The information conveyed by words in sentences

John Hale Cognitive Science Department
Johns Hopkins University

March 24, 2002

claim: reading time is related to
the reduction in uncertainty
about what the speaker means

plan of talk

1. formalizing "reduction in uncertainty"
   as entropy of parser state

2. application to processing phenomena

   (a) main-verb/reduced relative temporary ambiguity

   (b) noun phrase/zero vs. noun-phrase/sentence asymmetry

   (c) center-embedding

   (d) subject vs. object relative asymmetry

meaning $\equiv$ *PCFG derivation*

Grammar

(nontransformational) phrase structure
(Gazdar, Klein, Pullum & Sag 85) (Crain and Fodor 85)

probabilistic (Suppes 70) (Smith 73) (Jurafsky 96)

in correspondence with semantics (Steedman 2000)

rule choice in a PCFG

0.87 `NP` $\rightarrow$ `the boy`
0.13 `NP` $\rightarrow$ `the tall boy`

- Particular rule choices are alternative *outcomes*.

- Nonterminal symbols are *random variables*

entropy of rule choice

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

0.87 NP → the boy
0.13 NP → the tall boy

$$H(\text{NP}) = -[(0.87 \times \log_2 0.87) + (0.12 \times \log_2 0.12)]$$
$$\approx 0.55\text{bits}$$

derivations

All PCFGs have a 'start' symbol $\mathtt{S}$.

$H_G(\mathtt{S})$ quantifies *how hard it is, on average, to guess a derivation in* G.

abbreviate $H_G(\mathtt{S}|w_{0\ldots i})$ by H_i

formalized claim

word reading time is linearly related to entropy reduction

$$\mathsf{RT}(w_i) = \alpha \left[\mathsf{reduction}\left(H_{i-1}, H_i\right)\right] + \beta$$

deriving predictions

the entropy of a nonterminal XP is the sum of
the rule choice entropy
and the expected entropy of XP's children

$$h(\mathrm{XP}) = -\sum_{\mathrm{XP}\to\mathrm{X}'\ \mathrm{ZP}\in\mathrm{rules}(\mathrm{XP})} p_{\mathrm{XP}\to\mathrm{X}'\ \mathrm{ZP}} \log_2 p_{\mathrm{XP}\to\mathrm{X}'\ \mathrm{ZP}}$$

$$H(\mathrm{XP}) = h(\mathrm{XP}) + \sum_{\mathrm{XP}\to\mathrm{X}'\ \mathrm{ZP}\in\mathrm{rules}(\mathrm{XP})} p_{\mathrm{XP}\to\mathrm{X}'\ \mathrm{ZP}} \left[H(\mathrm{X}') + H(\mathrm{ZP})\right]$$

(Grenander 67)

solution to recursion relation

Assuming a well-defined probability model
and letting A be the *expectation matrix*, then

$$\begin{aligned} H &= h + AH \\ &= (I - A)^{-1} h \end{aligned}$$

(Grenander 67)

entropy of parser state

grammar	$\equiv$	a complete set of grammatical derivations
parser state	$\equiv$	a complete set of *partial* grammatical derivations

recursion relation applies to developing top-down analyses

closed form solution applies to unexpanded nonterminals

left recursion

! There are infinite number of possible partial derivations involving left-recursive rules such as `NP` $\rightarrow$ `NP PostModifier`

$\surd$ transform left recursive PCFGs into right recursive PCFGs

(Huang and Fu 1971)

Main-verb/reduced-relative 1/6

key probabilities estimated using the Penn Treebank

1.00	S	→	NP VP
0.88	NP	→	DT NN
0.12	NP	→	NP VP
1.00	PP	→	IN NP
0.50	VP	→	V PP
0.50	VP	→	V
1.00	DT	→	the
0.50	NN	→	horse
0.50	NN	→	barn
0.50	V	→	fell
0.50	V	→	raced
1.00	IN	→	past

Main-verb/reduced-relative 2/6

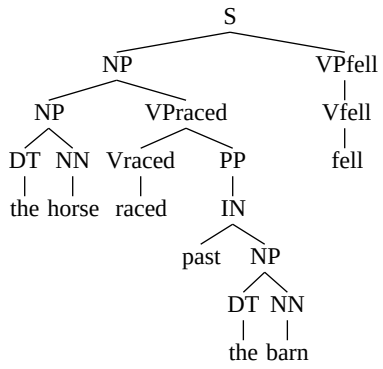
lexicalize so that "fell" cannot be a noun phrase modifier

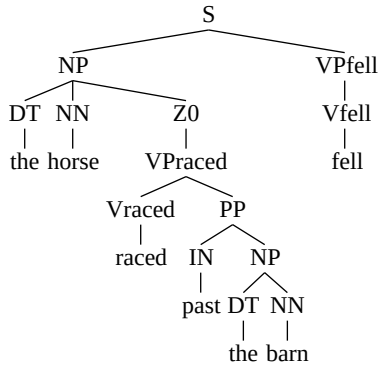
0.50	S	→	NP VPraced
0.50	S	→	NP VPfell
0.87	NP	→	DT NN
0.13	NP	→	NP VPraced
1.00	PP	→	IN NP
1.00	VPfell	→	Vfell
1.00	VPraced	→	Vraced PP
1.00	DT	→	the
0.50	NN	→	horse
0.50	NN	→	barn
1.00	Vfell	→	fell
1.00	Vraced	→	raced
0.50	IN	→	past
0.50	IN	→	in

Main-verb/reduced-relative 3/6

convert left recursion to right recursion;
add symbol Z0

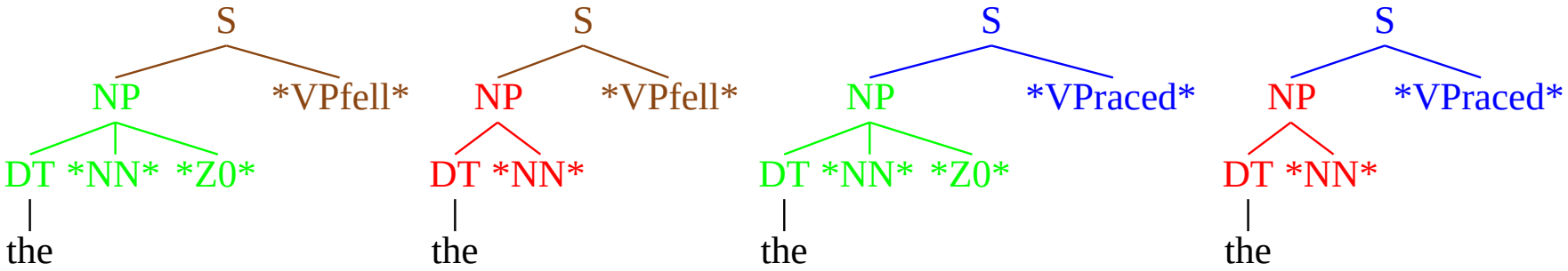
0.50	S	→	NP VPraced
0.50	S	→	NP VPfell
1.00	PP	→	IN NP
1.00	VPfell	→	Vfell
1.00	VPraced	→	Vraced PP
1.00	DT	→	the
0.50	NN	→	horse
0.50	NN	→	barn
1.00	Vfell	→	fell
1.00	Vraced	→	raced
0.50	IN	→	past
0.50	IN	→	in
0.87	NP	→	DT NN
0.13	NP	→	DT NN Z0
0.87	Z0	→	VPraced
0.13	Z0	→	VPraced Z0





Main-verb/reduced-relative 4/6

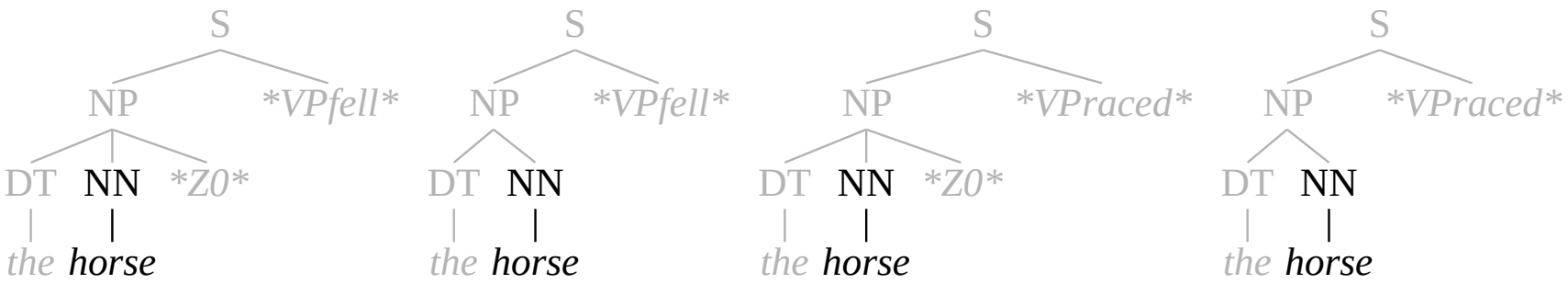
$$\{S \rightarrow NP\ VPraced, S \rightarrow NP\ VPfell\}$$
$$\times$$
$$\{NP \rightarrow DT\ NN, NP \rightarrow DT\ NN\ Z0\}$$



Main-verb/reduced-relative 5/6

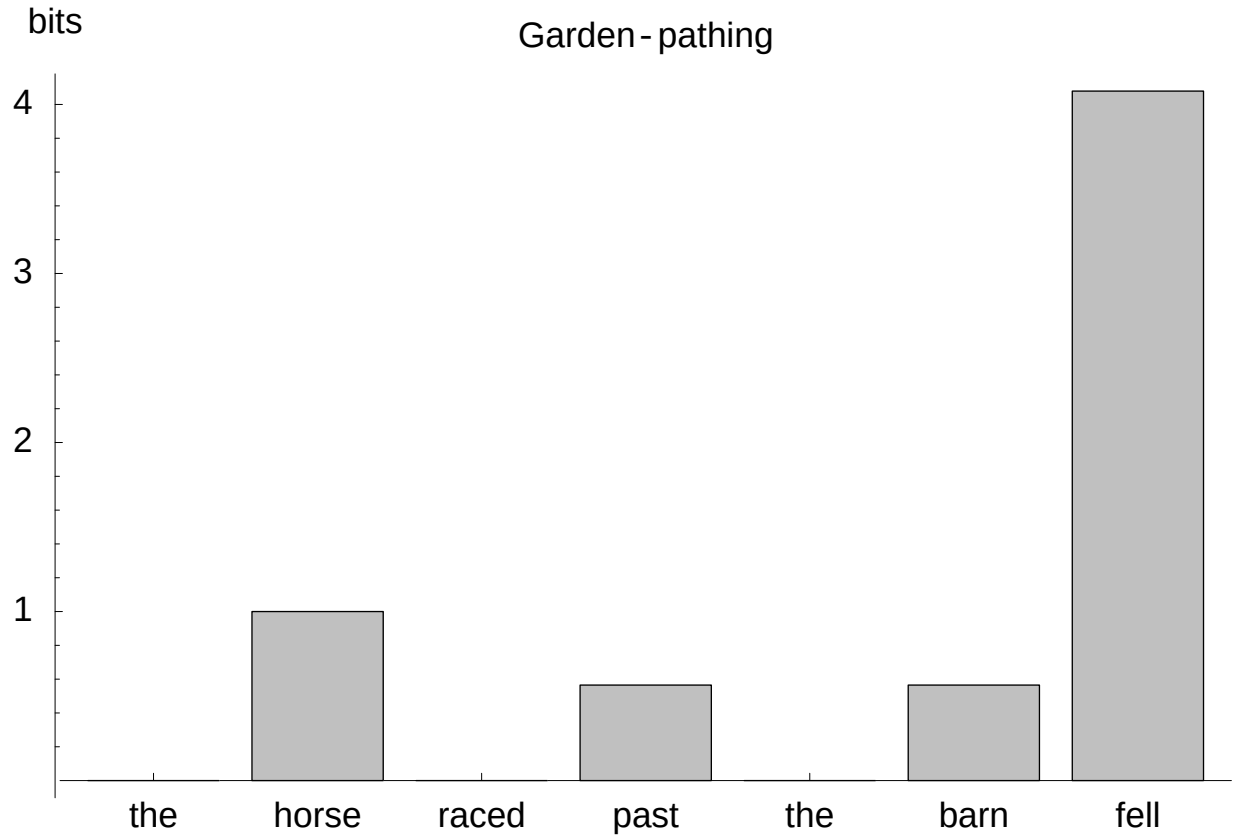
At "horse" the hearer gets 1 bit.

0.50 NN → horse
0.50 NN → barn



Main-verb/reduced relative 6/6

At "fell" the hearer gets 4 bits



NP/Z > *NP/S*

harder Before the woman visited the famous doctor...
$\rightsquigarrow$ had been drinking quite a lot
$\rightsquigarrow$ she verified that her health insurance was paid up

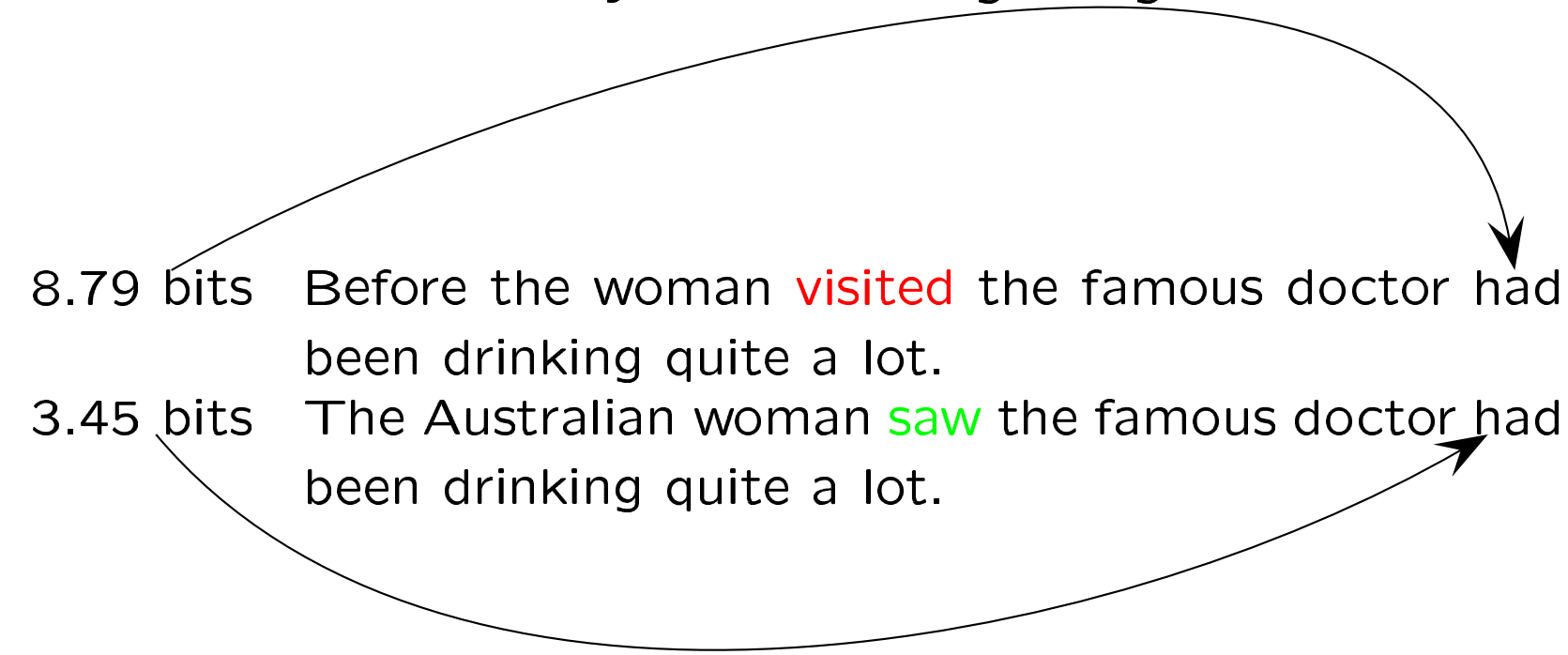
easier The Australian woman saw the famous doctor...
$\rightsquigarrow$ had been drinking quite a lot
$\rightsquigarrow$ about once a week after her bungee-jumping accident

The NP/Z ambiguity is harder than the NP/S ambiguity.

(Pritchett 88) (Sturt, Pickering and Crocker 99)

0.75	S	→	NP VP
0.25	S	→	PP SBAR
1.00	SBAR	→	NP VP
1.00	NP	→	SPECNP NBAR
1.00	SPECNP	→	DT
1.00	NBAR	→	N
0.25	VP	→	V[SUBCAT2] NP
0.25	VP	→	V[SUBCAT1]
0.25	VP	→	V[SUBCAT5] SBAR
0.25	VP	→	V[SUBCAT4,ASP] VBAR[PRP,COP]
1.00	VBAR[PRP,COP]	→	V[SUBCAT7,PRP,COP] VBAR[PRP]
1.00	VBAR[PRP]	→	V[SUBCAT2,PRP] ADVP
1.00	PP	→	PBAR SBAR
1.00	PBAR	→	P
1.00	P	→	before
1.00	ADVP	→	quite a lot
1.00	V[SUBCAT4,ASP]	→	had
1.00	V[SUBCAT7,PRP,COP]	→	been
1.00	V[SUBCAT2,PRP]	→	drinking
0.50	V[SUBCAT2]	→	visited
0.50	V[SUBCAT2]	→	saw
1.00	V[SUBCAT1]	→	visited
1.00	V[SUBCAT5]	→	saw
1.00	DT	→	the
0.50	ADJ	→	famous
0.50	ADJ	→	Australian
0.33	N	→	ADJ N
0.33	N	→	woman
0.33	N	→	doctor

more information conveyed at disambiguating word



8.79 bits Before the woman visited the famous doctor had been drinking quite a lot.

3.45 bits The Australian woman saw the famous doctor had been drinking quite a lot.

center-embedding

20.52 bits	the reporter disliked the editor
38.37 bits	the reporter $[_{S[+R]}$ who the senator attacked] disliked the editor
47.08 bits	the reporter $[_{S[+R]}$ who the senator $[_{S[+R]}$ who John met] attacked] disliked the editor

but

20.52 bits	the good good good good reporter disliked the editor

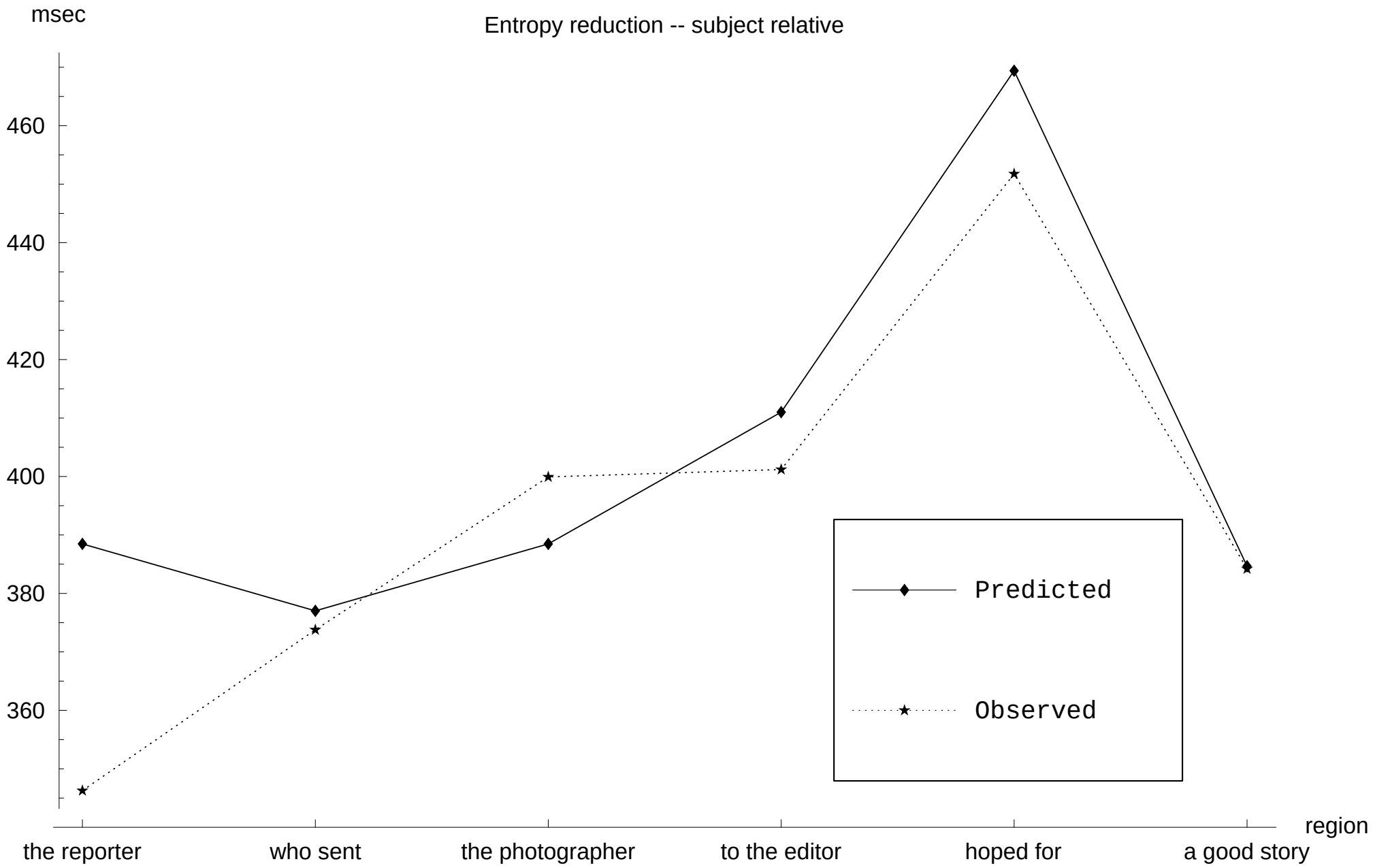
subject vs. object relative

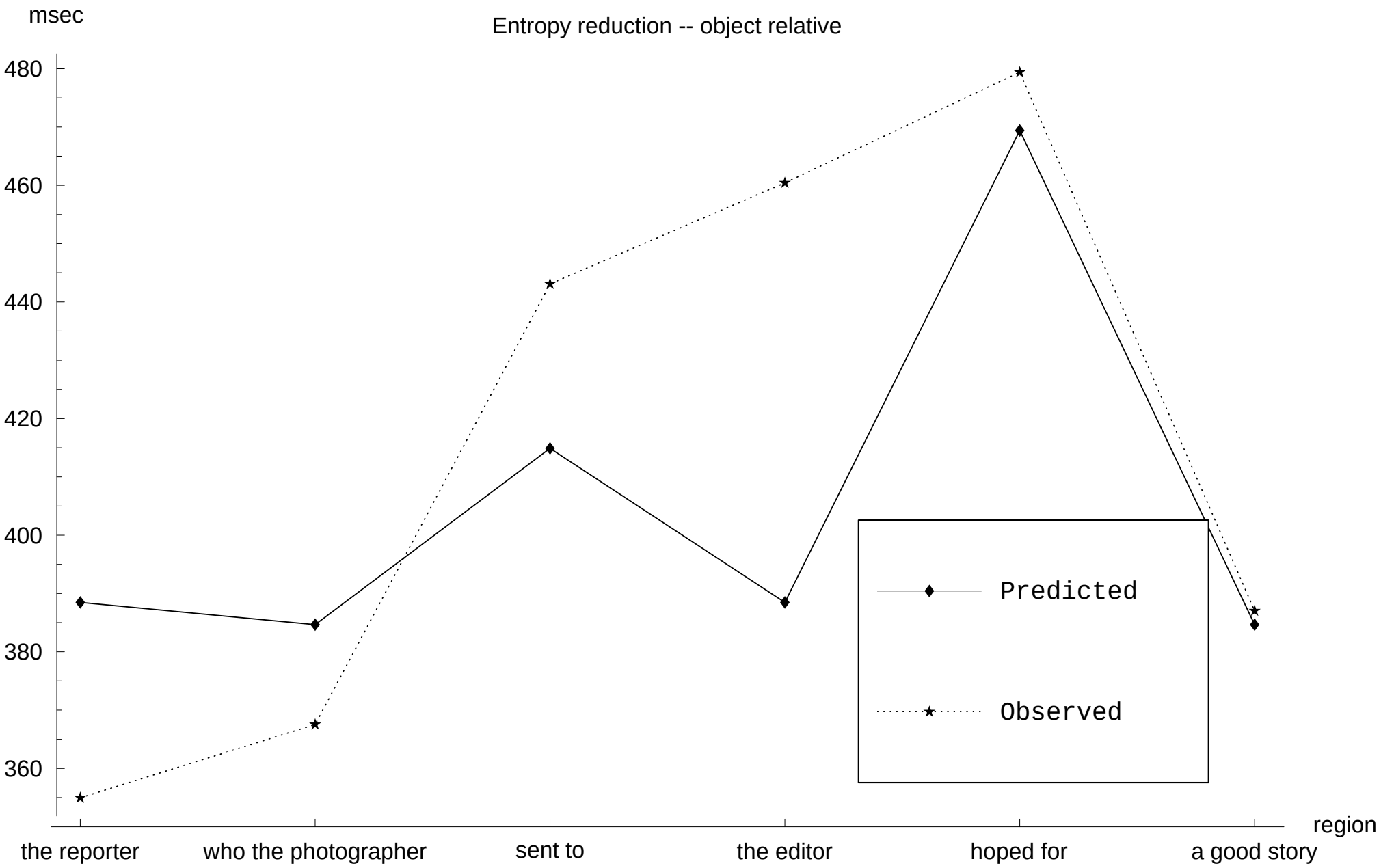
Subject relatives $<_{\text{easier}}$ object relatives

(Hawkins and Keenan 87)
(references in Gibson 98)

embedded verb RT↑ in object, but not subject

(Grodner, Watson and Gibson 2000)


msec
Entropy reduction -- subject relative
460
440
420
400
380
360
Predicted
Observed
region
the reporter
who sent
the photographer
to the editor
hoped for
a good story


msec
Entropy reduction -- object relative
480
460
440
420
400
380
360
Predicted
Observed
region
the reporter
who the photographer
sent to
the editor
hoped for
a good story

subject and object relative

$$\text{RT}(\textit{region}_{j\ldots k}) = \alpha\left(\text{average}_{i=j \text{ to } k}\left[\text{reduction}\left(H_{i-1}, H_i\right)\right]\right) + \beta$$

$\alpha = 7.38$
$\beta = 377$
$r^2 = 0.49, p < 0.01$

conclusion

Several types of sentence-processing phenomena

- main-verb/reduced relative temporary ambiguity
- noun phrase/zero vs. noun-phrase/sentence asymmetry
- center-embedding
- subject vs. object relative asymmetry

can be rationalized as instances of entropy reduction.