

# Structural Repetition in Dialogue?

Pat Healey,  
Matthew Purver,  
Christine Howes

Cognitive Science Research Group  
School of Electronic Engineering and Computer Science  
Queen Mary  
University of London

# Priming in Dialogue?

Convergence / Accommodation is the consequence of an “**automatic, resource-free priming mechanism** that underpins all successful human interaction”

(Pickering and Garrod, 2004).

“our analysis of dialogue demonstrates that **priming is the central mechanism** in the process of alignment and mutual understanding”

(Pickering and Garrod, 2004)

“*dialogue is **extremely repetitive***” and the evidence for structural alignment is “*extremely strong*” and “*ubiquitous*”

(Pickering and Ferreira, 2008)

# Question

How much do people repeat each others syntax in spontaneous face-to-face dialogue?

Existing work:

- focuses on specific constructions or subsets of constructions.
- doesn't always measure syntactic alignment independently of lexical alignment.

# Structural Repetition in Natural Dialogue?

Healey, Purver and Howes (2014):

- Compare real dialogues with artificially constructed 'chance' sequences of randomly paired turns.
- Calculate similarity of pairs of successive turns in Real vs. Chance.

# Real Dialogue

A: Are you going to go to all the day of the phonology lectures

B: I think I ought to do that

A: Yes. I think you had.Yeah

B: I mean I don't know how much I 'll take in

A: I think I 'll go to most of them. But I won't go to all of pragmatics the day before

# Baseline / Chance Dialogue

A: Are you going to go to all the day of the phonology lectures

A: Yes. I think you had.Yeah

A: I think I 'll go to most of them. But I won't go to all of pragmatics the day before

# Baseline / Chance Dialogue

A: Are you going to go to all the day of the phonology lectures

C: Yeah. Well I 'll write to him now

A: Yes. I think you had.Yeah

D: Uh do you remember the ones you took of Napoleon 's bedroom

A: I think I 'll go to most of them. But I won't go to all of pragmatics the day before

# Lexical Matching Calculation

Lexical similarity expressed via word pairs:

- number of matching word pairs between turns A and B =  $N_{AB}$
- similarity  $S_{lex} = \frac{N_{AB}}{\sqrt{N_{AA} \cdot N_{BB}}}$



# Syntactic Match Calculation

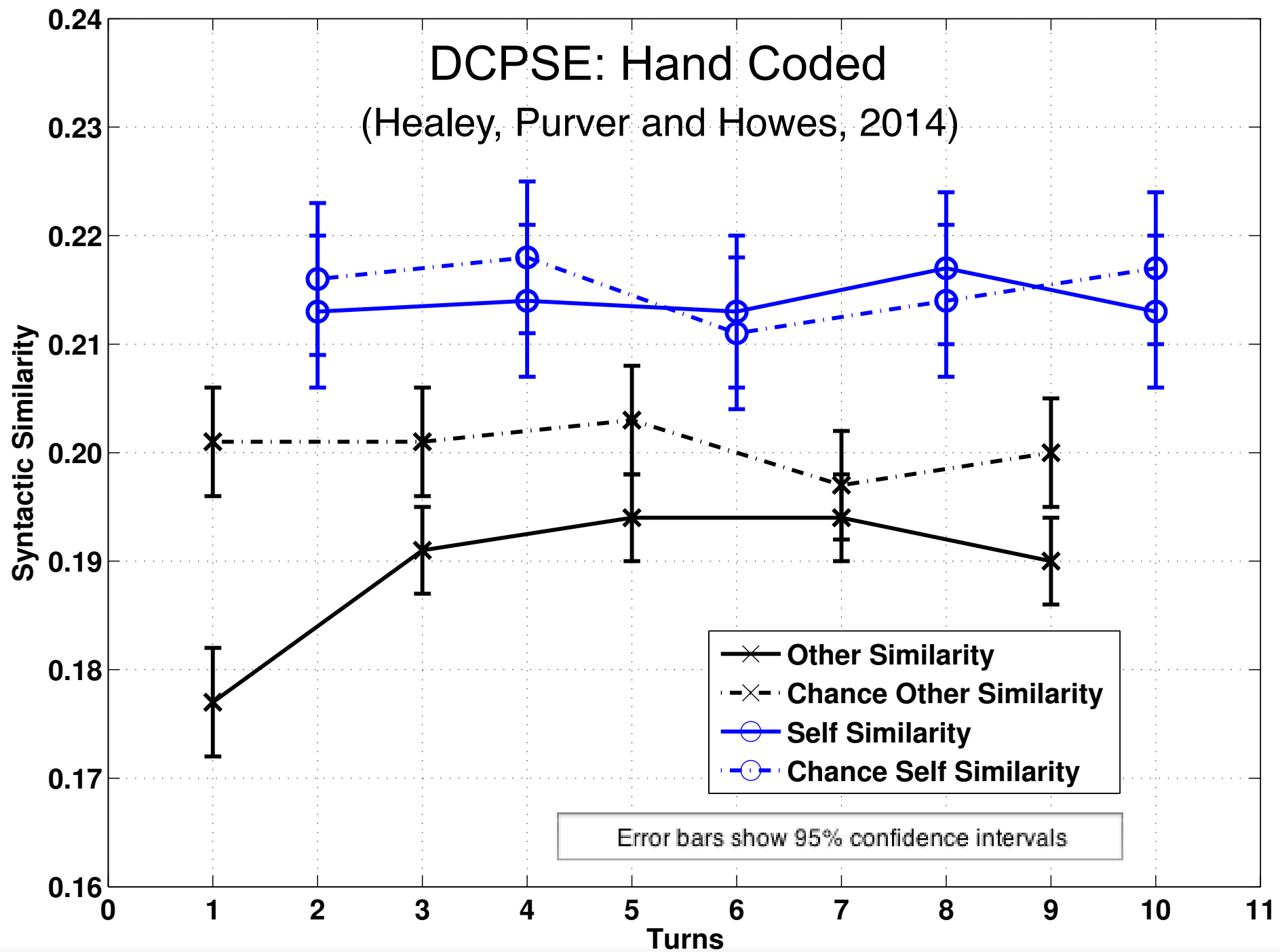
Syntactic similarity via tree kernel (variant of Moschitti, 2006):

- number of matching non-terminal syntactic rule pairs between turns A and B =  $N_{AB}$
- similarity  $S_{\text{syn}} = \frac{N_{AB}}{\sqrt{N_{AA} \cdot N_{BB}}}$

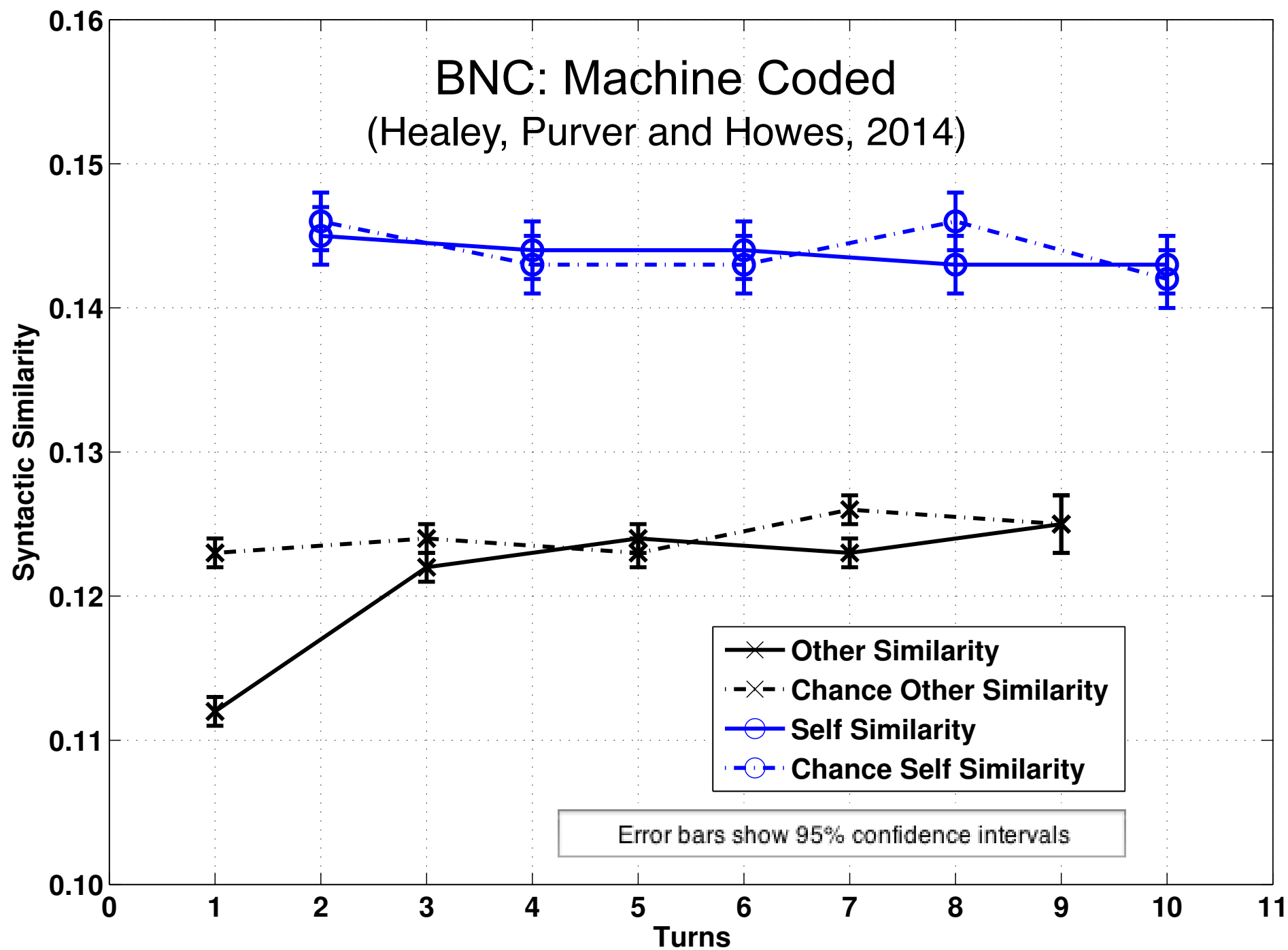
Self:	“turn 1”	Target (t)
Other:	“turn 2”	t+1
Self:	“turn 3”	t+2
Other:	“turn 4”	t+3
Self:	“turn 5”	t+4
Other:	“turn 6”	t+5
Self:	“turn 7”	t+6
Other:	“turn 8”	t+7
Self:	“turn 9”	t+8
Other:	“turn 10”	t+9
Self:	“turn 11”	t+10
Other:	“turn 12”	
Self:	“turn 13”	
Other:	“turn 14”	
Self:	“turn 15”	

# Resource-Free Priming Predictions:

1. **Cross-Speaker Priming:** repetition should be higher than chance.
2. **Cross-Level Priming** Lexical repetition increases syntactic repetition (and vice versa).
3. **Decaying Activation:** Levels of repetition should systematically decline with distance from the target.



# BNC: Machine Coded (Healey, Purver and Howes, 2014)



# Priming in Dialogue?

**‘Anti-Priming’ Effect:** other-repetition is *below* chance i.e. people *systematically diverge* in their use of syntactic constructions.

Why?

**Effects of Genre?:** Are ‘restricted’ dialogues inherently more repetitive?

**Rarity?** Low frequency constructions are more likely to be repeated?

# Corpora

## 1. **British National Corpus (BNC).**

Machine Coded Syntactic Trees.

Dyads only.

## 2. **Diachronic Corpus of Present-Day Spoken English (DCPSE)**

Hand Coded Syntactic Trees.

Dyads only.

## 3. **Switchboard Corpus (SWBD)**

Hand Coded Syntactic Trees.

Dyads only.

# Restricted vs. Unrestricted Dialogue

Are structural repetition effects specific to 'restricted' contexts?

- **BNC:** 'Educational' Dialogues vs. 'Demographic' Dialogues
- **DCPSE:** 'Broadcast Interviews' vs. 'Informal Conversations'
- **SWBD:** Partially restricted narratives on pre-defined topics.



# Average 'Other' Similarity Across Turns

Corpus	Real		Chance	
	Syntactic	Lexical	Syntactic	Lexical
<b>Switchboard</b>	0.21	0.07	0.28	0.07
<b>Broadcast Interview (DCPSE)</b>	0.37	0.12	0.36	0.08
<b>Unrestricted Dialogue (DCPSE)</b>	0.21	0.08	0.19	0.05
<b>Educational Dialogue (BNC)</b>	0.23	0.15	0.22	0.12
<b>Unrestricted Dialogue (BNC)</b>	0.12	0.07	0.12	0.04

# Results: Syntactic 'Other' Similarity

GLMM Analyses adjusted for lexical similarity:

Main Effects of Context: Restricted vs. Unrestricted

- (DCPSE:  $F_{(1;251)} = 152$ ;  $p < 0:001$ , BNC:  $F_{(1;780)} = 65:0$ ;  $p < 0:001$  )

Main Effect of Lexical Repetition

- (DCPSE:  $F_{(1;251)} = 102$ ;  $p < 0:001$ ; BNC:  $F_{(1;780)} = 173$ ;  $p < 0:001$ )

Main Effect of Conversation: Real vs. Chance

- (DCPSE:  $F_{(1;251)} = 20:7$ ;  $p < 0:001$  BNC:  $F_{(1;780)} = 5:04$ ;  $p = 0:02$  )

No Conversation X Genre Interaction

- (DCPSE:  $F_{(1;251)} = 0:23$ ;  $p = 0:63$ , BNC:  $F_{(1;780)} = 0:17$ ;  $p = 0:68$ )

# Structural Divergence

Systematic structural *divergence* in all three corpora:

Corpus	Real	Chance	p
<b>DCPSE</b>	0.26	0.28	p < 0:001
<b>BNC</b>	0.15	0.16	p < 0:02
<b>SWBD</b>	0.21	0.28	p < 0:001

Syntactic 'Other' Similarity: Estimated Marginal Means

# Rarity?

Low frequency constructions more likely to be repeated (Jaeger and Snider, 2013)?

- dative alternation most common and most reliable contrast in corpus and experimental studies.

Compare 10 most common syntactic rules (which account for 60% of observations) vs. all others.

# Effects of Rarity: Switchboard

## Main effects

- Lexical Repetition ( $F_{(1;5190)} = 841; p < 0:001$ )
- Conversation (Real vs. Chance:  $F_{(1;5190)} = 129; p < 0:001$ ).
- Frequency (Common vs. Rare) ( $F_{(1;5190)} = 58; 895; p < 0:001$ )

Conversation X Frequency interaction ( $F_{(1;5190)} = 154; p < 0:001$ )

	<b>Real</b>	<b>Chance</b>
Common Constructions	0.28	0.36
Rare Constructions	0.02	0.02

Estimated Marginal Means: Syntactic Other Similarity

# Effects of Rarity: DCSPE

## Main effects

- Lexical Repetition ( $F_{(1;503)} = 314$ ;  $p < 0:001$ )
- Conversation (Real vs. Chance:  $F_{(1;503)} = 8.93$ ;  $p = 0:003$ ).
- Frequency (Common vs. Rare) ( $F_{(1;503)} = 2,793$ ;  $p < 0:001$ )

Conversation X Frequency interaction ( $F_{(1;503)} = 6.14$ ;  $p = 0:014$ )

	<b>Real</b>	<b>Chance</b>
Common Constructions	0.26	0.34
Rare Constructions	0.04	0.04

Estimated Marginal Means: Syntactic Other Similarity

# Construction Types: Switchboard

Most Common Constructions	Least Common Constructions
NP:PRP	NP:NN:CD:CC:CD:CC:CD
S:NP:VP	VP:X:ADJP
INTJ:UH	S:INTJ:CC:PRN:NP:ADVP
PP:IN:NP	VP:ADVP:VB:NP:ADVP:NP
ADVP:RB	NP:NNP:CC:NNP:NNP
NP:DT:NN	SQ:INTJ:NP:PRN:VB:RB:NP:ADVP:VP
VP:VB:NP	SBARQ:RB:PRN:WHADVP:SQ
VP:VB	NP:DT:INTJ:JJ:JJ:NN:NN
S:VP	S:INTJ:NP:X:VP:VP
NP:NN	FRAG:PRN:CC:NP

# Construction Types: DCPSE

Most Common Constructions	Least Common Constructions
NP:PRON	PREDEL:AVP:AVP:VP:CL
VP:V	CL:CONNEC:AVP:PP:PP:NP:V P:NP:CL
AVP:ADV	CL:VP:PP:REACT
PP:PREP:NP	CL:CONNEC:NP:NP:VP:PP
DTP:ART	CL:SUBP:VP:NP:PP
NP:DTP:N	CL:CONNEC:NP:VP:NP:AVP:C L:PP
NP:N	CL:NP:VP:CL:NP:CL:CL
VP:AUX:V	NP:DTP:AJP:N:CL:PP
AJP:ADJ	NP:DTP:N:CL:NP
DTP:PRON	CL:INTERJEC:NP:NP:VP:NP



# Conclusions:

- Restricted contexts promote repetition.
- Rare constructions appear more likely to be repeated.
- But dialogue is **not** repetitive
  - people repeat some of each others words but in systematically different syntactic contexts.
    - to pose and answer questions
    - to make and evaluate proposals
    - to complain and remedy
    - to clarify and repair
  - .....to move conversation forward.

Priming processes are overwhelmed by the active, constructive engagement required for ordinary dialogue.