

Crosslinguistic patterns of phrasal word order

support efficient prediction

A single memory-based pressure favors consistency, harmony, and locality



1 The typological puzzle

Across languages, phrase order repeatedly shows the same three biases.

CONSISTENCY

Category order is stable

D Num Adj N

those three cute cats

Same template across lexical choices, not a new order for every phrase.

HARMONY

Heads tend toward an edge

harmonic

medial

N A B C

A N B C

or

C B A N

C B A N

Most dependent types fall on the same side of the head.

LOCALITY

Stronger associations are closer

small cute cat

cute small cat

preferred

marked / less natural

Can one general notion of efficiency explain all three?



Paper



Code

Robustness: the memory–surprisal tradeoff gives the same pattern; surprisal variance (UID) does not.
Limits: artificial source distributions, small vocabularies, and single-word dependents.

2 The proposal: Efficient Prediction

Prediction should be possible with only a small memory of the past.

Short contexts should predict almost as well as long contexts

I saw those three cute cats and then ...

small memory window

Store only what improves prediction of what comes next.

Predictive information (PI)

PI = I[past ; future]

Uncertainty (bits)

Context Window Size

Lower PI = less information must be retained for optimal prediction.

Toy simulation

Noun + up to 5 adjectives

3 adjective classes A B C

Zipfian vocabularies

100 random sources

Compare linguistically motivated policies that vary in H, C, and L.

N A B C

0.68 bits

0.29 bits

0.15 bits

association with head noun

Example policies

N A B C

vs.

b N c a

most natural

least natural

3 Results

Orders with all three properties require the least predictive memory.

Predictive information

bits • lower is better

order	properties	1.9	2.3	2.7	3.1
NABC	H C L	1.96 best			
NCBA	H C -	2.08			
ANBC	~H C L	2.11 (Romance-type)			
Nabc	H - L	2.32			
Ncba	H - -	2.42			nonharmonic
BNAC	- C L	2.82			
BNCA	- C -	2.85			
bNac	- - L	3.01			
bNca	- - -	3.06			

H = harmonic C = consistent L = local

best has ~36% lower PI than the worst policy

Why it works

Consistency

B → B / C / end

The last category is enough to predict the next category.

Harmony

... A B C N |

The noun becomes a simple, reliable phrase-edge marker.

Locality

A — N

Strong noun-dependent links fit inside short windows.

ONE PRESSURE

→

H C L

Low predictive information jointly favors the major crosslinguistic regularities of phrasal order.