

# Mythbusting: “Words”

Omer Preminger



UMD Department of Linguistics &  
Maryland Language Science Center



Commencement, May 2017

## *Hello!*

- My name is Omer Preminger
- I'm a faculty member here in the Department of Linguistics at the University of Maryland
- Among the courses I teach are undergraduate *Syntax II*, as well as the course in *Linguistic Field Methods*
- I work on syntax, mostly in languages that are understudied / underdocumented



## Doing science with humans on language

- We are all experienced users of language
- This both *helps* and *hinders* research on language
- It *helps*, because we have readily available “data generators” (people!) whenever we want to explore something or test a hypothesis
- It *hinders*, because as language users, we have all kinds of preconceptions about language
  - which are just that — **preconceptions**(For the sake of comparison, imagine if everyone who had a working heart thought themselves to be an instant cardiologist.)
- ➔ Part of what we try to teach our students in the LING program is to think critically about human language
  - and to distinguish facts from preconceptions

## A little bit about what we do here *(cont.)*

- In that vein, I would like to invite you all to think critically with me, about something we all grew up thinking is central to language:

*words*

## *Do you know any “words”?*

- Everybody who’s sitting here probably thinks they know a whole bunch of words!
- It might interest you, then, that *nobody in the history of human thought has ever come up with a working definition for what a “word” is.*
- To give you a taste of the problem, let’s review a few common-sense attempts at an answer to the question, *What is a “word”?*

# Some attempts

(i) word  $\equiv$  “the thing we write between spaces/punctuation”

- PROBLEM 1:

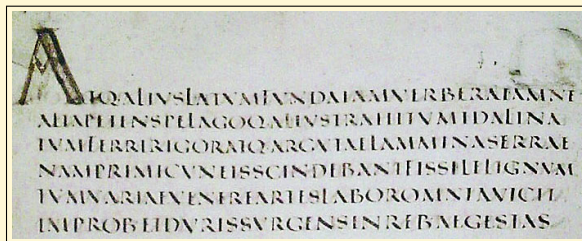


- doesn't know how to read or write  
⇒ *doesn't know words?!*

# Some attempts

(i) word  $\equiv$  “the thing we write between spaces/punctuation”

- PROBLEM 2:



[<http://anglandicus.blogspot.com/2013/06/irish-scribal-habits-scriptio-continua.html>]

- many writing systems, incl. early Latin & Greek, did not use spaces (a.k.a. “scriptio continua”)
  - $\Rightarrow$  *the authors of these texts didn't know words?!*



## Some attempts

(i) word  $\equiv$  “the thing we write between spaces/punctuation”

- PROBLEM 2:

- this is not just about ancient scripts, by the way —
  - modern-day Chinese script lacks anything we would identify as a “word boundaries”
  - modern-day Vietnamese script has spaces, but they are between every pair of syllables
    - meaning that the written units are smaller than what we’d identify as “words”

(and there are many other examples of this sort)

## Some attempts

(i) word  $\equiv$  “the thing we write between spaces/punctuation”

- PROBLEM 3:



*“me again...”*

“When I finally learn to write, I am going to have a good deal of trouble learning where spaces go and where they don’t.”

“This, despite the fact that I allegedly already ‘knew words’ before learning to write.”

# Some attempts

(i) word  $\equiv$  “the thing we write between spaces/punctuation”

- PROBLEM 4:



- *today* used to be “two words” (*to-day*) in the 19th century, and is now “one word” (one of many such examples)
  - but it is not clear that anything has changed about this expression, other than how it’s written

## Some attempts

(i) **word**  $\equiv$  “the thing we write between spaces/punctuation”

- PROBLEM 5:

**Wa-hake-natar-a-kwétar- $\Lambda$ -'. MOHAWK  
FACT-MS/ISO-bread- $\emptyset$ -cut-BEN-PUNC  
'He cut the bread for me.'**

[Baker 1996:27]

- There are languages in which there's no natural counterpart to our Anglo-centric (or Indo-European-centric) notion of “word”
  - and this doesn't seem to be a problem for the, um, language-ness of these languages

## Some attempts (*cont.*)

So let's try something else...

(ii) **word**  $\equiv$  “the smallest meaningful unit of language”

- This initially looks quite promising
    - *exam* means something, but *-xa-* doesn't seem to
  - But this approach runs into problems pretty quickly:
    - *believable* is a “word,” right?
      - but we can't really say that *believable* doesn't have meaningful subparts —
        - namely, *believ(e)* + *-able*  
(the latter meaning something like “*able to be X'ed*”)
- $\Rightarrow$  “words” (can) have meaningful subparts

## Some attempts (*cont.*)

(ii) word  $\equiv$  “the smallest meaningful unit of language”

- In fact, things get even worse for this definition —
  - you might try to save the previous example by saying something like: “*believe* is itself a word; that’s why *believable* is a complex entity”
- but what about the *atroc(i)-* part in *atrocious* and *atrocitiy*?
  - clearly, there is a shared, meaningful subpart among these two “words”  
(after all, we wouldn’t want to say that it’s an accident that they both start with *atroc(i)-*)
  - but since *atroc(i)-* is not a “word” —
    - **we now must admit that there are meaningful subparts to “words”**

## Some attempts (*cont.*)

- (ii) word ≡ “the smallest meaningful unit of language”
- On the flip side, there are things we would have to call “words” that don’t seem to have any meaning at all...!
    - *the whole kit and caboodle*
    - *to and fro*
    - *if I had my druthers*
    - *run the gamut*

[Harley 2006:10]

## Some attempts (*cont.*)

(ii) word  $\equiv$  “the smallest meaningful unit of language”

- In fact, this phenomenon seems to occur both above and below the “word” level:
  - *complet(e)- + -ion*  $\Rightarrow$  *completion*  
but:
    - *compet(e)- + -ion*  $\not\Rightarrow$  *competition*
- Instead:
  - *compet(e)- + -it + -ion*  $\Rightarrow$  *competition*
- ➔ But what is this “-it”?
  - **ANSWER:** (Harley 2006)  
It is a *caboodle* — a bit of language that is only meaningful in the context of the larger expression it is part of



## A few more relevant observations

- If all we have is these strings of meaningful little elements (with or without spaces in between them) —
  - maybe what's special about “words” is that they are those strings of meaningful pieces that we've encountered before
    - or that we are used to encounteringlike: *breakable*, *blackboard*, *commencement*, and so forth
- **nope:**
  - *any friend of yours is a friend of mine*
  - *use only as directed*
  - *may the force be with you*... and so forth

[Jackendoff 1997; Marantz 2001]

## A few more relevant observations (cont.)

- maybe what's special about “words” is that they can combine in creative/unexpected ways
  - *kick the ball* means what we expect it to mean
    - given the pieces, *kick* + *the* + *ball*
  - but *kick the bucket* does not
    - creative/unexpected
- **nope:**
  - *horrif(y)- + -ic*  $\Rightarrow$  *horrific*
    - as we would expect, given its parts
  - but *terrif(y)- + -ic*  $\Rightarrow$  *terrific*
    - creative/unexpected!

# Consequences

The bottom line is that in an expression like “*the commencement*”:

- There’s no qualitative difference between —
  - how *-ment* relates to *commence*and —
  - how *the* relates to *commencement*

Or, to put it in the parlance of our times: **#NoWords**

(Except in the obvious sense of written text, in some languages.)

## Consequences *(cont.)*

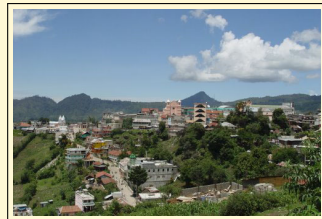
*Why is this important?*

Once we don't presuppose that language must be shoehorned into the preconceived template of “words” — interesting things happen:

- linguistic communities that happened to develop writing systems different from ours (e.g. Vietnamese) start looking less ‘exotic’
- removing these illusory differences can reveal hidden commonalities
  - *for example...*

## Consequences *(cont.)*

- Q'anjob'al is a Mayan language – spoken in Santa Eulalia (Guatemala) (among other places)



- Kaqchikel is another Mayan language – spoken in Patzún (Guatemala) (among other places)



## Consequences *(cont.)*

- Here's how you say “She saw you” in Q'anjob'al:

*maxach yila'*

- And here's how you say “She saw you” in Kaqchikel:

*xarutz'et*

⇒ *Looks different, right...?*

- Not if you break these down into their parts, ignoring so-called “words”:

(1) max-ach y-ila'  
PAST-you he/she-see

(2) x-a-ru-tz'et  
PAST-you-he/she-see

•➡ **same pieces, in the same order**

- and note: the writing systems for these languages are very late, colonial/post-colonial additions

## Take-home message

- As language users, we have all kinds of preconceptions about what language consists of
  - In particular, as speakers of Western European languages (like English):
    - we might think that the notion of *word* is central to what it means to know a language
  - But when examined critically, there seems to be nothing that *systematically* corresponds to where we put “spaces.”
- ⇒ **It is important to subject our common-sense notions about language to scientific investigations —**
- which sometimes yield results that are very different from what our initial, common-sense notions might suggest



- The same is true in all parts of linguistics that our students study —
  - **syntax & morphology** (the parts we just talked about)
  - **semantics & pragmatics**: the study of meaning, how the meaning of complex expressions arises from the meanings of their parts, and how language interacts with context
  - **phonetics & phonology**: the study of linguistic sounds and sound systems
  - **language acquisition**: the study of how children learn their native language(s)
  - **psycholinguistics**: the study of how language is processed and produced in real time
  - **neurolinguistics**: the study of how the human brain processes and produces language
  - **computational linguistics**: the formal and quantitatively precise modeling of linguistic knowledge and linguistic processes

❁ THANK YOU ❁

*And congratulations to our  
newly-minted graduates!!!*