Cross-linguistic variation in a grammaticality illusion

Nick Huang

National University of Singapore

15 February 2022

Native speakers' intuitions about (un)grammaticality

Native speakers can usually easily tell whether a sentence is a grammatical sentence of their language.

1. The **key is** on the table. OK
2. *The **key are** on the table. not OK

But for certain ungrammatical sentences, speakers seem much less aware

1. The key is on the table. OK
2. *The key are on the table. not OK
3. ?The key to the cabinets are on the table …

Native speakers of English frequently …

- Rate (3) as relatively acceptable in speeded judgment tasks
- Fail to detect the agreement mismatch in (3) while reading
- Require more time and attention to notice the agreement error in (3).

→ **“Grammaticality illusion”**

Illusions as a window to how the parser works

Illusions are cases of **parser failure**.

- What kind of parsing mechanisms are involved? Why do they fail?
- Why do we get a perception of relative acceptability?

Today: a puzzle involving **cross-linguistic differences** in an illusion.

- Why does the parser fail in some languages but not others?

→An opportunity to learn **why the parser generally succeeds across languages.**

Outline for today's talk

1. A grammaticality illusion involving double centre-embedding sentences
- Cross-linguistic variation in the illusion: *Why*?

2. Introduce a Mandarin "Missing NP" illusion (joint work with Colin Phillips)

3. Experiments and implications for theories of this illusion
- A similarity-based interference approach

4. Future directions

What is centre-embedding?

Single centre-embedding	The patient [$_{RC}$ the nurse saw] called.
Double centre-embedding	The patient [$_{RC}$ the nurse [$_{RC}$ the clinic hired] saw] called. **Grammatical (but difficult to understand)**

Classic case of grammaticality-acceptability mismatch.
(Chomsky & Miller 1963)

Typical explanation: centre-embedding requires us to maintain three incomplete clauses in memory – **too demanding**.

What is centre-embedding?

Double centre-embedding	The patient [$_{RC}$ the nurse [$_{RC}$ the clinic hired] saw] called. **Grammatical (but difficult to understand)**
Not double centre-embedding	The clinic hired the nurse who saw the patient who called. **Grammatical and semantically identical** **Much easier to understand**

Another way to improve acceptability: Make the sentence ungrammatical by deleting a VP

Double centre-embedding	The patient [$_{RC}$ the nurse [$_{RC}$ the clinic hired] saw] called. **Grammatical (but difficult to understand)**
"Missing VP"	The patient [$_{RC}$ the nurse [$_{RC}$ the clinic hired]] called. **Ungrammatical**

Ungrammatical missing VP sentences are surprisingly OK in English (& French, Spanish)

Double centre-embedding	The patient [$_{RC}$ the nurse [$_{RC}$ the clinic hired] saw] called. **Grammatical (but difficult to understand)**
"Missing VP"	The patient [$_{RC}$ the nurse [$_{RC}$ the clinic hired]] called. **Ungrammatical** **but relatively easy to process (rated more highly, read faster)** **(Grammaticality illusion)**

Frazier 1985 (citing Janet Fodor); Gibson & Thomas 1999; Frank et al. 2016; Frank & Ernst 2019; Huang & Phillips 2021; Gimenes et al. 2009 (French); Paneda & Lago 2020 (Spanish)

But not in German (or Dutch)

German examples from Vasishth et al. 2010

Double centre-embedding

Der Anwalt, [$_{RC}$ den der Zeuge, [$_{RC}$ den der Spion betrachtete], schnitt], überzeugte den Richter.
the lawyer who the witness who the spy watched avoided convinced the judge

Grammatical (Easier to process)

"Missing VP"

Der Anwalt, [$_{RC}$ den der Zeuge, [$_{RC}$ den der Spion betrachtete],], überzeugte den Richter.
the lawyer who the witness who the spy watched convinced the judge

Ungrammatical

(harder to process; error detected in online reading – no illusion)

Vasishth et al. 2010; Frank et al. 2016; Frank & Ernst 2019; but see Bader 2016, Haeussler & Bader 2015

But why cross-linguistic variation?

Null hypothesis: parsing mechanisms and memory resources are the same across languages.

The German/Dutch/French/English sentences have the **same word order**.

Our default prediction is **cross-linguistic uniformity**.

“Language statistics” accounts of cross-linguistic variation in the illusion

German/Dutch/French/English vary more generally in their word order.

- Differences in word order distributions affect parsing.

- Vasishth et al. (2010): differences in statistics help German speakers ***maintain predictions in memory more easily***.
- Frank et al. (2016): greater familiarity with certain constructions → ***easier processing***
- Futrell et al. (2020): distributional differences **affects *predictions of upcoming structures***

Two features of double centre-embedding sentences

1. Verb sequences

The patient [$_{RC}$ the nurse [$_{RC}$ the clinic **hired**] **saw**] **called**.

2. Long dependencies

The patient [$_{RC}$ the nurse [$_{RC}$ the clinic **hired**] **saw**] **called**.

Relative frequency of verb sequences and long dependencies in English vs. German

	English	**German**
Verb sequences	Infrequent	Frequent

English: “They have **met** the man who **loves** Kim.”
(Pseudo-)German: “They have the man who Kim **loves met**.”

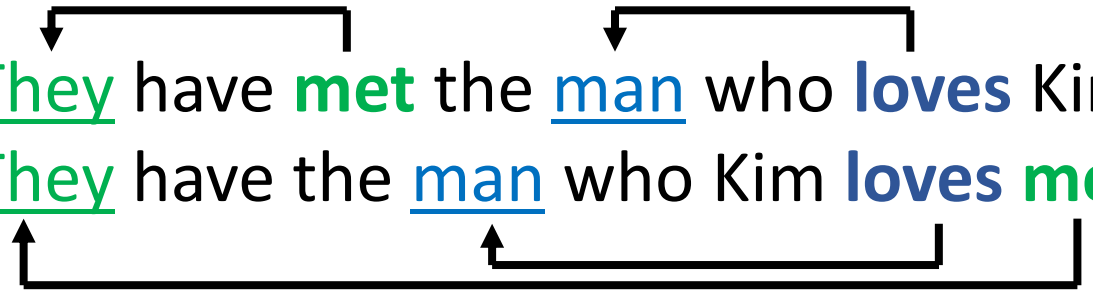
Sequence of 2 verbs

Relative frequency of verb sequences and long dependencies in English vs. German

	English	German
Verb sequences	Infrequent	Frequent
Long dependencies	Infrequent	Frequent

English: “They have **met** the man who **loves** Kim.”

(Pseudo-)German: “They have the man who Kim **loves** **met**.”



Lang. statistics accounts: German speakers are more "prepared" for processing double centre-embedding

	English	German
Verb sequences	Infrequent	Frequent
Long dependencies	Infrequent	Frequent

1. Verb sequences

The patient [$_{RC}$ the nurse [$_{RC}$ the clinic **hired**] **saw**] **called**.

2. Long dependencies

The patient [$_{RC}$ the nurse [$_{RC}$ the clinic **hired**] **saw**] **called**.

Vasishth et al. 2010; Frank et al. 2016; Frank & Ernst 2019; but see Bader 2016, Haeussler & Bader 2015

The language statistics account makes a testable prediction for Mandarin Chinese

- But first, some background on Chinese word order and center-embedding

Two key word order facts of Mandarin Chinese

Fact 1: SVO language

S	V	O
Zongli	jiejianle	buzhang.
PM	met	minister

Fact 2: Modifiers always come before nouns (*de* is a particle, not unlike English *who/that*)

		V	NP		NP
[$_{RC}$	ceng	zebeiguo	faguan	de]	[$_{Noun}$ buzhang]
	previously	rebuked	judge	DE	minister

"the minister who previously rebuked the judge" (subject relative clause)

Centre-embedding in Mandarin

Single centre-embedding

V V NP NP

Zongli jiejianle [$_{RC}$ ceng zebeiguo faguan de] buzhang.

PM met previously rebuked judge DE minister

“The prime minister met the minister who previously rebuked the judge.”

Double centre-embedding

V V V NP NP NP

Zongli jiejianle [$_{RC}$ ceng zebeiguo [$_{RC}$ gang jiefa tanwu-an de] faguan de] buzhang.

PM met previously rebuked just expose corruption-case DE judge minister

“*The prime minister met the minister who previously rebuked ___ exposed the corruption case.”

Mandarin double centre-embedding

Grammatical double centre-embedding (DCE)

	V		V		V	NP		NP		NP
Zongli	jiejianle	[$_{RC}$ ceng	zebeiguo	[$_{RC}$ gang	shenli	tanwu-an	bujiu	de]	faguan haojici	de] buzhang.
PM	met	previously	rebuked	just	hear	corruption-case	recently	DE	judge a.few.times	DE minister

"The prime minister met the minister who previously rebuked a few times the judge who just recently exposed the corruption case."

Ungrammatical Missing NP sentence

	V		V		V	NP			NP
Zongli	jiejianle	[$_{RC}$ ceng	zebeiguo	[$_{RC}$ gang	shenli	tanwu-an	bujiu	de]	] buzhang.
PM	met	previously	rebuked	just	hear	corruption-case	recently	DE	minister

"*The prime minister met the minister who previously rebuked ___ just recently heard the corruption case."

Are Mandarin grammatical sentences easier to process than the ungrammatical variant, as in German?

Huang & Phillips, 2021

Language statistics accounts predict Mandarin grammatical DCE should be easier

1. **Modifier-*de*-Noun sequences common**

Zongli jiejianle [$_{RC}$ ceng zebeiguo faguan haojici de] buzhang.

PM met previously rebuked judge a.few.times DE minister

“The prime minister met the minister who rebuked the judge a few times.”

Mandarin double centre-embedding sentences also have these sequences.

[$_{RC}$ ceng zebeiguo [$_{RC}$ gang shenli tanwu-an bujiu de] faguan haojici de] buzhang.

PM met previously rebuked just hear corruption-case recently DE judge a.few.times DE minister

“The prime minister met the minister who previously rebuked a few times the judge who just recently heard the corruption case.”

Language statistics accounts predict Mandarin grammatical DCE should be easier

2. Mandarin has long **verbal dependencies.**

Zongli **jiejianle** [$_{RC}$ ceng **zebeiguo** faguan haojici de] buzhang.

PM met previously rebuked judge a.few.times DE minister

"The prime minister **met** the minister who **rebuked** the judge a few times."

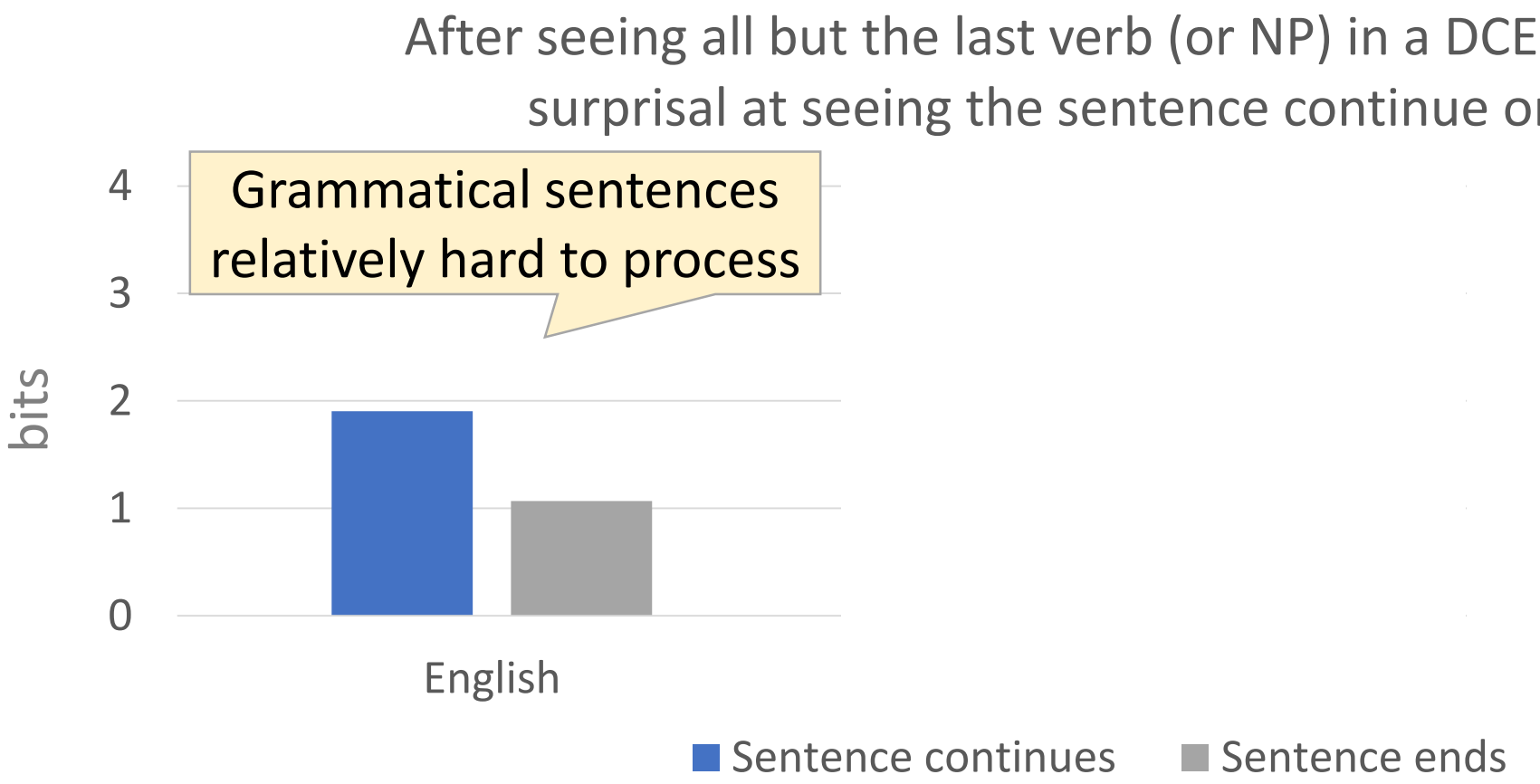
Mandarin double centre-embedding sentences have long verbal dependencies.

Zongli jiejianle [$_{RC}$ ceng zebeiguo [$_{RC}$ gang shenli tanwu-an bujiu de] faguan haojici de] buzhang.

PM met previously rebuked just hear corruption-case recently DE judge a.few.times DE minister

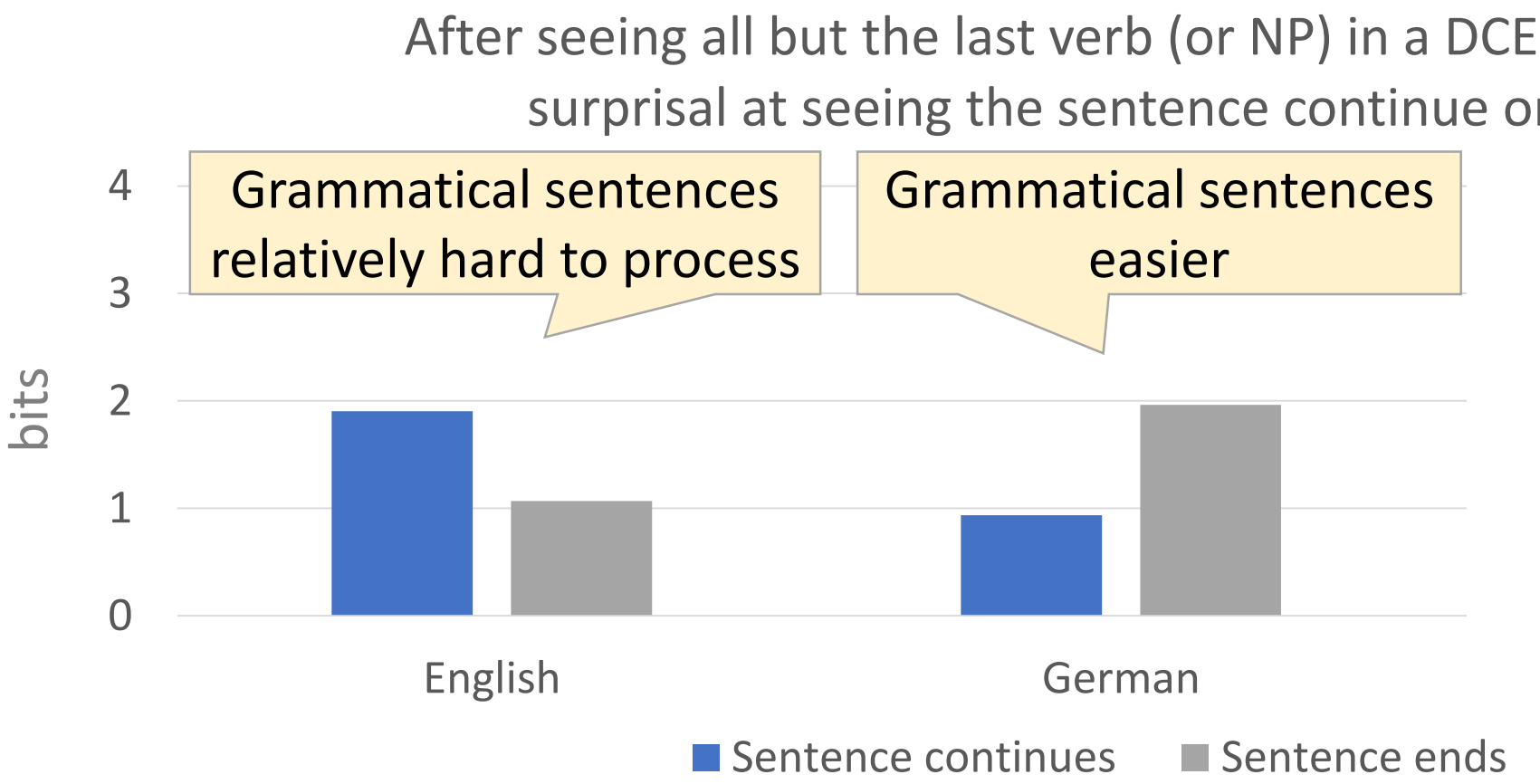
"The prime minister met the minister who previously rebuked a few times the judge who just recently heard the corruption case."

A computational model of language statistics accounts



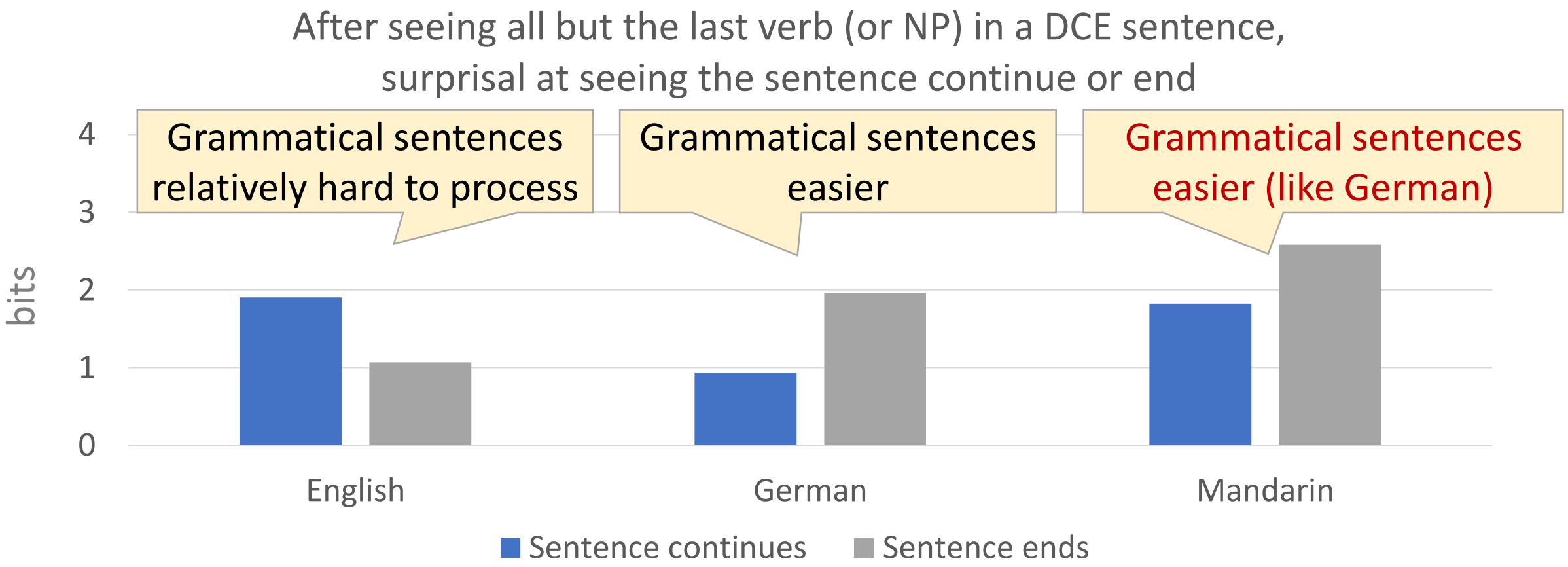
Huang & Phillips 2021, Exp. 4, also see Futrell et al. 2020, Futrell & Levy 2017

A computational model of language statistics accounts



Huang & Phillips 2021, Exp. 4, also see Futrell et al. 2020, Futrell & Levy 2017

A computational model of language statistics accounts



Huang & Phillips 2021, Exp. 4, also see Futrell et al. 2020, Futrell & Levy 2017

For thematic reasons, ungrammatical Mandarin Missing NP sentences should be worse

English Missing VP sentences (baseline case)

NP	NP	NP	V	V
The patient [$_{RC}$	the nurse [$_{RC}$	the clinic	hired]	] called.

- 1 transitive verb (*hired*) and 1 intransitive verb (*called*): **3 arguments needed.**
- There are **3 NPs present.**

For thematic reasons, ungrammatical Mandarin Missing NP sentences should be worse

Mandarin Missing NP sentence

	V		V		V	NP			NP
Zongli	jiejianle	[$_{RC}$ ceng	zebeiguo	[$_{RC}$ gang	shenli	tanwu-an	bujiu	de]	] buzhang.
PM	met	previously	rebuked	just	hear	corruption-case	recently	DE	minister

"*The prime minister met the minister who previously rebuked ___ just recently heard the corruption case."

- 3 verbs, all transitive:
  6 arguments needed.
- There are **only 3 NPs present**, including the subject.

→ Missing NP sentences don't have enough arguments to go around.

Interim summary

Mandarin Chinese double centre-embedding + Missing NP sentences ≈ English/German double centre-embedding + Missing VP sentences.

Predictions for Mandarin

1. Language statistics accounts: Mandarin should pattern like German (also Dutch).
2. From an thematic relations perspective, Mandarin Missing NP sentences are incomplete.

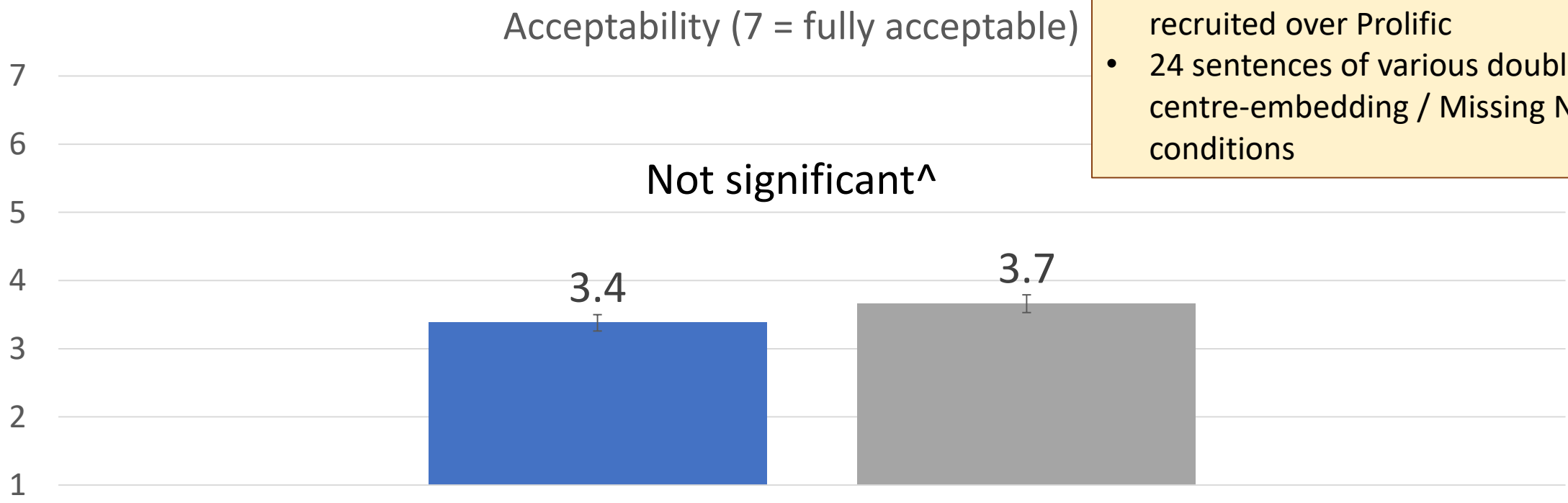
→**Grammatical double centre-embedding should be easier to process, i.e. no illusion in Mandarin.**

Double centre-embedding sentences not easier to process

Experiment 1 (acceptability judgments)

Experiment details

- 60 native speakers of Mandarin recruited over Prolific
- 24 sentences of various double centre-embedding / Missing NP conditions



Illusion exists in Mandarin (like English, not German).

Huang & Phillips 2021, Exp. 1

Maybe the illusion is an artifact of the materials

Mandarin Missing NP sentence

	V		V		V	NP				NP
Zongli	jiejianle	[$_{RC}$ ceng		zebeiguo	[$_{RC}$ gang	shenli	tanwu-an	bujiu	de]	] buzhang.
PM	met	previously		rebuked	just	hear	corruption-case	recently	DE	minister

"*The prime minister met the minister who previously rebuked ___ just recently heard the corruption case."

Maybe *jiejianle* "met" and *zebeiguo* "rebuked" were analyzed as **conjoined**?

- "The prime minister **met and previously rebuked** the minister …"

Experiment 1a: blocking a conjunction reading in Missing NP sentences

Mandarin Missing NP sentence

V V V NP NP

Zongli jiejianle [$_{RC}$ ceng zebeiguo [$_{RC}$ gang shenli tanwu-an bujiu de]] buzhang.

PM met previously rebuked just hear corruption-case recently DE minister

"*The prime minister met the minister who previously rebuked ___ just recently heard the corruption case."

Add **demonstrative articles + classifiers between the verbs.**

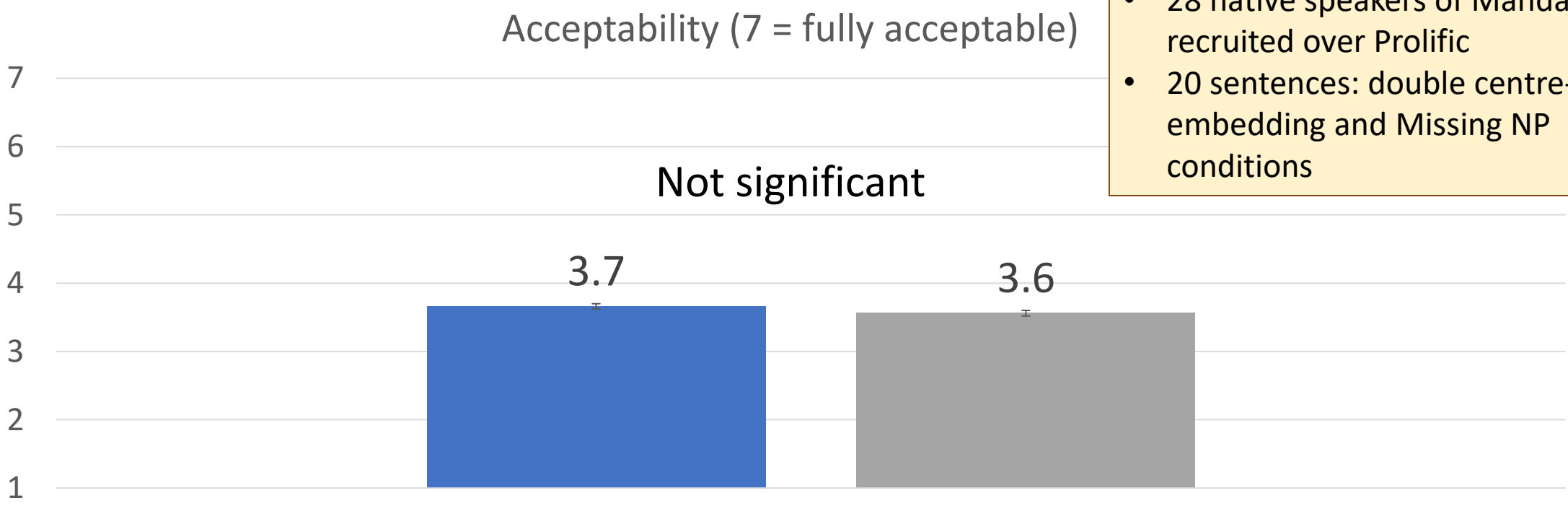
- Signals clearly that the material after the verb is an object NP.

Double centre-embedding sentences not easier to process

Experiment 1a (acceptability judgments)

Experiment details

- 28 native speakers of Mandarin recruited over Prolific
- 20 sentences: double centre-embedding and Missing NP conditions



Grammatical DCE sentences still not easier to process.

Maybe acceptability ratings are misleading / provide an incomplete picture?

Previous work on the illusion:

- German/Dutch speakers detect the ungrammaticality (incompleteness) of Missing VP sentences in **real-time ("online") reading**.
- English speakers do not.

Vasishth et al. 2010; Frank et al. 2016

Maybe acceptability ratings are misleading / provide an incomplete picture?

A scenario:

Mandarin speakers are like German/Dutch speakers in real-time processing: they can detect the ungrammaticality (incompleteness) of Missing NP sentences.

However, Mandarin speakers might just find it easier to repair Missing NP sentences.
→ **Relatively high acceptability ratings.**

Do Mandarin speakers detect the ungrammaticality of Missing NP sentences in real time?

Experiment 2 (self-paced reading)

															NP1	
Zongli	daodi	jiejianle	na-ming	[$_{RC}$	ceng		xiezhuguo	zhe-ge	[$_{RC}$	gang	jiefa		tanwu-an	
PM	actually	met	that-CL		previously		helped	this-CL		just	expose		corruption-case	
ADV1	*de*	*NP2*	*ADV2*		*de*	*NP3*	*or*	*or+1*		*or+2*				
bujiu	de]	lüshi	yi-liang-ci		de]	yiyuan	haishi	zhe-ming		xiaoxinjinshen	de	dashi?		
recently	DE	lawyer	one-two-time		DE	legislator	or	this-CL		cautious		DE	ambassador	

"Did the prime minister meet that legislator who … or this cautious ambassador?"

Double centre-embedding (Missing NP) structure

Spillover region (after detecting ungrammaticality, people usually slow down)

If Mandarin speakers can detect the ungrammaticality of Missing NP sentences…

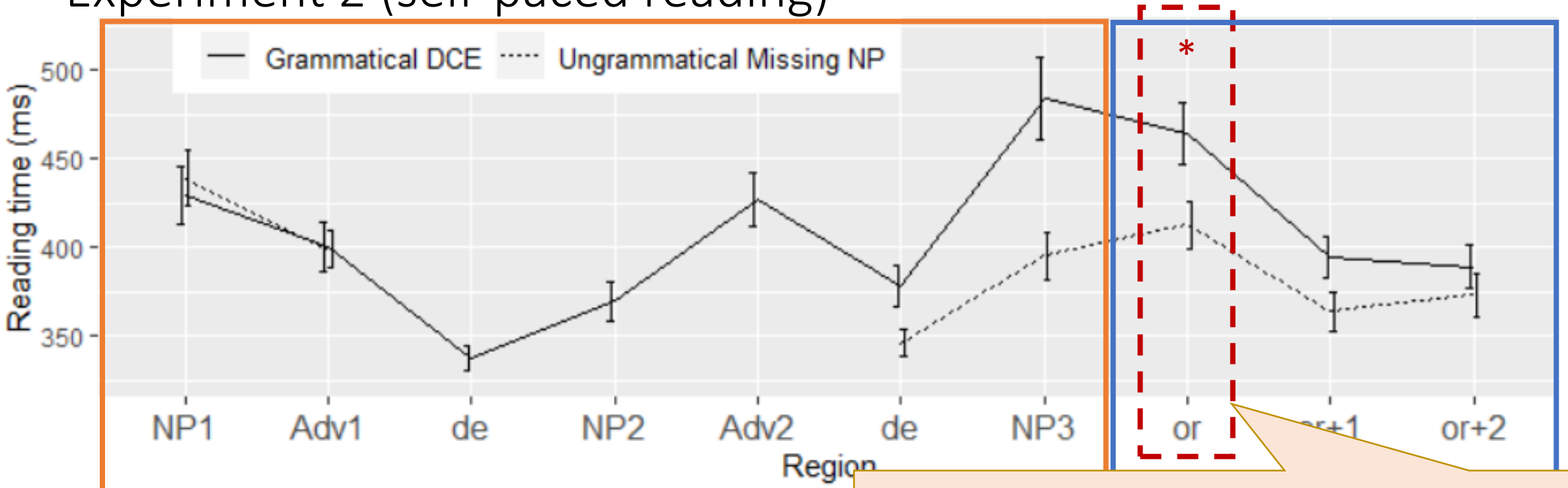
"Did the prime minister meet that ...NP ADV de NP

Missing NP structure

Missing NP condition:
Only 2 NPs present –
structure is still incomplete.

But speakers don't detect the ungrammaticality of Missing NP sentences in real time

Experiment 2 (self-paced reading)



Grammatical / Missing NP struct

Experiment details

- 40 native speakers Mandarin speakers on Prolific
- 20 DCE/Missing NP sentences

No slowdown for Missing NP sentences.
In fact, slowdown seen for grammatical DCE sentences → harder to process

Mandarin has a centre-embedding illusion, like English, unlike German

Grammatical sentences are **not** easier to process.

- Contrary to language statistics accounts.

In fact, Mandarin speakers don't seem to realize immediately that Missing NP sentences are ungrammatical.

An addition to our understanding of cross-linguistic variation in the illusion.

- Mandarin: a typologically-different language, with rather different word order.

But why are Missing NP sentences relatively acceptable?

Missing NP sentences should be thematically incomplete – more verbs than arguments.

Shouldn’t that affect processing or acceptability?

Missing NP sentences have 3 transitive verbs. Do they have enough NP arguments?

Mandarin Missing NP sentence

	V		V		V	NP			NP
Zongli	jiejianle	[$_{RC}$ ceng	zebeiguo	[$_{RC}$ gang	shenli	tanwu-an	bujiu	de]	] buzhang.
PM	met	previously	rebuked	just	hear	corruption-case	recently	DE	minister

- Does “met” have a theme argument (“object”)?
- Does “rebuked” have agent and theme arguments (“subject”, “object”)?
- Does “hear” have an agent argument?

Experiment 3: what verb-NP pairings are there?

Mandarin Missing NP sentence

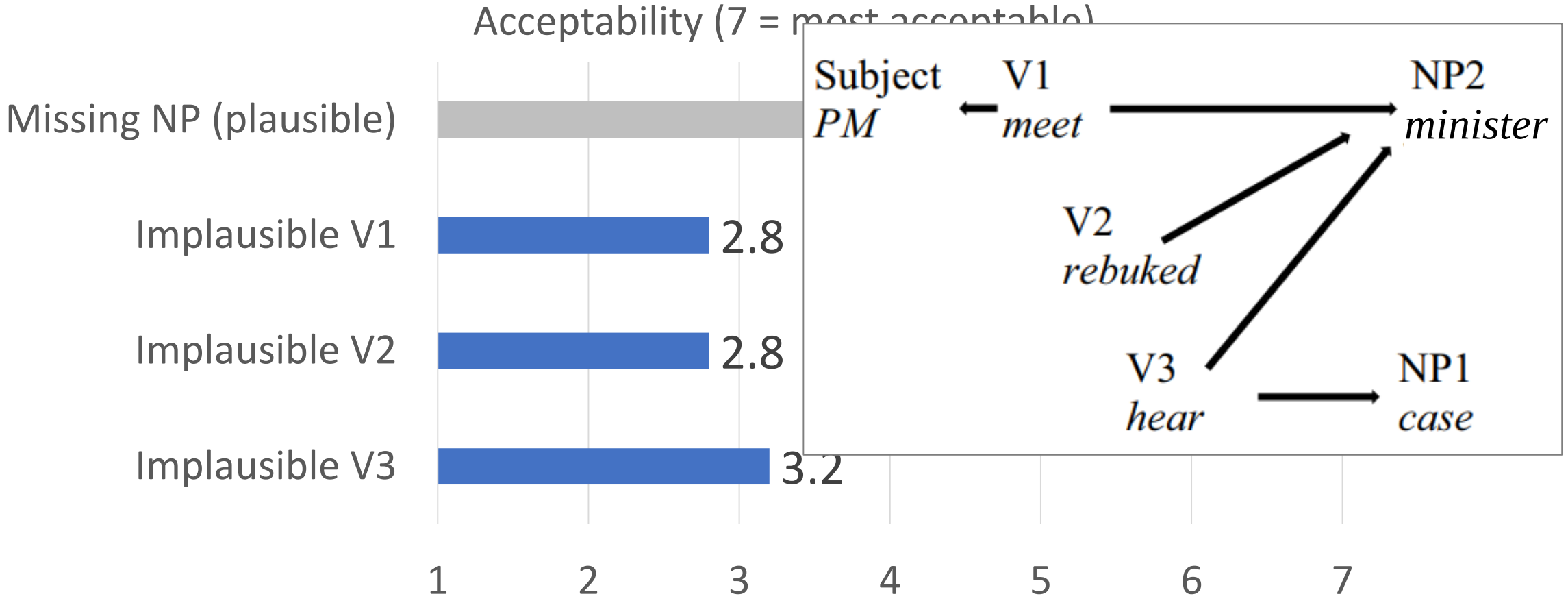
	V		V		V	NP			NP
Zongli	jiejianle	[$_{RC}$ ceng	zebeiguo	[$_{RC}$ gang	shenli	tanwu-an	bujiu	de]	] buzhang.
PM	met	previously	rebuked	just	hear	corruption-case	recently	DE	minister

E.g. Assume that “rebuked” does have a theme argument – “minister”.

- **How can we tell?** Change the “rebuked” to “reorganized”
- “Reorganized” requires an inanimate theme argument.
- The pairing “reorganized … the minister” is semantically odd: **lower acceptability**.

All verbs are linked thematically to NP2 “minister”

Experiment 3 (acceptability judgments)



Huang & Phillips 2021, Exp. 1

Interim summary

Grammatical double centre-embedding (DCE) sentences are distinctly easier to process in German, but **not** in English or Mandarin.

Acceptability ratings suggest that in Mandarin Missing NP sentences, all the verbs are linked to some noun: no verb is truly without an argument.

- Would explain the relative acceptability of Missing NP sentences.

Discussion: why cross-linguistic variation?

Intuition: Double centre-embedding sentences in German and English have the same word order but are structurally very different.

- German: main clause subject is in a topic-like position, and main verb is in V2 position.
- English: main clause subject is not topic-like, and no V2.

These abstract syntactic differences have psycholinguistic consequences.

Bader 2016

Our proposal: a similarity-based interference account

We incorporate this observation in a **cue-based retrieval** parsing account.

When parsing sentences, the parser builds a detailed syntactic representation.

German

				V		**V**		**V**	
Der Anwalt, [$_{RC}$	den	der Zeuge, [$_{RC}$	den	der Spion	**betrachtete**],	**schnitt**],	**überzeugte**	den Richter.
the lawyer		who the witness		who the spy	watched	avoided	convinced	the judge

Main clause verb position is structurally distinct

→ **greater parsing accuracy, sentence easier to process**

English

				V	**V**	**V**		
The patient [$_{RC}$	the nurse [$_{RC}$	the clinic	**hired**]	**saw**]	**called**.			

Main clause verb position is not structurally distinct

→ **parsing error more likely, sentence harder to process**

Extending the account to Mandarin

In Mandarin, the analogue to the English/German verbs is the NPs.

But Mandarin doesn't syntactically distinguish NPs inside main and relative clauses – **analogous to English**, not German.

							NP				**NP**			**NP**
Zongli	jiejianle	[$_{RC}$	ceng	zebeiguo	[$_{RC}$	gang	shenli	**tanwu-an**	bujiu	de]	**faguan**	haojici	de]	**buzhang**.
PM	met		previously	rebuked		just	hear	**corruption-case**	recently	DE	**judge**	a.few.times	DE	**minister**

"The prime minister met the minister who previously rebuked a few times the judge who just recently exposed the corruption case."

NP positions are not structurally distinct →
parsing error more likely, sentence not easier to process

Further implications

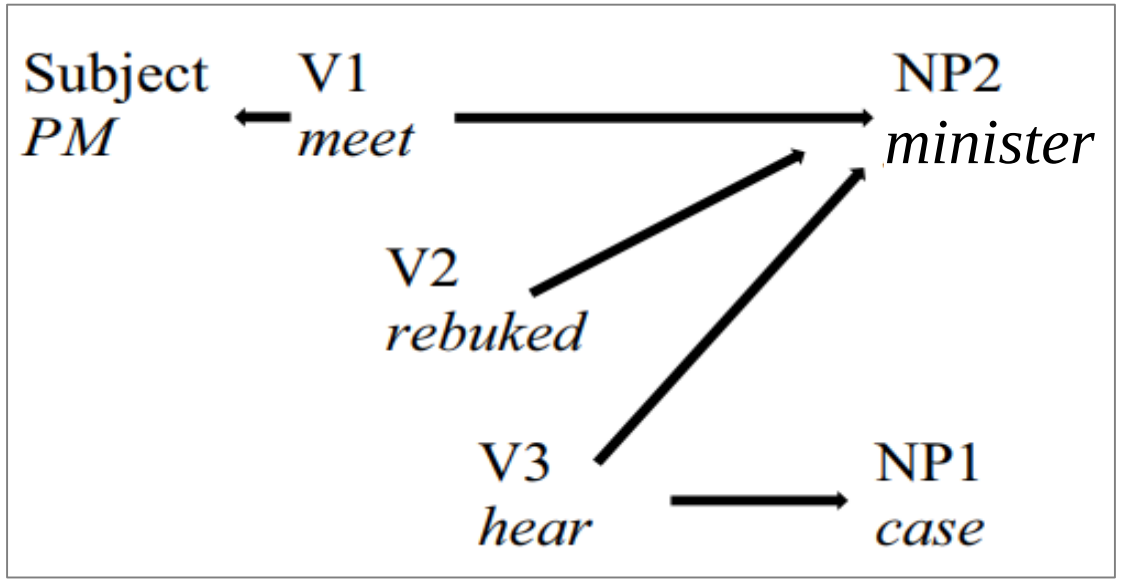
The parser builds detailed syntactic representations consistent with the grammar of a language.

- German: V2 structure; English: no V2 structure
- Mandarin: no special structure for NP objects
- Contrast with an alternative scenario: speakers only build shallow representations (e.g. a flat sequence of nouns and verbs), given the complexity of these sentences.

Grammatical differences
→ Representations are syntactically different
→ Cross-linguistic differences in parsing outcomes

Future research

Acceptability ratings have shown how Missing NP sentences are interpreted **offline**.



- How are these thematic relations actually constructed in real time?
  How are Missing NP sentences interpreted in real time?

Future research

What are the equivalent facts like in other languages?

E.g. strictly head-final languages like Korean?

- Language statistics accounts predict that Korean DCE should be easier than missing VP sentences (?)
- I'd love to chat more!

Conclusion

Cross-linguistic variation in centre-embedding illusions

Highlighted the case of Mandarin “Missing NP” sentences.

- Mandarin presents challenges for language statistics approaches to cross-linguistic variation.
- Results more easily accounted for under an interference approach

Much more that can/should be learned:

- Time course of processing these illusions
- The range of cross-linguistic variation

Thank you

Special thanks to:

- Organisers of today's talk
- Scott Kaplowitz (UMD) and Koh Rui En (NUS) for research assistance.

- Various aspects of this research was supported by NSF Grant #DGE-1449815.