Syntactic Dependency Length Shaped by Strategic Memory Allocation

Weijie Xu, Richard Futrell

University of California, Irvine

The 6th Workshop on Research in Computational Linguistic Typology and Multilingual NLP (SIGTYP)

Efficient use of bounded working memory

Nonlocal Syntactic Dependency

Efficient use of bounded working memory

Nonlocal Syntactic Dependency [1-2]

memory retrieval

The **monster** who stayed in the tower **approached** the princess…

Efficient use of bounded working memory

Nonlocal Syntactic Dependency [1-2]

memory retrieval

The **monster** **who stayed in the tower** **approached** the princess…

memory interference

Efficient use of bounded working memory

Dependency Locality Principle [3]

The **monster approached** the princess…

Efficient use of bounded working memory

Dependency Locality Principle [3]

The **monster approached** the princess…

…a more detailed characterization of WM?

Efficient use of bounded working memory

Dependency Locality Principle

…a more detailed characterization of WM?

Strategic Allocation of Working Memory

Strategic Memory Allocation

novel and **unpredictable** information is prioritized [4-5]

Strategic Memory Allocation

novel and **unpredictable** information is prioritized [4-5]

⬇

more difficult to encode, but more robust against interference

Strategic Memory Allocation

novel and unpredictable information is prioritized [4-5]

⬇

more difficult to encode, but more robust against interference

….How does SMA influence dependency locality?

Informativity Effect on Dependency Locality

Novel information is more robust against interference

Informativity Effect on Dependency Locality

Novel information is more robust against interference

low info load

The evil **monster**

Informativity Effect on Dependency Locality

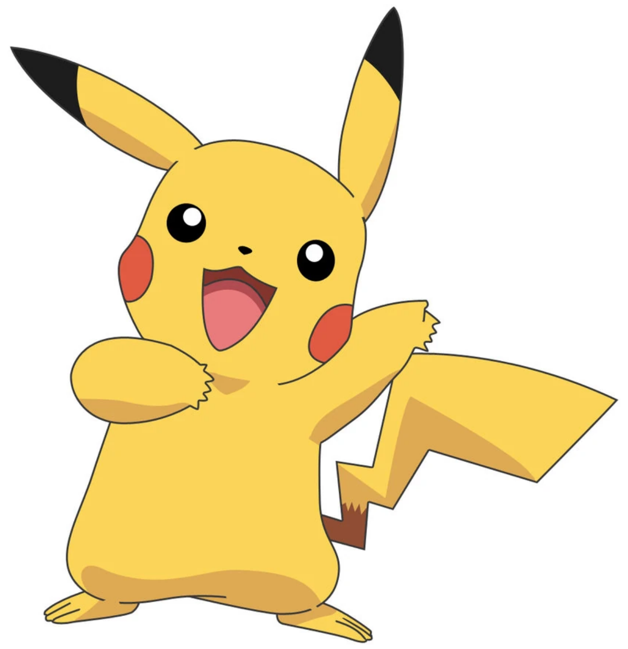
Novel information is more robust against interference

low info load

The evil **monster**

high info load

The cute **monster**



Informativity Effect on Dependency Locality

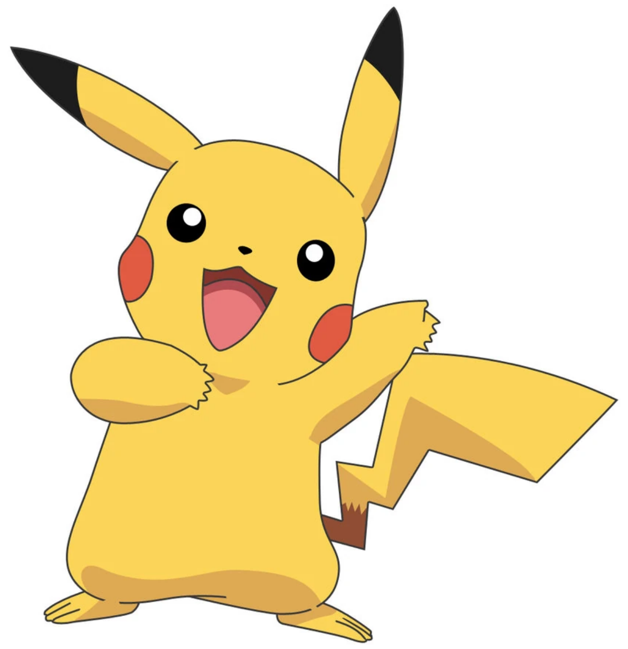
Novel information is more robust against interference

low info load

The evil **monster**

high info load

The cute **monster**



Informativity Effect on Dependency Locality

Novel information is more robust against interference

low info load

The evil **monster** in the tower **approached**…

$L = 3$

high info load

The cute **monster** who stayed in the tower near the castle **approached**…

$L = 8$

Informativity Effect on Dependency Locality

Novel information is more robust against interference

The evil **monster** in the tower **approached**…

$L = 3$

The cute **monster** who stayed in the tower near the castle **approached**…

$L = 8$

L **positively correlates with the informativity of the antecedent**

Informativity Effect on Dependency Locality

Novel information is more robust against interference

The evil **monster** in the tower **approached**…

$L = 3$

The cute **monster** who stayed in the tower near the castle **approached**…

$L = 8$

L **positively correlates with the informativity of the antecedent**

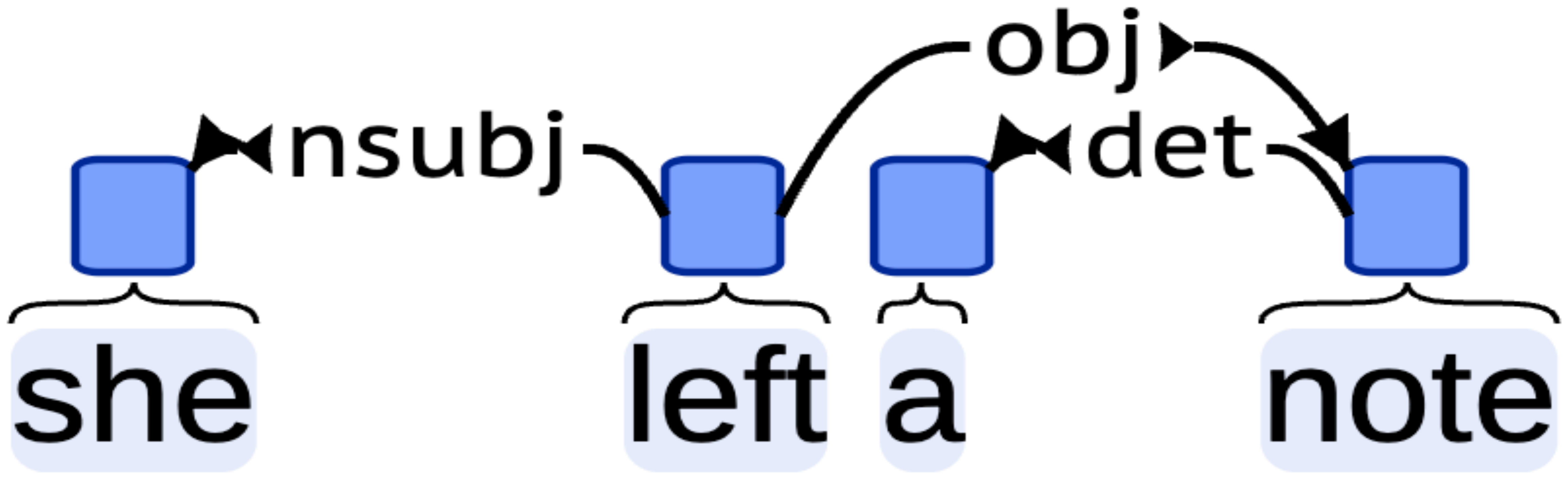
Surprisal $-\ln p(w_i \mid w_1 \ldots w_{i-1})$

Corpus Analysis

L **positively correlates with the informativity of the antecedent**

Corpus Analysis

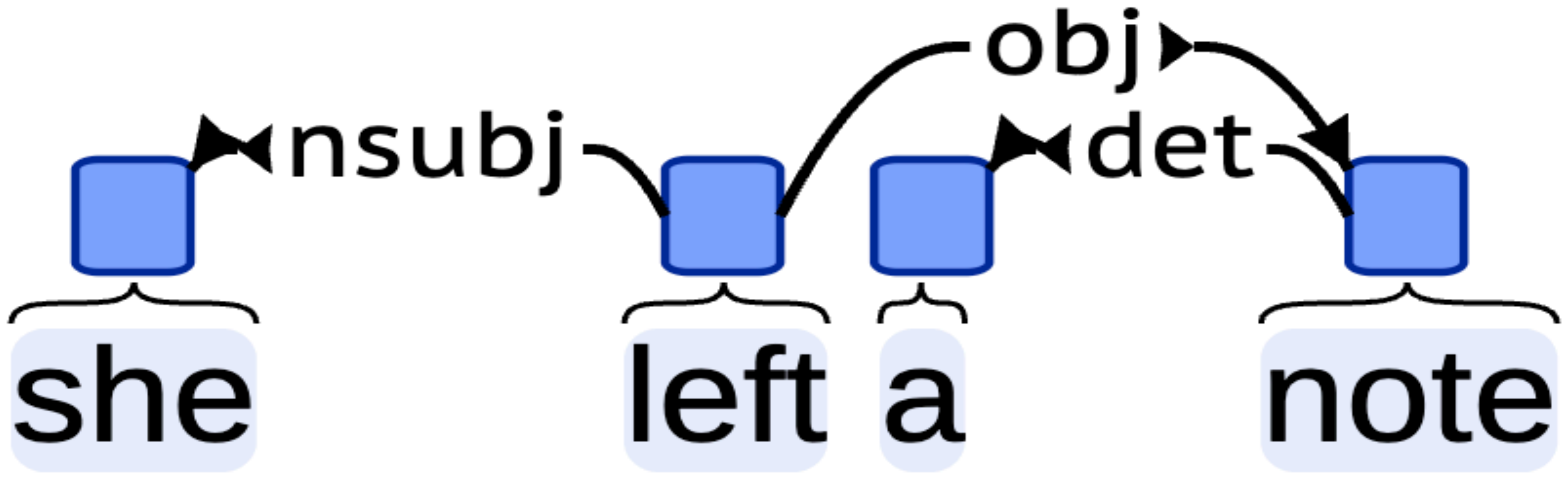
L **positively correlates with the informativity of the antecedent**



Universal Dependencies (UD) [6]

Corpus Analysis

L **positively correlates with the informativity of the antecedent**

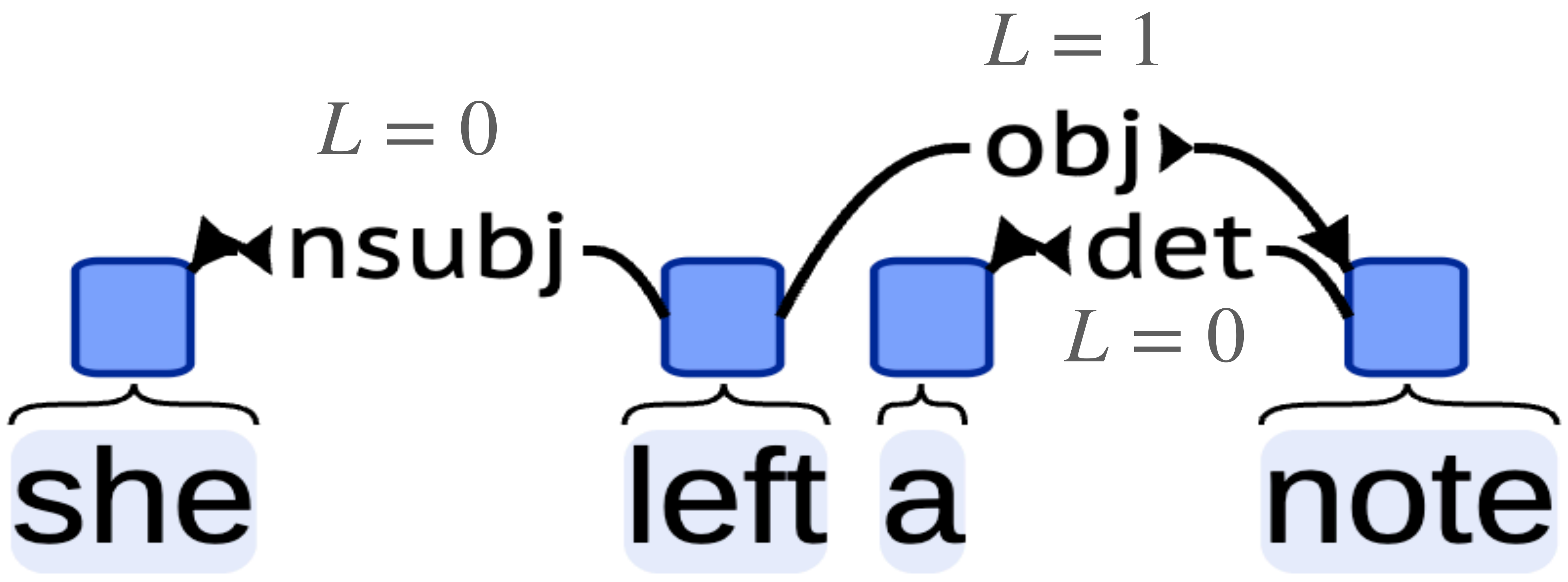


$-\ln p(\text{she}) \quad -\ln p(\text{left} \mid \text{she}) \quad -\ln p(\text{a} \mid \text{she left}) \quad -\ln p(\text{note} \mid \text{she left a})$

GPT-3 Language Model [7]

Corpus Analysis

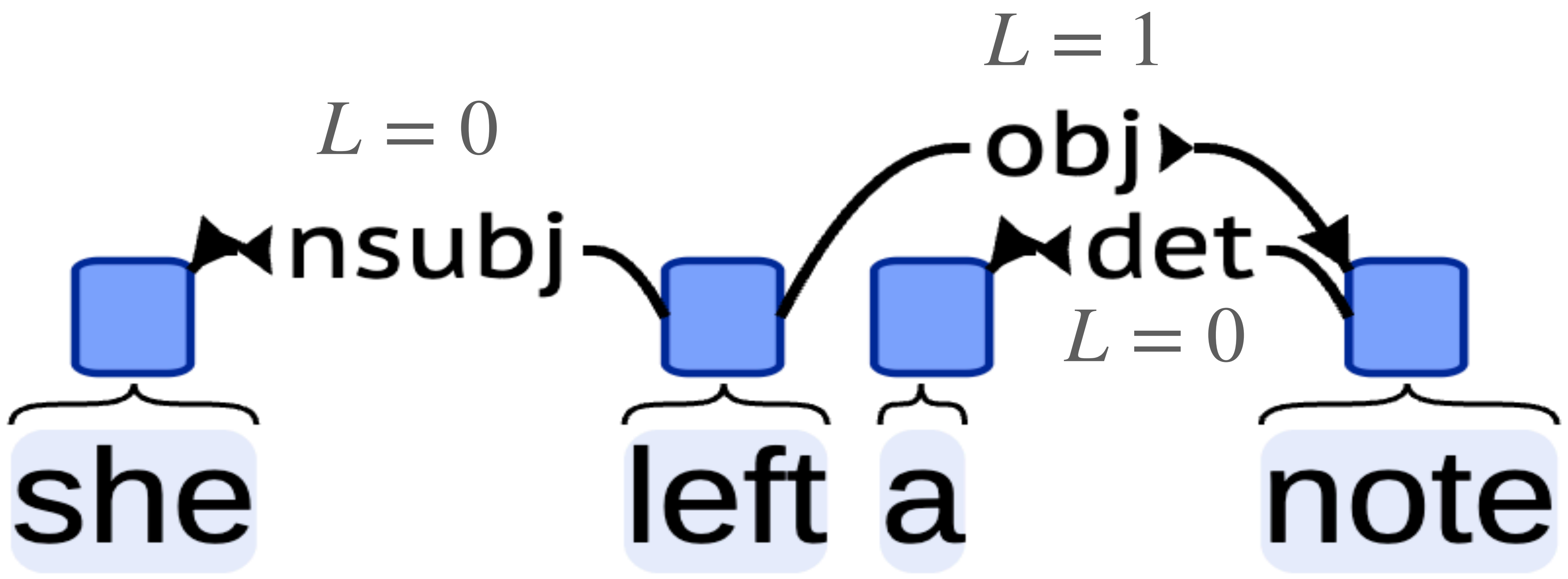
L **positively correlates with the informativity of the antecedent**



$-\ln p(\text{she})$ $-\ln p(\text{left} \mid \text{she})$ $-\ln p(\text{a} \mid \text{she left})$ $-\ln p(\text{note} \mid \text{she left a})$

Corpus Analysis

L **positively correlates with the informativity of the antecedent**



$-\ln p(\text{she})$ $\quad -\ln p(\text{left} \mid \text{she})$ $\quad -\ln p(\text{a} \mid \text{she left})$ $\quad -\ln p(\text{note} \mid \text{she left a})$

L ~ antecedent surprisal

Corpus Analysis

Data

Universal Dependencies (UD) [8]

	Corpus Name	Corpus Type	# Tokens
Amharic	ATT	doc-by-doc	12,682
Danish	DDT	sent-by-sent	80,378
English	GUM	doc-by-doc	126,530
German	GSD	sent-by-sent	268,404
Italian	ISDT	doc-by-doc	294,430
Japanese	GSD	sent-by-sent	168,333
Korean	Kaist	doc-by-doc	296,446
Mandarin	GSDSimp	sent-by-sent	98,616
Russian	SynTagRus	doc-by-doc	1,206,302
Spanish	AnCora	doc-by-doc	469,366
Turkish	BOUN	sent-by-sent	103,627

Corpus Analysis

Preprocessing data

Original Corpus → $-\ln p(w \mid c)$ from GPT-3 for each token → Extract sents w/ $L \geq 5$ → Remove puncts and tokens w/ $-\ln p(w \mid c) > 20$ → Data Set 1 (full dataset)

Corpus Analysis

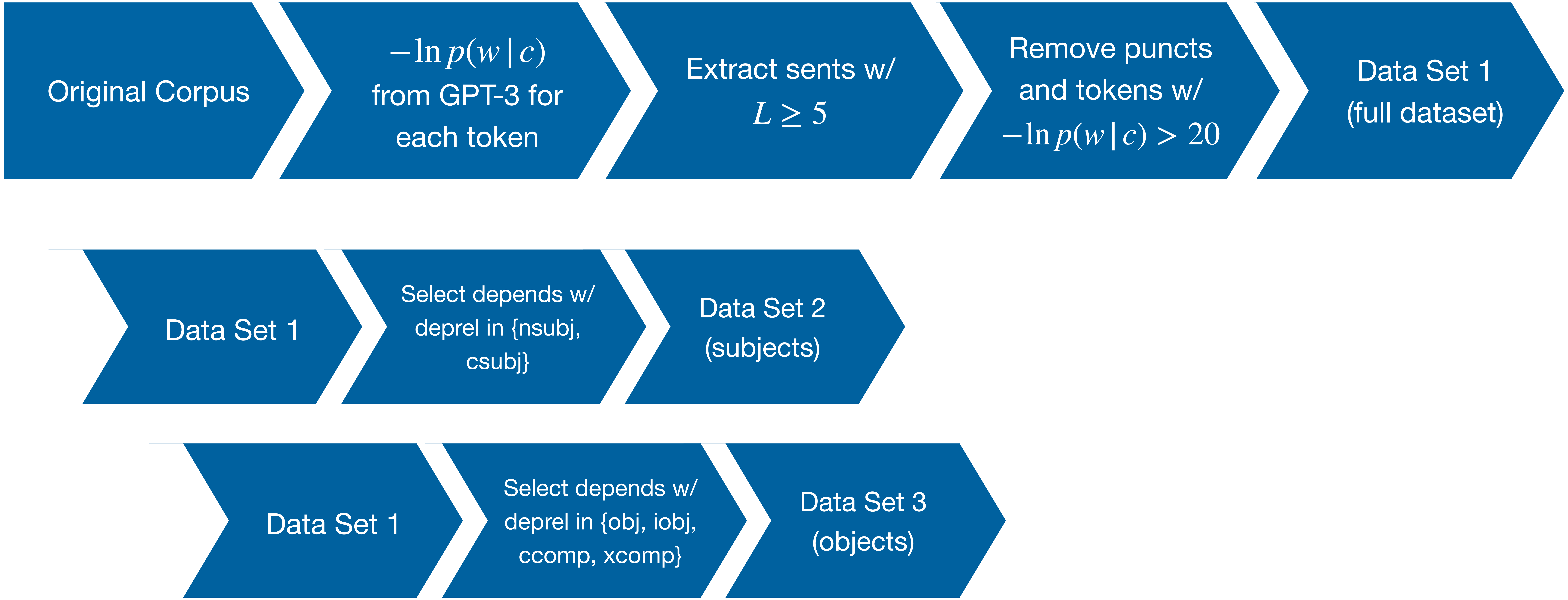
Preprocessing data

Original Corpus → $-\ln p(w \mid c)$ from GPT-3 for each token → Extract sents w/ $L \geq 5$ → Remove puncts and tokens w/ $-\ln p(w \mid c) > 20$ → Data Set 1 (full dataset)

Data Set 1 → Select depends w/ deprel in {nsubj, csubj} → Data Set 2 (subjects)

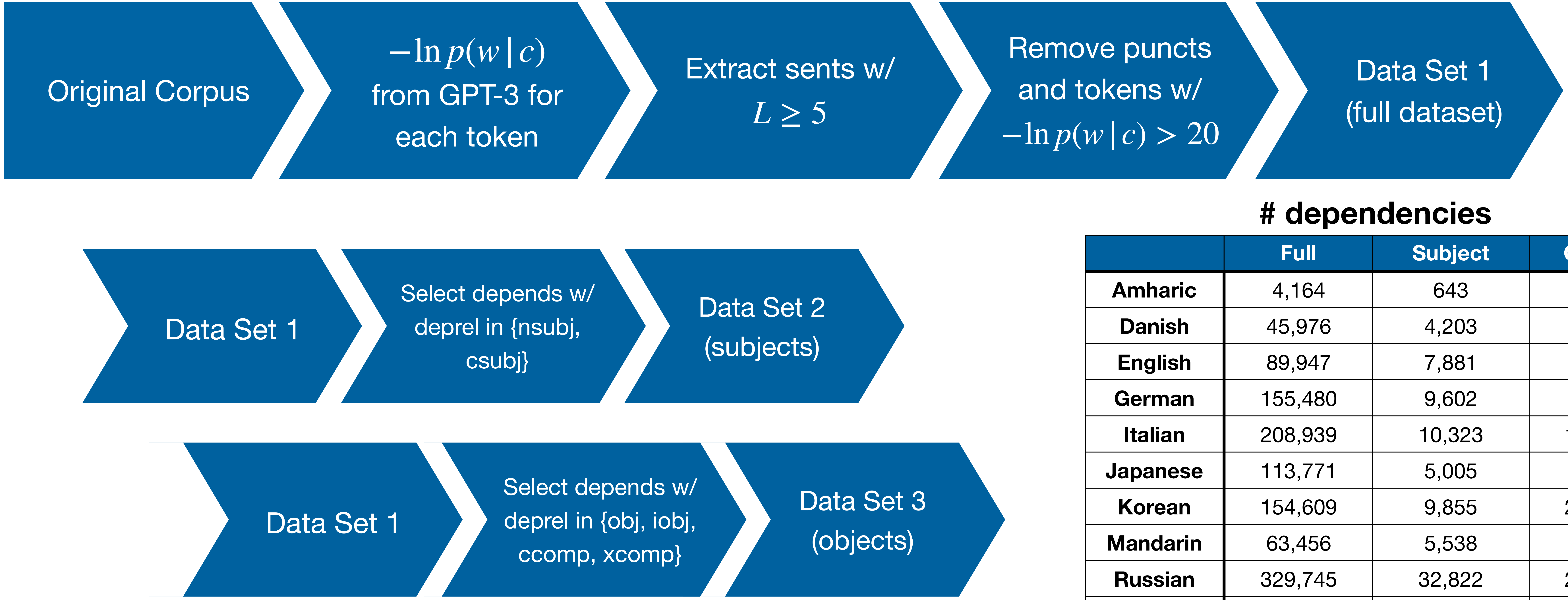
Corpus Analysis

Preprocessing data



Corpus Analysis

Preprocessing data



 # dependencies

	Full	Subject	Object
Amharic	4,164	643	525
Danish	45,976	4,203	3,963
English	89,947	7,881	7,296
German	155,480	9,602	8,474
Italian	208,939	10,323	11,735
Japanese	113,771	5,005	4,018
Korean	154,609	9,855	24,690
Mandarin	63,456	5,538	7,576
Russian	329,745	32,822	25,065
Spanish	333,728	21,472	31,143
Turkish	45,914	3,861	4,680

Corpus Analysis

Two measures of dependency length L

The evil **monster** in the tower **approached**…

$L = 3$

Corpus Analysis

Two measures of dependency length L

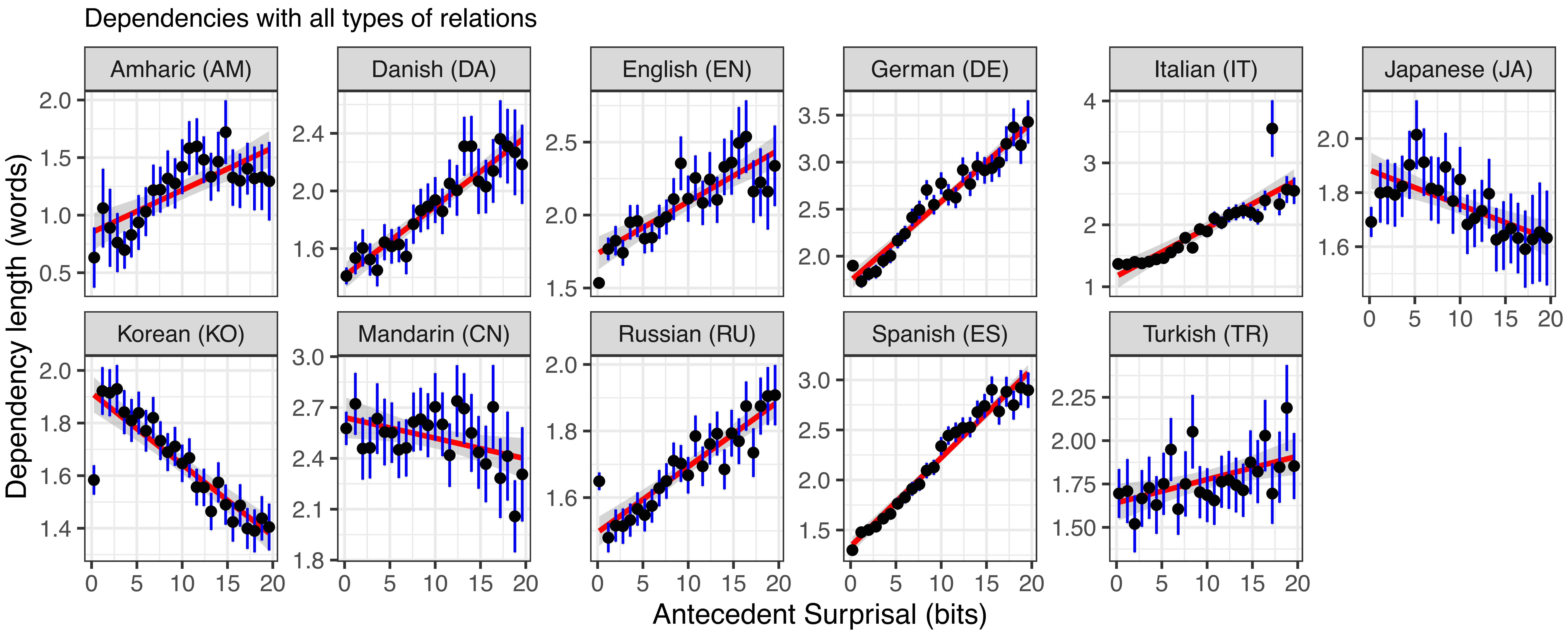
The evil **monster** in the tower **approached**…

$L = 3$

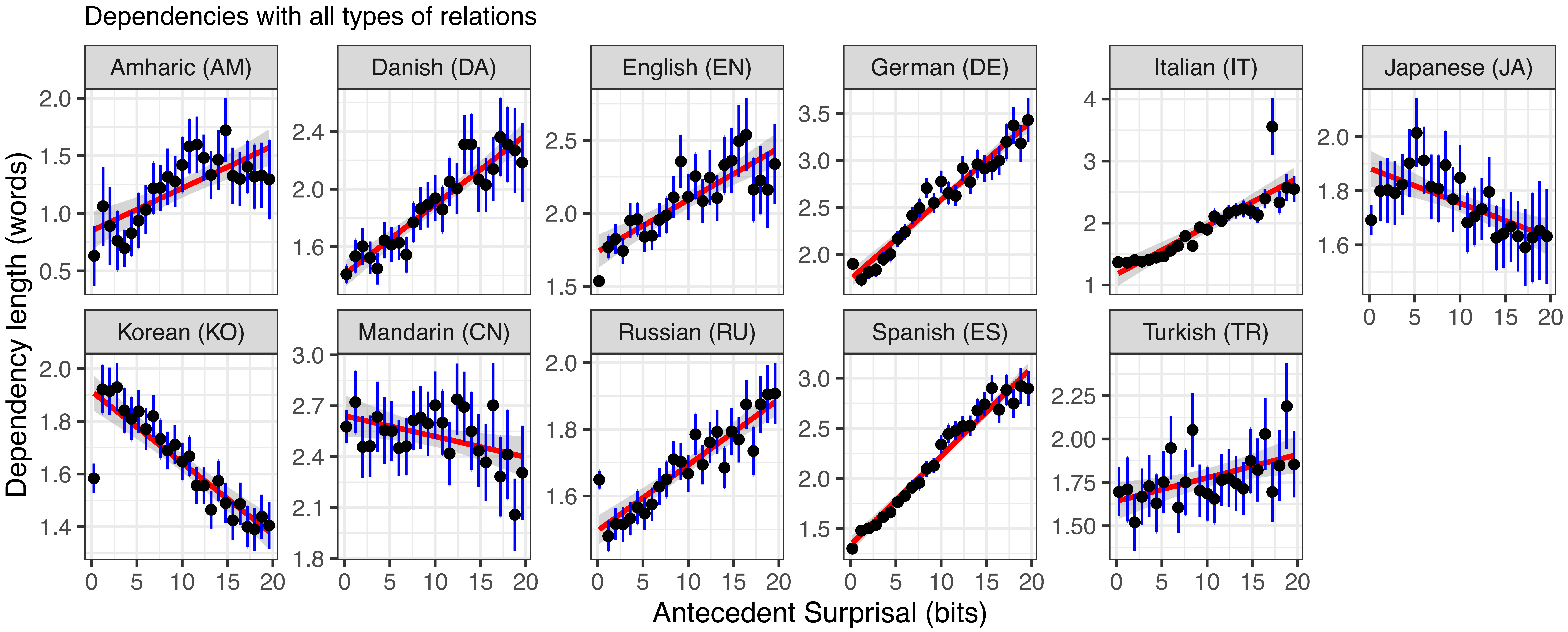
The evil **monster** in the tower **approached**…

$L = -\left(\ln p(\text{'in'}) + \ln p(\text{'the'}) + \ln p(\text{'tower'})\right)$

Results: Full Dataset (L as word counts)

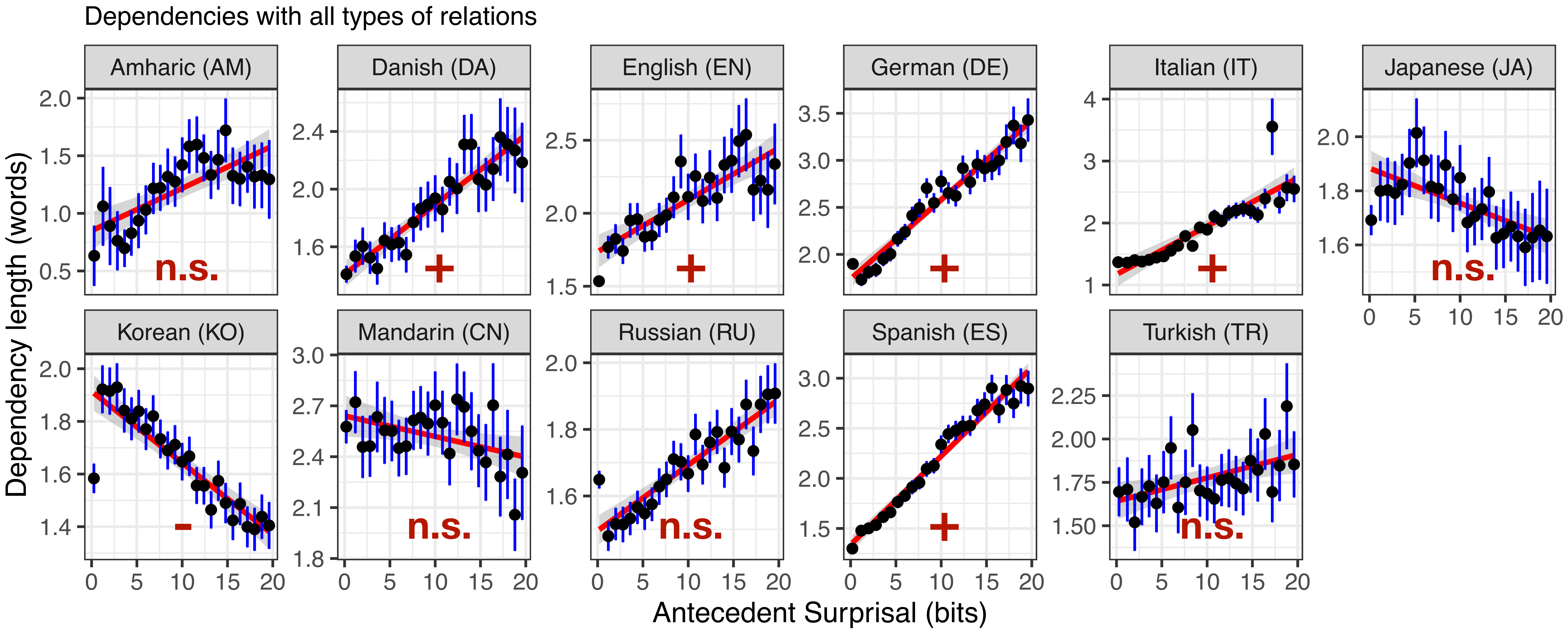


Results: Full Dataset (L as word counts)



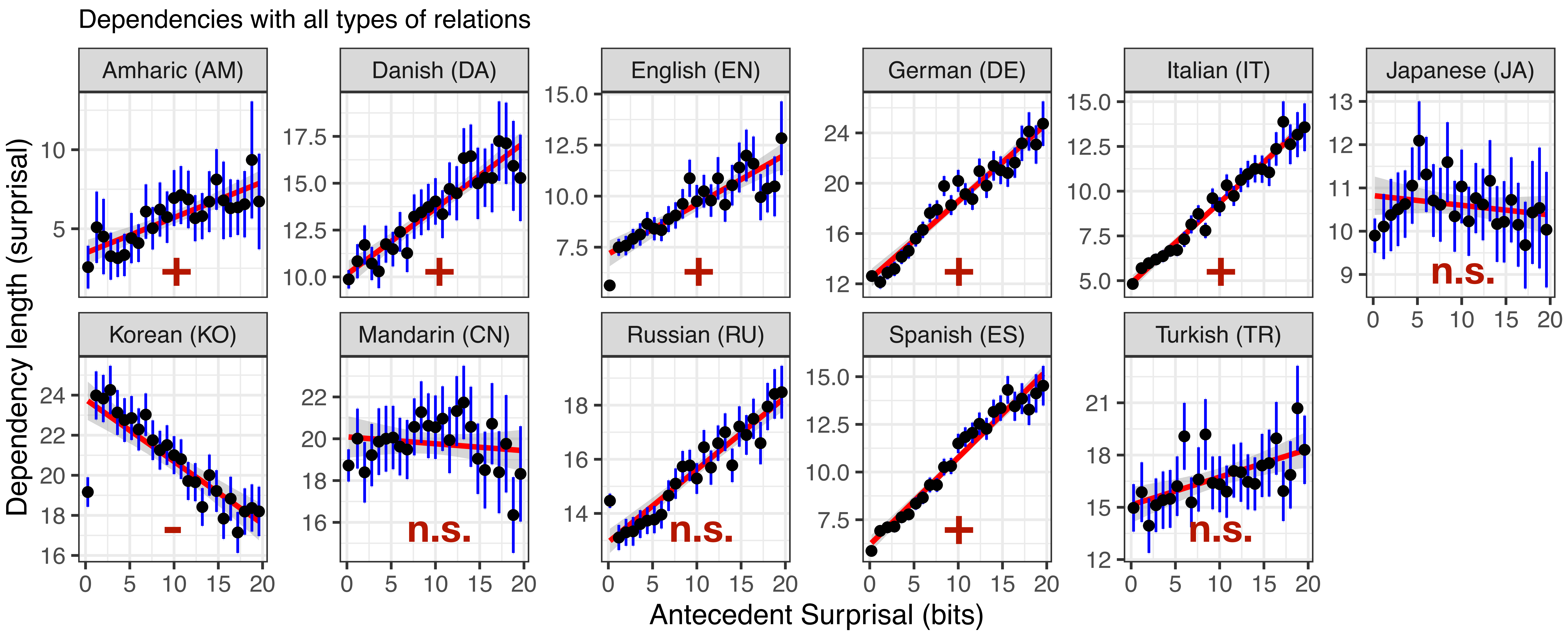

```
L ~ sent_position + sent_length + antec_postion + antec_surprisal + (1+antec_surprisal|dep-type)
```

Results: Full Dataset (L as word counts)



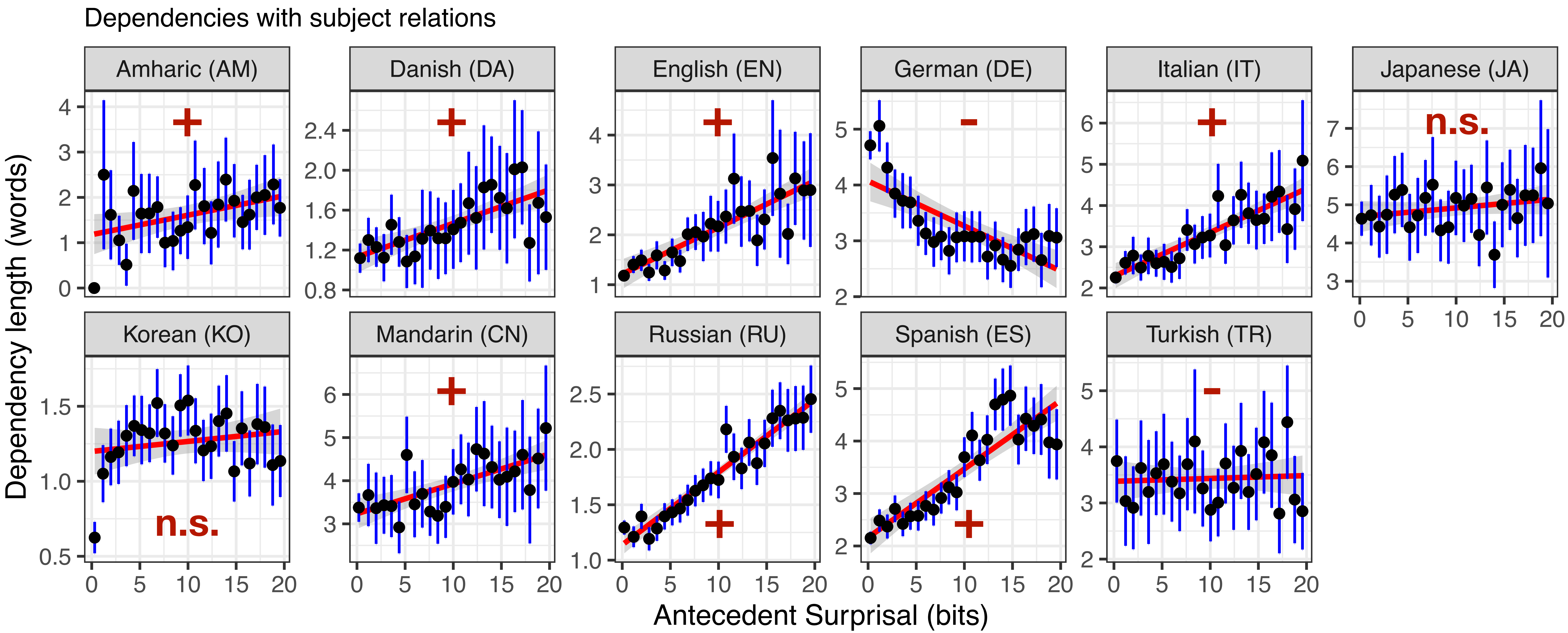

```
L ~ sent_position + sent_length + antec_postion + antec_surprisal + (1+antec_surprisal|dep-type)
```

Results: Full Dataset (L as surprisal)



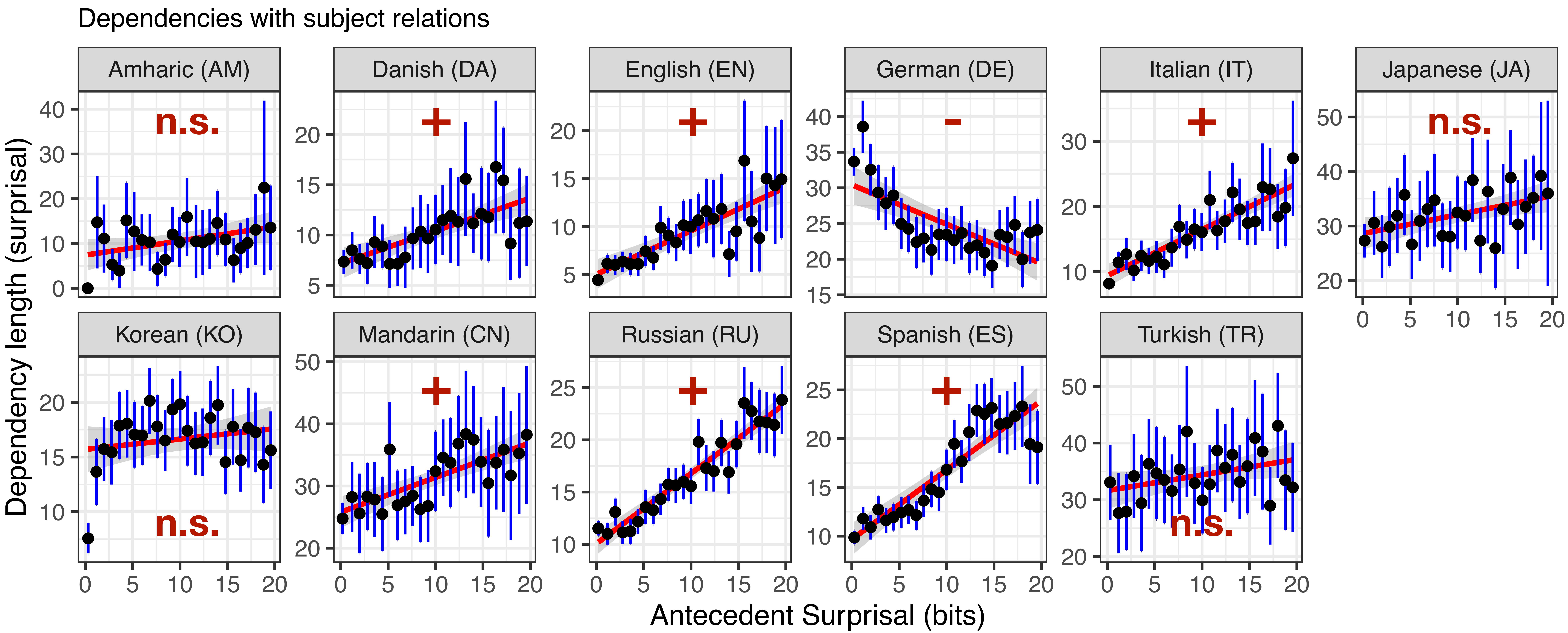

```
L ~ sent_position + sent_length + antec_postion + antec_surprisal + (1+antec_surprisal|dep-type)
```

Results: Subject Relations (L as word counts)



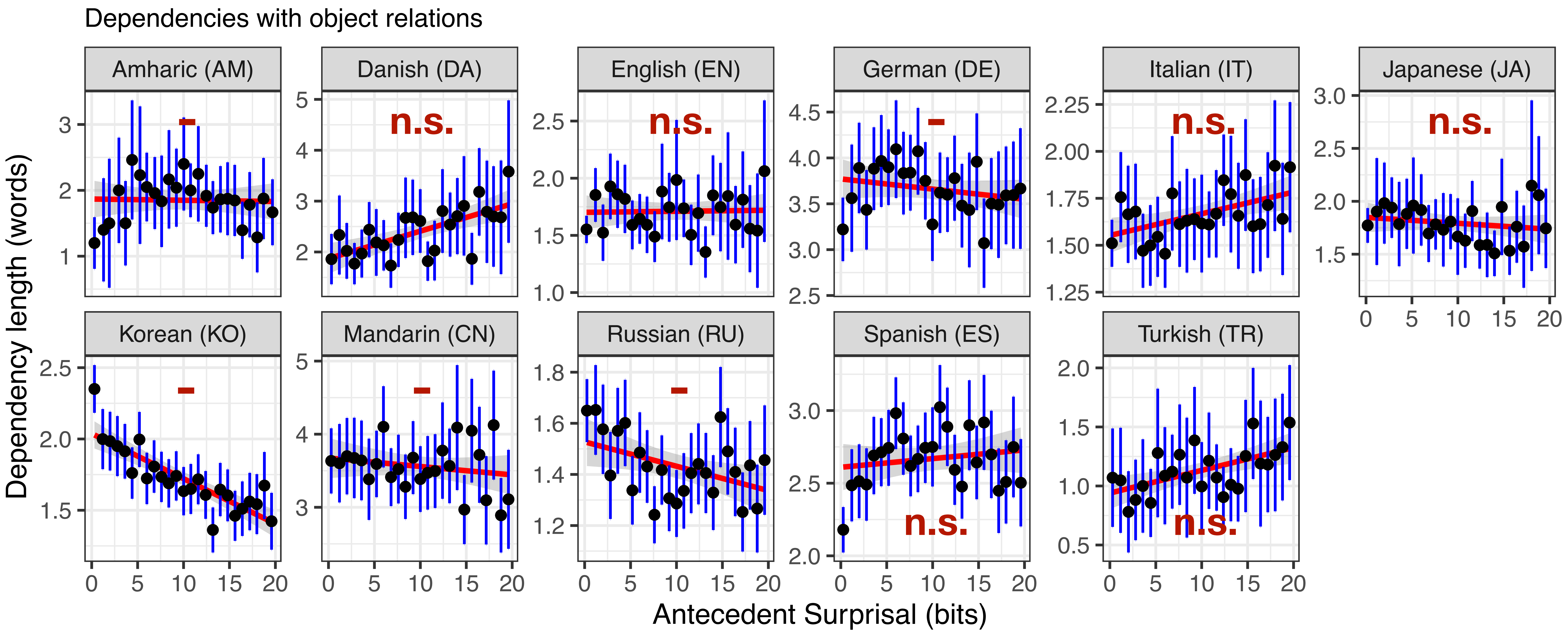

```
L ~ sent_position + sent_length + antec_postion + antec_surprisal
```

Results: Subject Relations (L as surprisal)



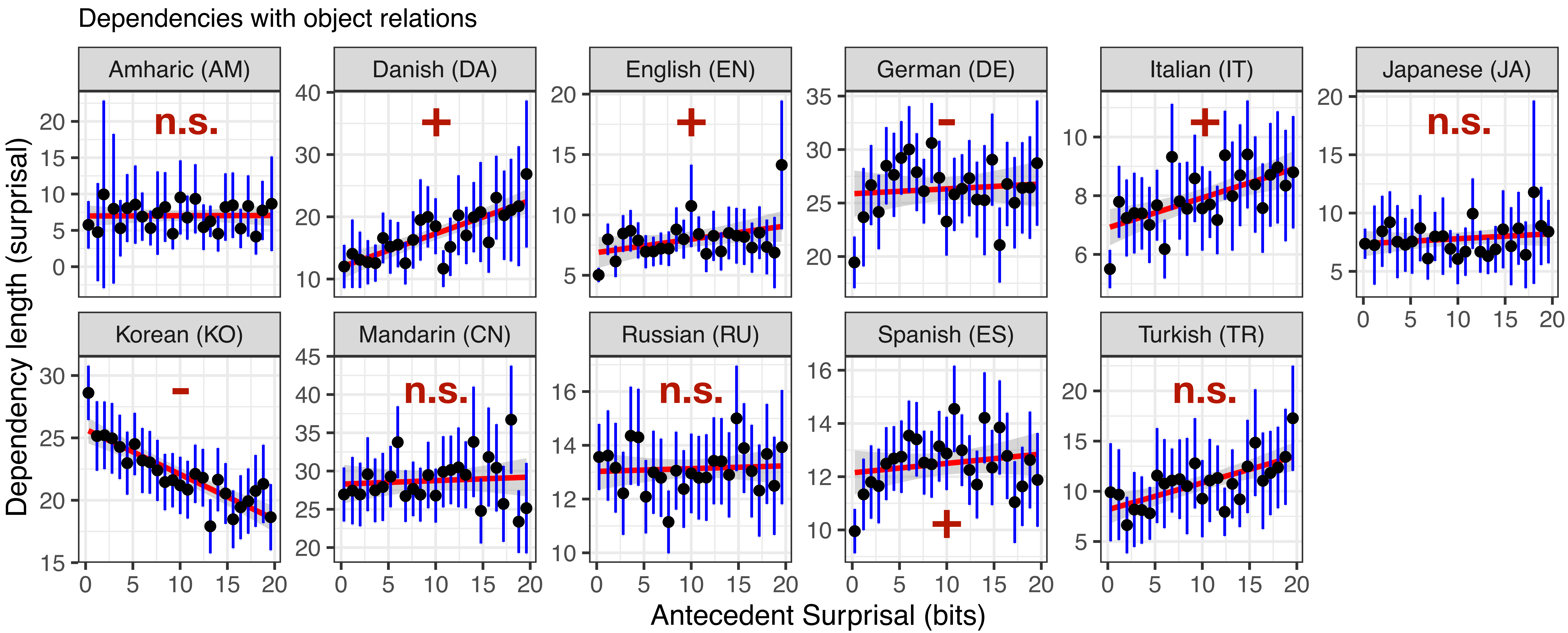

```
L ~ sent_position + sent_length + antec_postion + antec_surprisal
```

Results: Object Relations (L as word counts)



`L ~ sent_position + sent_length + antec_postion + antec_surprisal`

Results: Object Relations (L as surprisal)




```
L ~ sent_position + sent_length + antec_postion + antec_surprisal
```

Results

	Full Dataset		Subject Relations		Object Relations	
Language	L (words)	L (surprisal)	L (words)	L (surprisal)	L (words)	L (surprisal)
Amharic	$p = 0.175$	$+$	$+$	$p = 0.186$	$-$	$p = 0.876$
Danish	$+$	$+$	$+$	$+$	$p = 0.447$	$+$
English	$+$	$+$	$+$	$+$	$p = 0.743$	$+$
German	$+$	$+$	$-$	$-$	$-$	$-$
Italian	$+$	$+$	$+$	$+$	$p = 0.093$	$+$
Japanese	$p = 0.416$	$p = 0.775$	$p = 0.088$	$p = 0.985$	$p = 0.21$	$p = 0.94$
Korean	$-$	$-$	$p = 0.072$	$p = 0.156$	$-$	$-$
Mandarin	$p = 0.062$	$p = 0.331$	$+$	$+$	$-$	$p = 0.359$
Russian	$p = 0.395$	$p = 0.050$	$+$	$+$	$-$	$p = 0.454$
Spanish	$+$	$+$	$+$	$+$	$p = 0.058$	$+$
Turkish	$p = 0.161$	$p = 0.784$	$-$	$p = 0.59$	$p = 0.384$	$p = 0.083$

Results

Language	Full Dataset L (words)	Full Dataset L (surprisal)	Subject Relations L (words)	Subject Relations L (surprisal)	Object Relations L (words)	Object Relations L (surprisal)
Amharic	$p = 0.175$	$+$	$+$	$p = 0.186$	$-$	$p = 0.876$
Danish	$+$	$+$	$+$	$+$	$p = 0.447$	$+$
English	$+$	$+$	$+$	$+$	$p = 0.743$	$+$
German	$+$	$+$	$-$	$-$	$-$	$-$
Italian	$+$	$+$	$+$	$+$	$p = 0.093$	$+$
Japanese	$p = 0.416$	$p = 0.775$	$p = 0.088$	$p = 0.985$	$p = 0.21$	$p = 0.94$
Korean	$-$	$-$	$p = 0.072$	$p = 0.156$	$-$	$-$
Mandarin	$p = 0.062$	$p = 0.331$	$+$	$+$	$-$	$p = 0.359$
Russian	$p = 0.395$	$p = 0.050$	$+$	$+$	$-$	$p = 0.454$
Spanish	$+$	$+$	$+$	$+$	$p = 0.058$	$+$
Turkish	$p = 0.161$	$p = 0.784$	$-$	$p = 0.59$	$p = 0.384$	$p = 0.083$

Results

Language	Full Dataset		Subject Relations		Object Relations	
	L (words)	L (surprisal)	L (words)	L (surprisal)	L (words)	L (surprisal)
Amharic	$p = 0.175$	$+$	$+$	$p = 0.186$	$-$	$p = 0.876$
Danish	$+$	$+$	$+$	$+$	$p = 0.447$	$+$
English	$+$	$+$	$+$	$+$	$p = 0.743$	$+$
German	$+$	$+$	$-$	$-$	$-$	$-$
Italian	$+$	$+$	$+$	$+$	$p = 0.093$	$+$
Japanese	$p = 0.416$	$p = 0.775$	$p = 0.088$	$p = 0.985$	$p = 0.21$	$p = 0.94$
Korean	$-$	$-$	$p = 0.072$	$p = 0.156$	$-$	$-$
Mandarin	$p = 0.062$	$p = 0.331$	$+$	$+$	$-$	$p = 0.359$
Russian	$p = 0.395$	$p = 0.050$	$+$	$+$	$-$	$p = 0.454$
Spanish	$+$	$+$	$+$	$+$	$p = 0.058$	$+$
Turkish	$p = 0.161$	$p = 0.784$	$-$	$p = 0.59$	$p = 0.384$	$p = 0.083$

Limitations

Corpus annotation quality

Language models for understudied languages

Conclusions

Conclusions

Empirically, less predictable antecedents can tolerate longer dependency length

Conclusions

Empirically, less predictable antecedents can tolerate longer dependency length

Dependency locality pressure can be further modulated by informativity

Conclusions

Empirically, less predictable antecedents can tolerate longer dependency length

Dependency locality pressure can be further modulated by informativity

Strategic memory allocation prioritizes unexpected linguistic units for WM resources, resulting in more robust memory representation against interference

Thanks for your listening!

Selected Bibliography

[1] Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1-76.

[2] Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science*, *29*(2), 261-290.

[3] Futrell, R., Levy, R. P., & Gibson, E. (2020). Dependency locality as an explanatory principle for word order. *Language*, *96*(2), 371-412.

[4] Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes*, *26*(3), 376-405.

[5] Hahn, M., Futrell, R., Levy, R., & Gibson, E. (2022). A resource-rational model of human processing of recursive linguistic structure. *Proceedings of the National Academy of Sciences*, *119*(43), e2122602119.

[6] Nivre, J., de Marneffe, M. C., Ginter, F., Hajic, J., Manning, C. D., Pyysalo, S., ... & Zeman, D. (2020, May). Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 4034-4043).

[7] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877-1901.