength of NS overrides NP causing a prediction of VS
Dependency

Sleator & Temperley (1993)

the boy.n the girl.n chases.v loves the cat.n

• How do we know the “boy” links to “loves” and not to “chases” which is closer?

• Dependency grammar parsers typically generate possible linkages and use rules (e.g. planarity) and cost functions to select between linkages

• Model assumes that heads negate the predictions of dependants and vice versa and thus avoid search (~80% weights negative)
Hierarchical Phrase Structure

• sentences are constructed from phrases

• several tests e.g. what can stand alone as the answer to a question

• Where did you go?
  – to the cinema
  – the cinema
  – to the*
Miniature Corpus

SS we saw EE
SS we went there EE
SS we went to Paris EE
SS we went to the cinema EE
we: we 92 went 4
gone: went 87 to 4 we 3 saw 2 the 2
to: to 77 the 7 there 6 went 4 saw 2 Paris 2
the: the 62 Paris 14 to 8 cinema 5 there 4 saw 2
cinema: cinema 62 Paris 14 the 5 there 5 saw 2 to 2
NNP -> {mark, simon, paul, luke, john, jim, sue}
NN -> {dog, cat, bird, snowman, lawyer, fireman}
VB -> {loves, detests, believes, sees, knows}
DT -> {a, the}
WP -> who
NP -> NNP | DT NN | DT NN WP VP
VP -> VB | VB NP
S -> NP VP
A Linguist in a Net

Linear Classifier

the who Jim detests snowman ...
Hierarchical Phrase Structure
Training on a real corpus

- New York Times newswire from Gigaword corpus
- 1,000,000 sentences
- Vocabulary 10,017
- Words outside vocabulary converted to xxx possibly with affixes (e.g. xxxs, xxxed, xxxation)
- ~80% of weights negative
- Used cross entropy to index sentence complexity
Subject versus Object
Extracted Relative Clauses

- The reporter who attacked the senator admitted the error. (1.96)
- The reporter who the senator attacked admitted the error. (2.07)
Subject versus Object

Extraction Questions

• Who did the senator say likes the reporter? (1.84)

• Who did the senator say the reporter likes? (1.76)

Juffs & Harrington (1994)
Singly versus Doubly Nested Relative Clauses

• The intern who the nurse supervised had bothered the administrator who lost the medical reports (3.06)

• The administrator who the intern who the nurse supervised had bothered lost the medical reports (3.31)

Gibson (1998)
Embedding of Clausal Modifiers

• If the mother gets upset when the baby is crying, the father will help, so the grandmother can rest easy. (2.07)

• If when the baby is crying, the mother gets upset, the father will help, so the grandmother can rest easily. (3.24)

• Because if when the baby is crying, the mother gets upset, the father will help, the grandmother can rest easily. (4.0)

Gibson (1998)
The Psychological Reality of Propositions

- Ratcliff & McKoon (1978)
  - The mausoleum that enshrined the tzar overlooked the square.
- Probe with **square** then with **mausoleum**
  - priming effect 111ms
- Probe with **square** then with **tzar**
  - priming effect 91ms.
- Subjects are faster again if the first probe is from the same proposition as second
A closer look at priming data

- Priming results may not be bi-directional

  The chauffeur jammed the clutch when he parked the truck
  
  clutch -> truck  583ms
  truck -> clutch  564ms  (equivalent to the within priming)

  The mausoleum that enshrined the tzar overlooked the square
  
  tzar -> square  614ms
  square -> tzar  553ms  (equivalent to the within priming)
Results using Cloze Priming

![Graph showing Mean Accuracy (Proportion Correct) and Mean Reaction Time (ms) for First and Second Conditions with Coord, Anaphor, and Relative conditions.](image-url)
Are propositions necessary?

syntagmatic

paradigmatic

Before

After
SS what do children drink? milk EE
SS what do children drink? water EE
SS what do cats drink? milk EE
SS what do cats drink? water EE
SS water is drunk by cows. EE
SS water is drunk by horses. EE
SS milk is produced by cows. EE
SS milk is produced by cows. EE
SS milk is produced by cows. EE
SS eggs are produced by chickens. EE
SS eggs are produced by chickens. EE
SS eggs are produced by chickens. EE
SS wool is produced by sheep. EE
SS wool is produced by sheep. EE
SS wool is produced by sheep. EE
A world without propositions

SS: SS 0.99
what: what 0.95 cats 0.02
do: do 0.89 cats 0.04 children 0.02
horses: cats 0.63 children 0.33
drink: drink 0.90 cats 0.04 children 0.02
?: ? 0.95 cats 0.02
water: water 0.72 milk 0.23 cats 0.03 children 0.02
EE: EE 0.99

Correctly answers question about horses despite the fact that has never been asked about horses before.
Captures a kind of transformational regularity
Inference by Coincidence

- Tennis news from ATP website
- 377 questions
- Overall 67% correct
- Who won the match between Sampras and Agassi? Sampras
- Sampras wins 5\textsuperscript{th} US Open
What do cows drink?

SS: SS 0.99
what: what 0.90 milk 0.05 cats 0.02
do: do 0.76 milk 0.14 cats 0.04 children 0.02
cows: cats 0.63 children 0.33
drink: drink 0.80 milk 0.09 cats 0.04 children 0.03
?: ? 0.93 milk 0.02 cats 0.02
water: milk 0.79 water 0.20
EE: EE 0.99

Paradigmatic Weights
milk: milk 3.02 water 1.37 wool 1.33 cats -0.44 children -0.44 cows -0.42
water: water 3.93 milk 1.27 cats -0.48 children -0.48 do -0.41 what -0.40

Syntagmatic Weights
cows: milk 2.75 wool -1.85 cows -1.79 drunk 0.99 is 0.97 sheep -0.86
Discourse Processing

• Model can be applied directly to discourse level phenomena with linear computational cost

• Provides a method for generating context representations that explains discontinuities (c.f. drift)

• Exponential decay ensures graceful shift from local order dependent phenomena at short time scales to more global order information at longer time scales

• Will look like Latent Semantic Analysis (LSA) at paragraph level

• Explains how function words are excluded from content considerations
Topics Corpus

theatre -> play | stage | audience | shakespeare | actor
football -> coach | football | team | score | game
printing -> press | image | graphic | type | paper
science -> hypothesis | experiment | scientist | data | explanation
law -> judge | trial | jury | guilty | lawyer
school -> study | test | class | teacher | math

Five words sampled from each topic with replacement with an ss marker separating topics
Summed Topics Probability
Pronoun Resolution

• Greene, McKoon & Ratcliff (1992) proposed that a pronoun be seen as a cue which is used to retrieve referents in the current discourse context
• Model instantiates this theory
• Explains how pronoun could be present to be associated to referent in the first instance
Pronoun Resolution

mary: she-0.27 he-0.07 num-0.05 mother-0.04 xxx-0.04 one-0.03
believes: asked-0.28 lou-0.08 called-0.08 not-0.06 call-0.05
john: xxx-0.04 nodded-0.02 xxs-0.02 it-0.02 alone-0.01 asked-0.01

• When “she” recurs Mary will be automatically reinstated by paradigmatic association
• Because John is bound to whole paradigmatic pattern resolution sensitive to factors such as number, focus, the reinstatement of topic, etc.
• Note “they” is not part of paradigmatic pattern
• Note does rely exclusively on pronouns (see “mother”)
Conclusion

• Psycholinguistic evidence suggests association plays a role in sentence processing

• An associative model of sentence processing can account for:
  – center embedding phenomena exhibited by people
  – hierarchical phrase structure representations
  – propositional information

• The model is unsupervised and computationally tractable

• Prima facie case for use in discourse processing and pronoun resolution
Sentence processing involves the resolution of syntagmatic and paradigmatic constraints:

- Syntagmatic associations occur between words that appear together (e.g. run fast)
- Paradigmatic associations occur between words that fill the same slot. (e.g. deep shallow)

your mother is a hampster and your father smells of elderberries

father  gangster  sister  eggs
sister  kiwi  brother  cologne
lover  sheep  cousin  fish
Inference by Coincidence

• Who won the match between Carlsen and Keifer? Carlsen
  – Kafelnikov now meets Kenneth Carlsen of Denmark in the second round.

• Who won the match between Keifer and Safin? Safin
  – Safin and Kafelikov surge towards hometown showdown

• Who won the match between Srichaphan and Lapentti? Srichaphan
  – Srichaphan has now won two titles in four finals this year.
How to estimate word distribution?

- Stack based probabilistic parser?
- Simple recurrent network?
- String edit model?
- Maximum Entropy model?
- Multinomial logistic regression model?