The Building-Blocks of Language

Omer Preminger
omer@lingsite.org

Background

At the center of much of the 20th century discussion of language —

the SIGN

ENG: /ˈtej.bl̩/

SPA: /ˈme.sa/

HEB: /ʃul.ˈxan/

ENG: /ˈtej.bl̩/

SPA: /ˈme.sa/

HEB: /ʃul.ˈxan/

Much has been made about the often arbitrary nature of the relation between the meaning of a SIGN and its form (Saussure 1916, Hjelmslev 1943)

As contrasted with, e.g., onomatopoeia, iconicity, etc.

when Wilhelm von Humboldt says language makes "infinite use of finite means" —

one thing that language certainly seems to have finitely many of is this kind of arbitrary, non-decomposible SIGNs.

ENG: /ˈtej.bl̩/

SPA: /ˈme.sa/

HEB: /ʃul.ˈxan/

It's also far from clear that humans are the only animal that can use arbitrary signs.

Cf. Seyfarth & Cheney (1980), and subsequent work, on vervet monkey alarm calls.

⇒ This invites the inference that the "secret sauce" of human language is the combinatorics —

Humboldt's "infinite use"

Chomsky's "Strong Minimalist Thesis"

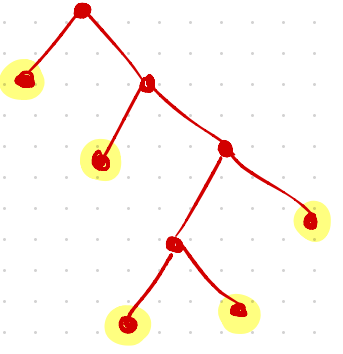
Chomsky's "Strong Minimalist Thesis" (SMT)

Chomsky (1995, 2007, i.a.)
Hauser, Chomsky & Fitch (2002)

SMT – The only linguistically-proprietary cognitive capacity is "MERGE"

↓

The ability to recursively assemble objects into hierarchical structures



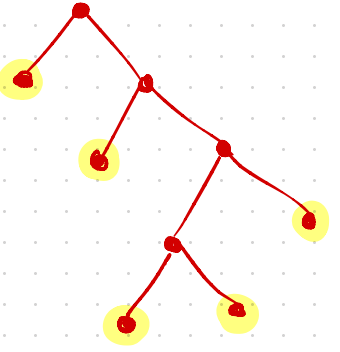
Chomsky's "Strong Minimalist Thesis" (SMT)

Chomsky (1995, 2007, i.a.)
Hauser, Chomsky & Fitch (2002)

SMT ⇒ Everything beyond this capacity is not linguistically proprietary, from a cognitive standpoint.

Instead, it relies on properties of other cognitive systems (e. g. motor systems, perceptual systems, non-linguistic thought), plus general principles of computation.

TODAY'S TALK:



An argument against SMT, based on the nature of linguistic atoms.

Specifically: an argument that the atoms themselves are linguistically proprietary,

and are unlike anything that could have existed outside the linguistic system.

TODAY'S TALK:

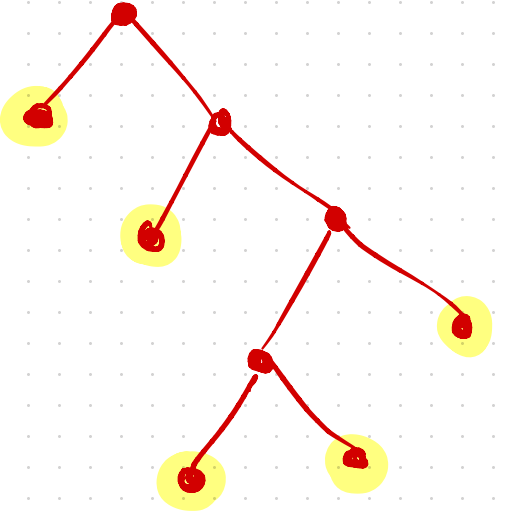
In other words:
the atoms are also cognitively special

⇓

"MERGE" is not the only linguistically proprietary cognitive capacity

⇓

The SMT is false.



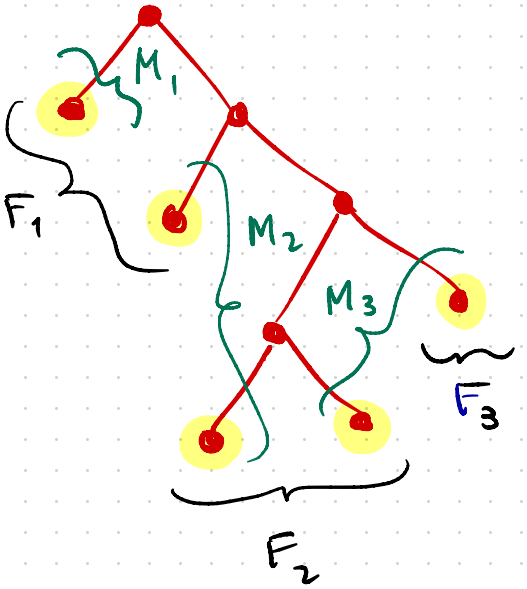
SNAPSHOT OF THE CLAIM:

- syntactic terminals don't "have forms" and they don't "have meanings"

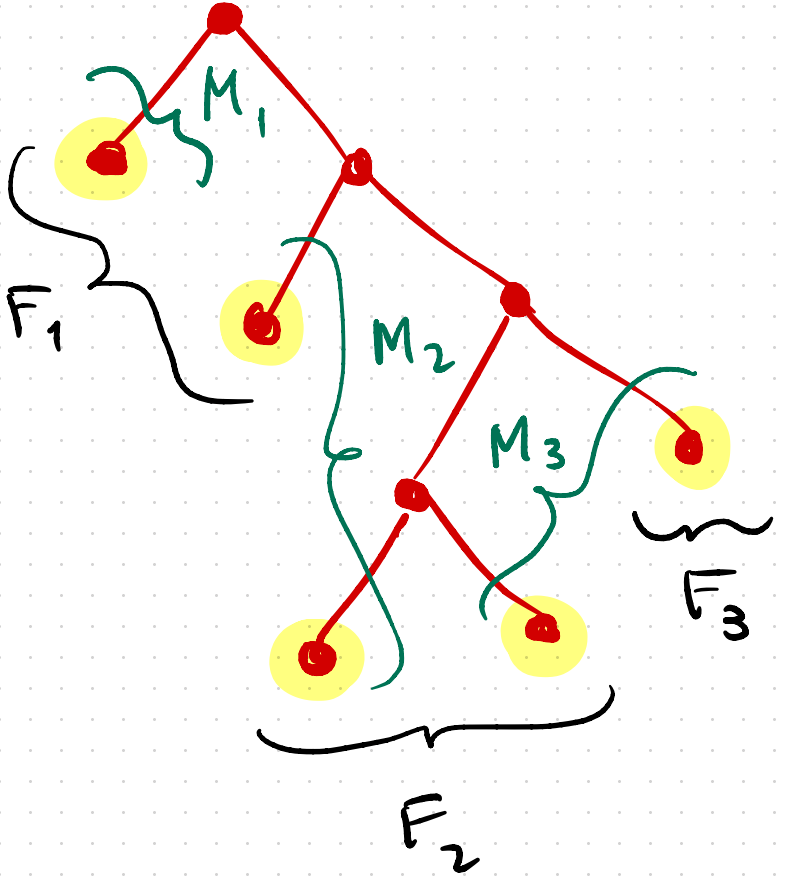
- they are, instead, fully abstract

- they come to be associated with FORM via many-to-one rules from syntactic terminals to exponents

- they come to be associated with MEANING via many-to-one rules from syntactic terminals to listed meanings



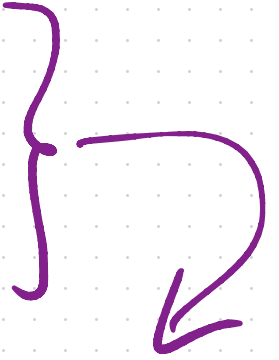
NB:
contiguity


M1
F1
M2
M3
F3
F2

FURTHER CONSEQUENCES:

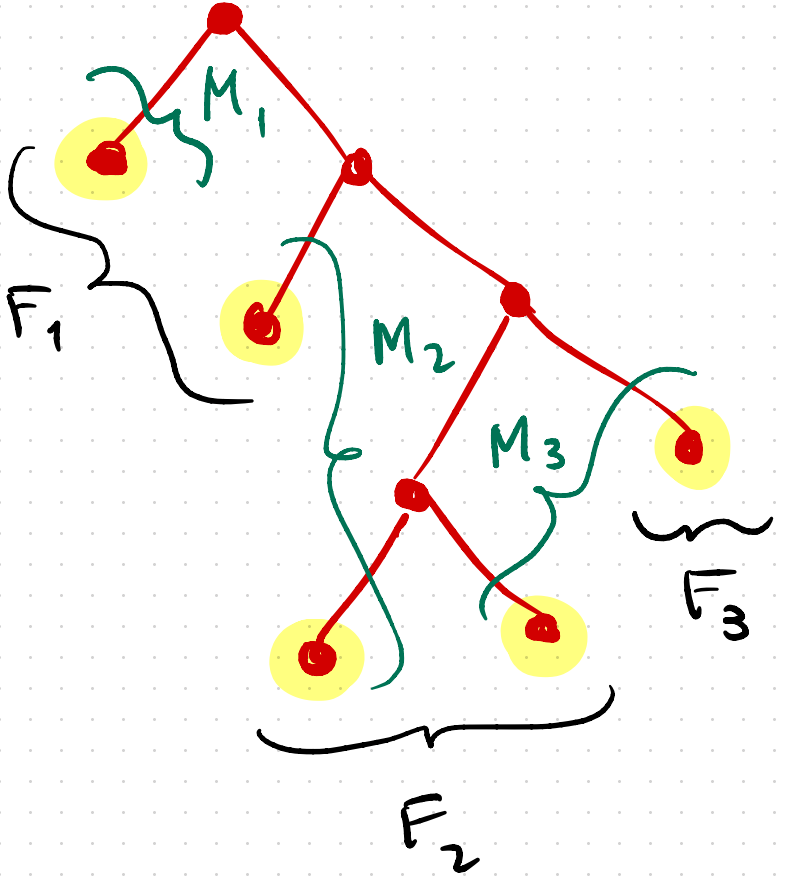
"What does the word(/morpheme) *w* mean?"

"How do speakers (of this language) pronounce the meaning *m*?"



NOT, STRICTLY SPEAKING, COHERENT QUESTIONS!

(because words/morphemes aren't interpreted, and meanings aren't pronounced)


M1
F1
M2
M3
F3
F2

PRELIMINARIES:

(I) The term "word" is not useful in the context of FORM-MEANING relations. *(Marantz 2001, i.a.)*

PRELIMINARIES:

(I) The term "word" is not useful in the context of FORM-MEANING relations. *(Marantz 2001, i.a.)*

(a) "phonoligical word" is not suited to serve as the relevant notion of "word"

There's (probably) such a thing as "phonological words" —

but phonological words can correspond to composed meanings:

[ðə.ˈdɔg]
“the dog”

they need not even be constituents:

[ðə.ˈskajd.bij.gɹej]
“The sky’d be grey.”

PRELIMINARIES:

(I) The term "word" is not useful in the context of FORM-MEANING relations. *(Marantz 2001, i.a.)*

✓ (a) "phonoligical word" is not suited to serve as the relevant notion of "word"

(b) and neither is "orthographic word"

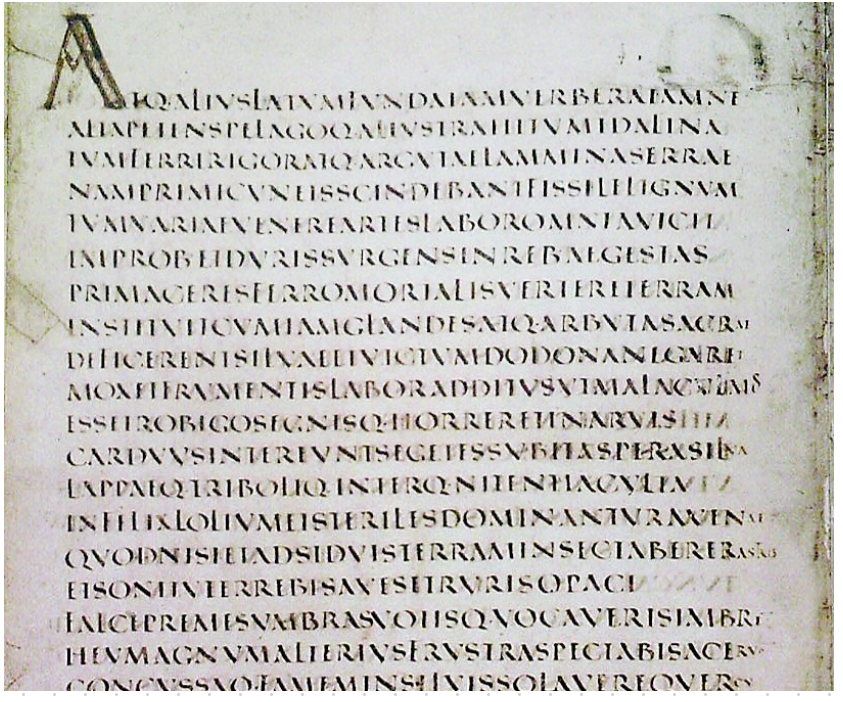
There are (sometimes) such things as orthographic words…

BUT:

- speaks English
- doesn't know how to read/write

⇒ doesn't know "words"?

⇒ doesn't know units of FORM-MEANING correspondence?



"scriptio continua"

Many writing systems (incl. early Latin & Greek) lacked spaces altogether

⇒ no "words"...?

Furthermore:

- The writing system for modern-day Vietnamese, for example, has spaces – but they individuate ~syllable-sized units

  > smaller than anything that could realistically be called "word" in the language (Noyer 1998)

- and, of course, not every natural language even has a writing system

PRELIMINARIES:

(I) The term "word" is not useful in the context of FORM-MEANING relations. *(Marantz 2001, i.a.)*

✓ (a) "phonoligical word" is not suited to serve as the relevant notion of "word"

✓ (b) and neither is "orthographic word"

At this juncture, one typically launches a final attack on any remaining, "intuitive" notion of word (see, e.g., Marantz 2001).

MEANING: "chew the fat" (cf. *chew*, *the*, *fat*)
"believable" (cf. *believe*, *-able*)
"terrific" (cf. *terrify*, *-ic*)

FORM: "went" (cf. *go*)
"ownership" (cf. *owner*, *-ship*)
"cat" (cf. *cap*, *hat*, …)

But I've come to believe that this is completely unnecessary —

> *In science, we do not need to refute intuitive, nebulous "proto-theories" based on folk-scientific notions.*

Unless & until someone presents an explicit, non-phonological non-orthographic definition of "word" that is not post-hoc...

PRELIMINARIES:

✓ (I) The term "word" is not useful in the context of FORM-MEANING relations. *(Marantz 2001, i.a.)*

- ✓ (a) "phonoligical word" is not suited to serve as the relevant notion of "word"
- ✓ (b) and neither is "orthographic word"

PRELIMINARIES:

✓ (I) The term "word" is not useful in the context of FORM-MEANING relations. *(Marantz 2001, i.a.)*

(II) Morphological exponents cannot serve as units of FORM-MEANING mapping, either.

(Aronoff 1976, i.a.)

(a) Just like "chew the fat" requires X-MEANING mapping where X > "word"...

it also requires X-MEANING mapping where X > morphological exponent

(b) And so does "terrific" (cf. *terrify, -ic*).

(c) suppletion:

go – went —— What's the FORM side of the FORM-MEANING mapping, here?

Anishinaabemowin (Algonquian); Sigwan Thivierge, p.c.:

miskomin-**ag**	ni-gii-amw-aa-**ag**
raspberry.ANIM-ANIM.PL	1-PST-eat.TA-DIR-ANIM.PL
'I ate raspberries.'	
miin-**an**	ni-gii-miiji-n-Ø-**an**
blueberry.INAN-INAN.PL	1-PST-eat.TI-TI3-INAN.PL
'I ate blueberries.'	

(d) forms without meaning:

complete ~ completion
compete ~ *competion (cf. *competition*)

⟹ What is this "extra" *-ti/-it*?
In particular: what does it MEAN?

"Just morphology"…? Not quite…

(Harley 2006)

(d’) in cahoots

short shrift

spick and span

(cf. competition)

(Noyer 1998, Harley 2006)

PRELIMINARIES:

✓ (I) The term "word" is not useful in the context of FORM-MEANING relations. *(Marantz 2001, i.a.)*

✓ (II) Morphological exponents cannot serve as units of FORM-MEANING mapping, either. *(Aronoff 1976, i.a.)*

A Methodological Note:

The discussion of MEANING so far has mostly been about open-class items.

Whereas most formal semantics these days is about closed-class items.

$\Longrightarrow$ *Problem…?*

No.

The focus on closed-class items in formal semantics is merely a *heuristic* choice.

CENTRAL IDEA:
Open-class items (dog, beauty) will involve the same principles & mechanisms as closed-class items (every, the). But we have a better guess for what the latter mean…

⟹ Thus, by parity of reasoning:

If we're able to learn something about interpretation & meaning from open-class items —

> It should be taken to be general, as well, and apply to closed-class items too.

Key Data

"go off" ~ explode, be triggered

"go"$_{\text{NONPAST}}$ ~ "went"$_{\text{PAST}}$

"went off" ~ exploded, was triggered

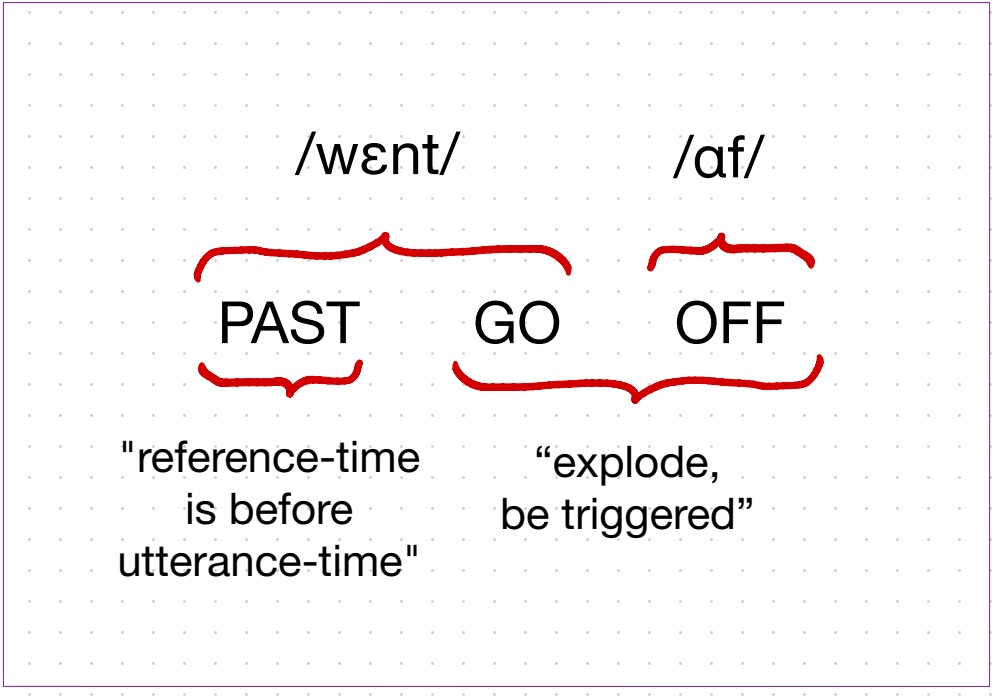
syntactic elements – at minimum:

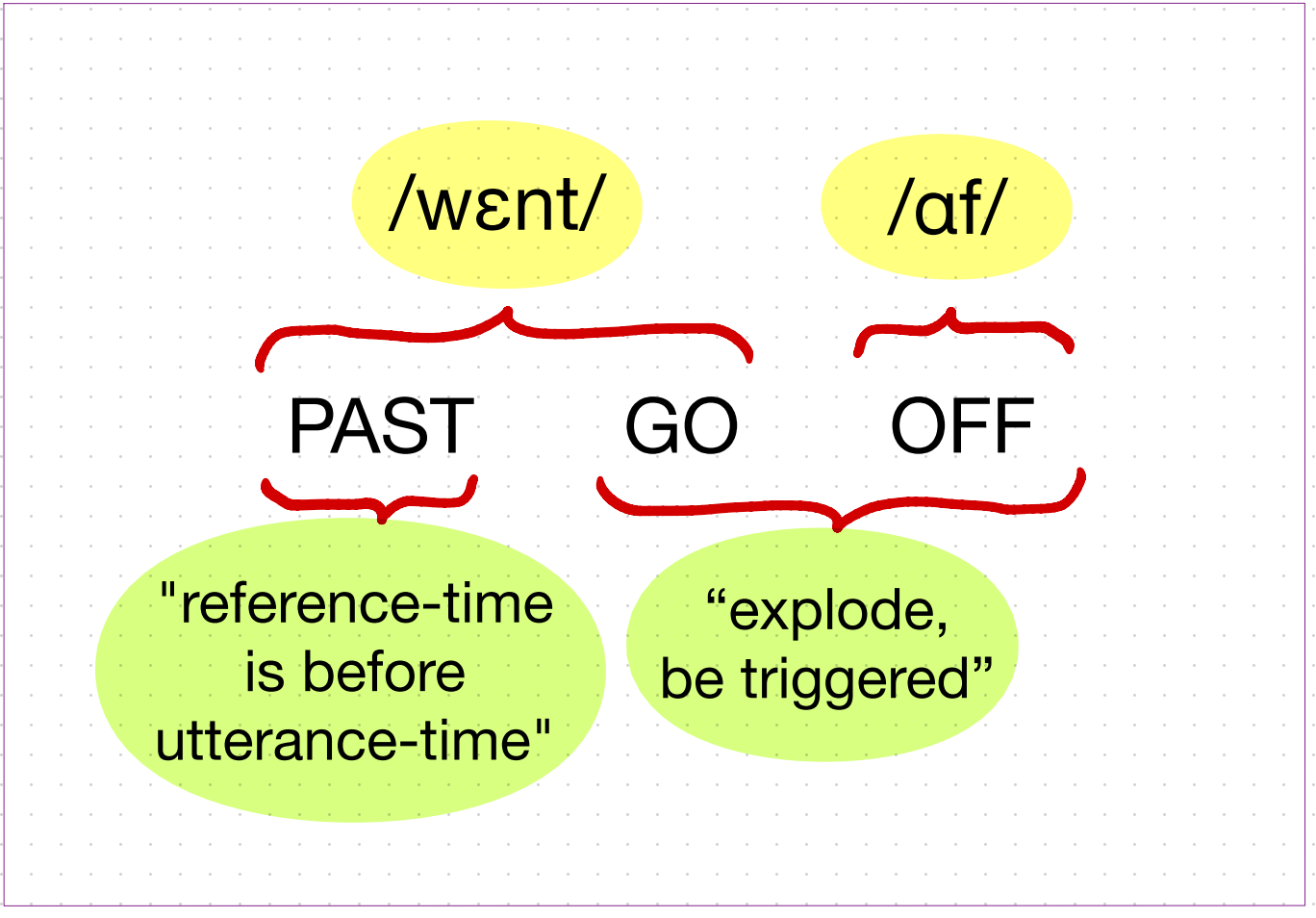
PAST ~ T or Infl or … bearing [+PAST] features

GO ~ whatever it is that distinguishes the verb "go" from "run", "dance", etc.

OFF ~ whatever it is that distinguishes the preposition/particle "off" from "on", "up", "in", etc.

mappings from syntax to FORM and to MEANING:

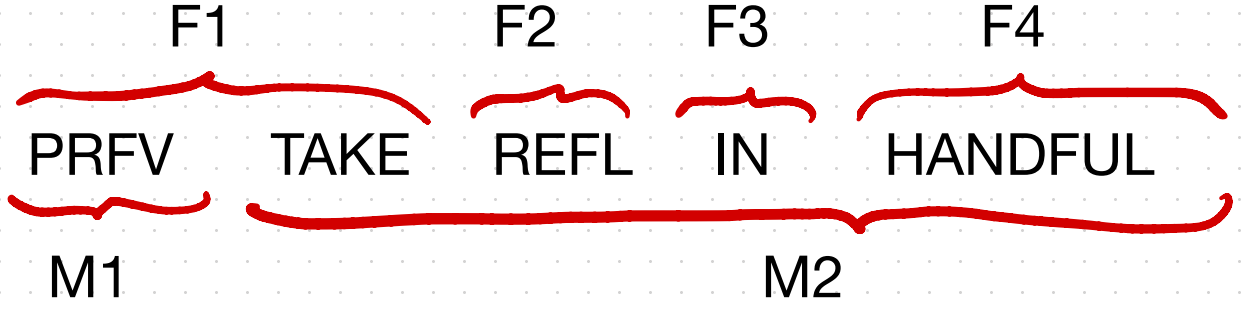



/wɛnt/
/ɑf/
PAST
GO
OFF
"reference-time is before utterance-time"
“explode, be triggered”

Polish (Slavic);
Asia Pietraszko, p.c.:

a. Bierz się w garść!
take.IMPF.IMP.2SG REFL in handful
'Pull yourself together *(imperfective)*!'

b. Weź się w garść!
take.PRFV.IMP.2SG REFL in handful
'Pull yourself together *(perfective)*!'



German (Germanic);
Hagen Blix, p.c.:

a. mit jemand-em auf gut-em Fuß stehen
with someone-DAT on good-DAT foot stand
'to get along well with someone'

b. Wir standen damals auf bess-er-em Fuß als heute.
we stand.PAST.3PL back.then on good.CMPR-CPMR-DAT foot than today
'Back then, we got along better than today.'

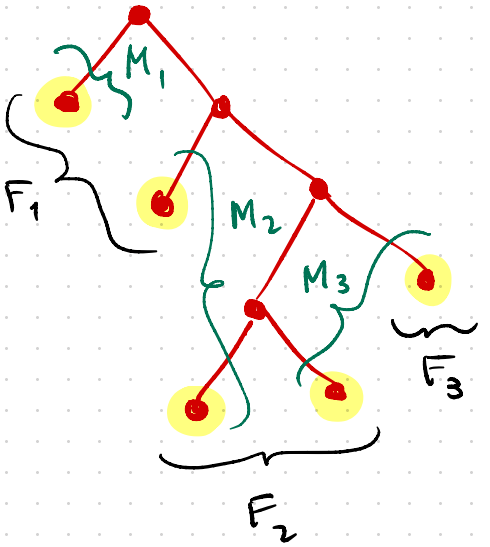
F5 F6

Architecture

(1) fully abstract syntactic atoms (e.g. PAST, STAND, IN, etc.)

(2) many-to-one rules from sets of nodes in (1) to units of FORM

(3) many-to-one rules from sets of nodes in (1) to units of MEANING



NB:
contiguity

What is "lexical acquisition" on this type of model?

traditionally: the child learns a "word" — its form(s), its meaning(s), and its syntactic properties

What is "lexical acquisition" on this type of model?

traditionally: the child learns a "word" — its form(s), its meaning(s), and its syntactic properties

That's not a thing…

⟹ what does “learning /ˈtej.bl̩/”

or “learning  ”

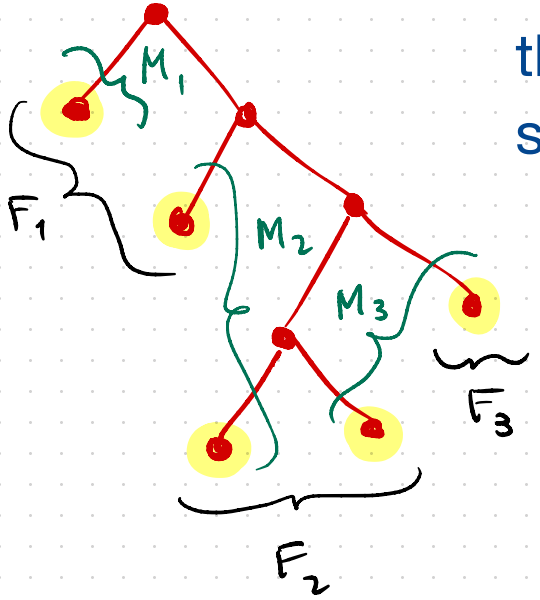
or …

amount to, in the proposed architecture?

Let's make the simplifying assumption that the child has successfully done "morphological segmentation"—

i.e., division of the incoming speech stream into morphological exponents

This means the child has successfully identified that they heard the sequence "F1 F2 F3"



But this still radically under-determines the *structure* that could have spelled out —

and even more so the meanings that this structure could have been associated with

Cognition. 2001 Sep;81(2):B33-44. doi: 10.1016/s0010-0277(01)00122-6.

The role of exposure to isolated words in early vocabulary development

M R Brent [1], J M Siskind

Affiliations + expand

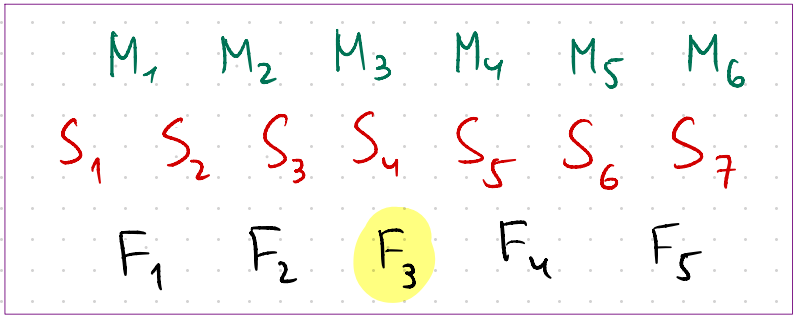
PMID: 11376642 DOI: 10.1016/s0010-0277(01)00122-6

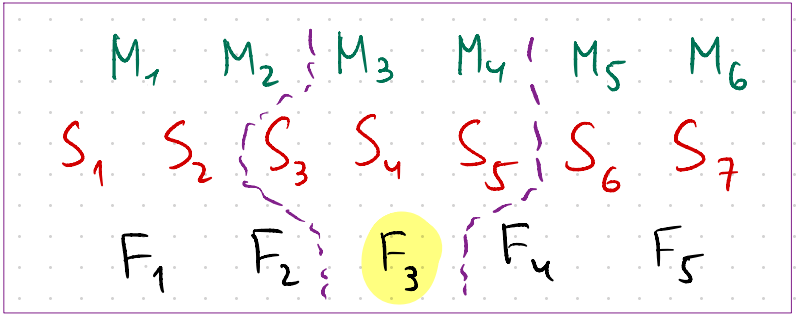
Abstract

Fluent speech contains no known acoustic analog of the blank spaces between printed words. Early research presumed that word learning is driven primarily by exposure to isolated words. In the last decade there has been a shift to the view that exposure to isolated words is unreliable and plays little if any role in early word learning. This study revisits the role of isolated words. The results show (a) that isolated words are a reliable feature of speech to infants, (b) that they include a variety of word types, many of which are repeated in close temporal proximity, (c) that a substantial fraction of the words infants produce are words that mothers speak in isolation, and (d) that the frequency with which a child hears a word in isolation predicts whether that word will be learned better than the child's total frequency of exposure to that word. Thus, exposure to isolated words may significantly facilitate vocabulary development at its earliest stages.

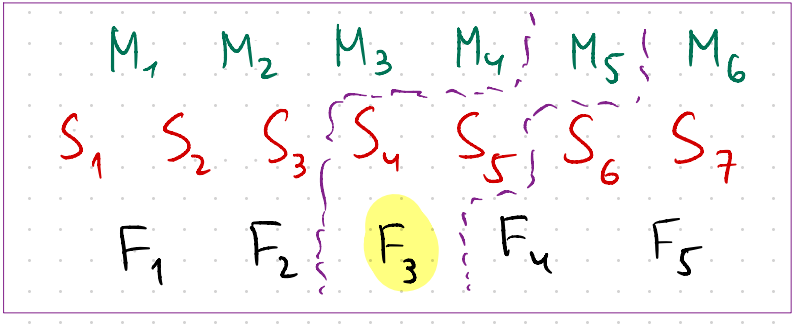
a variety of word types, many of which are repeated in close temporal proximity, (c) that a substantial fraction of the words infants produce are words that mothers speak in isolation, and (d) that the frequency with which a child hears a word in isolation predicts whether that word will be learned better than the child's total frequency of exposure to that word. Thus, exposure to isolated words may significantly facilitate vocabulary development at its earliest stages.

Why?

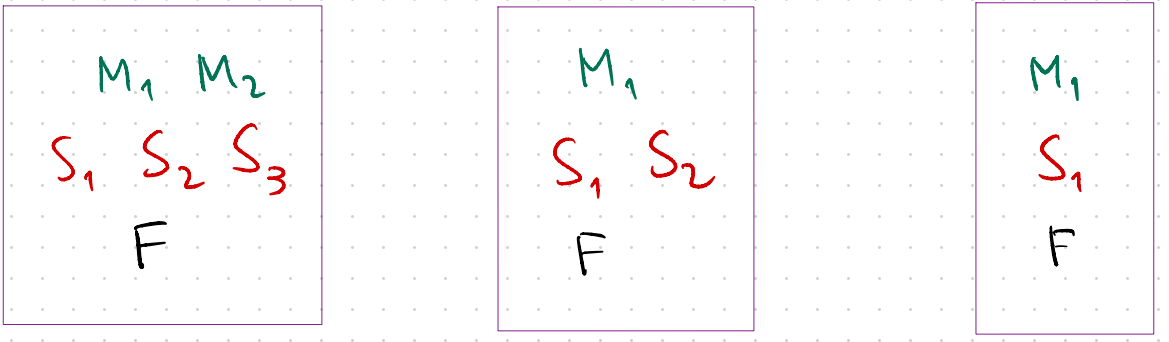

M1 M2 M3 M4 M5 M6
S1 S2 S3 S4 S5 S6 S7
F1 F2 F3 F4 F5

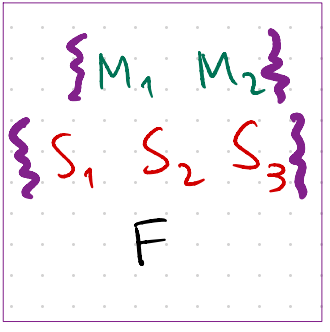

M1 M2 M3 M4 M5 M6
S1 S2 S3 S4 S5 S6 S7
F1 F2 F3 F4 F5

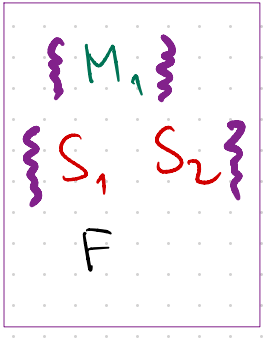
?

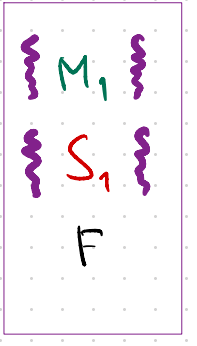

M1 M2 M3 M4 M5 M6
S1 S2 S3 S4 S5 S6 S7
F1 F2 F3 F4 F5

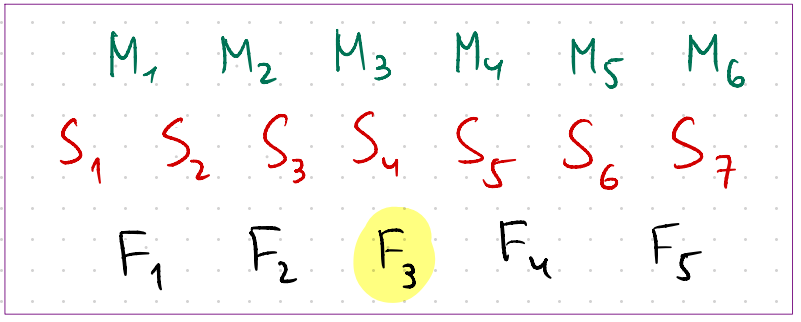
?

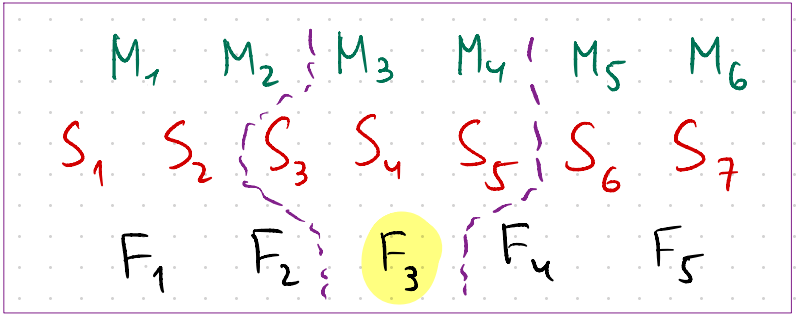

M1 M2
S1 S2 S3
F
M1
S1 S2
F
M1
S1
F


M1 M2
S1 S2 S3
F

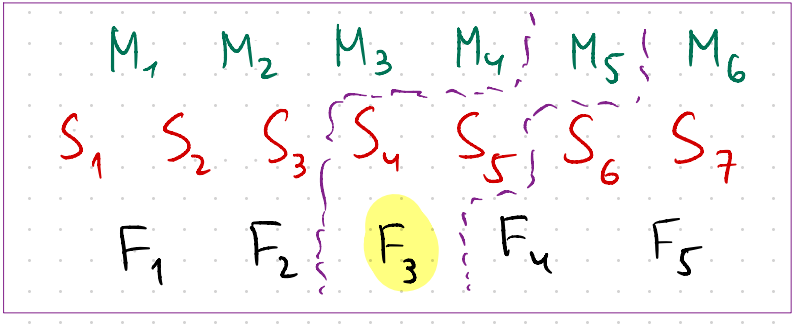

M1
S1 S2
F


M1
S1
F


M1 M2 M3 M4 M5 M6
S1 S2 S3 S4 S5 S6 S7
F1 F2 F3 F4 F5


M1 M2 M3 M4 M5 M6
S1 S2 S3 S4 S5 S6 S7
F1 F2 F3 F4 F5

?


M1 M2 M3 M4 M5 M6
S1 S2 S3 S4 S5 S6 S7
F1 F2 F3 F4 F5

?

Child Dev. 2018 Jul;89(4):1247-1267. doi: 10.1111/cdev.12731. Epub 2017 Feb 1.

Quantitative Linguistic Predictors of Infants' Learning of Specific English Words

Daniel Swingley [1], Colman Humphrey [1]

Affiliations + expand

PMID: 28146333 PMCID: PMC5538897 DOI: 10.1111/cdev.12731

Table 5

Regression Coefficients and Descriptive Statistics of Significant Predictors in the Word-Saying Analysis

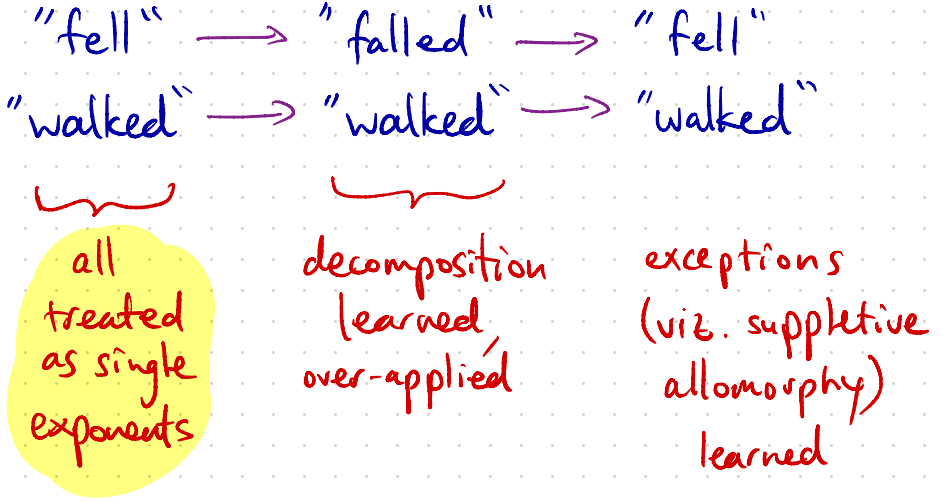
Predictor	Coef	Exp(coef)	IQR	90–10R	*p* Value
Total frequency.c	0.2754	1.3171	2.77	4.01	.0005
Isolated freq.c	0.5197	1.6815	0.00	1.10	.0005
MLU.c	−0.1147	0.8917	1.00	3.00	.0527
Duration ratio.c	0.2922	1.3393	0.53	2.42	.1233
Class(closed)	−0.7239	0.4848	na	na	.1709
Class(pred.)	−1.5665	0.2088	na	na	.0031

Note. Coef refers to the estimated beta coefficient. Exp(coef) provides the number by which the odds of saying a word should be multiplied given an increase of 1 in the predictor's value. IQR (interquartile range) is the difference in value between the 75th and 25th percentiles for values of the numerical predictors. 90–10R is like the IQR but uses the 90th and 10th percentiles. MLU = mean length of utterance.

⟹ Learners attempt to "penetrate" this massive many-to-many-to-many mapping problem by establishing single-exponent (or low-number-of-exponent) foot-holds

As evinced by their over-reliance on fragmentary ("one-word") utterances.

In essence, this is the single-item bias familiar from well-known developmental trajectories like the following:



Many-to-one mappings: rare?

At this juncture, a potential concern:
are we reducing-to-the-worst-case based on a handful of "unusual" examples?

(1)	a. /k-b-ʃ/ + CaCuC	kvuʃim ‘pickles’	(Hebrew)
	b. /k-b-ʃ/ + CCiC	kviʃ ‘road’	
	c. /k-b-ʃ/ + Ci(C)CuC	kibuʃ ‘conquest’	*Aronoff 2007*

(2)	a. /x-ʃ-b/ + CaCaC	xaʃav ‘think’
	b. /x-ʃ-b/ + CiC(C)eC	xiʃev ‘calculate’
	c. /x-ʃ-b/ + hiCCiC	hixʃiv ‘consider’

NB1: Every instance of composition that is not *exclusively phonological* or *exclusively semantic* is **syntactic**.

NB2: NB1 is not an "assumption" — it's the only game in town (unless & until someone comes up with a working, cross-linguistic definition of "word"… *don't hold your breath!*)

⟹ Pretty much every open-class item in Semitic involves a *joint* mapping

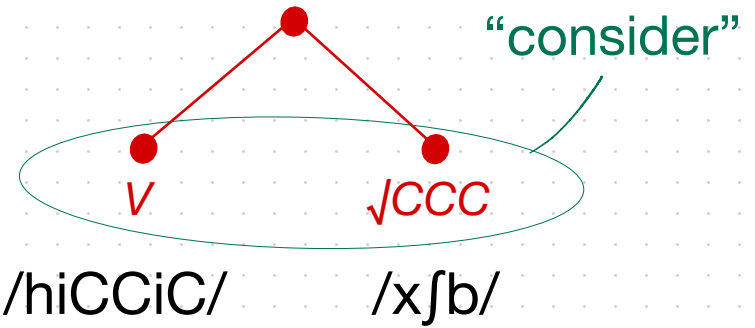
from at least two syntactic terminals – the √CCC root, and the *n*/*v*/etc. associated with the template – to a meaning

AND REMEMBER:

Thus, by parity of reasoning:

If we're able to learn something about interpretation & meaning from open-class items —

> It should be taken to be general, as well, and apply to closed-class items too.



More evidence: gaps, gaps, gaps

in cahoots
newfangled
short shrift
huckleberry
spick and span

cf.:
* s-cahoot in
* shrift short
* spick span and

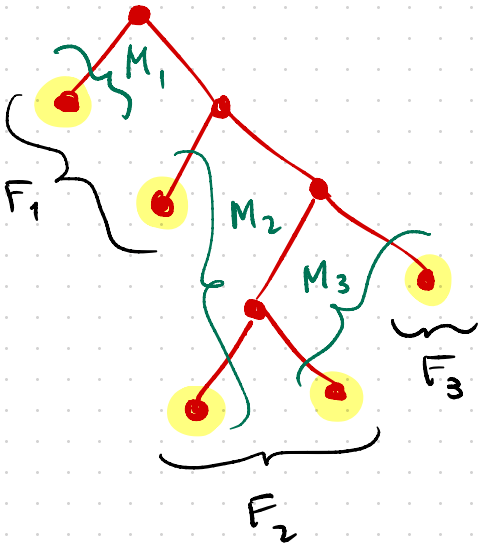
(Noyer 1998, Harley 2006)

Architecture

(1) fully abstract syntactic atoms (e.g. PAST, STAND, IN, etc.)

(2) many-to-one rules from sets of nodes in (1) to units of FORM

(3) many-to-one rules from sets of nodes in (1) to units of MEANING



NB:
contiguity

{ √CAHOOT } → ✗

{ *n*, √CAHOOT } → ✗

-
-
-

{ IN, D[-def], Num[pl], *n*, √CAHOOT } → “engaged in a conspiracy” ✓

Conclusions:

- There are no "words" (in any non-phonological, non-orthographic sense of the term) *(Marantz 2001, i.a.)*
- Morphological exponents don't map onto units of meaning *(Aronoff 1976, i.a.)*
- Instead, the architecture of human language involves...

(1) fully abstract syntactic atoms
(e.g. PAST, STAND, IN, etc.)

(2) many-to-one rules from sets of
nodes in (1) to units of FORM

(3) many-to-one rules from sets of
nodes in (1) to units of MEANING

None of (1)/(2)/(3) are anything that even could have existed outside of/prior to human language

(cf., for example, vervet monkey calls)

⟹ Chomsky's "Strong Minimalist Thesis" (SMT) —

the claim that MERGE is the only
linguistically-proprietary cognitive capacity

— is demonstrably false.

Thank You!

Muito Obrigado!