COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 40 (2016) 1382–1411

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12274

Uncertainty and Expectation in Sentence Processing: Evidence From Subcategorization Distributions

Tal Linzen,[a] T. Florian Jaeger[b]

[a]Department of Linguistics, New York University
[b]Department of Brain and Cognitive Sciences, Department of Computer Science, and Department of Linguistics, University of Rochester

Received 9 October 2014; received in revised form 28 April 2015; accepted 4 May 2015

a significant predictor of processing difficulty: Greater reduction in uncertainty was correlated with increased reading times (RTs). We additionally replicated previously observed effects of expectation violation (surprisal), orthogonal to the effect of uncertainty. This suggests that both surprisal and

Uncertainty About the Rest of the Sentence

John Hale

Department of Linguistics and Germanic, Slavic, Asian and African Languages, Michigan State

Received 26 March 2005; received in revised form 24 November 2005; accepted 5 Decembe

Abstract

A word-by-word human sentence processing complexity metric is presented. This metri the intuition that comprehenders have more trouble on words contributing larger amounts of about the syntactic structure of the sentence as a whole. The formalization is in terms of the entropy of grammatical continuations, given the words that have been heard so far.

To calculate the predictions of this metric, Wilson and Carroll's (1954) original entro idea is extended to infinite languages. This is demonstrated with a mildly context-sensitive l includes relative clauses formed on a variety of grammatical relations across the Accessibili of Keenan and Comrie (1977).

Predictions are derived that correlate significantly with repetition accuracy results obtai tence-memory experiment (Keenan & Hawkins, 1987).

derstand. In the cognitive science of language, such metrics have long held out hope available experimental measures to computational theories. Miller and Chomsky instance, expressed this hope in writing: "It might be useful, therefore, to develop

Alison Kim
Hanna Hubarava

Computational Psycholinguistics SS23

Uncertainty About the Rest of the Sentence

John Hale

Dep[illegible] Linguistics and Germanic, Slavic, Asian and African Languages, Michigan State University

Received 26 March 2005; received in revised form 24 November 2005; accepted 5 December 2005

Abstract

A word-by-word human sentence processing complexity metric is presented. This metric formalizes the intuition that comprehenders have more trouble on words contributing larger amounts of information about the syntactic structure of the sentence as a whole. The formalization is in terms of the conditional entropy of grammatical continuations, given the words that have been heard so far.

To calculate the predictions of this metric, Wilson and Carroll's (1954) original entropy reduction idea is extended to infinite languages. This is demonstrated with a mildly context-sensitive language that includes relative clauses formed on a variety of grammatical relations across the Accessibility Hierarchy of Keenan and Comrie (1977).

Predictions are derived that correlate significantly with repetition accuracy results obtained in a sentence-memory experiment (Keenan & Hawkins, 1987).

Keywords: Linguistics; Computer science; Psychology; Syntax; Language understanding; Information; Mathematical modeling; Computer simulation; Relative clauses; Probabilistic grammars; Entropy reduction; Minimalist grammars; Accessibility hierarchy

Intuition: Humans have more difficulty processing words that contribute more *information* about the syntactic structure of a sentence as a whole.

Previous work: Formalization via *entropy reduction (ER)* on the word level for finite-state grammars (Wilson & Carroll, 1954)

Question: Can this intuition be formalized via *ER*

- On the sentence level?
- For grammars other than context-free ones?

Answer: Yes.

Contribution: ERH holds beyond PCFGs and beyond word level

- Demonstrated with a mildly context-sensitive (MCS) language constructed with relative clauses (RC)
- RCs built on grammar rules across *Accessibility Hierarchy* (Keenan & Hawkins, 1987)

See Jupyter notebook

`04-2_surprisal_entropy_LM.ipynb`

Introduction to information theory

Questions 4-2: 1.1-3, 5-6

Entropy in PCFGs

Question 4-2: 1.4

Probabilistic context-free grammar (PCFG): production rules

- govern a *finite* number of nonterminals and terminals
- are each assigned a probability

0.5	S	$\rightarrow$	john loves Mary
0.5	S	$\rightarrow$	john sleeps

Fig. 1. A very simple PCFG.

0.5	S	$\rightarrow$	john thinks CP
0.5	S	$\rightarrow$	john sleeps
1.0	CP	$\rightarrow$	that S

Fig. 2. Recursive PCFG.

word #	1	2	3	4	5	6	7	8	...
	john	sleeps							
	john	thinks	that	john	sleeps				
	john	thinks	that	john	thinks	that	john	sleeps	
					⋮				

Fig. 3. Sentences generated by recursive PCFG.

4. Why is it a bit tricky to apply the original formula for entropy to language (i.e., to the rules of a PCFG)?

$$H(X) = -\sum_{x \in X} P(x) \log_2 P(x)$$

PCFG rules are applied **recursively** until we are left only with terminals.

With simple PCFG (Fig. 1), given nonterminal *S*, the possible derivations are enumerable. With PCFG in Fig. 2, the possible outcomes of nonterminals are more difficult to enumerate.

Entropy in PCFGs

(cont.)

See: *Hale (2006)* pp. 646-647

Solution to recursion challenge: *Grenander's theorem (1967)*

We are given a PCFG *G*. Let $\Pi(\xi_i)$ be the subset of *G* that governs nonterminal ξ_i. For rule *r* and its probability *P(r)*, the entropy of nonterminal ξ_i is given by

$$H(\xi_i) = -\boxed{\sum_{r \in \Pi(\xi_i)} P(r) \log_2 P(r)} + \boxed{\sum_{r \in \Pi(\xi_i)} P(r) \left[H(\xi_{j_1}) + H(\xi_{j_2}) + \ldots \right]}$$

(Probability of rule *r*) times (log probability of rule *r*), summed over all rules *r* governing nonterminal ξ_i

(Probability of rule *r*) times (sum of entropies of child nonterminals ξ_{j1}, ξ_{j2}, … produced by parent nonterminal ξ_i), summed over all rules *r* governing nonterminal ξ_i

Entropy reduction hypothesis

Entropy reduction: amount of disambiguation work performed at word w_i

$$ER(w_i) = max(0, H_{i-1} - H_i)$$

- w_i performs disambiguation relative to $w_{i-1} \rightarrow ER(w_i) > 0$
- w_i introduces uncertainty relative to $w_{i-1} \rightarrow ER(w_i) = 0$

Entropy reduction hypothesis: ER positively related to processing difficulty

- Sentence-level processing difficulty = ∑ difficulties of each word (Hale, 2003b)

SUBJECT ⊃ DIR. OBJECT ⊃ INDIR. OBJECT ⊃ OBLIQUE ⊃ GENITIVE ⊃ OCOMP

Fig. 7. The Accessibility Hierarchy of relativizable grammatical relations.

Accessibility hierarchy (AH)

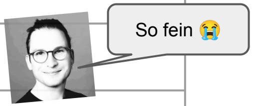
See: *Keenan & Comrie (1977),* Wiki page “Relative clause” (2023)

Accessibility hierarchy generalizes RC formation

- Every language occupies some position on the hierarchy
- Relativization can occur *at and to the left of* the occupied position

Example 🥑: DO-relativizing language

Allowed	**Forbidden**
DO: The guacamole *[that Hanna and Alison made yesterday]*	**IO:** The guacamole *~~[that Hanna and Alison gave to Patrick]~~*
Sub.: The guacamole *[that tastes divine]*	**Obl.:** The guacamole *~~[for which Patrick gave Hanna and Alison a good grade]~~*
	Gen.: The guacamole *~~[whose taste brought Patrick to tears]~~*
	Ob. comp.: The guacamole *~~[that tasted better than Martin Volk's]~~*



AH and language comprehension

See: *Keenan & Hawkins (1987)*

SUBJECT ⊃ DIR. OBJECT ⊃ INDIR. OBJECT ⊃ OBLIQUE ⊃ GENITIVE ⊃ OCOMP

Fig. 7. The Accessibility Hierarchy of relativizable grammatical relations.

Psycholinguistic experiment with native EN-speaking adults and children

1. Stimulus sentence recording, followed shortly by 8-digit sequence
2. Write down 8 digits from memory
3. Write down stimulus sentence

Grammatical Relation:	SU	DO	IO	OBL	GenS	GenO
Repetition Accuracy:	406	364	342	279	167	171
errors ($= R.A._{max} - R.A.$)	234	276	298	361	471	469

Fig. 8. Results from Keenan and Hawkins (1987).

Result: repetition accuracy *decreases* as the RC is formed *lower* on the AH

Conclusion: the lower the position occupied by an RC, the more difficult it is to comprehend-produce that RC

…but we don't know *why* 🤡

SUBJECT ⊃ DIR. OBJECT ⊃ INDIR. OBJECT ⊃ OBLIQUE ⊃ GENITIVE ⊃ OCOMP

Fig. 7. The Accessibility Hierarchy of relativizable grammatical relations.

Experiment

Hypothesis

Question: Is ER an adequate explanation for this drop in comprehensibility-production as one goes down the AH (L→R)?

Hypothesis: Yes.

- **If** *person's difficulty on each word* $\propto$ *word's surprisal*
- **Then** *person's difficulty on* ***sentence*** = ∑ *difficulties of each word*
- This holds not only for CFGs, but also for MCSGs

Reasoning

- People generally perform information-processing work $\propto$ ER brought about each word
- On a **sentence** level, this work increases when the position on the AH decreases (L→R)

MCSGs derived from Keenan & Hawkins (1987) stimuli

Adjunction analysis (Chomsky, 1977)

- **Adjunct:** structurally dispensable part of a sentence, clause, etc.
- Full sentences and relative clauses are related by the movement of WH words (interrogative words, e.g. *who*, *what*, etc.)
- Example

  Which guacamole did Patrick enjoy?

  The ~~which~~ guacamole, *[which Patrick enjoyed]*...

Promotion analysis (Brame, 1976; Kayne, 1994)

- It is not only the WH-word that moves in relativization
- Evidence seen in e.g. idiomatic expressions
- Example: *to make headway*

  We made headway.

  👎 The headway was satisfactory.

  👍 The ~~made~~ headway *[made by the team]* was satisfactory.

Experiment

Data

subject
they had forgotten that the boy who told the story was so young
the fact that the girl who paid for the tickets is very poor doesn't matter
I know that the girl who got the right answer is clever
he remembered that the man who sold the house left the town

direct object
they had forgotten that the letter which Dick wrote yesterday was so long
the fact that the cat which David showed to the man likes eggs is strange
I know that the dog which Penny bought today is very gentle
he remembered that the sweets which David gave Sally were a treat

indirect object
they had forgotten that the man who Ann gave the present to was old
the fact that the boy who Paul sold the book to hates reading is strange
I know that the man who Stephen explained the accident to is kind
he remembered that the dog which Mary taught the trick to was clever

oblique
they had forgotten that the box which Pat brought with apples in was lost
the fact that the girl who Sue wrote the story with is proud doesn't matter
I know that the ship which my uncle took Joe on was interesting
he remembered that the food which Chris paid the bill for was cheap

genitive subject
they had forgotten that the girl whose friend bought the cake was waiting
the fact that the boy whose brother tells lies is always honest surprised us
I know that the boy whose father sold the dog is very sad
he remembered that the girl whose mother sent the clothes came too late

genitive object
they had forgotten that the man whose house Patrick bought was so ill
the fact that the sailor whose ship Jim took had one leg is important
I know that the woman whose car Jenny sold was very angry
he remembered that the girl whose picture Clare showed us was pretty

Fig. 9. Stimuli from Hawkins and Keenan (1974/1987).

Experiment

Data (cont.)

See: Hale (2006), *Figure 14*

Keenan & Hawkins (1987) stimuli covered by two MCSGs (adjunction vs. promotion)

Calculating entropy requires **probability distribution**

1. H&K's stimuli viewed as mini treebank
2. Use relative frequency estimation technique (Chi, 1999) to derive CF derivation rule probabilities
3. 24 stimuli = (6 AH positions) x (4 relativization examples)
4. Weigh derivations by Brown corpus counts (Kučera & Francis, 1967)

count	grammatical relation	definition
1430	subject	co-indexed trace is the first daughter of S
929	direct object	co-indexed trace is immediately following sister of a V-node
167	indirect object	co-indexed trace is part of a PP not annotated as benefactive, locative, manner, purpose, temporal or directional
41	oblique	co-indexed trace is part of a benefactive, locative, manner, purpose, temporal or directional PP
34	genitive subject	WH word is *whose* and co-indexed trace is first daughter of S
4	genitive direct object	WH word is *whose* and co-indexed trace is immediately following sister of a V-node

Fig. 13. Counts from Brown portion of Penn Treebank III.

Experiment

Procedure

See: Hale (2006), *3.5 Conditional entropy*; Lang (1974); Billot & Lang (1989)

Calculate ERs of MCSG-generated sentences

Step 1: intersecting a grammar *G* with a finite-state language (Lang, 1974; Billot & Lang, 1989)

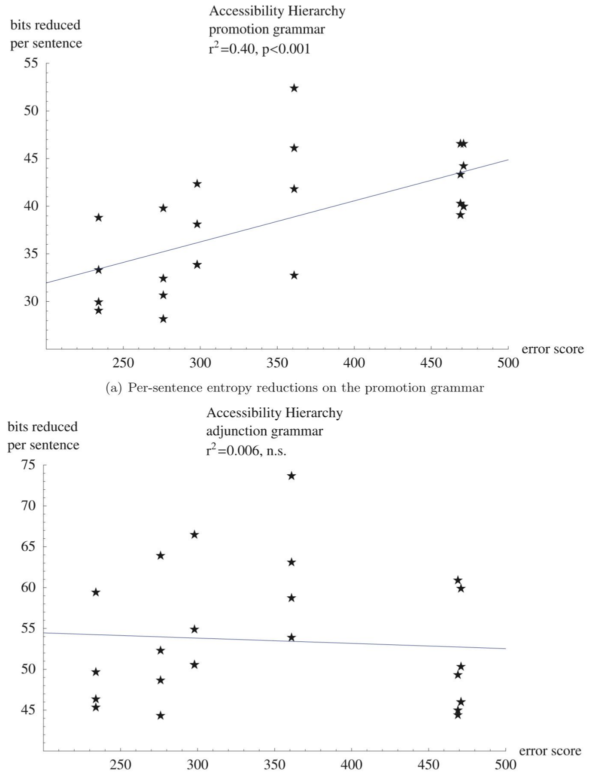
1. Consider initial substring $w = w_1, \ldots, w_n$ as finite-state language *w(?)**
2. Define set of grammatical derivations that start with *s*
   a. Intersect *w(?)** with language generated by MCSG
   b. Obtain a chart of parses that generate *w*
3. Use Grenander's theorem (1967) to calculate entropy of *S*

$$H(\xi_i) = -\sum_{r \in \Pi(\xi_i)} P(r) \log_2 P(r) + \sum_{r \in \Pi(\xi_i)} P(r) \left[H(\xi_{j_1}) + H(\xi_{j_2}) + \ldots \right]$$

Step 2: interpret chart entropies as uncertainty between each *word* in each stimulus sentence

Step 3: word ERs summed for each sentence to measure total *sentence* difficulty

$$ER(w_i) = max(0, H_{i-1} - H_i)$$



(a) Per-sentence entropy reductions on the promotion grammar

(b) Per-sentence entropy reductions on the adjunction grammar

Fig. 15. Predictions of two probabilistic MGs in conjunction with the ERH.

Independent variable (x): error scores from Keenan & Hawkins (1987) digit-interference/sentence recall experiment

Dependent variable (y): ER (bits)

Figure 15a: promotion grammar
Figure 15b: adjunction grammar

Key takeaway: ER correlates with sentence difficulty (operationalized by AH position) for promotion grammar only

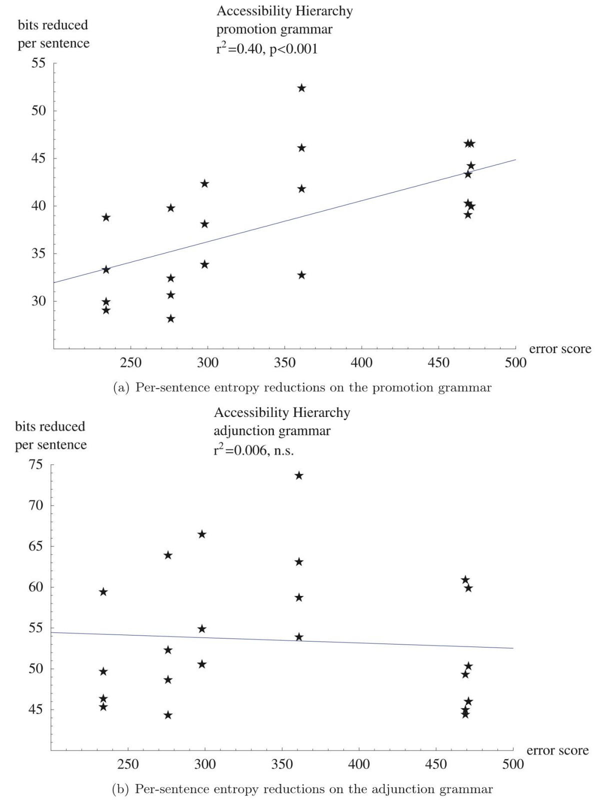


Fig. 15. Predictions of two probabilistic MGs in conjunction with the ERH.

Why not for adjunction grammar?

One idea: among adjunction-promoted RCs, GEN- and OCOMP-relativized RCs decrease in entropy

Entropy increases when a distribution is not informative (think back to Question 4.2.1.4 about uniform distributions)

GEN- and OCOMP-RCs constructed with adjunction grammar (WH-movement) might have fewer continuation possibilities at a given point

Fewer continuation possibilities ?= decreased entropy

Decreased processing difficulty from OBL → GEN/OCOMP

WHAT?

Successfully extended former work (Wilson & Carroll, 1954) from word level in CFGs to sentence level in infinite grammars (e.g. MCSGs containing RCs)

HOW?

Provide ERH-based explanation of existing result that comprehension-production of RCs increases when relativization occurs lower on the AH (Keenan & Hawkins, 1987)

SO WHAT?

Formalize what used to simply be *intuition* about sentence processing difficulty

YET…

The ERH extension holds for MCSGs whose RCs are constructed via promotion grammar but not adjunction grammar

Can we say that an MCSG represents all classes of infinite grammars?

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 40 (2016) 1382–1411

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12274

Uncertainty and Expectation in Sentence Processing: Evidence From Subcategorization Distributions

Tal Linzen,[a] T. Florian Jaeger[b]

[a]*Department of Linguistics, New York University*

[b]*Department of Brain and Cognitive Sciences, Department of Computer Science, and Department of Linguistics, University of Rochester*

Received 9 October 2014; received in revised form 28 April 2015; accepted 4 May 2015

What is human sentence processing based on?

Human sentence processing is expectation based:
As we read a sentence, we use our statistical experience with a language to generate predictions about upcoming syntactic structure.

How is human sentence processing affected by the readers' (un)certainty about those expectations?

?

1. What was the goal of Linzen Jaeger (2016)'s study?

Doubts among the researchers about the viability of entropy reduction as a predictor of processing

⇩

Need for verification of competing theories

"The goal of the current study is

to assess the effects of entropy and entropy reduction

in the same materials,

within the same syntactic framework,

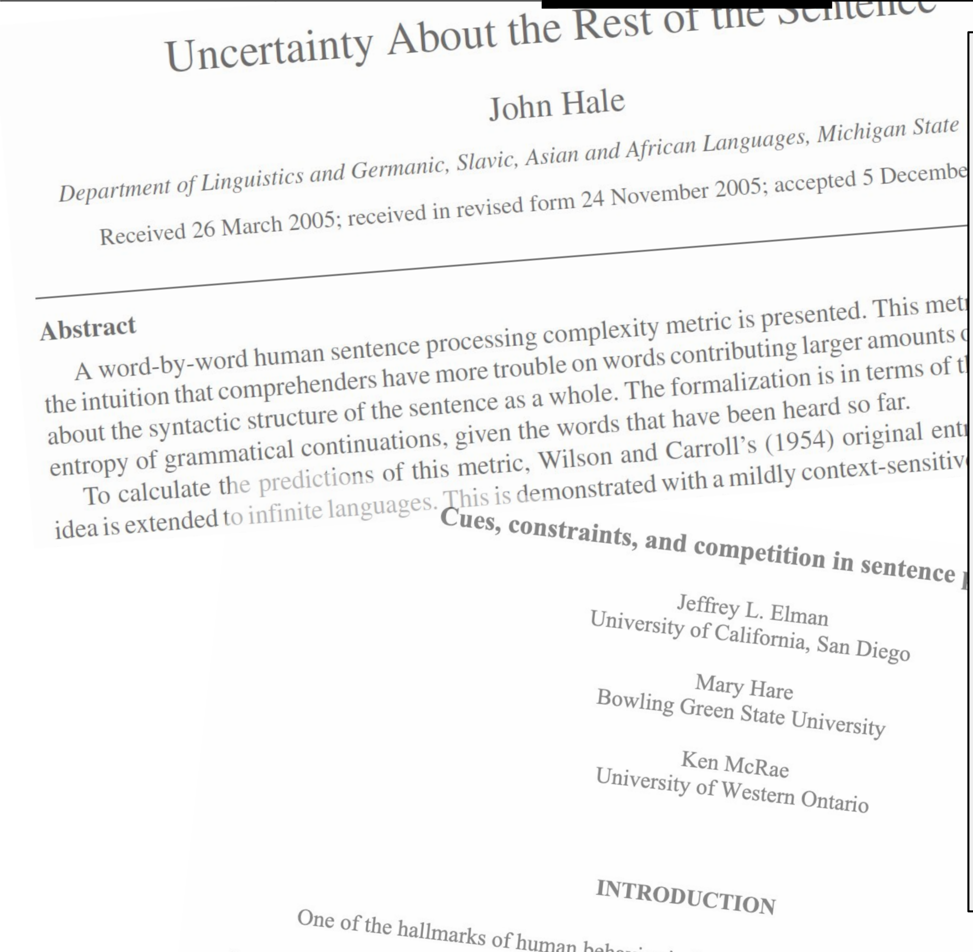
while avoiding structural confounds."

Key findings:

- uncertainty about the full structure, but not about the next step, is a significant predictor of processing difficulty
- greater reduction in uncertainty is correlated with increased reading times (RTs).
- previously observed effects of expectation violation (surprisal) are replicated
- both surprisal and uncertainty appear to affect human RTs.

Comparison of **two types of uncertainty:**

- uncertainty about the **verb's complement** (reflects the next prediction step)
- uncertainty about the **full sentence** (reflects an unbounded number of prediction steps)


Uncertainty About the Rest of the Sentence

John Hale

Department of Linguistics and Germanic, Slavic, Asian and African Languages, Michigan State

Received 26 March 2005; received in revised form 24 November 2005; accepted 5 Decembe

Abstract

A word-by-word human sentence processing complexity metric is presented. This metr
the intuition that comprehenders have more trouble on words contributing larger amounts o
about the syntactic structure of the sentence as a whole. The formalization is in terms of t
entropy of grammatical continuations, given the words that have been heard so far.
To calculate the predictions of this metric, Wilson and Carroll's (1954) original entr
idea is extended to infinite languages. This is demonstrated with a mildly context-sensitiv

Cues, constraints, and competition in sentence p

Jeffrey L. Elman
University of California, San Diego

Mary Hare
Bowling Green State University

Ken McRae
University of Western Ontario

INTRODUCTION

One of the hallmarks of human behavi

2 hypotheses about the potential role of uncertainty in parsing

1. **the competition hypothesis**:

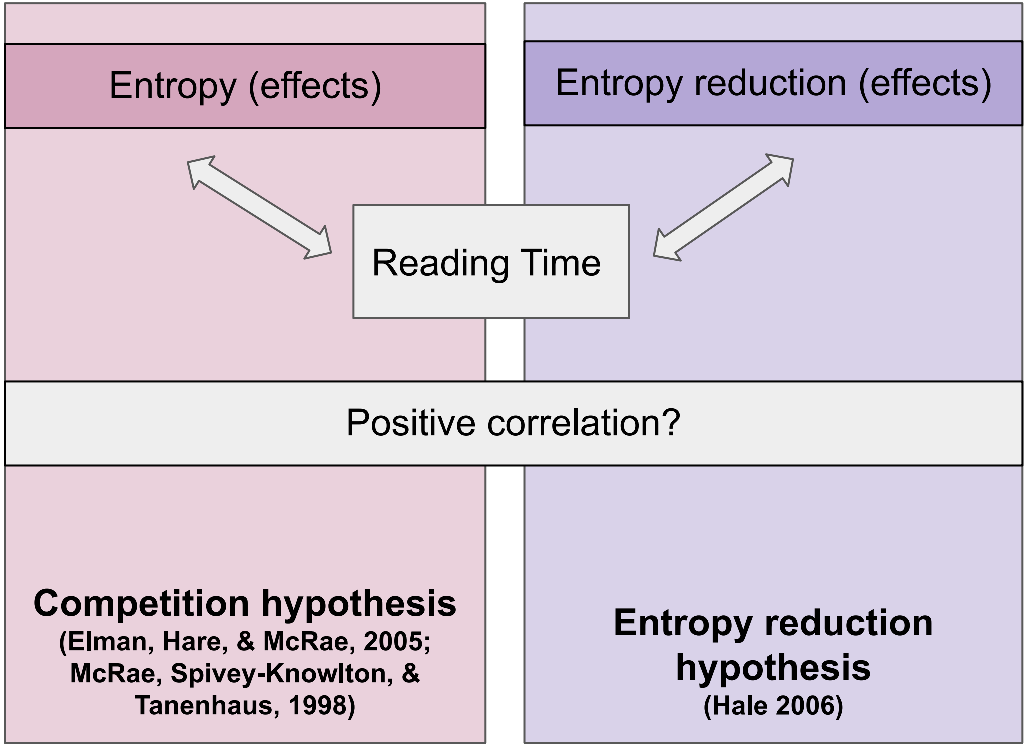
   higher uncertainty should result in the activation of multiple structures that compete with each other, thereby slowing down processing
   (Elman et al., 2005; McRae et al., 1998)

2. **the entropy reduction hypothesis:**

   processing is slowed down by any word that reduces uncertainty
   (Hale, 2006)

Distinction 1

"We emphasize two distinctions that are not always clearly highlighted in the literature. First, we distinguish ***entropy effects from entropy reduction effects****."*



Following Hale (2006), we define entropy reduction at a word to be 0 when the entropy after the word is greater than before it.

Distinction 2

Second, we compare uncertainty in single-step prediction and uncertainty in full prediction. The entropy reduction hypothesis (as formulated in Hale, 2006) predicts only that the latter should be correlated with RTs. [...] The results of the current study show that uncertainty in single-step prediction and uncertainty in full prediction can differ dramatically and exhibit qualitatively different correlations with RTs.

Uncertainty in single-step prediction

Uncertainty in full prediction

Current study examines both types of uncertainty, Showing that they can differ dramatically and exhibit different correlations with RT.

Entropy reduction hypothesis
(Hale 2006)

Self-paced Reading Time Experiment

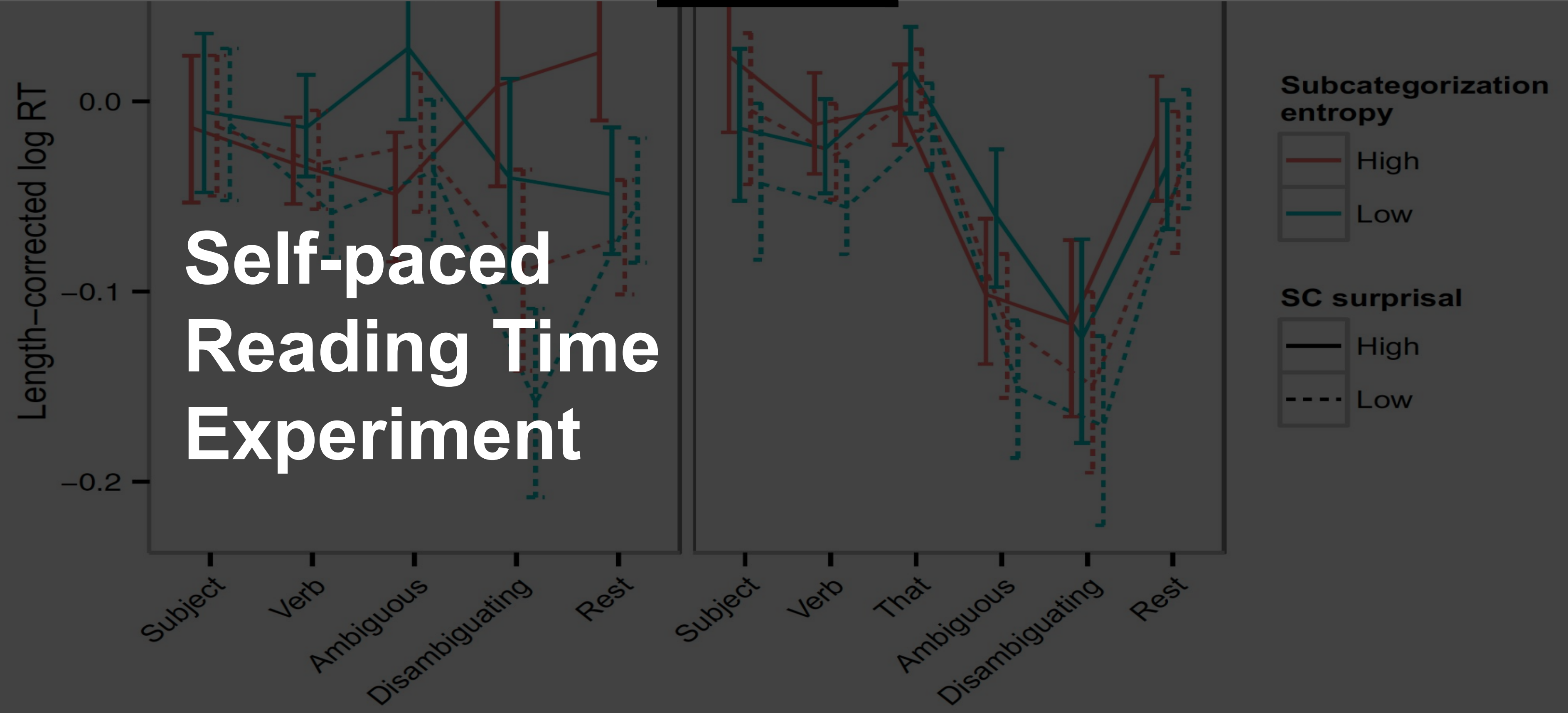
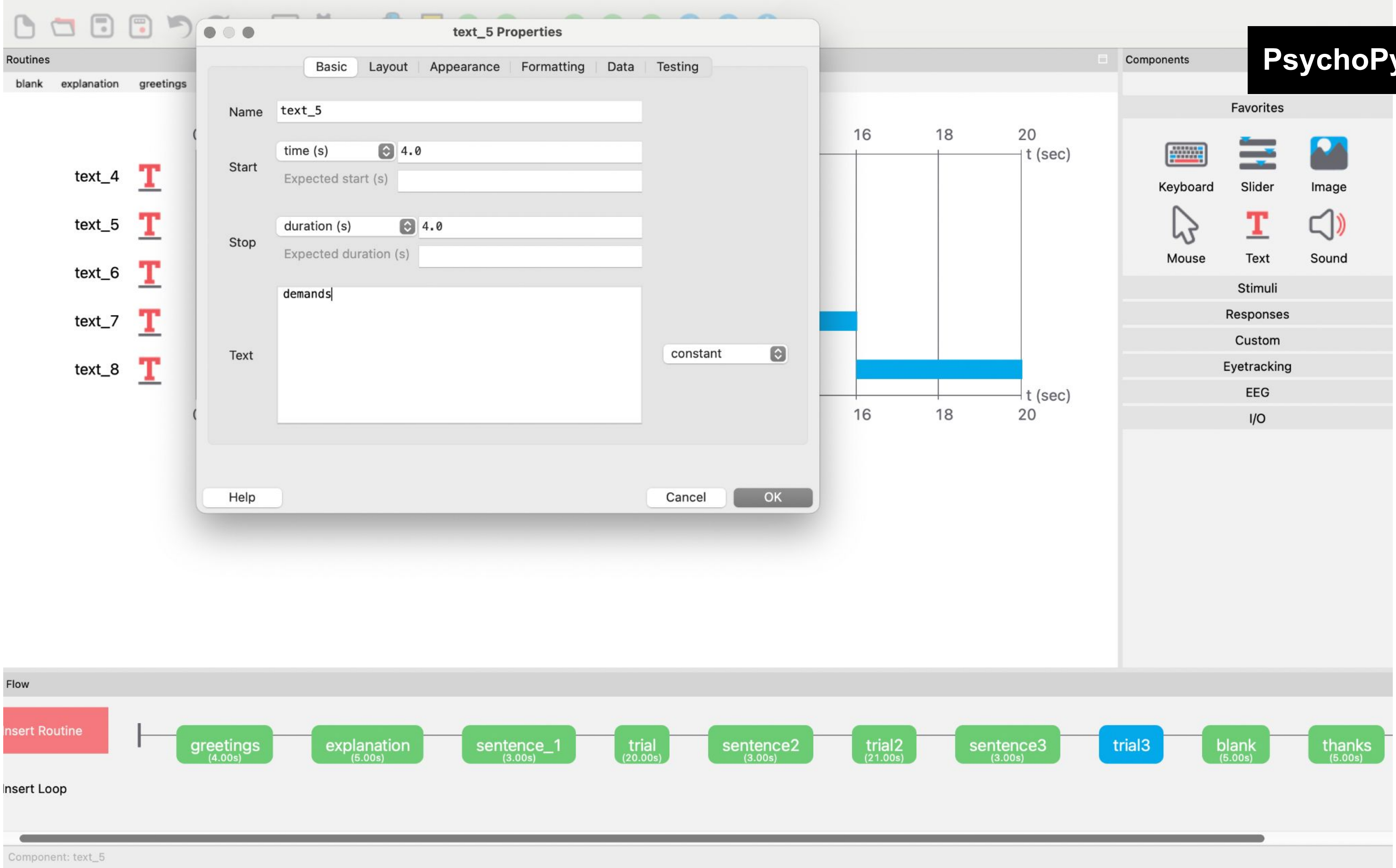


Fig. 4. Mean reading times (RTs). Error bars show bootstrapped 95% confidence intervals.

PsychoPy

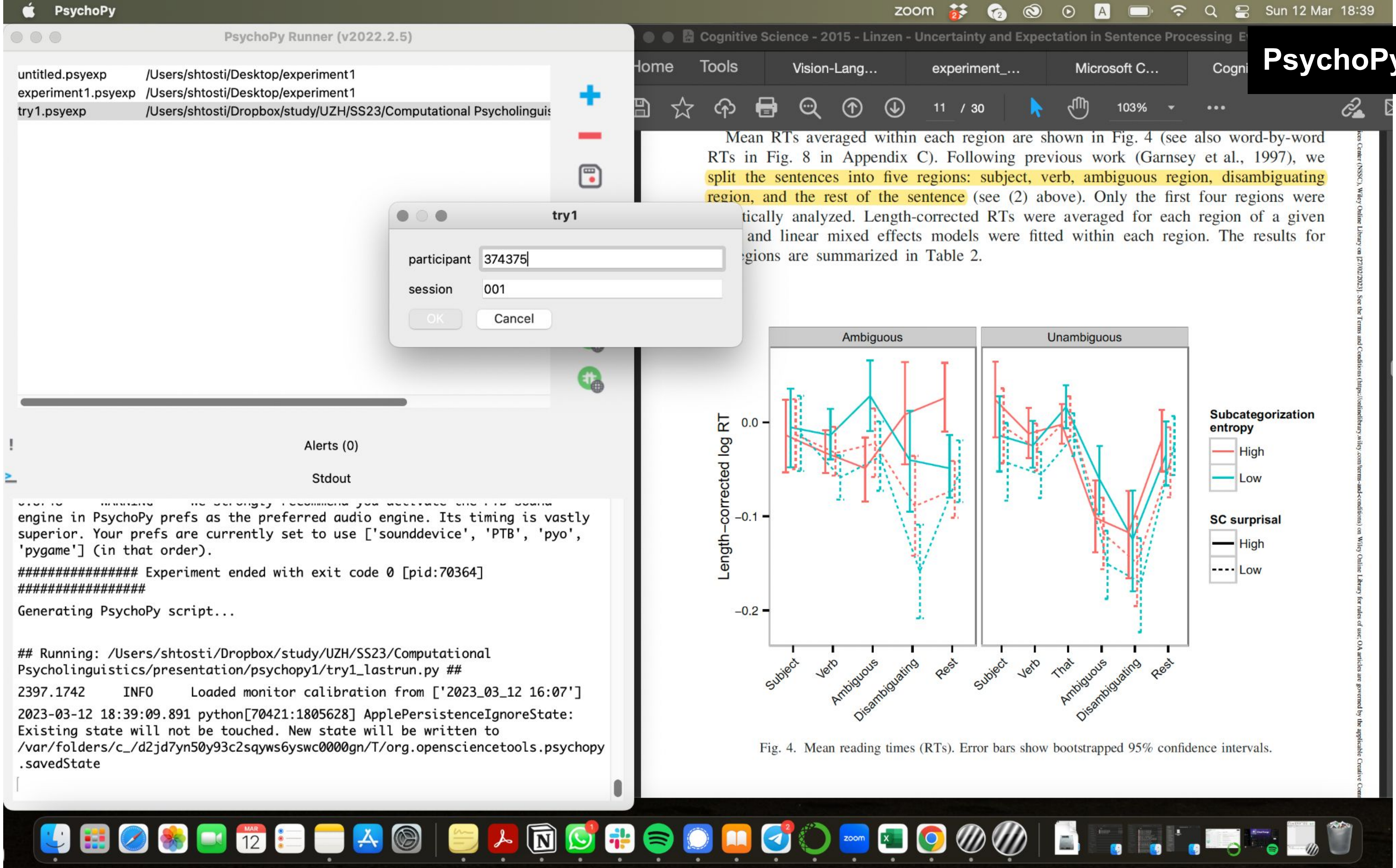
Routines
blank explanation greetings thanks trial routine_1 sentence2 sentence_1 trial2 sentence3 trial3
0 2 4 6 8 10 12 14 16 18 20
t (sec)
text_19
text_20
text_21
text_22
text_23
text_24
text_25
Components
Favorites
Keyboard
Slider
Image
Mouse
Text
Sound
Stimuli
Responses
Custom
Eyetracking
EEG
I/O
Flow
Insert Routine
Insert Loop
greetings (4.00s)
explanation (5.00s)
sentence_1 (3.00s)
trial (20.00s)
sentence2 (3.00s)
trial2 (21.00s)
sentence3 (3.00s)
trial3
blank (5.00s)
thanks (5.00s)

PsychoPy



text_5 Properties
Basic
Layout
Appearance
Formatting
Data
Testing
Name
text_5
Start
time (s)
4.0
Expected start (s)
Stop
duration (s)
4.0
Expected duration (s)
Text
demands
constant
Help
Cancel
OK
Routines
blank
explanation
greetings
text_4
text_5
text_6
text_7
text_8
16
18
20
t (sec)
Components
Favorites
Keyboard
Slider
Image
Mouse
Text
Sound
Stimuli
Responses
Custom
Eyetracking
EEG
I/O
Flow
Insert Routine
Insert Loop
greetings (4.00s)
explanation (5.00s)
sentence_1 (3.00s)
trial (20.00s)
sentence2 (3.00s)
trial2 (21.00s)
sentence3 (3.00s)
trial3
blank (5.00s)
thanks (5.00s)
Component: text_5

PsychoPy



PsychoPy
zoom
Sun 12 Mar 18:39
PsychoPy Runner (v2022.2.5)
untitled.psyexp /Users/shtosti/Desktop/experiment1
experiment1.psyexp /Users/shtosti/Desktop/experiment1
try1.psyexp /Users/shtosti/Dropbox/study/UZH/SS23/Computational Psycholinguis
try1
participant 374375
session 001
OK
Cancel
Alerts (0)
Stdout
engine in PsychoPy prefs as the preferred audio engine. Its timing is vastly superior. Your prefs are currently set to use ['sounddevice', 'PTB', 'pyo', 'pygame'] (in that order).
################ Experiment ended with exit code 0 [pid:70364] ################
Generating PsychoPy script...
Running: /Users/shtosti/Dropbox/study/UZH/SS23/Computational Psycholinguistics/presentation/psychopy1/try1_lastrun.py
2397.1742 INFO Loaded monitor calibration from ['2023_03_12 16:07']
2023-03-12 18:39:09.891 python[70421:1805628] ApplePersistenceIgnoreState: Existing state will not be touched. New state will be written to /var/folders/c_/d2jd7yn50y93c2sqyws6yswc0000gn/T/org.opensciencetools.psychopy.savedState
Cognitive Science - 2015 - Linzen - Uncertainty and Expectation in Sentence Processing
Home
Tools
Vision-Lang...
experiment_...
Microsoft C...
11 / 30
103%
Mean RTs averaged within each region are shown in Fig. 4 (see also word-by-word RTs in Fig. 8 in Appendix C). Following previous work (Garnsey et al., 1997), we split the sentences into five regions: subject, verb, ambiguous region, disambiguating region, and the rest of the sentence (see (2) above). Only the first four regions were ...tically analyzed. Length-corrected RTs were averaged for each region of a given ... and linear mixed effects models were fitted within each region. The results for ...gions are summarized in Table 2.
Ambiguous
Unambiguous
Length−corrected log RT
0.0
−0.1
−0.2
Subject
Verb
Ambiguous
Disambiguating
Rest
Subject
Verb
That
Ambiguous
Disambiguating
Rest
Subcategorization entropy
High
Low
SC surprisal
High
Low
Fig. 4. Mean reading times (RTs). Error bars show bootstrapped 95% confidence intervals.

PsychoPy

Welcome to our
Self-Paced Reading
Experiment!

PsychoPy

This is a short demo of implementing a self-paced reading task in PsychoPy.

PsychoPy

Sentence 1

PsychoPy

Patrick

PsychoPy

demands

PsychoPy

a

PsychoPy

nice

presentation.

PsychoPy

Sentence 2

PsychoPy

Patrick

PsychoPy

demands

PsychoPy

that

PsychoPy

the

PsychoPy

presentation

PsychoPy

is

PsychoPy

nice.

PsychoPy

Sentence 3

PsychoPy

Patrick

PsychoPy

demands

PsychoPy

the

PsychoPy

presentation

PsychoPy

to

PsychoPy

be

PsychoPy

nice.

PsychoPy

RT Experiment

2. Describe the experimental design of their experiment (i.e., what did the stimuli look like and why?).

The experiment is modeled after Garnsey et al. (1997).

Self-paced reading experiment paradigm

1. Half of the sentences read by a given participant include the complementizer “**that**”, as in (a) (unambiguous sentences)
2. half do not, as in (b) (ambiguous sentences).

two factors are manipulated:

- the subcategorization entropy of the main verb (high vs. low)
- the surprisal of an SC given the verb (high vs. low)

unambiguous ⟹	a.	The men *subject*	discovered *verb*	that *that*	the island *ambiguous*	had been invaded *disambiguating*	by the enemy. *rest*
ambiguous ⟹	b.	The men *subject*	discovered *verb*		the island *ambiguous*	had been invaded *disambiguating*	by the enemy. *rest*

RT Experiment

2 x 2 x 2 design:

Ambiguity x subcategorization entropy x SC surprisal

Competition hypothesis					
Larger uncertainty (entropy)	**High** subcategorization entropy conditions.	→	Slower processing	→	Longer reading time RT at the verb region
Smaller uncertainty (entropy reduction)	**Low** subcategorization entropy conditions				
Entropy reduction hypothesis					

RT Experiment

Participants	128
Platform	Mechanical Turk
Payment for participation	$1.75
Average completion time	17 minutes
# experimental lists	Eight experimental lists were created: The 32 items were randomized such that sets of four consecutive items had one item of each condition (with fillers interspersed).
# participants per experimental list	16

32 verbs

Classification of subcategorization frames into six categories:

- transitive (Klaus adore cookies)
- Intransitive (we watched attentively)
- quote (he said "that's fine by me")
- finite sentential complement (Trent yelled (that) the road was in sight)
- infinitival complement (she wanted to share her insight with others),
- "Other."

Step 1

32 sentence pairs
One for each verb (64 sentences in total).
Each pair contained:

- one version of the sentence with the complementizer that after the verb
- one without it

Step 2

Table 1
Lexical variables

Condition	Subcategorization Entropy	SC-Surprisal	Frequency (log)
Low entropy/low surprisal	1.13 (0.08)	1.52 (0.55)	3.64 (1.56)
Low entropy/high surprisal	1.09 (0.18)	4.15 (1.03)	4.31 (2.01)
High entropy/low surprisal	1.7 (0.12)	1.58 (0.45)	3.6 (1.24)
High entropy/high surprisal	1.68 (0.17)	3.86 (0.79)	3.85 (1.25)

Notes. Mean subcategorization entropy, SC surprisal, and log-transformed frequency of the main verb in each of the conditions of the factorial design. We use Entropy to refer to high versus low subcategorization frame entropy (i.e., single-step entropy at the verb) and Surprisal to refer to high versus low sentential complement surprisal. Standard errors are shown in parentheses.

List of Materials

A.1. Low subcategorization entropy, low SC surprisal

The men discovered (that) the island had been invaded by the enemy.
The women revealed (that) the secret had been exposed by the officials.
The man noticed (that) the mistake had not happened due to negligence.
The woman assumed (that) the blame might have belonged to the driver.
They all indicated (that) the problem might not bother the entire team.
Two people found (that) the equipment should be reported stolen right away.
Some people sensed (that) the conflict should be resolved quickly and peacefully
Many people guaranteed (that) the loan would be paid off on time.

A.2. Low subcategorization entropy, high SC surprisal

The woman determined (that) the estimate had been inflated by the accountant.
Two people heard (that) the album had been criticized in the magazine.
Some people understood (that) the message had not meant much to foreigners.
They all read (that) the newspaper might be going out of business.
The women worried (that) the parents might have become quite restless recently.
Many people advocated (that) the truth should be made public without delay.
The man taught (that) the children should be sheltered from all harm.
The men projected (that) the film would not gross enough in cinemas.

A.3. High subcategorization entropy, low SC surprisal

They all claimed (that) the luggage had been stolen from the hotel.
Some people regretted (that) the decision had been reached without any discussion.
The men remembered (that) the appointment had not changed since last week.
The women warned (that) the drivers might have drunk too much vodka.
Many people feared (that) the future might not hold hope for them.
The man proposed (that) the idea should be abandoned for financial reasons.
Two people suggested (that) the scene should be filmed right before sunset.
The woman announced (that) the wedding would be postponed until late August.

A.4. High subcategorization entropy, high SC surprisal

The men forgot (that) the details had been worked out in advance.
The man observed (that) the patient had been sent home too early.
The woman recalled (that) the speech had not gone over very well.
The women answered (that) the questions might be discussed during the meeting.
Some people added (that) the numbers might have decreased since last year.
Two people wrote (that) the interview should be conducted over the phone.
Many people advised (that) the president should be considering further budget cuts.
The men begged (that) the judge would not treat the defendant harshly.

Full Entropy Analysis

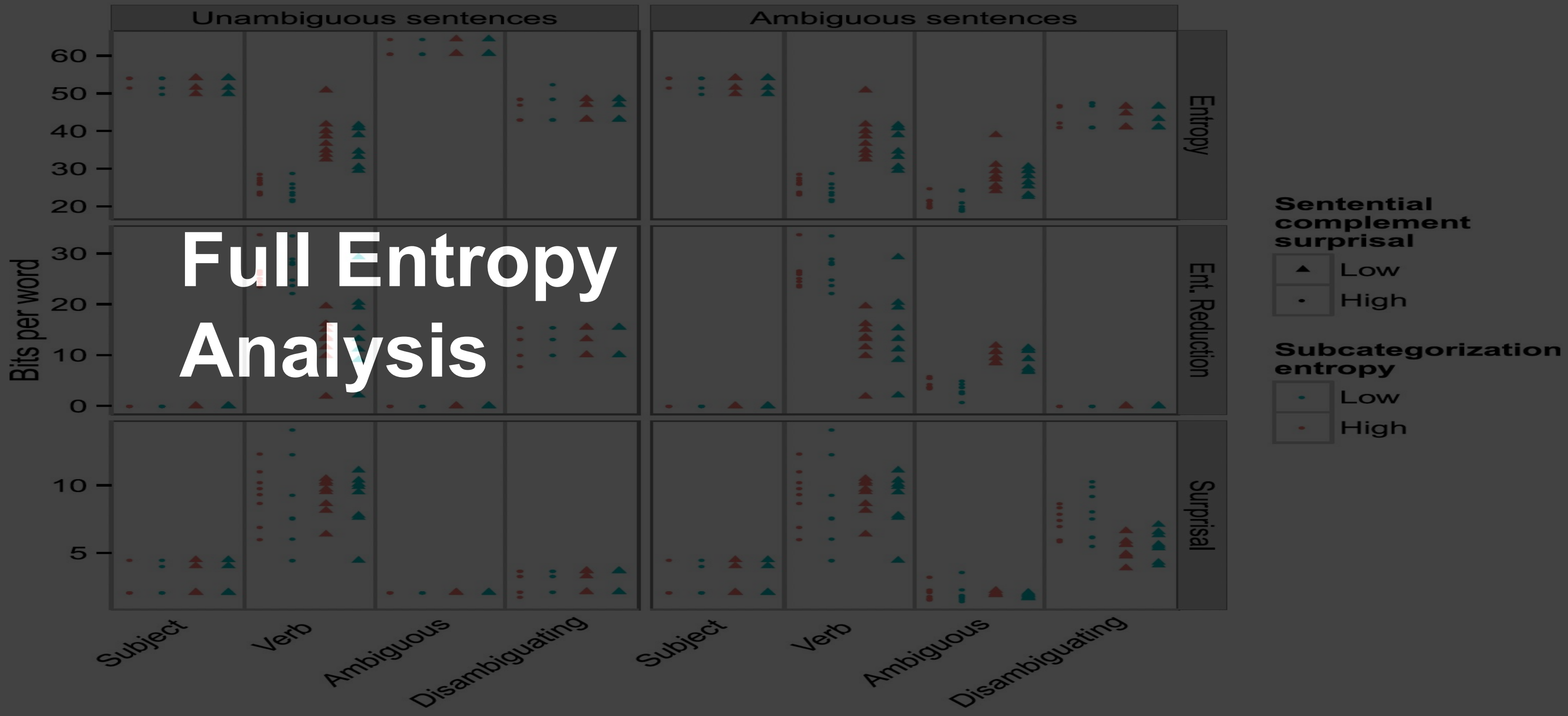
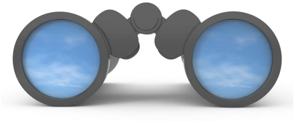


Fig. 7. Variability across items of PCFG-derived variables (entropy, entropy reduction, and surprisal). The values plotted are the mean of each variable across the region. Following Hale (2006), entropy reduction at a word is defined to be 0 when entropy after that word is greater than entropy before it.

Full Entropy Analysis

What is full entropy?



What is the **lookahead distance** n that humans use in sentence processing?
The current study has evaluated the two ends of the spectrum.

n = 1 (single-step entropy) **n = ∞ (full entropy)**

Depth of Analysis

Full Entropy Analysis

Predicting syntactic structure an **unlimited number of derivation steps ahead**:

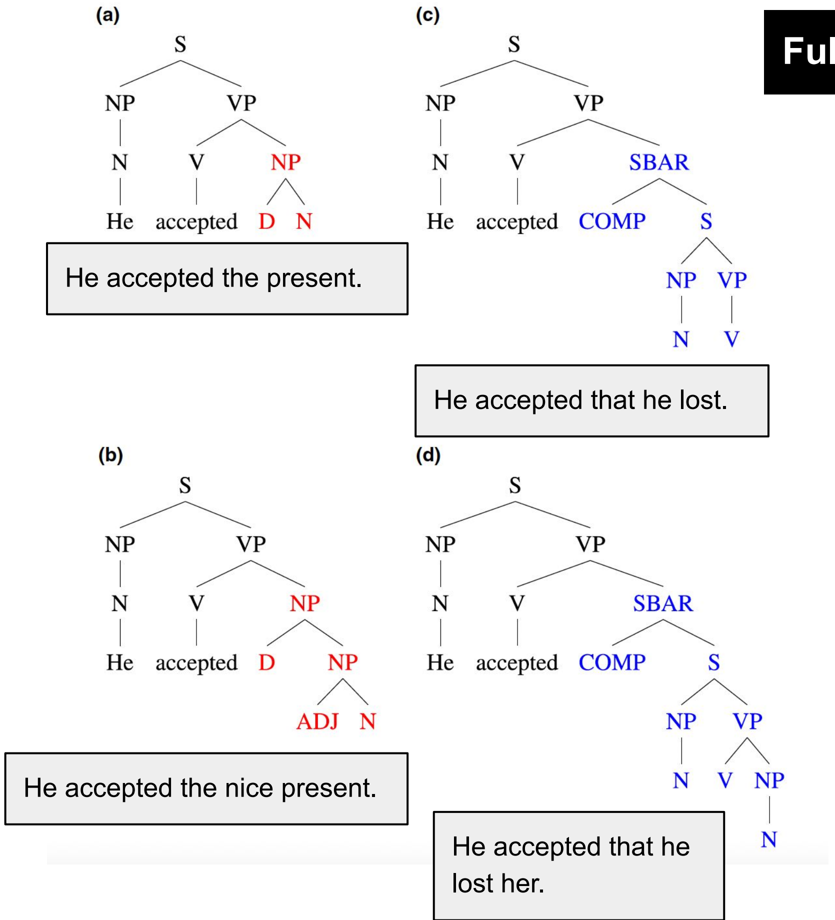
In addition to predicting the category of the complement (NP or SC), the reader predicts its internal structure.

(a) Prediction of a simple determiner + noun phrase;

(b) prediction of a noun phrase with an adjective;

(c) prediction of an SC with an intransitive verb;

(d) prediction of an SC with a transitive verb.



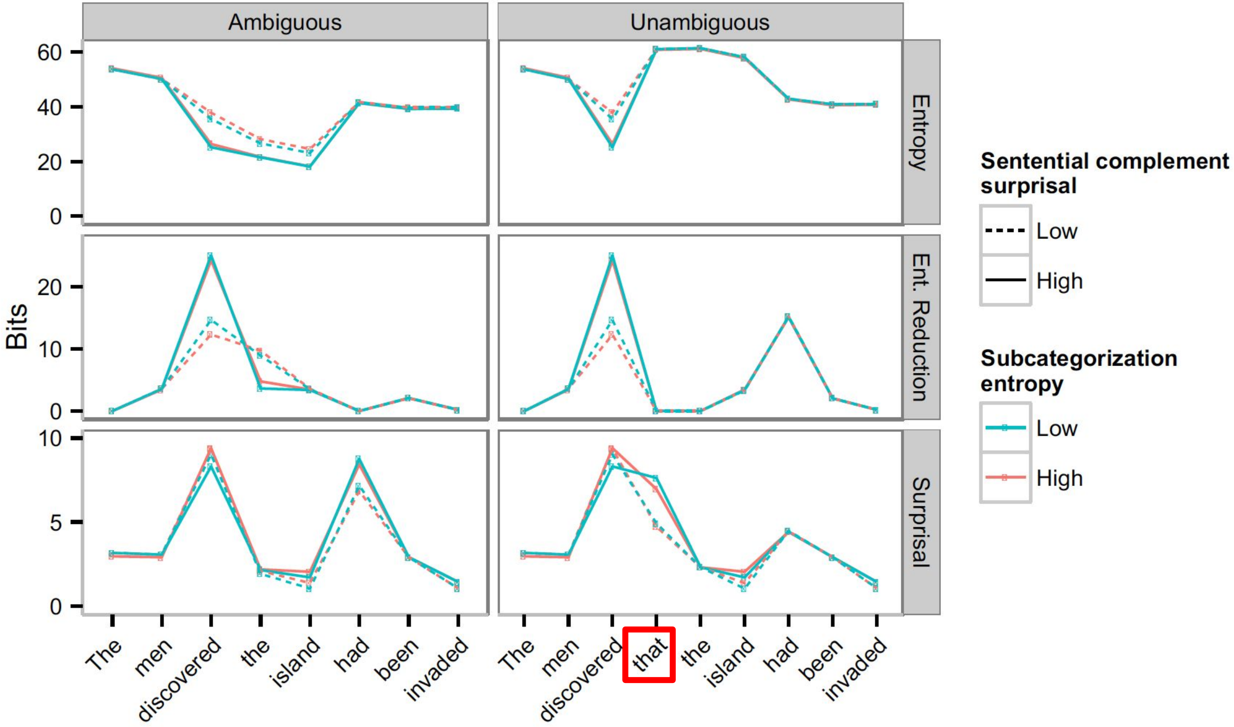


Fig. 5. Word-by-word entropy, entropy reduction, and surprisal predictions derived from the probabilistic context-free grammar based on the Penn Treebank. Predictions are averaged within each of the four conditions of the factorial design. Following Hale (2006), we define entropy reduction at a word to be 0 when entropy after the word is greater than entropy before it.

Full Entropy Analysis

"In order to assess the effect of the **full uncertainty** that a comprehender might experience during incremental sentence understanding, we derived full entropy estimates from a probabilistic context-free grammar (**PCFG**) based on the Penn Treebank."

"The **full entropy** estimates derived from the grammar take into account **not only the uncertainty about the next syntactic node but also the uncertainty about the internal structure of that node.**"

RT Experiment

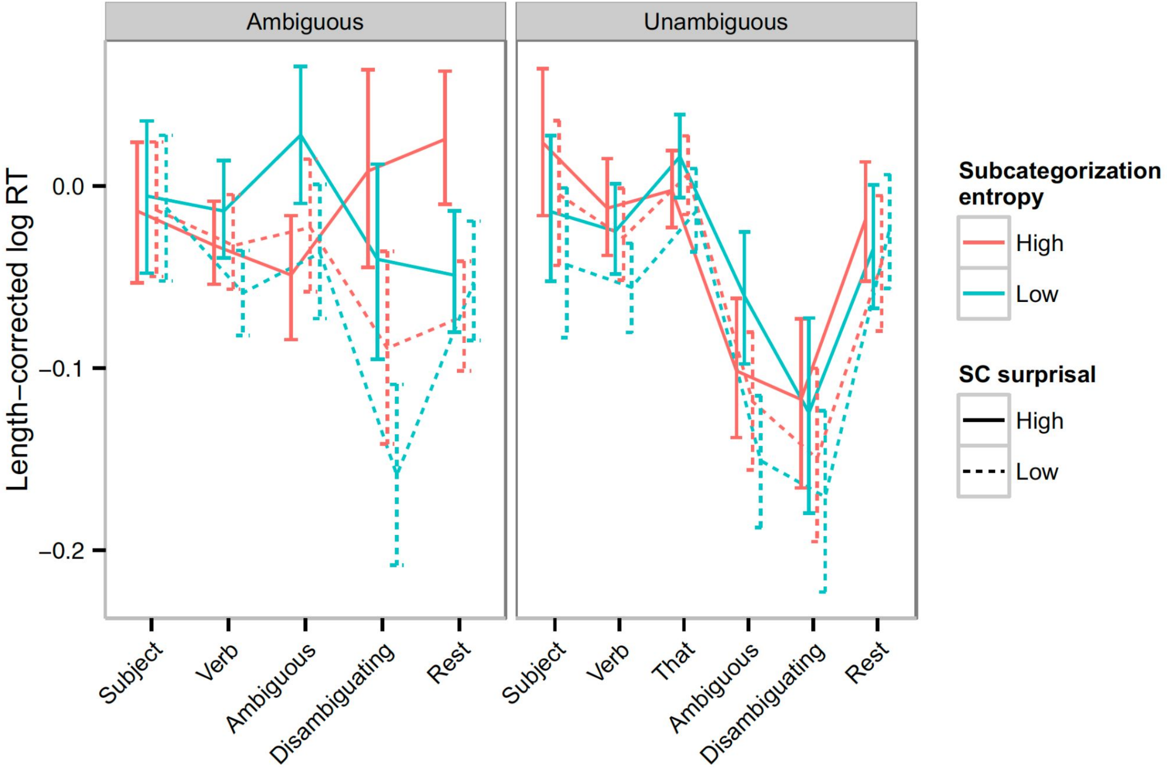


Fig. 4. Mean reading times (RTs). Error bars show bootstrapped 95% confidence intervals.

Reading Times (self-paced)	
Subject region	- All effects below significance (**p > 0.1**)
Verb region	- Subcategorization entropy: effect below significance (**p > 0.5**) - Unexpectedly, low SC surprisal verbs were read more slowly, though this difference was only marginally significant (**p = 0.09**)
Ambiguous region	- The effect of subcategorization entropy not significant (**p > 0.3**). - The three-way interaction between surprisal, entropy, and ambiguity did not approach significance (**p > 0.5**)
Disambiguating region	- The subject of the embedded clause (the “ambiguous region”) was read faster in unambiguous sentences than in ambiguous sentences (**p < 0.001**). - A marginal effect of SC surprisal (**p = 0.1**). - The three-way interaction between surprisal, entropy, and ambiguity did not approach significance (**p > 0.5**)

Full Entropy Analysis

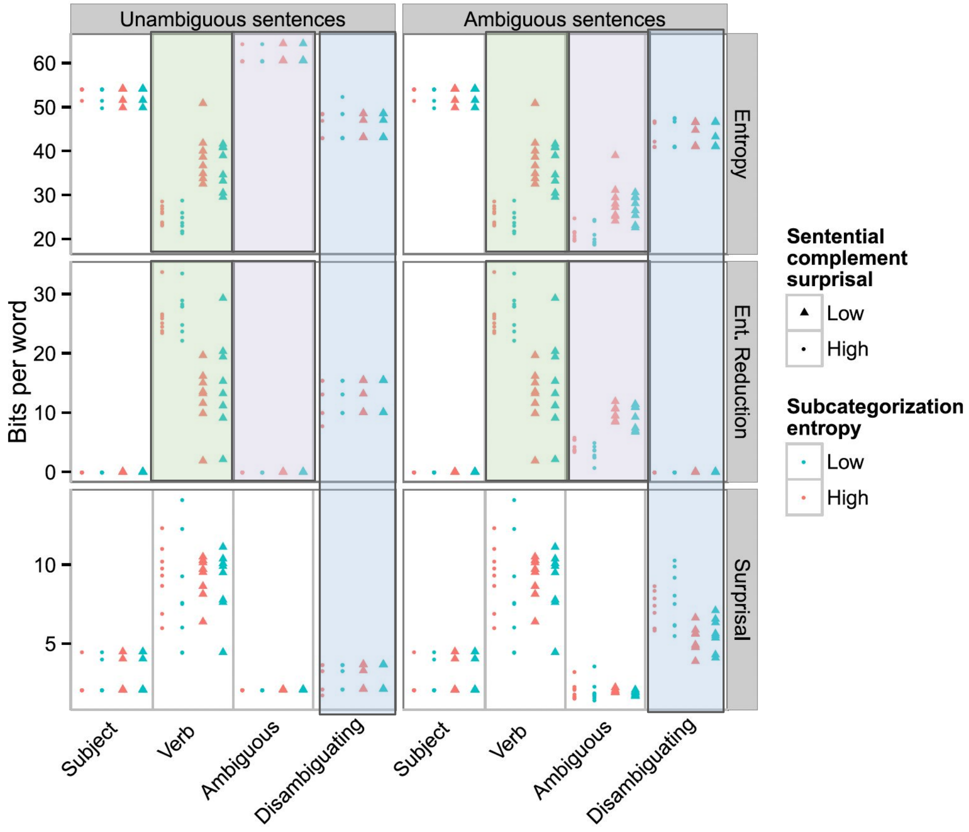


Fig. 7. Variability across items of PCFG-derived variables (entropy, entropy reduction, and surprisal). The values plotted are the mean of each variable across the region. Following Hale (2006), entropy reduction at a word is defined to be 0 when entropy after that word is greater than entropy before it.

Subject region	- Linear mixed-effects models did not yield any significant effects in this region (all **$p > 0.3$**).
Verb region	- The entropy reduction analysis found a positive effect of entropy reduction on RTs at the verb (**$p = 0.047$**). - The entropy analysis found a marginal negative effect of entropy (**$p = 0.08$**). - Neither analysis revealed any other effects (**$p > 0.5$**)
Ambiguous region	- Entropy reduction correlated with increased RTs (**$p < 0.001$**) - entropy correlated with decreased RTs (**$p < 0.001$**). - No other effects reached significance in either analysis (**$p > 0.5$**)
Disambiguating region	- Surprisal had a significant effect in both the entropy analysis (**$p = 0.01$**) and the entropy reduction analysis (**$p = 0.03$**) - None of the other predictors was significant (**$p > 0.1$**)

Full Entropy Analysis

a significant effect of entropy reduction on RTs at the verb:

higher entropy reduction at the verb correlated with **longer RTs**

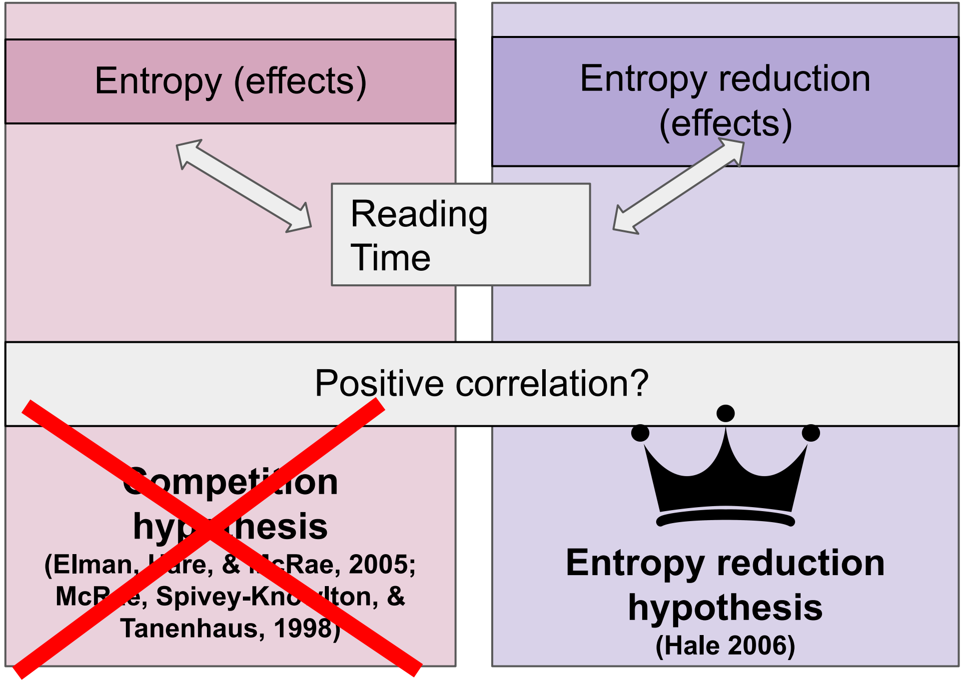
RTs on the verb and ambiguous region:
the observed pattern is **consistent with the predictions of the entropy reduction hypothesis** and **inconsistent with the predictions of the competition hypothesis**.

Table 3
Predictions made by uncertainty-based theories (based on full entropy)

	Verb	Ambiguous	Disambiguating
Competition	Lower SC surprisal, Higher subcat entropy → Higher full entropy → Longer RTs ✗	Longer RTs in unambiguous than in ambiguous sentences ✗	
Entropy reduction	Lower SC surprisal, Higher subcat entropy → Higher full entropy → Lower full entropy reduction → Shorter RTs ✓	Shorter RTs in unambiguous than in ambiguous sentences ✓	Longer RTs in unambiguous than ambiguous sentences ⊘
Within ambiguous sentences only:			
Competition		Lower SC surprisal, Higher subcat entropy → Longer RTs ⊘	
Entropy reduction		Lower SC surprisal, Higher subcat entropy → Longer RTs ⊘	

Notes. Predictions made by the competition and entropy reduction hypotheses for the three main regions of the materials (the hypotheses do not make any predictions for the subject region). Predictions are shown both for the whole data set, focusing on the comparison between ambiguous and unambiguous sentences, and specifically for ambiguous sentences. Cells are shaded whenever a hypothesis does not predict any RT difference in a region. Predictions confirmed by the results are marked with ✓; predictions not found confirmed are marked with ⊘; predictions rejected by the results are marked with ✗.

4. Summarize the results of the study.



WHAT? **syntactic expectations** induced by individual lexical items

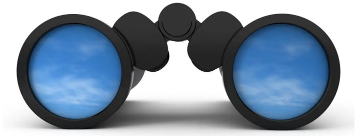
WHY? To examine the role of **uncertainty over expectations** in parsing

SO? The results lend some support to the **entropy reduction hypothesis** (Hale, 2006).

YET, The entropy reduction hypothesis failed to predict RTs when in conflict with the surprisal hypothesis → **surprisal and uncertainty** both play a role in explaining processing difficulty.

PLUS, Uncertainty was not a significant predictor of RTs when only the syntactic category of the verb's complement was considered, and became significant only when the internal complexity of the complement was taken into account (**full entropy**)

Determining the appropriate **lookahead distance**



The finding that RTs can be predicted by full but not single-step entropy supports the idea that human parsing during reading involves lookahead of at least **several derivation steps**.

References

Billot, S., & Lang, B. (1989). The structure of shared forests in ambiguous parsing. In *Proceedings of the 1989 Meeting of the Association for Computational Linguistics*.

Brame, M. K. (1976). In Conjectures and refutations in syntax and semantics (pp. 125–143). Amsterdam: North-Holland.

Chi, Z. (1999). Statistical properties of probabilistic context-free grammars. *Computational Linguistics*, 25, 131–160.

Chomsky, N. (1977). On Wh-movement. In P. Culicover, T. Wasow, & A. Akmajian (Eds.), *Formal syntax* (pp. 71–132). New York: Academic.

Grenander, U. (1967). *Syntax-controlled probabilities* (Tech. Rep.). Providence, RI: Brown University, Division of Applied Mathematics.

Hale, J. (2003b). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 32, 101–123.

Kayne, R. S. (1994). The antisymmetry of syntax. Cambridge, MA: MIT Press.

Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8, 63–99.

Keenan, E. L., & Hawkins, S. (1987). The psychological validity of the accessibility hierarchy. In E. L. Keenan (Ed.), *Universal grammar: 15 essays* (pp. 60–85). London: Croom Helm.

Lang, B. (1974). Deterministic techniques for efficient non-deterministic parsers. In J. Loeckx (Ed.), *Proceedings of the 2nd Colloquium on Automata, Languages and Programming* (pp. 255– 269). Saarbrücken, Germany: Springer.

Relative clause. (2023, 2023, March 10). In *Wikipedia*. https://en.wikipedia.org/wiki/Relative_clause#Accessibility_hierarchy

Taylor, W. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 30, 415–433.

Wilson, K., & Carroll, J. B. (1954). Applications of entropy measures to problems of sequential structure. In C. E. Osgood & T. A. Sebeok (Eds.), *Psycholinguistics: A survey of theory and research* (pp. 103–110). Bloomington: Indiana University Press.

Q & A