Course summary, and some new simulation results

In the form of a dialogue between Proserpina, who favors p-words, and Antigone, who thinks them unnecessary (or, Tevye's monologue but with phonological theories instead of sons-in-law). These two proseminar participants sum up the quarter's findings.[1]

1. Proserpina

We can agree that we need to be able to refer to something like a word. E.g.

- "final" devoicing—at the *fin* of what?
- phonotactic restrictions—don't seem to hold across units larger than a word

But I additionally claim that

- the p-word is not identical to the syntactic word
- p-word-formation depends, at least in part, on purely phonological factors

Together, these claims mean that **the phonological grammar determines where the p-word boundaries go**—taking into account nonphonological information like morpheme boundaries, maybe syntactic structure (remember *Hiawatha*), maybe processing outcomes.

2. Antigone

If the p-word is the same as the syntactic word, it's the syntax's job to identify words, and there's nothing specifically phonological going on.

E.g., Embick & Noyer 2001's[2] *morphosyntactic word* (MWd) =
"X^0 [that] is the highest segment of an X^0 not contained in another X^0" (p. 574).

Depending on the architecture, we might also get a natural division between lexical and postlexical phonology.

3. *Proserpina*

But what if that morphosyntactic structure results from movement sensitive to phonological factors?

E&N discuss the Latin clitic *–que* 'and', which attaches after first MWd of second conjunct:

[bon+ī puer+ī] [bon+ae-**que** puell+ae]
good boys good-and girls
'good boys and good girls' (p. 575)

But when the second conjunct begins with a preposition, its syllable count matters:

[1] I didn't put bibliographic information for references previously cited this quarter. See bibliography on web page.
[2] David Embick & Rolf Noyer (2001). Movement operations after syntax. *Linguistic Inquiry* 32: 555-595.

circum-**que** ea loca	in rēbus-**que**
around-and those places	in things-and
'and around those places'	'and in things'
contrā-**que** lēgem	dē prōvinciā-**que**
against-and law	from province-and
'and against the law'	'and from the province' (p. 576)

E&N propose that local dislocation—merger that occurs during or after vocabular insertion, and thus can be sensitive to phonological content—combines monosyllabic prepositions with their complements.

So at least under this architecture (Distributed Morphology)—and possibly under any architecture that can handle the Latin facts—the notion "word" is not phonology-free.

We've seen other examples of phonological word ≠ syntactic word

phonological word < syntactic word
Dutch syllabifies C with a following V within stem+suffix (except *–achtig*), but not across prefix-stem or stem-stem boundary (Booij 2002):

(on).+(aar.d+ig)	'unkind'	
(loods).-(pet)	'sea-captain's cap'	
(lan[t]).-(a.del)	'landed gentry'	< land-adel

Italian has one main stress per word (as diagnosed by length, possibility of lax mid vowels), but compounds and certain prefixes (possibly analyzable as compounding) can have more than one main stress (Nespor & Vogel, Peperkamp):

(t[ò]sta-tóre)	'toaster'	< t[ɔ́]sta
(t[ɔ́]sta)-(páne)	'bread toaster' (toast bread)	
([ɛ́]kstra)-(coniugále)	'extramarital'	

4. *Antigone*

We could also analyze these cases as level-ordering: add the cohering affixes early, compound and add non-cohering late.

5. *Proserpina*

That can result in some bracketing paradoxes, though.

Dutch *on-grammatical-iteit* 'ungrammaticality' = $[[\text{on } [\text{grammaticaal}]_A\ ]_A \text{ iteit}]_N$

We want *on-* to attach late, but it's low in the morphological tree.

If I spoke Dutch, I might be able to concoct even worse bracketing paradoxes, where an enclitic has to attach before a prefix:

(ont).-(er.v-e.n't) 'disinherit it'?? (I'm sure that is wrong, because it got 0 Google hits)

A level-ordering explanation also leaves unexplained phonological regularities in what belongs to which level.

Dutch V-initial suffixes (except *–achtig*) are "early", and so are those that lack a full vowel, but other C-initial suffixes are "late":

(klach**t**-je) 'complaint' *t* is obligatorily deleted (*je* lacks a full V)
(zich**t**)-(baar) 'visible' *t* is only optionally deleted
(Booij 1995)

6. *Antigone*

But do those phonological regularities need to be captured in the phonological grammar? What if affixes that in earlier levels got there, diachronically, because they are (somehow) more difficult for learners to parse out?

7. *Proserpina*

We've also seen cases of **phonological word > syntactic word**

Dutch syllabifies enclitics (but not proclitics) with their hosts (Booij 1995):

/ət/ (Zij) (koch.t-e.n **'t**) (buk) If /n/ were p-word final, it would be able to delete.
(Ik) (merkte **'t**) (direkt) > [mɛrktət] or [mɛrktənət]
(wilde-[n]-**ik**) 'wanted I' vs. (**je** [ʔ](eet)) (lekker) 'you eat nicely'

8. *Antigone*

I guess those are harder. But we could be talking differences in clitic-group structure, not differences in p-word structure.

I wonder if a Bybeean approach could work in these cases: clitic/host combinations that are frequently used together develop a phonologically reduced representation (in the Dutch case, no glottal stop).

That would work if the host+enclitic sequences are more frequent than proclitic+host, or if we introduce an additional bias (somehow) towards beginning a new token, for purposes of lexicon-updating, whenever we hit a new content word—or, in the Dutch case, whenever we hit a stressed syllable (see Cutler & Norris 1988[3] on English): I bet most Dutch content-word tokens (and even types?) begin with a stressed syllable.

Let's tackle the second claim now. **If p-word-formation doesn't depend on phonological factors**—for example if it's all derivable from lexical access—then there's nothing specifically phonological going on.

[3] Anne Cutler & D. Norris (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14: 113-121.

For example, we can say that morpheme boundaries within strings that tend to get accessed as wholes get erased (*tosta+tore* > *tostatore*) or downgraded. Then the phonology applies, without prosodic structure.

9. *Proserpina*

Here are some cases we've seen in which phonological factors seem to play a role in setting p-word boundaries:

Minimality: Italian disyllabic prefixes seem to be separate, bearing their own primary stress (([ɛ́]kstra)-(coniugále)), but not monosyllabic prefixes (*pr[ɛ́]-X).

10. Antigone

But maybe there is a processing explanation here. Shorter affixes should be harder to identify as such because they're more likely to occur as accidental substrings (see Baroni's 2000[4] distributional learner of morphology).

If you've ever gone through a wordlist by hand to identify words containing certain prefixes, you'll recall the relief you feel when you get to a prefix like *anti-* (you can count on almost all words beginning with that string containing the prefix), and the despair when you get to *co-* (you have to exclude *cool*, *compare*, *cot*, ...).

11. *Proserpina*

Linear order (left/right asymmetries): In Dutch, left edge of root obligatorily starts new syllabification domain, and right edge only optionally so: *(stem)(stem)*, and *(prefix (stem))* no matter what the prefix's prosodic shape, but *(stem suffix)* or *((stem) suffix)*, depending on the suffix's prosody. Also *(clitic (host))* vs. *(host clitic)*.

12. Antigone

Because the speech signal proceeds forward in time, there could a parsability advantage to morphemes that occur early in the string (like prefixes). I'm not sure that this explains why compounds should be parsed *stem+stem*, though.

13. *Proserpina*

Sounds pretty vague to me, and as though it could easily not work—the prefix has an advantage in a prefixed word, but so does the stem in a suffixed word; in both cases, the complex parse is promoted. Have you implemented this?

14. Antigone

Not really. But see my Matcheck simulations below, with left-anchoring vs. both-edge-anchoring privileged.

[4] Marco Baroni (2000). Distributional Cues in Morpheme Discovery: A Computational Model and Empirical Evidence. UCLA Ph.D. dissertation.

15. *Proserpina*
Boundary signals: Raffelsiefen 1997 shows that loans into English from French—which presumably all came in with final stress—got parsed as complex only if their stress pattern was incompatible with a simple parse
refúse was fine (cf. native *forgíve*), but *reconvéne* > *rè-convéne*

16. Antigone
But that's not the phonological grammar determining p-word structure! Raffelsiefen's saying that the learner uses knowledge of the phonological grammar to make guesses about the structure of loans.

We could easily say that the learner is deciding where the morpheme boundaries are, not where the prosodic boundaries are.

17. *Proserpina*
Syllabification: compare Dutch V-initial suffixes with full vowels (incorporated into p-word to avoid hiatus) to C-initial suffixes with full vowels (separate) (Booij 1993—see also van Oostendorp 1994[5]).

(wan.de.l-aar) 'walker'
(vijf.)-(ling) 'quintuplet'

(zicht)-__ en (taast)-(bar) 'visi(ble) and tangible'
*absurd-__ en (banal-iteit) 'absurd(ity) and banality'

18. Antigone
If that's a real phenomenon, it's harder for me to account for.
But let me try attacking the evidence.
For syllabification, say that resyllabification after addition of a morpheme needs to resolve hiatus, but doesn't have to maximize onsets—this is van Oostendorp's proposal.
As for ellipsis, say that it's not allowed to break up a syllable in the original word.

19. *Proserpina*
I want to go back to what you said about using (erasable) morphological boundaries instead of prosodic structure.

It's hard to see how that would work for the Japanese compounding facts discussed by Itô & Mester (2003).

First, recall that rendaku happens only when (but not always when) the stem to the left lack accent:

[5] Marc van Oostendorp (1994). Affixation and integrity of syllable structure in Dutch. In *Linguistics in the Netherlands 1994*. Amsterdam: John Benjamins.

(yoshino **z**ákura) vs. (nihón)(**s**akura+mátsuri)

If we take deaccenting of *stem*$_1$ to indicate that *stem*$_1$ and *stem*$_2$ are together in some prosodic domain, then rendaku occurs only when the two stems are both in that domain. But rendaku moreover occurs only when there is a morpheme boundary!

20. Antigone

So there are boundaries of different strengths. Rendaku requires a weak boundary. Or, say that the domain of deaccenting is bigger, like the p-phrase. I'm not necessarily arguing against *all* prosodic structure here.

21. *Proserpina*

I mentioned above some of the tendencies we've seen in what determines p-word construction. How about what p-word structure then does?

Raffelsiefen proposed that truly segmental rules don't actually care about p-words. P-words determine only prosodic properties:

- metrical groupings, as in Hayes 1989
- stress and footing
  - English (índolent) vs. (ìn)(décent)
  - Bengali (ɔ́)-(fɔ̀l) vs. (ʃɔ́ɹol)
  - Kwara'ae metathesis (if we take the foot to be the domain of metathesis, and I know Jeff has argued that feet are not necessary there)
  - Guugu Yimidhirr vowel length
- syllabification
  - whether initial geminates get to surface in Bengali (ɔ́-ʃ.ʃos̪t̪i̪t̪e) vs. (bína)(ʃɔs̪t̪i̪t̪e)
  - English onset vs. coda allophones (ou[t̚])-(ride) vs. (ou.[t^{h}]ré)
  - hiatus resolution (or lack thereof) in various languages

But not truly segmental rules.

- We saw that Turkish vowel harmony domains don't line up well with stress facts.
- Raffelsiefen shows that English /n/ assimilation (*ir*regular) doesn't correlate with any other properties.
- But what about the Dutch rule that affects long vowels before (heterosyllabic) [r], indicated by underlining: (k**eu**.r-ing) 'test' vs. (k**eu**)-(ring) 'cue ring'. Presumably the syllabification is the same in both cases, yet the rule applies differently.
- And what about Sanskrit nati, whose trigger ({ṣ, r, ṛ, ṝ}) and target (n) have to be in the same p-word (Selkirk 1980)—syllabification doesn't seem to play a role there.

(stem suffix)$_\omega$	ka**r**ma**n**+**ā**	>	karma**ṇ**ā
	dū**ṣ**+a**n**am	>	dūṣa**ṇ**am
	b**ṛ**mh+a**n**am	>	bṛmh+a**ṇ**am
(stem)$_\omega$(stem)$_\omega$	b**r**aḥma**n** - **y**aḥ	>	braḥma**n**yaḥ
	k**ṣ**ip - **nu**ḥ	>	kṣip**n**uḥ

What other cases have we seen of rules that look segmental?

22. Antigone

Maybe we can call the Dutch and Sanskrit cases lexical rules, that apply before compounding.

23. *Proserpina*

Anything else you want to add?

24. Antigone

Yes. I promised a simulation earlier, so here it is. The idea is to use Matcheck to generate parses (simple or complex) for words, and see how the winning parse compares to the putative prosodic structure. This gives a non-phonological baseline against which claims of phonological involvement can be compared.

(1) Matcheck review

(Baayen et al. 2000, Baayen & Schreuder 2000)

Matcheck models the timecourse of word recognition.

- When a lexical entry reaches an activation threshold (actually, when its share of the total activation, $p(w, t)$, reaches a threshold), it is copied into a memory buffer.
- When a set of lexical entries that span the target word (e.g. *be*, *stel*, *auto*) exists in the buffer, that parse of the target becomes available (*be*+*stel*+*auto*).
- First full parse to become available is assumed to have "won" (but we could also assign interpretations to other information, such as how long it took the first parse to become available, and which other parses become available when).

What determines activation?

Resting activation is just frequency. At item$_i$'s activation increases at each timestep until it reaches threshold or, if it$_i$'s not edge-aligned, until t is greater than it$_i$s similarity to the target. Then activation gradually falls back to the resting level.

Amount of increase is a function of the item's length, resting activation, and how much of the target it matches (and some free parameters).

(2) Example

English lexicon used was BNC, unlemmatized conversation. Affix frequencies were just the sums of the frequencies of the words that contain those affixes (by my unscrutinized intuition.)

Caution 1: the representations used are orthographic! (and indeed, Matcheck was developed as a model of visual word recognition).

Caution 2: the lexicon is somewhat limited, so a lot of words couldn't be parsed at all.

Parameters:

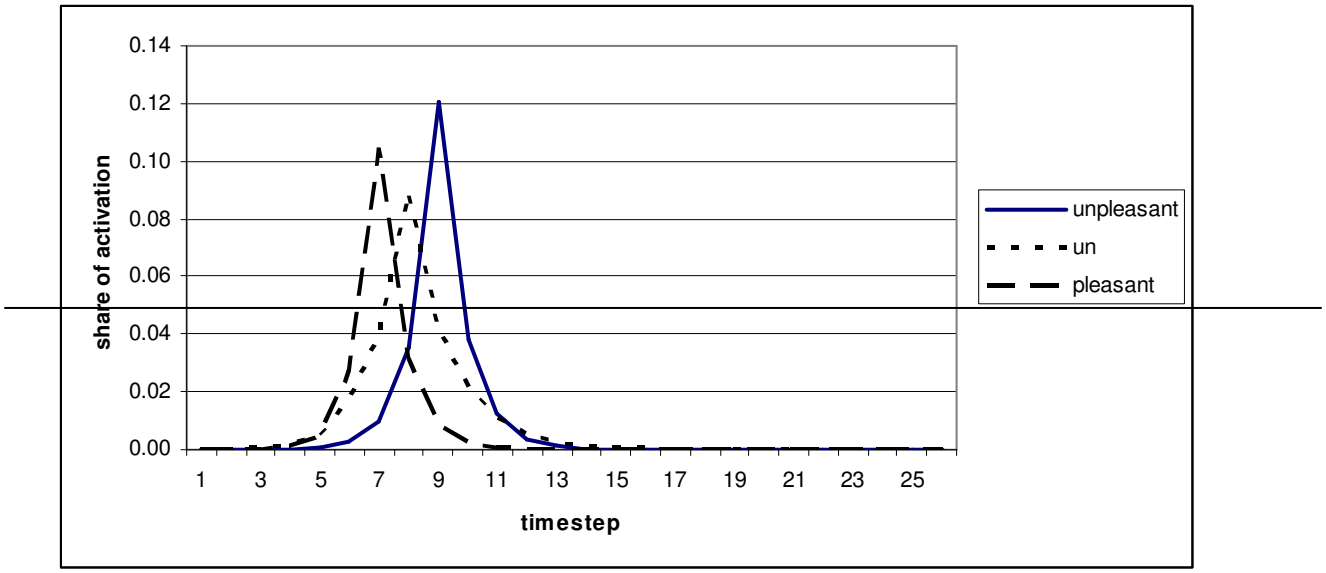
- θ = 0.05 (threshold—an item must grab at least 5% of the total activation to get copied to the buffer.
- α = 1.2 (spike parameter: how much of a difference does length make)
- δ = 0.3 (basic decay rate)
- ζ = 4 (forest-vs.-trees: how much of an advantage does an item that covers more of the target get)
- ε = 0 (system noise—didn't use)
- κ = 0 (summed activation threshold—didn't use)
- σ = 1 (information extraction rate—didn't use)

I made one big change to the algorithm: an item counts as edge-aligned only if it is aligned to the *left* edge of the target (stripping off any items already in the buffer). In original Matcheck, alignment to either edge counts.

Sample output, with 100-timestep run

```
target unpleasant
that reached threshold 0.05 at time 2
last reached threshold 0.05 at time 4
please reached threshold 0.05 at time 5
pleasant reached threshold 0.05 at time 7
un reached threshold 0.05 at time 8
spanning un+pleasant at 8
unpleasant reached threshold 0.05 at time 9
spanning unpleasant at 9
u reached threshold 0.05 at time 17
```

Tracking just the activation of *un*, *pleasant*, and *unpleasant* over time (with horizontal line through threshold):



(3) Baroni's prefixed words

Recall that use of [s] is taken to be the signature of a complex parse—will this line up with Matcheck results? If so, it's no shock—Baroni found that frequency played an important role—but it will reassure us that the simulation works reasonably well.

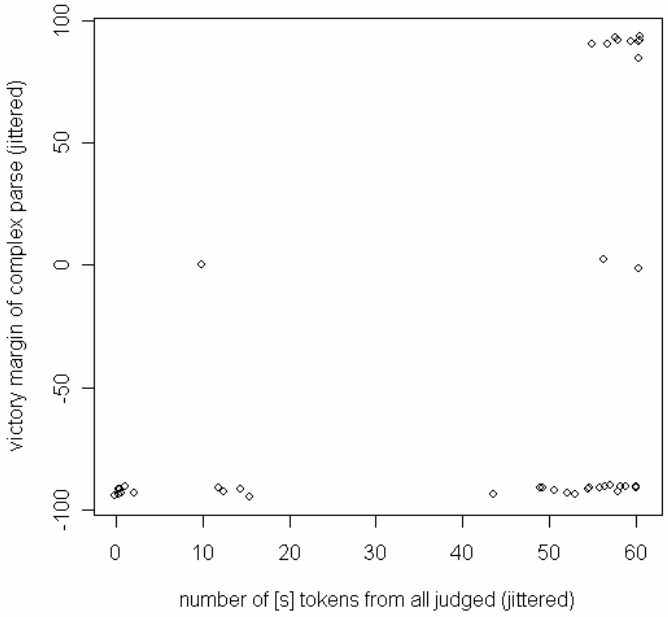
Wordlist used was *Lessico di frequenza dell'italiano parlato* of De Mauro et al.[6]; same procedure as above for estimating bound-morpheme frequencies. Same parameter settings.
Adapted from Baroni's appendix (plus Matcheck results).

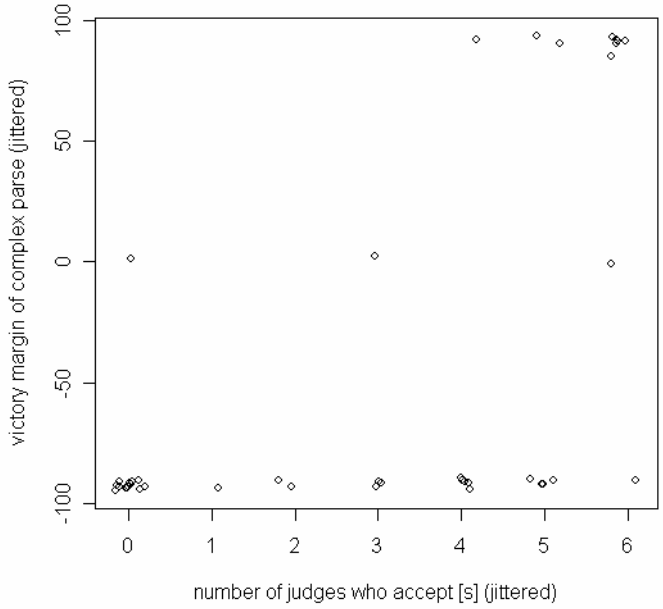
Word	# of (s) realizations for highly correlated judges	total # of (s) realizations	# of judges who accept [s]	spannings	complex parse's margin of victory (treat absent parse as if at timestep 100)
presalario	50	60	6	pre+salario at 7	93
bisezione	48	58	5	bi+sezione at 7	93
bisessuale	50	60	6	bi+sessuale at 8	92
risollevato	50	60	6	ri+sollevato at 8	92
parasimpatico	47	57	6	para+simpatico at 10	90
resuscitare	46	55	5	re+suscitare at 10	90
preselezione	49	59	6	pre+selezione at 9	91
asociale	50	60	6	a+sociale at 15	85
risoluzione	46	56	3	ri+soluzione at 7 risoluzione at 9	2
presentimento	10	10	0	pre+sentimento at 9 presentimento at 10	1
bisettimanale	50	60	6	bisettimanale at 9 bi+settimanale at 10	-1
desacralizzare	50	60	6	desacralizzare at 10	-90
prosecuzione	48	58	4	prosecuzione at 10	-90
proseguimento	48	57	5	proseguimento at 10	-90
Risorgimento	47	57	5	risorgimento at 10	-90
risolutivo	50	60	4	risolutivo at 9	-91
resurrezione	49	59	4	resurrezione at 9	-91
risoluto	47	56	2	risoluto at 9	-91
riserva	50	55	3	riserva at 9	-91
preservare	44	49	3	preservare at 9	-91
riservatezza	44	49	4	riservatezza at 9	-91
risarcire	12	12	0	risarcire at 9	-91
presunzione	1	1	0	presunzione at 9	-91
presuntuoso	0	0	0	presuntuoso at 9	-91
resistente	0	0	0	resistente at 8 r+esistente at 37	-91
risentimento	48	58	4	ri+sentimento at 8	-92
presidio	46	54	5	presidio at 8	-92

[6] available at http://languageserver.uni-graz.at/badip/badip/home.php

risanamento	41	51	5	risanamento at 8	-92
desueto	14	14	0	desueto at 8	-92
presunto	0	0	0	presunto at 8	-92
preside	49	58	2	preside at 7	-93
riservato	47	52	3	riservato at 7	-93
residenza	11	13	0	residenza at 7	-93
desumere	2	2	0	desumere at 7	-93
desiderio	0	0	0	desiderio at 7	-93
Resistenza	0	0	0	resistenza at 7 r+esistenza at 37	-93
risolvere	48	53	4	risolvere at 6	-94
presidente	38	43	1	presidente at 6	-94
risalto	16	16	0	risalto at 6	-94
deserto	0	0	0	deserto at 6	-94
risultato	0	0	0	risultato at 6	-94
desistere	0	0	0	d+esistere at 17	
antisettico	50	60	6		
asessuato	50	60	6		
asettico	50	60	6		
asincrono	50	60	6		
bisecolare	50	60	6		
bisessuati	50	60	6		
bisillabo	50	60	6		
iposolfito	50	60	6		
paraselene	50	60	6		
polisillabo	50	60	6		
polisindeto	50	60	6		
prosegui	50	60	5		
risiede	50	60	5		
risommergere	50	60	6		
trasuda	50	60	6		
asimmetrico	49	59	6		
asintoto	50	59	6		
asintotica	49	59	6		
bisolfito	49	59	6		
presiede	49	59	5		
risalita	49	59	5		
risaputo	49	59	6		
bisecante	48	58	6		
risuona	48	58	6		
asepsi	47	57	6		
risucchio	47	57	6		
metasemia	50	56	6		
risonanza	46	56	5		
risorse	50	56	5		
risolubile	45	55	4		
polisemo	49	54	6		
trasognata	48	54	5		
asindeto	43	53	5		
preservativo	49	53	3		
risentito	45	53	5		
risorgive	43	53	6		
risorto	48	53	5		
risanabile	41	51	5		
presidiato	43	48	3		
antesignano	41	47	6		
parasanghe	37	42	5		
risacca	31	41	3		
coseno	31	36	4		

residenziale	27	32	0		
resipiscenza	23	32	2		
bisesto	21	31	2		
residuato	26	31	1		
bisettrice	26	28	1		
resiliente	16	21	1		
risalta	15	15	0		
risarcibile	13	13	0		
desinenza	2	2	0		
bisestile	1	1	0		
designato	1	1	0		
desolato	1	1	0		
presupposizione	1	1	2		
desinare	0	0	0		
presagio	0	0	0		
presuntivo	0	0	0		
resisti	0	0	0		





This tells us that Matcheck is doing something along the right lines.
Although there is over all too much of a bias against the complex parse, the words where the complex parse wins (top half of plots) are all treated by participants as complex (right side of plots).

But all this is no surprise: since these words all have the same morphology, their differing structure can be explained only by semantic or frequency factors (or number of syllables, but only one item with a disyllabic suffix succeeded in being parsed).

What would be really interesting is if we can derive the difference between, say, prefixed words (vary according to meaning/frequency/something) and compounds (vary only according to morphology).

(4) Italian prefixed words vs. compounds
With this small lexicon, most words from Peperkamp (1996) couldn't be parsed. Here are the results for what could be, with Peperkamp's prosodifications.

Matcheck does a not-bad job distinguishing the types of compounds

(root+root)—53% simple, 24% both, 24% complex
analogo at 7
autografo at 8
biblioteca at 7
biologo at 9
demo+cratico at 7, democratico at 7
dialogo at 6
filosofo at 7
fono+metria at 8
foto+grafia at 6, fotografia at 6
frigorifero at 7
glotto+logia at 10
grafo+logia at 7
morfo+logia at 9, morfologia at 9
pornografia at 8, porno+grafia at 10
stereotipo at 8
omonimo at 8
tele+crazia at 6

(root)+(stem) or (root+(stem))—"familiar" (not sure which are which, except that *eurosocialista* is familiar): 15% simple, 12% both, 73% complex
anti+fascismo at 8
anti+sciopero at 8
anti+sovietico at 8
filo+americano at 8
filo+drammatico at 9
filo+sovietico at 8
filo+tedesco at 6
multi+milionario at 9
onnipotente at 9
pro+sindaco at 12
uni+laterale at 7, unilaterale at 9
vicepresidente at 9, vice+presidente at 10
omo+sessuale at 9
neuro+linguistica at 12
pseudo+modello at 10
monopolare at 9
italo+tedesco at 8
anglo+americano at 11
catto+comunista at 9
euro+mercato at 6
auto+mobile at 6, automobile at 7
televisione at 7
motocicletta at 9
anarco+sindacalista at 12
auto+controllo at 7
euro+socialista at 8

(stem)+(stem)
porta+ombrelli at 8

```
senzatetto at 8
mezzo+giorno at 5, mezzogiorno at 7
(stem+(stem))—"familiar" (i.e., frequent)
copriletto at 7
pomodoro at 7
reggiseno at 8
```

- Two types of prefixes—not enough examples in Peperkamp (I ran only her examples).

```
(prefix(stem)): 85% complex
a+religioso at 21
a+sessuale at 18
a+sociale at 16
bi+mensile at 8
bi+sessuale at 9
co+detenuto at 9
co+ufficiale at 9
disonesto at 9
pre+avviso at 6
pre+finanziare at 9
pre+selezione at 9
in+elegante at 7
inutile at 6

(prefix)(stem)
arcivescovo at 9
inter+disciplinare at 10
extra+parlamentare at 12

lexical prefixes—not sure what structure proposed
col+laterale at 7, collaterale at 8
con+detenuto at 7
con+ufficiale at 7
irregolare at 9, ir+regolare at 11
indegno at 9
```

Over all: 19% simple, 10% both, 71% complex

- How does this compare to suffixes? Hard to interpret, since Peperkamp doesn't discuss suffixes as fully (not as many examples). But, the suffixes do get parsed complexly less often

```
(root+suffix): 23% simple, 46% both, 31% complex
videoteca at 9 (suffix derived from non-native root)
ginecologo at 7
mao+ismo at 11
famoso at 6, fam+oso at 9
tenera+mente at 9
leon+cino at 9
cas+ale at 4, casale at 8
sigaretta at 7, sigar+etta at 9
matemat+ica at 7, matematica at 7, ma+tematica at 8
cortese+mente at 9, cortesemente at 9
lineare at 9
umil+mente at 10
sud+ista at 5, sudista at 9
```

cf. (stem)clitic
compra+gli at 5
dar+gli at 3
perder+lo at 5
vender+lo at 5
vendi+gli at 5

(5) Raffelsiefen's boundary cues
Recall: Raffelsiefen proposes that a word has to get parsed complexly if it's not a well-formed p-word (unless the base is not an existing word).

In most cases this is trivial, because the very property that allows us to diagnose the word's prosodic status would do the same for the learner.
Where it gets interesting is in accounting for how words change over time (Raffelsiefen section 2.4).

For example, assuming that verbs borrowed from French originally had final stress, some would be parsable as wholes (even though the base exists):

(ensúre) disyllabic Verb—final stress expected
(represént) trisyllabic Verb ending in CC—final stress expected

but

decompóse > (dè)(compóse) trisyllabic Verb ending in V(C)—final stress not expected (*compóse* is fine because disyllabic)

incorréct > (ìn)(corréct) trisyllabic Adj. with final stress—even if ends CC, final stress not expected (*corréct* is fine because disyllabic)

Do we really need the phonology? Or the key instead something like word length, plus accidental properties of the trisyllabic verbs that happen to line up with ending in a CC or not?

See chart on next page: Raffelsiefen's story is looking pretty good. This parsing model, at least, is not able to do the work of the phonological prediction. Every group of words is about evenly split between the simple parse and the complex parse.

	disyllabic V—simple	trisyllabic V ending CC—simple	trisyllabic V ending V(C)—complex	trisyllabic A—complex	trisyllabic A but simple (maybe borrowed before base)	borrowed from Latin -> never had final stress—simple	bound base—simple
simple parse wins in Matcheck	refuse at 7 release at 8 defraud at 11 discount at 7 disguise at 9 disease at 8 recite at 9 restrain at 9 res+train at 11 decry at 11 deform at 11 depart at 9 depress at 9 enforce at 10	recommend at 8 disappoint at 10	subdivide at 11 decompose at 11 disobey at 10 disallow at 11	discontent at 10 immature at 9 indirect at 10 illegal at 8 immortal at 9 immobile at 11 disloyal at 11 irregular at 11 ir+regular at 17	infinite at 9	allocate at 10	reconcile at 10 comprehend at 10 contradict at 11
complex parse wins in matcheck	re+mark at 4 remark at 9 re+place at 4 replace at 7 dis+charge at 6 discharge at 10 dis+close at 5 disclose at 11 de+face at 4 re+turn at 4 return at 7 de+fault at 5 default at 11 ex+change at 5 exchange at 7 en+close at 5 enclose at 11 en+sure at 5 ensure at 8	re+collect at 6 re+present at 6 represent at 9 dis+concert at 7		in+correct at 7 incorrect at 9 dis+honest at 6 dishonest at 11 ab+normal at 6 abnormal at 9 in+decent at 6 indecent at 10			ref+er at 7 refer at 8
parse failed	submerge disclaim declaim defame degrade denote despoil exclaim depose subserve enchain enroll encrust	disaffect	reconvene reconsign recompose reconfer disengage rebaptize decentralize presignify	impolite indiscrete indistinct inexact incorrupt immodest infertile intolerant inelegant immoral incurious indocile illiberal	impotent infamous impious	recreate dislocate	resurrect

By the way, we don't see any explanation here either for why certain native coinages got fused (Raffelsiefen proposes that having a main stress that doesn't need to move under prosodic fusion makes it easier):

simple parse wins in Matcheck	`forever at 7`
	`necklace at 8`
	`shepherd at 8`
complex parse wins in Matcheck	`for+give at 4` `forgive at 8`
	`al+one at 5` `alone at 5`
	`cup+board at 5` `cupboard at 6`
parse failed	disintegrate vinyard

(6) Left vs. right edge

I haven't explored this much, but we can ask what's the effect, in Matcheck, of privileging both edges (or even just the right edge!) vs. left edges.

Compare what Matcheck does with the suffixed Italian words when both edges are privileged.

More words that weren't in the lexicon become parseable (presumably because the suffix gets a boost).

But when there is a whole-word competitor, both-edge alignment seems to make just about no difference to the outcome.

Why might that be?

- For forms whose complex parse wins in the left-edge model (e.g., *casale*), the suffix didn't need its alignment to reach threshold—presumably it got there on the merits of its similarity to the target, or because the stem reached threshold early enough that the suffix became left-aligned. So the suffix probably (see below) reaches threshold at the same point in the two-edge model.
- For forms whose simple parse wins...
    - the suffix may still reach threshold without being right-aligned (e.g., *famoso*), and thus there's no change in the two-edge model.
    - the suffix might be doomed no matter what—by a too-small decay rate, for instance.
    - the suffix might not be similar enough to reach threshold before its activation is deemed to have peaked, but its decay rate is big enough that if it can stay in the game longer (through its right-edge alignment), it will eventually reach threshold.
- And of course, we have to consider the effect that edge-alignment has on all the other items that are competing for a share of the total activation.

	only left edge counts	**complex advantage**	**both edges count**	**complex advantage**
ginecologo	ginecologo at 7	−93	ginecologo at 7 gineco+logo at 8	−1
videoteca	videoteca at 9	−91		
lineare	lineare at 9	−91	linea+re at 8 lineare at 9	−91
famoso	famoso at 6 fam+oso at 9	−3	famoso at 6 fam+oso at 9	−3
sigaretta	sigaretta at 7 sigar+etta at 9	−2	sigaretta at 7 sigar+etta at 9	−2
matematica	matemat+ica at 7 matematica at 7 ma+tematica at 8	0	matemat+ica at 7 matematica at 7 ma+tematica at 8 m+a+tematica at 26	0
cortesemente	cortese+mente at 9 cortesemente at 9	0	cortese+mente at 9 cortesemente at 9	0
casale	cas+ale at 4 casale at 8	4	cas+ale at 4 casale at 8	4
sudista	sud+ista at 5 sudista at 9	4	sud+ista at 5 sudista at 9 su+di+s+ta at 15 su+di+s.+a at 15 s.+di+s+ta at 15 s.+di+s.+a at 15	4
maoismo	mao+ismo at 11	89	ma+o+ismo at 6 mao+ismo at 11	89
umilmente	umil+mente at 10	90	umil+mente at 10	90
			bu+sino at 10	90
			coca+ina at 10 co+c+a+ina at 15	90
teneramente	tenera+mente at 9	91	tenera+mente at 9	91
leoncino	leon+cino at 9	91	leon+cino at 9	91
			blu+astro at 8	91
			hotel+accio at 8	91
			carita+tevole at 8	91
			vas+aio at 8	92
			ero+ina at 7	93
			libr+aio at 5	95
overall	**23% simple** **46% both** **31% complex**		**5% simple** **33% both** **62% complex**	