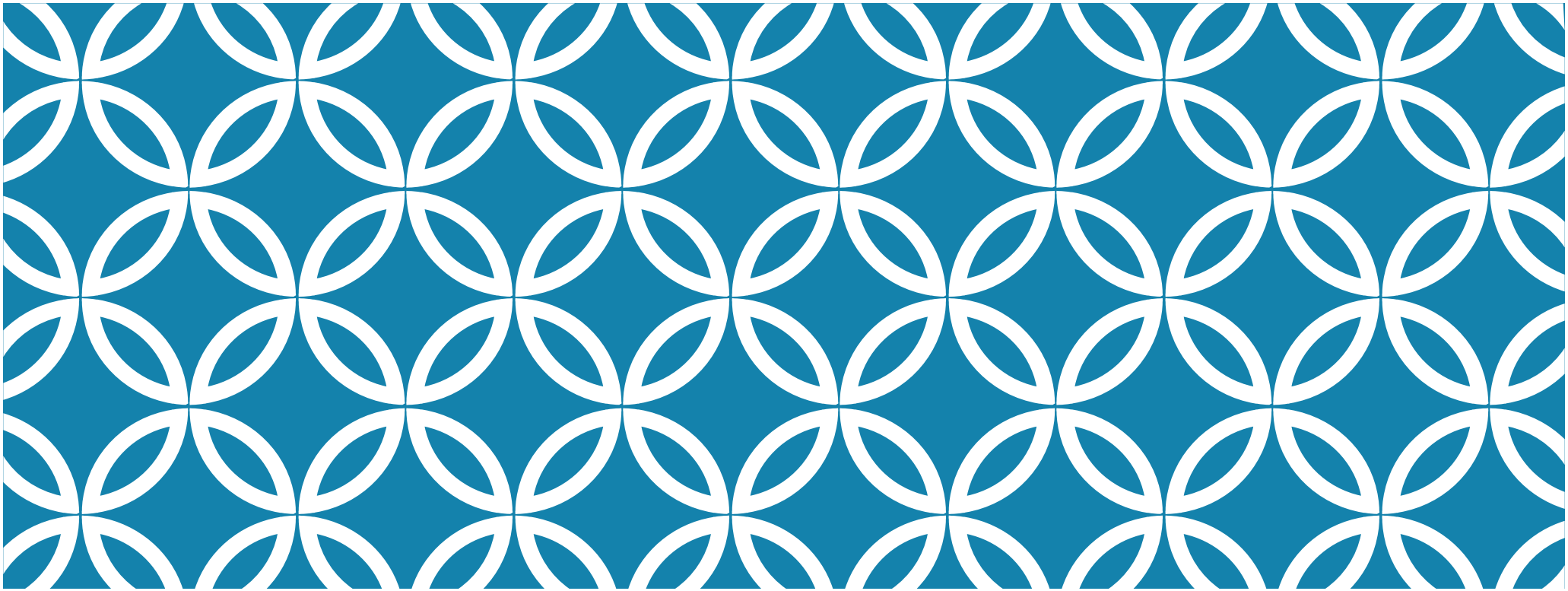
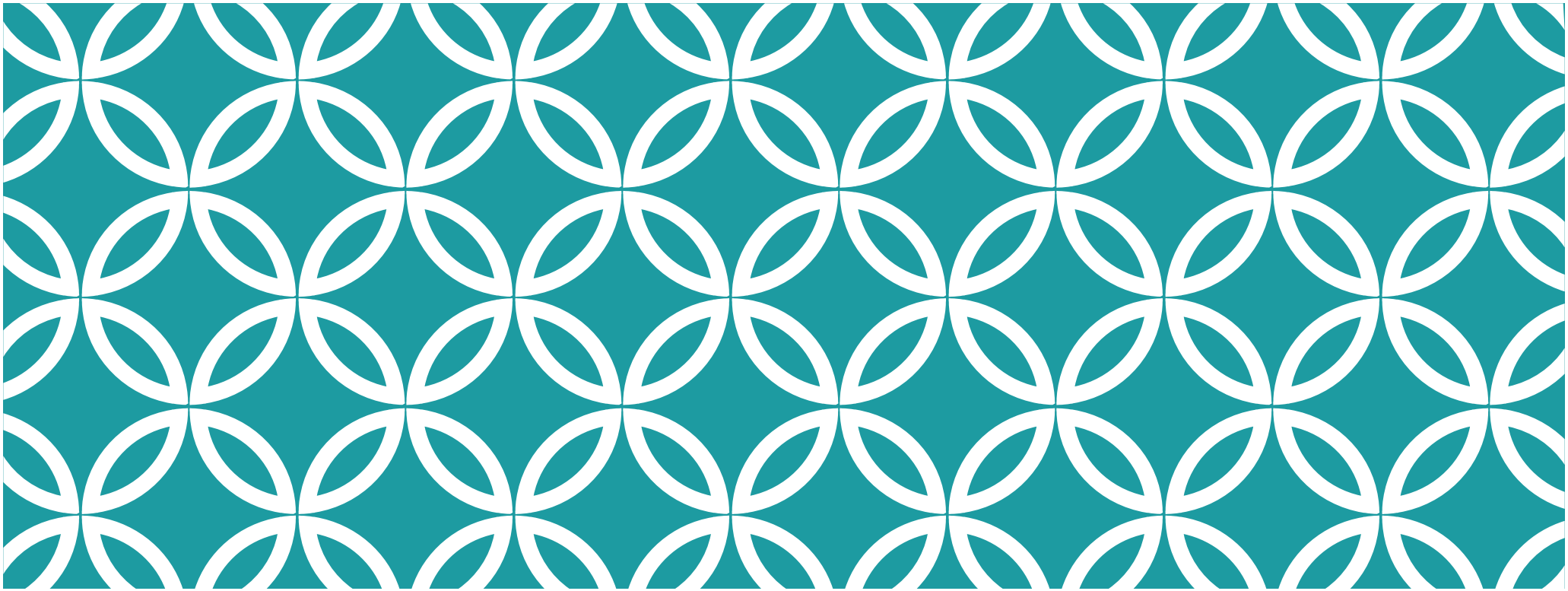


# DISCUSSION: PROCESSING AND DELPH-IN GRAMMARS

David Inman

# ROADMAP

1. Hawkins' theory of processing
2. Testable predictions
3. Discussion



# 1 HAWKINS' THEORY OF PROCESSING

# HAWKINS' THEORY OF PROCESSING

How did I get into this?

- I'm working with Minimalists who are testing a theory of word order change (Roberts 2019)[1] by using data from Indo-European and Austronesian (maybe Bantu)
- They wanted to compare this with a functionalist approach: I've been reading Hawkins' approach to word order (1979[2], 1982[3], 1990[4], 2004[5])

[1] Roberts, Ian G. 2019. *Parameter hierarchies and universal grammar*. Oxford University Press.
[2] Hawkins, John A. 1979. "Implicational universals as predictors of word order change." *Language*. 618–648.
[3] Hawkins, John A. 1982. "Language universals and the logic of historical reconstruction." *Linguistics*. 367–390.
[4] Hawkins, John A. 1990. "A parsing theory of word order universals." *Linguistic inquiry* 21.2. 223–261.
[5] Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford University Press.

# HAWKINS' THEORY OF PROCESSING

Why is it interesting?

- Hawkins (1982) makes a lot of predictions about left-or-right of Noun ordering in NPs. His reasoning is a little complex, but partially subsumed by later hypotheses in Hawkins (2004)
- E.g., if the language is strictly postpositional and the genitive is to the right of the noun, the relative clause will be right of the noun, too.

# HAWKINS' THEORY OF PROCESSING

Why is it interesting?

- Of 64 combinations (Adposition, Demonstrative, Numeral, Adjective, Genitive, Relative clause), Hawkins correctly predicts the 24 that are unattested in Grambank.
- Only 2/64 are predicted to be impossible and attested in more than 5 languages.

# HAWKINS' THEORY OF PROCESSING

General summary:

- Universal processing preferences are formalized into syntactic rules
  - Processing and generation likely share components
- Processing hypotheses are taken from experimental data and typology
- Processing constraints are universal
  - They apply at every stage of a language
  - They apply to every type of language

# 1. MINIMIZE DOMAINS (MiD)

"The human processor prefers to minimize the connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combination and/or dependency are processed. The degree of this preference is proportional to the number of relations whose domains can be minimized in competing sequences or structures, and to the extent of the minimization difference in each domain." (Hawkins 2004: 31)

✧Reformulation of Early Immediate Constituents (EIC) (Hawkins 1990)

# 1. MINIMIZE DOMAINS (MiD)

- Minimize the number of words needed to generate a given relationship (constituency, semantic relationship).
- Predicts Greenbergian word order constraints, heavy NP shift

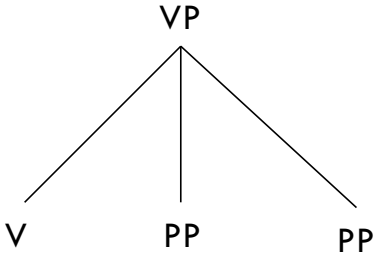
Terms:

- Immediate constituents (IC) → not necessarily ICs of HPSG/ERG trees!
- Constituent recognition domain (CRD): the number of words necessary to generate all ICs of some mother node

# 1. MINIMIZE DOMAINS (MiD)

1. John [**went** [**to London**]$_{PP}$ [**in** the late afternoon]$_{PP}$ ]$_{VP}$
2. John [**went** [**in the late afternoon**]$_{PP}$ [**to** London]$_{PP}$ ]$_{VP}$

Assumed tree:




✧(2) is worse than (1)

✧IC-to-word ratio: how many ICs are recognized per word

(1) 3/4 = 75% (VP)

(2) 3/6 = 50% (VP)

# 1. MINIMIZE DOMAINS (MiD)

✧On-line IC-to-word ratio: When each constituent is constructed, what is the ratio *at that point* of words that can construct an IC?

1. John [**went** [**to London**]$_{PP}$ [**in** the late afternoon]$_{PP}$ ]$_{VP}$

   1/1 2/3 3/4
   100% 66% 75% $=_{avg}$ 81%

2. John [**went** [**in the late afternoon**]$_{PP}$ [**to** London]$_{PP}$ ]$_{VP}$

   1/1 2/5 3/6
   100% 40% 50% $=_{avg}$ 63%

(Hawkins 2004: 50)

# 1. MINIMIZE DOMAINS (MiD)

✧On-line IC-to-word ratio predicts ordering effects even with identical CRDs

1. John [**went** [**to London**]$_{PP}$ [**in the late afternoon**]$_{PP}$ [**after** a long siesta]$_{PP}$ ]$_{VP}$

| 1/1 | 2/3 | 3/7 | 4/8 | |
|---|---|---|---|---|
| 100% | 66% | 43% | 50% | $=_{avg}$ 65% |

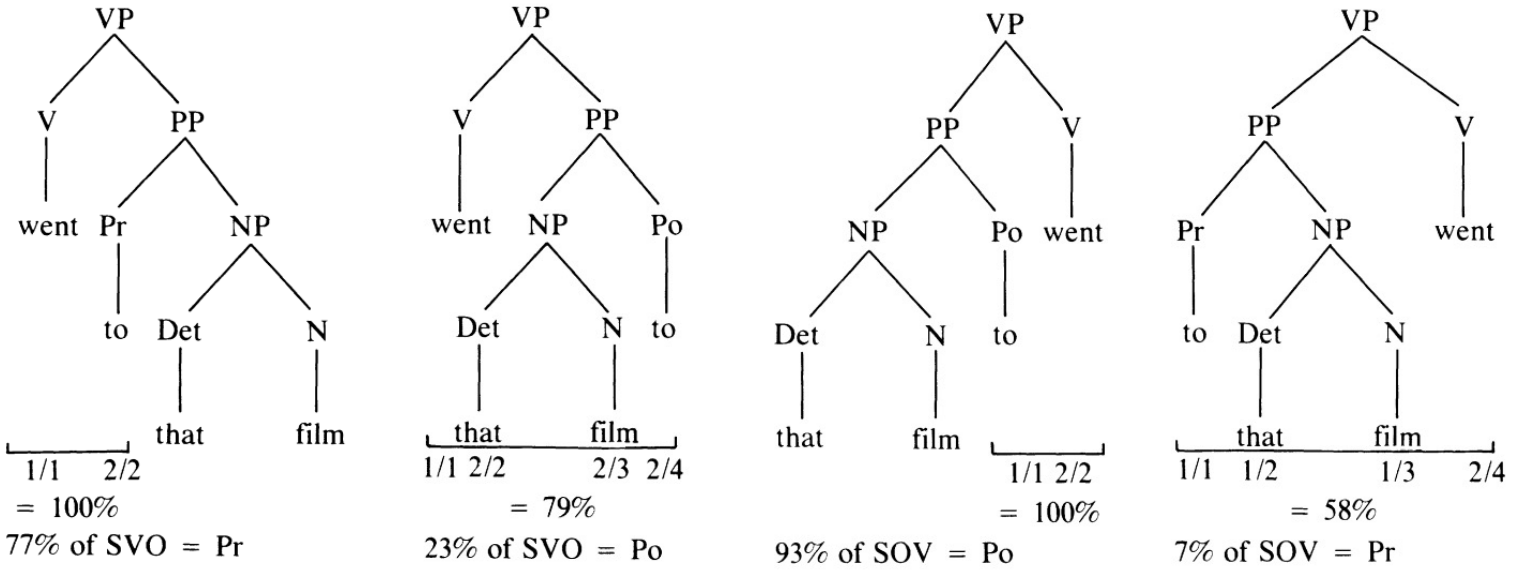
2. John [**went** [**in the late afternoon**]$_{PP}$ [**to London**]$_{PP}$ [**after** a long siesta]$_{PP}$ ]$_{VP}$

| 1/1 | 2/5 | 3/7 | 4/8 | |
|---|---|---|---|---|
| 100% | 40% | 43% | 50% | $=_{avg}$ 58% |

(Hawkins 2004: 50)

# 1. MINIMIZE DOMAINS (MiD)

- Consistent left- or right-headedness also minimizes domains.




Hawkins (1990: 238–239)

# 1. MINIMIZE DOMAINS (MiD)

- Heavy NP/S' shift and extraposition goes to the right in head-initial languages, to the left in head-final languages.
- Japanese:

Mary-ga kinoo John-ga kekkonsi-ta to it-ta.
Mary yesterday John married that said
'Mary said that John got married yesterday.'

a. $S_1$[ $_{NP}$[Mary-ga] $_{VP}$[ $_{S'}$[ $S_2$[kinoo John-ga kekkonsi-ta] to] it-ta]].

b. $S_2$[ $_{S'}$[ $S_1$[Kinoo John-ga kekkonsi-ta] to] $_{NP}$[Mary-ga] $_{VP}$[it-ta]].

- English:

(11) a. $_S$[ $_{S'}$[That Bill was frightened] $_{VP}$[surprised $_{NP}$[Mary]]].

b. $_S$[It $_{VP}$[surprised $_{NP}$[Mary] $_{S'}$[that Bill was frightened]]].

Hawkins (1990: 231)

# 1. MINIMIZE DOMAINS (MiD)

- This occurs at every domain: Minimize on-line IC-to-word ratio for S, VP, NP, ...
- The occurs at a semantic level: minimize domains that disambiguate semantic meanings, e.g. **The bucket was kicked by Pat.*
- Combinations that minimize domains are preferred: PostP + V-final, PrepP+ V-initial

# 2. MINIMIZE FORMS (MiF)

"The human processor prefers to minimize the formal complexity of each linguistic form F (its phoneme, morpheme, word, or phrasal units) and the number of forms with unique conventionalized property assignments, thereby assigning more properties to fewer forms. These minimizations apply in proportion to the ease with which a given property P can be assigned in processing to a given F." (Hawkins 2004: 38)

# 2. MINIMIZE FORMS (MiF)

- Put more work into context and inference and less into combinatoric semantics
- Zipfian distribution: most efficient to maximally reduce frequently used words and constructions
- "Inference is cheap"[6] — communication by context, inference is older (likely more efficient to process) than language
- Not as much to evaluate syntactically

[6] Levinson, Stephen C. *Presumptive meanings: The theory of generalized conversational implicature*. MIT press, 2000.

# 3. MAXIMIZE ON-LINE PROCESSING (MaOP)

"The human processor prefers to maximize the set of properties that are assignable to each item X as X is processed, thereby increasing O(n-line) P(roperty) to U(ltimate) P(roperty) ratios. The maximization difference between competing orders and structures will be a function of the number of properties that are unassigned or misassigned to X in a structure/sequence S, compared with the number in an alternative." (Hawkins 2004: 51)

# 3. MAXIMIZE ON-LINE PROCESSING (MaOP)

✧Don't backtrack, don't garden path. Greedy assignment of words to constituents should be minimally different from final constituent assignment.

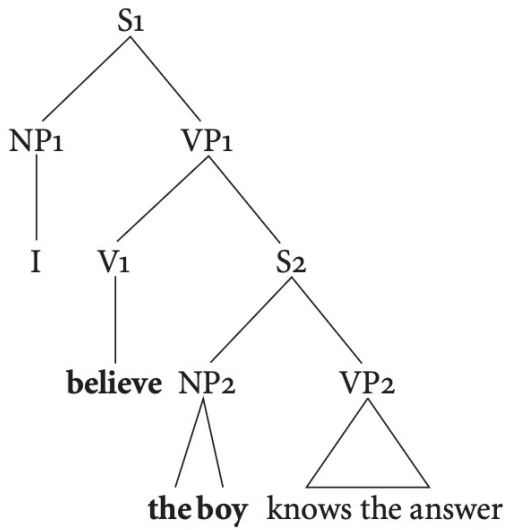
✧Minimize the sum of:

a. the number of words and phrases that are misassigned in on-line processing
b. the number of additional nodes introduced when correcting misassignments in (a)
c. the number of mother-daughter attachments that are temporarily misassigned in (a)
d. the number of relations (semantic) that are misassigned in (a)
e. the number of mother-daughter attachments that replace those misassigned in (c)
f. the number of relations (semantic) that replace those misassigned in (d)

Hawkins (2004: 53-54)

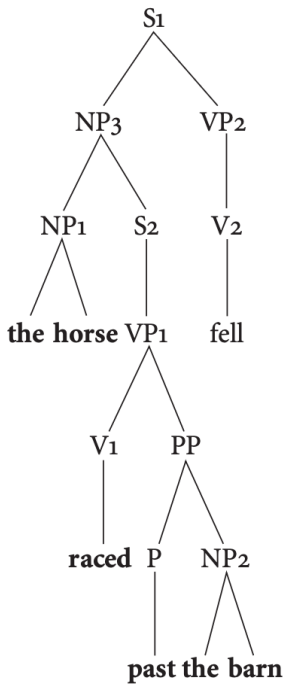
# 3. MAXIMIZE ON-LINE PROCESSING (MaOP)

(3.22)



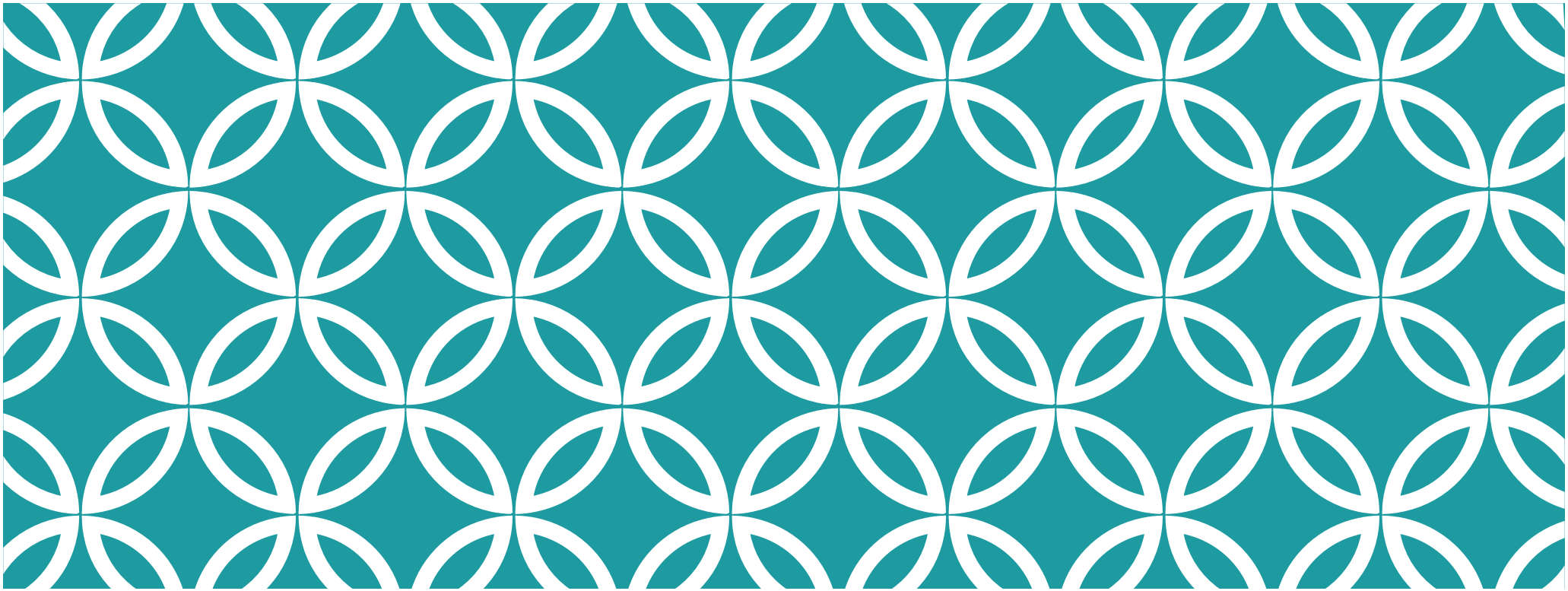

| | | |
|---|---|---|
| (a) | *the*, *boy*, NP2 | [3] |
| (b) | S2 | [1] |
| (c) | vp1[NP2] | [1] |
| (d) | NP2 = Accus of *believe*<br>NP2 = $\theta$-object of *believe* | [2] |
| (e) | s2[NP2], vp1[S2] | [2] |
| (f) | NP2 = Nomin of *knows*<br>NP2 = $\theta$-subject of *knows* | [2] |
| | TOTAL = | 11 |

(3.23)




| | | |
|---|---|---|
| (a) | *the*, *horse*, *raced*, *past*, *the*, *barn*, NP1, VP1, PP, NP2 | [10] |
| (b) | NP3, S2 | [2] |
| (c) | s1[NP1], s1[VP1] | [2] |
| (d) | VP1 = predication on NP1<br>VP1 = active VP<br>NP1 = Nomin of *raced*<br>NP1 = $\theta$-subject of *raced* | [4] |
| (e) | np3[NP1], s1[NP3], s2[VP1], np3[S2] | [4] |
| (f) | VP1 = rel clause modifying *horse*<br>VP1 = passive VP<br>Nomin of *raced* = gap<br>$\theta$-subject of *raced* = gap | [4] |
| | TOTAL = | 26 |

Hawkins (1990: 54-55)



# 2 TESTABLE PREDICTIONS

# TESTABLE PREDICTIONS

1. Sentences with multiple possible orders or equivalent constructions will, on average, prefer those that more conform to MiD and MaOP.

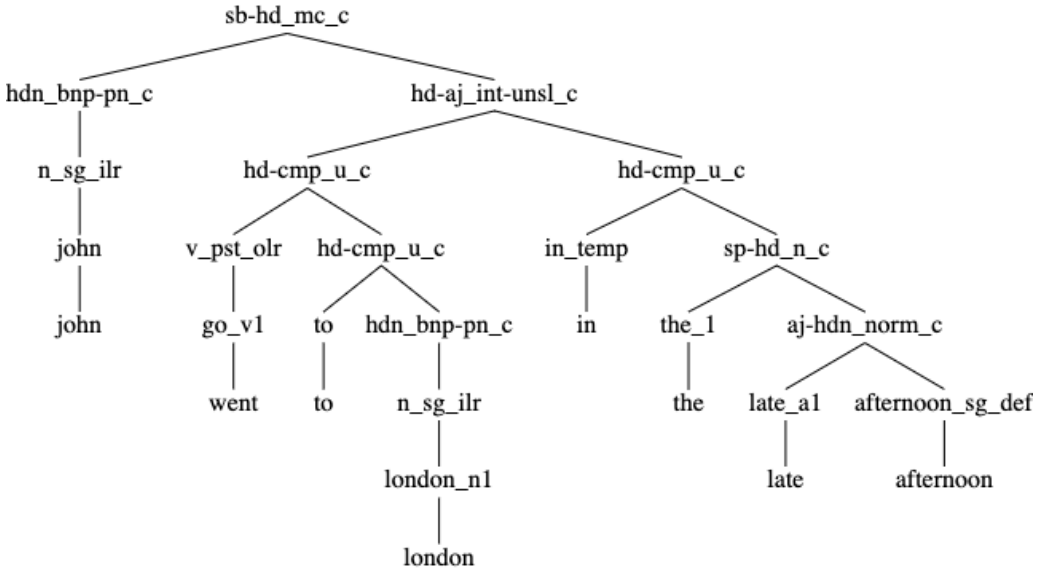
- ✧Take a DELPH-IN grammar plus a large corpus
- ✧Parse each sentence as given, calculate its MiD and/or MaOP score
- ✧Generate from the MRS, calculate alternative sentences' scores
- ✧Is there an on-average lower MiD and/or MaOP score for observed sentences? Is the preference comparable to the difference between the scores?

✧Hawkins has already made (much of) this algorithmic.

# TESTABLE PREDICTIONS

1. Sentences with multiple possible orders or equivalent constructions will, on average, prefer those that more conform to MiD and MaOP.

✧This will require some tree manipulations, e.g. “John went to London int he late afternoon”



# TESTABLE PREDICTIONS

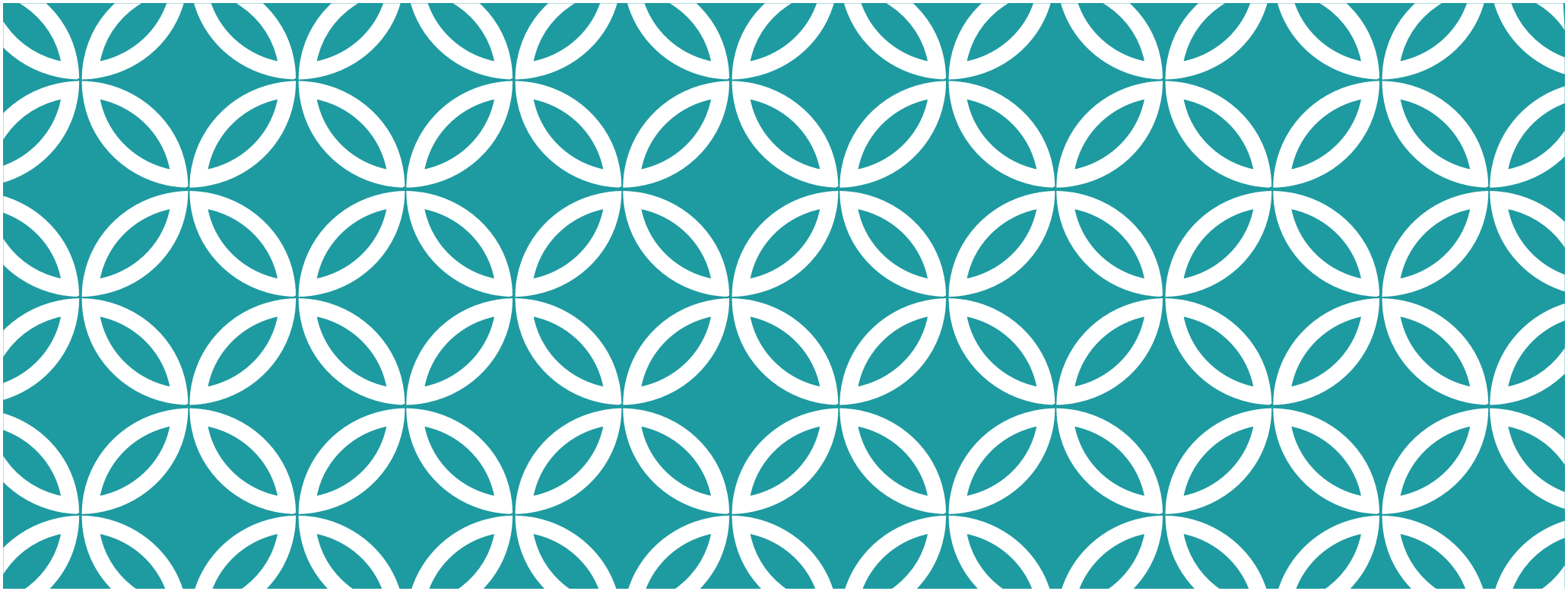
2. Common constructions will have lower MiD scores on average than uncommon ones.

- ✧MiD is calculated for each domain, grammaticalization is sensitive to frequencies.
- ✧In a corpus of natural text, it is expected that bad scores will be more common in uncommon constructions.

# TESTABLE PREDICTIONS

3. There will be a Zipfian distribution of MiD and MaOP values across natural corpora.

✧On a frequentist account, if we collect all the domains across all sentences in a large corpus, we should see processing pressure apply most strongly to common constructions.



# 3 DISCUSSION

# DISCUSSION

- Other *testable* predictions from this hypothesis?
- Could this be tested with the ERG or other grammars?
- What machinery would be needed?
  - Tree transformations?
  - Calculating multiple trees for MaOP?
  - Calculating MiD?
- Are others interested in pursuing this line of investigation?