

Parsing, fast and slow

A two-system approach to some
topics in linguistics

Jamie Y. Findlay

University of Oslo

ILN departmental linguistics seminar,

6 Oct. 2023

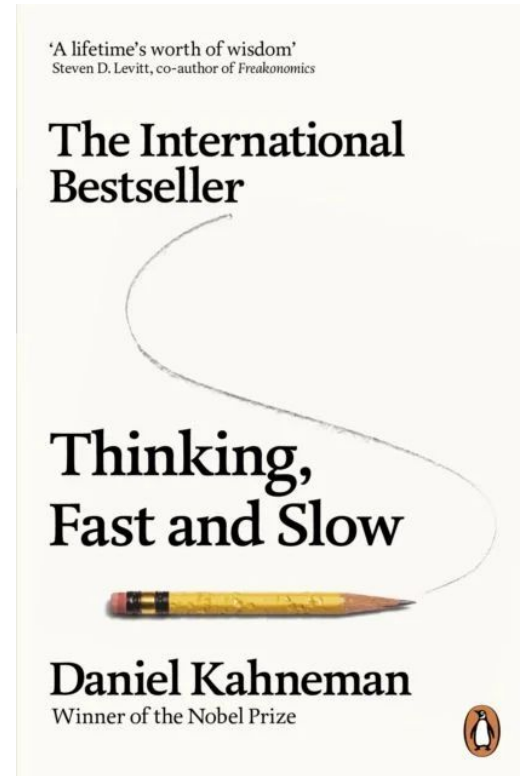


The dual process hypothesis

(Wason & Evans 1974)

System 1	System 2
<ul style="list-style-type: none">• Answer $2 + 2 = ?$• Detect that one object is more distant than another• Orient to the source of a sudden sound• Complete the phrase “bread and ...”• Drive a car on an empty road• Find a strong move in chess (if you are a chess master)	<ul style="list-style-type: none">• Answer $34 \times 79 = ?$• Look for a woman with white hair• Search memory to identify a surprising sound• Compare two washing machines for overall value• Fill in an official form• Check the validity of a complex logical argument

(Stanovich & West 2000: 658)



(Kahneman 2011)

The Linda problem

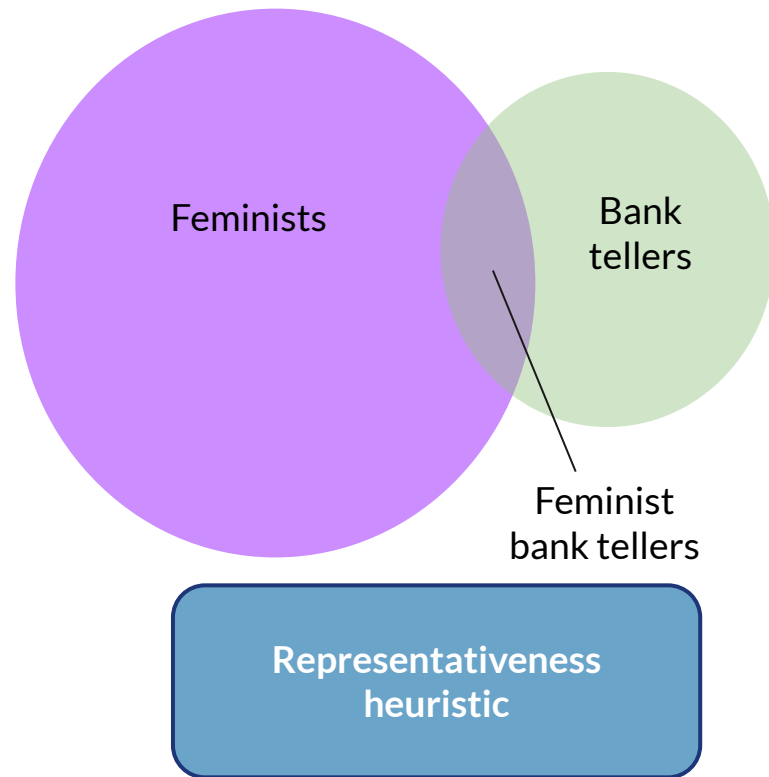
(Tversky & Kahneman 1982)

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Which is more likely?

- A. Linda is a bank teller. 10–15%
- B. Linda is a bank teller and active in the feminist movement. 85–90%

(For an interesting linguistic perspective on this kind of problem, see Asudeh & Giorgolo 2020: ch. 6.)



Letter estimates

(Tversky & Kahneman 1973)

If a random word is taken from an English text, is it more likely that the word starts with a K, or that K is the third letter?

Availability bias

An aside

T&K investigated 5 letters reported to have this property by Mayzner & Tresselt (1965):

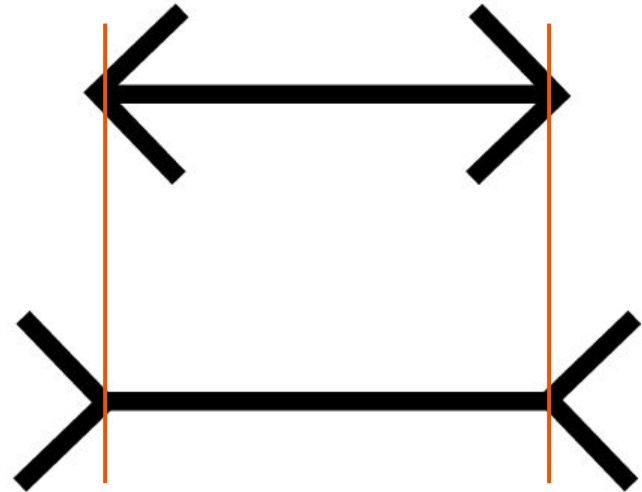
k, l, n, r, v

	DS9 (47,377 words)			Shakespeare (958,363 words)		
	1st	3rd		1st	3rd	
<i>k</i>	1.5%	1.1%	0.73x	1.1%	0.8%	0.73x
<i>l</i>	2.0%	2.4%	1.20x	3.3%	4.1%	1.24x
<i>n</i>	2.3%	4.2%	1.83x	3.0%	4.6%	1.53x
<i>r</i>	2.4%	6.4%	2.67x	1.6%	7.9%	4.94x
<i>v</i>	0.5%	2.0%	4.00x	0.7%	1.9%	2.71x

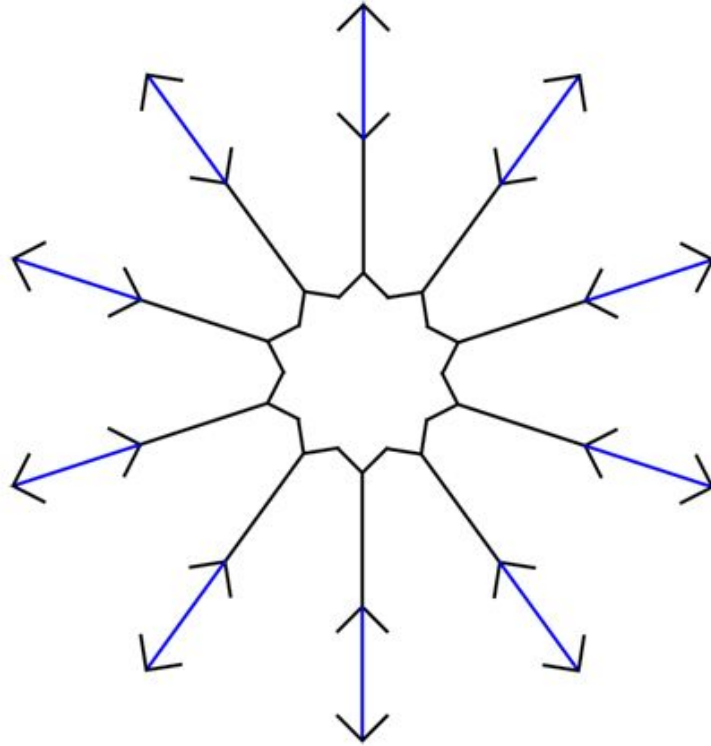
Heuristics and biases

- People systematically make the ‘wrong’ choices/judgements, because they make use of shortcuts.
- Evolutionary motivation: S2 is expensive!
 - We are “cognitive misers”.
(Fiske & Shelley 1984: 13)
- S2 can ‘take the reins’ and suppress S1.
 - But this capacity varies!

The Müller-Lyer illusion



Sarcone's Dynamic Müller-Lyer Illusion



This talk

Applying this contrast to various linguistic phenomena.

1. Psycholinguistics

2. Large Language Models (LLMs)

3. Literary language

4. Socially-motivated language change

Psycholinguistics

Language processing as System 1

The Stroop effect
(Stroop 1935)

Purple Red
Brown
Red Green
Blue

Linguistic illusions

The Moses illusion

How many animals of each kind did Moses take on the Ark?

(Erickson & Mattson 1981)



Grammaticality illusions

*** More Americans have been to Russia than I have.**

(Montalbetti 1984)



Negative illusions

No head injury is too trivial to be ignored.

→ All head injuries should be ignored, no matter how trivial.

No dictatorship is too benevolent to be condemned.

→ All dictatorships should be condemned, no matter how benevolent.

(Wason & Reich 1979)

'Good Enough' parsing

(Ferreira et al. 2002)

Misinterpretation of garden-path sentences

While Anna dressed the baby played in the crib.

1. Did the baby play in the crib?
2. Did Anna dress the baby?



(Christianson et al. 2001)

Misinterpretation of passive sentences

1. The man bit the dog.
2. The man was bitten by the dog.
3. The dog bit the man.
4. The dog was bitten by the man.

(Ferreira 2003)

1. NVN heuristic
2. Plausibility heuristic

'pseudo-parse' 'quick and dirty' 'fast and frugal'
Late Assignment of Syntax Theory (LAST)

(Bever 1970; Fodor et al. 1974; Garrett 2000; Townsend & Bever 2001; *i.a.*)

What about System 2?

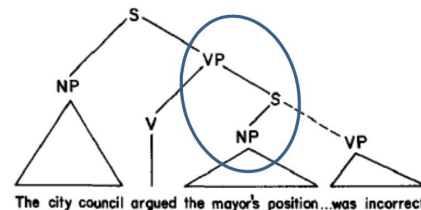
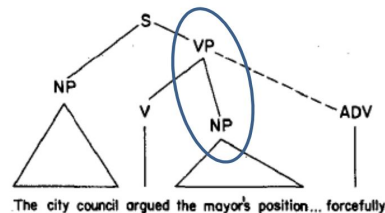
- Visual illusions like Müller-Lyer can be demonstrated by e.g. measurement.
- What is the equivalent for linguistic illusions?

Parser = S1?
Grammar = S2?

Cf. *Garden path* theory of processing.

**Minimal attachment/
Minimal everything**

(Frazier 1987; Fodor & Inoue 1998)



The farmer the nurse the lawyer represented saw died.

Apparent exceptions

“We find many situations where on-line language processing is highly sensitive to **detailed grammatical constraints** and immune to interference from inappropriate material. These findings make it difficult to maintain that speakers respond to the pressures of on-line sentence comprehension by resorting to superficial or ‘good enough’ representations.”

(Phillips et al. 2011)

Real-time Principle C

1. [[**His/her** managers] chatted amiably with some fans while the talented, young **quarterback** signed autographs for the kids], but **Carol** ...

Slowdown at quarterback for gender mismatch
– **binding syntactically licensed.**

2. [**He/she** chatted amiably with some fans while the talented, young **quarterback** signed autographs for the kids], but **Steve/Carol** ...

No slowdown at quarterback for gender mismatch
– **binding syntactically unlicensed.**

(Kazanina et al. 2007)

Real-time *wh*-islands

1. We like the **book/city** [that the author wrote __ unceasingly and with great dedication about __] while waiting for a contract.

Slowdown at *wrote* for implausible object (*city*)
– **gap posited.**

2. We like the **book/city** [that the author [who wrote unceasingly and with great dedication] saw __] while waiting for a contract.

No slowdown at *wrote* for implausible object
– **no gap posited.**

(Traxler & Pickering 1996)

S1 and S2 in interaction?

- Is it wrong to describe language processing as purely S1/heuristic-based, therefore?
- In other words, are we more impressed with Phillips et al.'s **“detailed grammatical constraints”** than we should be ...?

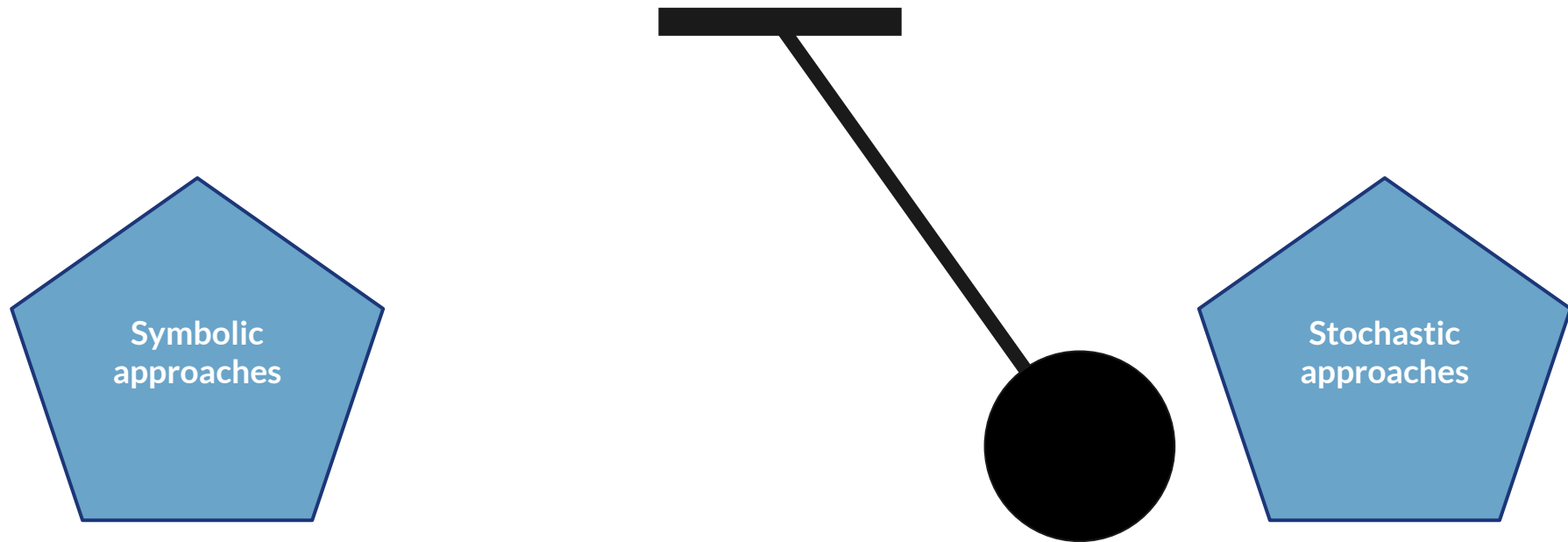
“In the past there were plenty of tasks – including writing essays – that we’ve assumed were somehow ‘fundamentally too hard’ for computers. And now that we see them done by the likes of ChatGPT we tend to suddenly think that computers must have become vastly more powerful – in particular surpassing things they were already basically able to do [...].

But this isn't the right conclusion to draw. [...] **[I]nstead what we should conclude is that tasks – like writing essays – that we humans could do, but we didn't think computers could do, are actually in some sense computationally easier than we thought.”**

(Stephen Wolfram, [www](#))

Large Language Models

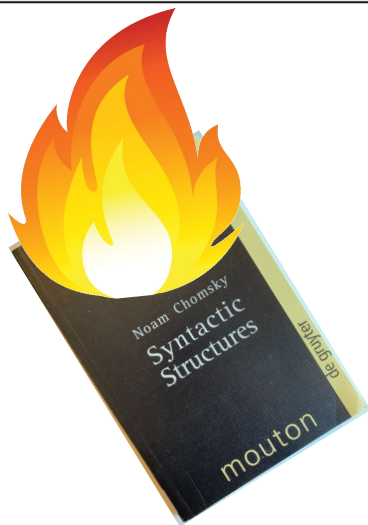
The AI pendulum



Responses to LLMs in linguistics

The end of the generative enterprise?

(“Modern language models refute Chomsky’s approach to language” – Piantadosi 2023)



‘Stochastic parrots’ 🦜?

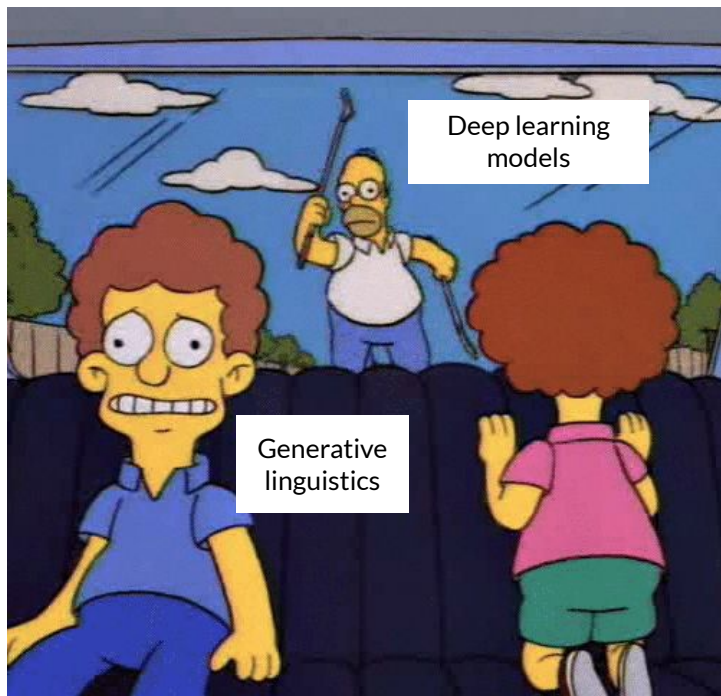
(“the language modeling task, because it only uses form as training data, cannot in principle lead to learning of meaning” – Bender & Koller 2020; Bender et al. 2021)



(Searle 1980)

(See also Katzir 2023; Kodner et al. 2023; Milway 2023; and Murphy 2023 for direct responses to Piantadosi 2023.)

A god of the (linguistic) gaps

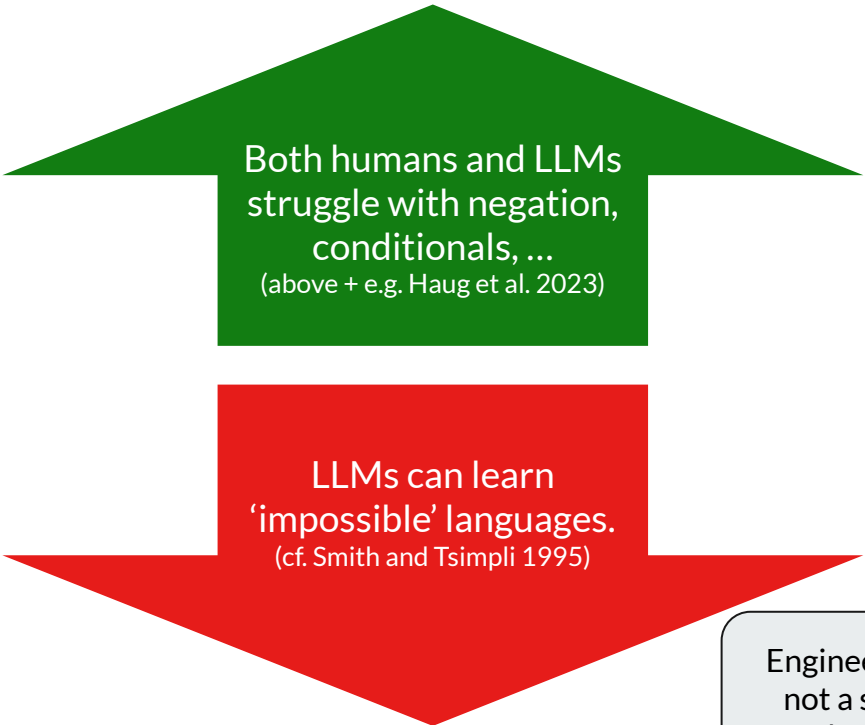


“There’s two ways in which a system can be deficient. One way is it’s not strong enough – [it] fails to do certain things. The other way is it’s too strong – it does what it shouldn’t do. [... The former] can in principle at least be fixed – you add another trillion parameters, double the number of terabytes and maybe do better. When a system is too strong it’s unfixable typically and that’s the problem with GPT and the other systems.”

(Chomsky, reported in Milway 2023)

A different AI alignment problem

1. Can LLMs accurately capture (all) linguistic generalisations?
2. Are they good models of the way *human minds* capture those generalisations?



Both humans and LLMs struggle with negation, conditionals, ...
(above + e.g. Haug et al. 2023)

LLMs can learn 'impossible' languages.
(cf. Smith and Tsimpli 1995)

Engineering tool,
not a scientific
theory ...?

If all you have is a hammer, ...



A two-system story

LLMs model S1	(Generative) Grammars model S2
<ul style="list-style-type: none">• Heuristic-based, 'good enough'.• Domain-general procedures	<ul style="list-style-type: none">• Detailed, explicit.• Domain-specific knowledge and rules

Competence vs. performance/System 2 vs. System 1

The mouse that the cat that the dog painted taught sang.

1. What did the mouse do?
2. What did the cat do?

RB

In the sentence "The mouse that the cat that the dog painted taught sang", what did the mouse do? And what did the cat do?



The sentence "The mouse that the cat that the dog painted taught sang" is a grammatically correct sentence, but it is semantically nonsensical because it contains a syntactical ambiguity that makes it difficult to determine its meaning.



The sentence can be interpreted in different ways depending on how one parses it. One possible interpretation is that the mouse taught someone or something to sing, and the cat painted a dog. Another possible interpretation is that the mouse sang a song that was taught to it by a cat that was painted by a dog.

Therefore, it is impossible to determine with certainty what the mouse or the cat did in this sentence without additional context or clarification.

RB

Try to think a bit longer and answer the questions again



I apologize, but even after careful consideration, I must reiterate that the sentence "The mouse that the cat that the dog painted taught sang" is still semantically nonsensical due to the syntactical ambiguity it presents. It is not possible to determine with certainty what the mouse or the cat did in this sentence without further clarification or context.



"But differently from humans, further time and resources are of no use to ChatGPT. When it misparses a sentence, this is because its knowledge is flawed, not because processing problems got in the way."

(Katzir 2023)

So is this just the competence-performance distinction?



Competence \subseteq Performance(?)

“No doubt, a reasonable model of language use will incorporate, as a basic component, the generative grammar that expresses the speaker-hearer’s knowledge of the language [...]”

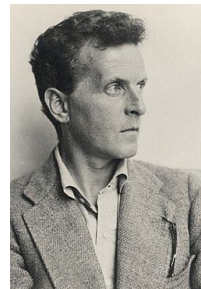
(Chomsky 1965: 9)

BUT

System 1 and System 2 are disjoint

The limits of my language (model) ...

LLMs and generative grammar need not be competing theories of the same thing; they simply describe different cognitive systems.



“Die Grenzen meiner
Sprache bedeuten die
Grenzen meiner Welt.”
(Wittgenstein 1922)

Literary language

Defamiliarization

Russian formalism's view of literature



“organized violence committed
on ordinary speech”
(Roman Jakobson)

“By ‘enstranging’ objects and complicating form, the device of art makes perception **long** and ‘**laborious**’. The perceptual process in art has a purpose all its own and ought to be extended to the fullest.”

(Shklovsky 1990 [1929]: 6)



S2 ✓

“[T]he very hallmark of the artistic [... is ...] an artifact that has been intentionally removed from the domain of **automatized perception**. It is ‘artificially’ created by an artist in such a way that the perceiver, pausing in his reading, dwells on the text. This is when the literary work attains its greatest and most long-lasting impact.”

(Shklovsky 1990 [1929]: 12)



S1 ✗

Defamiliarizing idioms

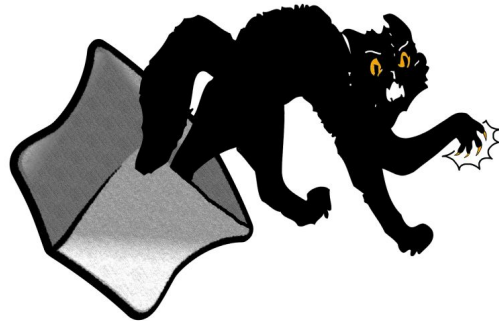
Inferred idioms (Pulman 1993)

1. Good gawd it's another porcine flyer. ([www](#))
2. Shit finally completes 29-month journey towards fan. ([www](#))



Extended idioms (Egan 2008; Bargmann et al. in prep)

1. The cat is out of the bag, claws fully exposed. (iWeb)
2. This month, Meriden City Council's PLA chickens came home to roost, and they laid a big stinky egg on Meriden taxpayers. ([www](#))
3. Sometimes [...] the person that you'd take a bullet for is behind the trigger. (Fall Out Boy, *Miss Missing You*)



(Findlay 2018, 2019: 45ff, 92ff.; Findlay et al. 2019; Bargmann et al. in prep)

Defamiliarizing idioms

Conjunction modification (Ernst 1981)

1. With the recession, oil companies are having to tighten their Gucci belts.
2. Malvolio deserves almost everything he gets, but ... there is that little stab of shame we feel at the end for having had such fun pulling his cross-gartered leg for so long.
3. Bruce, a shark, found it a part he could really sink his three rows of teeth into.



(www)

Culture jamming



Socially-motivated language change

Socially-motivated language change

- *Negro* → *colored* → *black* (→ *African-American*)
- *homosexual* → *gay*
- *cripple* → *handicapped* → *disabled*
- *policeman* → *police officer*
- *chairman* → *chair(person)*
- *stewardess* → *flight attendant*



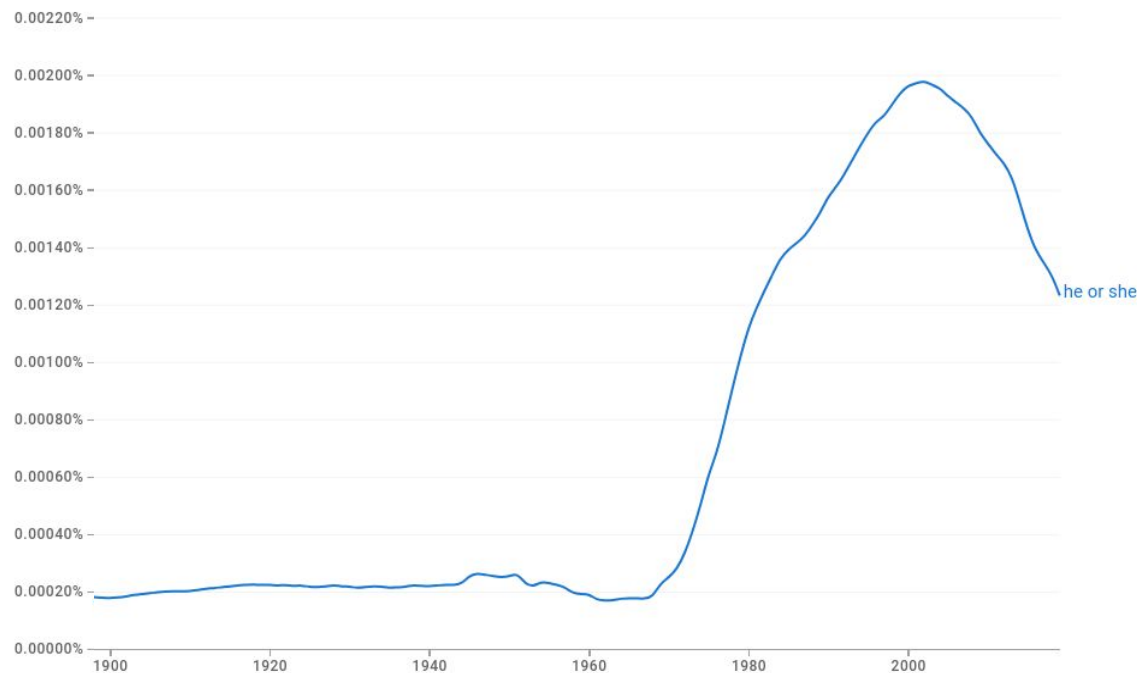
Inclusive Language

- $he_{[GENERIC]} \rightarrow he \text{ or } she / they_{[3SG, GENERIC]}$
- $they_{[3SG, GENERIC]} \rightarrow they_{[3SG, SPECIFIC]}$

“no Person shall be a Representative ...
who shall not, when elected, be an
inhabitant of that state in which **he** shall
be chosen”
(US Constitution, art. I sec. 2)

1. Everyone_i should do their_i best.
2. %Chris_i did their_i best.

Replacing generic *he*



“words importing the masculine gender shall include females”

(UK Interpretation Act 1889, sec. 1)

“(a) words importing the masculine gender include the feminine;
(b) words importing the feminine gender include the masculine”

(UK Interpretation Act 1978, sec. 6)

But I can't control my System 1 thinking!

"It is realistic to hope to change only those linguistic uses of which speakers themselves can be made aware, as they use them. One chooses, in speaking or writing, more or less consciously and purposefully among nouns, adjectives and verbs; one does not choose among pronouns in the same way."

(Lakoff 2004 [1975]: 71)

"I don't want to offend anyone. But it's a bit much to expect me to start saying things that are clearly and decisively ungrammatical according to my own internalized grammar. I'll do my best, but it will be a real struggle."

(Geoffrey K. Pullum [on Language Log](#), Dec. 4 2017)

Literature on prejudice and bias shows that S2 *can* overcome S1.

(Fiske 1989; Fiske & Neuberg 1990; Strack 1992; Thompson et al. 1994; Richeson et al. 2003; Burgess et al. 2007; but see FitzGerald et al. 2019 for a more skeptical view)

"I don't want to offend anyone. But it's a bit much to expect me to **stop treating women poorly when my unconscious bias clearly and decisively tells me they are the lesser sex.** I'll do my best, but it will be a real struggle."



Conclusions

What's the point of all this?



But also ...

it can be illuminating to draw connections between disparate fields of inquiry.

- Dissolves false tensions
- Offers new perspectives
- Brings in original insights

Questions?

References

- Asudeh, Ash & Gianluca Giorgolo. 2020. *Enriched meanings*. Oxford University Press.
- Bargmann, Sascha, Jamie Y. Findlay, & Manfred Sailer. In preparation. Defrosting frozen metaphors: on the extensibility and manipulability of idioms.
- Bender, Emily M. & Alexander Koller. 2020. Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198. Association for Computational Linguistics.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, & Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? 🦜. In *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. Association for Computing Machinery.
- Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In John R. Hayes (ed.), *Cognition and the development of language*, 279–362. Wiley.
- Burgess, Diana, Michelle van Ryn, John Dovidio, & Somnath Saha. 2007. Reducing racial bias among health care providers: Lessons from social-cognitive psychology. *Journal of General Internal Medicine* 22(6). 882–887.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. MIT Press.
- Christianson, Kiel, Andrew Hollingworth, John F. Halliwell, & Fernanda Ferreira. 2001. Thematic roles assigned along the garden path linger. *Cognitive Psychology* 42. 368–407.
- Egan, Andy. 2008. Pretense for the complete idiom. *Noûs* 42(3). 381–409.
- Erickson, Thomas D. & Mark E. Mattson. 1981. From words to meaning: a semantic illusion. *Journal of Verbal Learning and Verbal Behavior* 20(5). 540–551.
- Ernst, Thomas. 1981. Grist for the linguistic mill: idioms and ‘extra’ adjectives. *Journal of Linguistic Research* 1(3). 51–68.
- Ferreira, Fernanda, Karl G. D. Bailey, & Vittoria Ferraro. 2002. Good-enough representations in language comprehension. *Current Directions in Psychological Science* 11(1). 11–15
- Ferreira, Fernanda. 2003. The misinterpretation of noncanonical sentences. *Cognitive Psychology* 47(2). 164–203.

References

- Findlay, Jamie Y. 2018. When the cat you let out of the bag has claws: The role of metaphor in understanding idioms. Talk given at the Linguistics and English Language Research Seminar, University of Manchester. [\[slides\]](#)
- Findlay, Jamie Y. 2019. *Multiword expressions and the lexicon*. D.Phil. thesis, University of Oxford.
- Findlay, Jamie Y., Sascha Bargmann, & Manfred Sailer. 2019. Why the butterflies in your stomach can have big wings: combining formal and cognitive theories to explain productive extensions of idioms. Talk given at the International Conference of the Europäische Gesellschaft für Phraseologie (EUOPHRAS 2019): Productive Patterns in Phraseology, Santiago de Compostela, Spain. [\[slides\]](#)
- Fiske, Susan T. 1989. Examining the role of intent: Towards understanding its role in stereotyping and prejudice. In J. S. Uleman & J. A. Bargh (eds.), *Unintended thought*. Guilford.
- Fiske, Susan T. & Steven L. Neuberg. 1990. A continuum of impression formation, from category-based to individuating processes: influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology* 23. 1–74.
- Fiske, Susan T. & Taylor, Shelley E. 1984. *Social cognition*. McGraw-Hill.
- FitzGerald, Chloë, Angela Martin, Delphine Berner, & Samia Hurst. 2019. Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: A systematic review. *BMC Psychology* 7(29).
- Fodor, Janet D. & Atsu Inoue. 1998. Attach anyway. In Janet D. Fodor and Fernanda Ferreira (eds.), *Reanalysis in sentence processing*, 101–141. Kluwer.
- Frazier, Lyn. 1987. Sentence processing: a tutorial review. In Max Coltheart (ed.), *Attention and performance XII: the psychology of reading*, 559–586. Lawrence Erlbaum Associates.
- Garrett, Merrill. 2000. Remarks on the architecture of language processing systems. In Yosef Grodzinsky, Lewis P. Shapiro, & David Swinney (eds.), *Language and the brain: Representation and processing*, 31–69. Foundations of Neuropsychology. Academic Press.
- Haug, Dag T. T., Jamie Y. Findlay, & Ahmet Yıldırım. 2023. The long and the short of it: DRASTIC, a semantically annotated dataset containing sentences of more natural length. In *Proceedings of the 4th International Workshop on Designing Meaning Representations*, 89–98. Association for Computational Linguistics.

References

- Kahneman, Daniel. 2011. *Thinking, fast and slow*. Penguin.
- Katzir, Roni. 2023. Why large language models are poor theories of human linguistic cognition. A reply to Piantadosi (2023). Unpublished manuscript, [lingbuzz/007190](#).
- Kazanina, Nina, Ellen F. Lau, Moti Lieberman, Masaya Yoshida & Colin Phillips. 2007. The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language* 56(3). 384–409.
- Kodner, Jordan, Sarah Payne, & Jeffrey Heinz. 2023. Why linguistics will thrive in the 21st century: A reply to Piantadosi (2023). Unpublished manuscript, [lingbuzz/007190](#).
- Lakoff, Robin. 2004 [1975]. *Language and woman's place: text and commentaries* (revised and expanded edition – edited by Mary Bucholtz). Oxford University Press.
- Mayzner M. S. & M. E. Tresselt. 1965. Tables of single-letter and bigram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements* 1(2). 13–32.
- Milway, Daniel. 2023. A response to Piantadosi (2023). Unpublished manuscript, [lingbuzz/007264](#).
- Montalbetti, Mario. 1984. *After binding: on the interpretation of pronouns*. MIT dissertation.
- Murphy, Elliot. 2023. Notes on large language models and linguistic theory. Blog post.
- Phillips, Colin, Matthew W. Wagers, & Ellen F. Lau. 2011. Grammatical illusions and selective fallibility in real-time language comprehension. In Jeffrey T. Runner (ed.), *Experiments at the interfaces*, 147–180. *Syntax and Semantics* 37. Brill.
- Piantadosi, Steven. 2023. Modern language models refute Chomsky's approach to language. Unpublished manuscript, [lingbuzz/007180](#).
- Pulman, Stephen G. 1993. The recognition and interpretation of idioms. In Cristina Cacciari & Patrizia Tabossi (eds.), *Idioms: processing, structure, and interpretation*, 249–270. Lawrence Erlbaum.

References

- Richeson J. A., Baird A. A., Gordon H. L., Heatherton T. F., Wyland C. L., Trawalter S., & Shelton J. N. 2003. An fMRI investigation of the impact of interracial contact on executive function. *Nature Neuroscience* 6. 1323–1328.
- Searle, John. 1980. Minds, brains and programs. *Behavioral and Brain Sciences* 3. 417–57.
- Shklovsky, Viktor. 1990 [1929]. *Theory of prose* (2nd edn.). Translated by Benjamin Sher. Dalkey Archive Press.
- Smith, Neilson Voyne & Ianthi-Maria Tsimpli. 1995. *The mind of a savant: Language learning and modularity*. Blackwell Publishing.
- Stanovich, Keith E. & Richard F. West. 2000. Individual differences in reasoning: implications for the rationality debate. *Behavioral and Brain Sciences* 23. 645–726.
- Strack, Fritz. 1992. The different routes to social judgments: Experiential versus informational strategies. In L. L. Martin & A. Tesser (eds.), *The construction of social judgments*, 249–275. Lawrence Erlbaum Associates, Inc.
- Strack, Fritz & Roland Deutsch. 2004. Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review* 8(3). 220–247.
- Stroop, John Ridley. 1935. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* 18(6). 643–662.
- Thompson, Erik P., Robert J. Roman, Gordon B. Moskowitz, Shelly Chaiken, & John A. Bargh. 1994. Accuracy motivation attenuates covert priming: The systematic reprocessing of social information. *Journal of Personality and Social Psychology* 66(3). 474–489.
- Traxler, Matthew J. & Martin J. Pickering. 1996. Plausibility and the processing of unbounded dependencies: an eye-tracking study. *Journal of Memory and Language* 35(3). 454–475.
- Tversky, Amos & Daniel Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5(2). 207–232.
- Tversky, Amos & Daniel Kahneman. 1982. Judgments of and by representativeness. In Daniel Kahneman, Paul Slovic & Amos Tversky (eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.
- Wason, Peter C. 1968. Reasoning about a rule. *Quarterly Journal of Experimental Psychology* 20(3). 273–281.

References

- Wason, Peter C. & Jonathan St. B. T. Evans. 1974. Dual processes in reasoning? *Cognition* 3(2). 141–154.
- Wason, Peter C. & Shuli S. Reich. 1979. A verbal illusion. *Quarterly Journal of Experimental Psychology* 31(4). 591–597.
- Wittgenstein, Ludwig. 1922. *Logisch-Philosophische Abhandlung*.