LSA.360   The Phonological Status of Morphologically Complex Words

Class coordinates
Time: Mondays and Thursdays, 3:45PM-5:30 PM
Classroom: Education 208 (near the Green Library, the Meyer Library, and the Clock Tower)
Web page: log in to coursework.stanford.edu
Instructor coordinates

Name: Kie Zuraw [ˈkʰaɪ ˈzɝˌכ]
E-mail: kie@ucla.edu
Web page: www.linguistics.ucla.edu/people/zuraw (if I get around to it, I'll post copies of
handouts and assignments [but not readings] here for those who are not officially enrolled)
Office: 460-040E (Margaret Jacks Hall)
Office hours: Wednesdays 9-11 AM or by appointment (I can change this to another time on
Wednesdays if you prefer—I'll take a poll on the first day of class)

Catalog description
A word's morphological structure influences its phonological behavior, including alternations,
stress assignment, and subjectability to phonotactic restrictions. Accounts of this influence have
invoked boundary symbols, derivational interleaving of phonology and morphology, prosodic
structure, and the possibility that some morphologically complex words are represented in the
lexicon as monomorphemic. We will explore how distributional factors--word frequency, affix
frequency, etc.--influence whether (part of) a word behaves as simple or complex. Students will
investigate corpora, including optionally corpora of their own choosing. We will look at models
of competition between whole-word and decomposed lexical access routes (e.g., Baayen &
Hay's), and explore how the outcome of processing could become fossilized in listed
pronunciations. We will also ask whether distributional facts alone account for the phonological
behavior of morphologically complex words, or if phonological structure--e.g., prosodic words,
boundary symbols--is needed. The course should be suitable for students at all levels.

Prerequisite: A previous course in phonology. Basic Unix skills would be helpful, but can be
picked up in the course.

Course outline

		topics	*readings*	*Assignment*
Class 1	Thurs, July 5	intro and background; case	Peperkamp (1995)	compare the distribution
		study; relationship between	Baroni (2001)	of two affixes with
Class 2	Mon, July 9	distributional/usage factors		different phonological
		and grammar		behavior (due July 12)
Class 3	Thurs, July 12	outer vs. inner affixes;	Bybee (2002)	compare your affixes'
		productivity; affixation vs.	Hay (2003), chapter	productivity
		compounding	6	OR compare affix and
Class 4	Mon, July 16			common compounding
				element (due July 19)
Class 5	Thurs, July 19	parsing model (based on	Baayen & Hay	run some items with your
		Hay, Baayen), with	(2002)	affixes through the
Class 6	Mon, July 23	implementation and some	Pierrehumbert	parsing model (due July

		simulations	(2002)	26)
Class 7	Thurs, July 26	case study of processing model vs. grammar: various Tagalog morphophonology		

Readings
The Baayen & Hay reading is not available on the course webpage (but the rest are). I will give you a link, or you can easily find it online yourselves.

Baayen, Harald & Jennifer Hay (2002). Parsing and productivity. In G. E. Booij and J. v. Marle (eds.) *Yearbook of Morphology 2001*. Dordrecht: Kluwer.
http://www.ling.canterbury.ac.nz/jen/documents/hayandbaayen.pdf

Baroni, Marco 2001. The representation of prefixed forms in the Italian lexicon: Evidence from the distribution of intervocalic [s] and [z] in northern Italian. In Geert Booij and Jaap van Marle (eds.), Yearbook of Morphology 1999, Dordrecht: Springer. 121-152.
http://polorovereto.unitn.it/~baroni/publications/szjoined.pdf

Bybee, Joan. 2002. Phonological Evidence for Exemplar Storage of Multiword Sequences. SSLA 24. 215-221.
http://www.unm.edu/~jbybee/BybeeExemplarMultiword.pdf

Hay, Jennifer (2003). Causes and consequences of word structure. Routledge. Chapter 6.

Peperkamp, Sharon (1995). Prosodic constraints in the derivational morphology of Italian. Yearbook of Morphology 1994, 207-244.
http://www.unice.fr/dsl/tobweb/interface/Peperkamp_95%20-
%20Prosodic%20constraints%20in%20the%20derivational%20morphology%20of%20Italian.pdf

Pierrehumbert, Janet (2002) Word-specific phonetics . Laboratory Phonology VII, Mouton de Gruyter, Berlin, 101-139.
http://www.ling.northwestern.edu/~jbp/publications/word_specific.pdf

Assignments
Directions will be posted on class webpage.

Class 1/2: intro and background; case study

1   Intro/background: phonology is affected by morphology
(1)   Some phonology occurs only at morpheme boundaries

Classic example from Kiparsky: Finnish[1]

Both *t* and *s* are possible before *i*:

tila	'room'	paa**si**	'boulder'
æi**ti**	'mother'	**si**næ	'you (sg.)'
sil**ti**	'however'	kuu**si**	'six'
val**ti**on	'public'		

But, t → s / __ + i (except geminate *tt*)

infinitive	*Let him/her X!*	*'active instructive II infinitive'*	*she/he was Xing*	
halu**t**+a	halu**t**+koon	halu**t**+en	halu**s**+i	'want'
noe**t**+a	noe**t**+koon	noe**t**+en	noke**s**+i	'smudge (?)'
pie**t**+æ	pie**t**+køøn	pie**t**+en	pike**s**+i	'pitch'
filma**t**+a	filma**t**+koon	filma**t**+en	filma**s**+i	'film'
cf.				
oll+a	ol+koon	oll+en	ol+i	'be'
aja+a	aja+koon	aja+en	ajo+i	'go'
puhu+a	puhu+koon	puhu+en	puhu+i	'speak'

(2)   Or, more generally, only in (morphologically?) derived environments

e → i / __ # (for some stems)

jok**e**+na	'river' essive sg.	jok**i**	'river' nom. sg.
mæk**e**+næ	'river' essive sg.	mæk**i**	'hill' nom. sg.
uks**e**+na	'door' essive sg.	uks**i**	'door' nom. sg.
æit**i**+næ	'mother' essive sg.	æit**i**	'mother' nom. sg.
kahv**i**+na	'coffee' essive sg.	kahv**i**	'coffee' nom. sg.

Vowel raising feeds *t* → *s* (but underlying *ti* doesn't change)

ve**te**+næ	'water' essive sg.	ve**si**	'water' nom. sg.
kæ**te**+næ	'hand' essive sg.	kæ**si**	'hand' nom. sg.
yh**te**+næ	'one' essive sg.	yk**si**	'one' nom. sg.

(3)   Possibly, some phonology is blocked by morpheme boundaries

Here it's harder to find good examples (phonotactics that hold strongly or totally within morphemes but very weakly or not at all across morpheme boundaries).[2]

---

[1] with additional data from verbix.com

Labial Attraction in Turkish: The sequence ɑCu is more common than ɑCɯ when C is labial and voiced, and the reverse otherwise (data from Inkelas et al.'s TELL, http://socrates.berkeley.edu:7037/).

C	*aCɯ*	*aCu*	*C*	*aCɯ*	*aCu*
v	2	50	dʒ	34	5
b	18	36	tʃ	15	5
m	8	28	d	21	6
			g	0	1
p	31	13	ɟ	83	4
f	7	2	h	9	12
dʒ	34	5	ʒ	0	3
tʃ	15	5	k	73	18
d	21	6	c	0	0
g	0	1	l	85	9
ɟ	83	4	n	43	9
h	9	12	r	113	10
ʒ	0	3	s	46	14
k	73	18	ʃ	55	2
c	0	0	t	54	16
l	85	9	j	63	2
n	43	9	z	34	8

But, as far as I know, the quality of C makes no difference in the rounding of a following suffix high vowel, which simply follows that of the previous vowel: *kol-un* 'arm-gen.sg.' vs. *baʃ-un* 'head-gen.sg.'.

(4)   Not all morpheme boundaries are equal

English, from Allen (1978): *n+C* sequences undergo much more fusion in *in+C* than in *un+C*:

lawful	unlawful	legal	illegal
balanced	u{n,m}balanced	balance	imbalance
rationalized	unrationalized	rational	irrational
grateful	u{n,ŋ}grateful	gratitude	i{n,ŋ}gratitude
(p. 15: These are Allen's examples, but not in all cases her transcriptions.)			
natural	u[nn]atural	numerable	i[n]umerable

(5)   Analysis under lexical phonology and morphology (e.g., Siegel 1974, Kiparsky 1982)

---

[2] The relative rarity of this situation is reflected in *SPE*'s convention that a + symbol in a rule's structural description matches only to forms that contain a morpheme boundary in the corresponding position, but a form's morpheme boundaries can be ignored when matching it to a part of structural description without any +.

Some affixes attach too late for certain phonology to apply:

stem	balance	balance
add "Class I" affix	--	in+balance
strong assimilation	--	imbalance
add "Class II" affix	balance+d	--
add "Class II" affix	un+balance+d	--
weak assimilation	u{n,m}balanced	--

(6) Domains larger than the morpheme—all examples from Nespor & Vogel 1986

- Rule that applies across **whole utterance** (≈ sentence): Gorgia Toscana, in Tuscan Italian

$$
\begin{Bmatrix} p \\ t \\ k \end{Bmatrix} \rightarrow \begin{Bmatrix} \phi \\ \theta \\ h \end{Bmatrix} / \quad [-\text{cons}] \_\_ [-\text{cons}]
$$

[lo 'sai **h**o'm ɛ dif'fiʃile // **h**o'noʃʃe k'kweste '**h**ɔse] (// represents "filled pause")
    /k/   /k/   /k/
   it know how is difficult   know   these   things
'You know how difficult it is to know these things.'

- Rule that applies to **small groups of words**: Northern Italian stress reduction (retract stress on first word in case of clash)

   (Sára   státa ammazzata) (la vipera) < *sará*
   will.have been killed   the adder
   'the adder has probably been killed.'

   (le cittá) (mólto nórdiche) (non mi piacciono) *(*cítta)*
    the cities very nordic   not me please
   'I don't like very Nordic cities.'

   (le cítta   nórdiche) (non mi piacciono)
    the cities nordic   not me please
   'I don't like Nordic cities.'

FYI, under N&V's analysis, the groups are formed right to left: take a constituent containing a lexical head X and group it with preceding material until you hit a constituent containing a lexical head outside of X's maximal projection (or the beginning of the sentence). Optionally, if X's complement (or adjunct, in later terms) forms a non-branching group to the right of X, it can join X's group, as in (le cítta nórdiche).
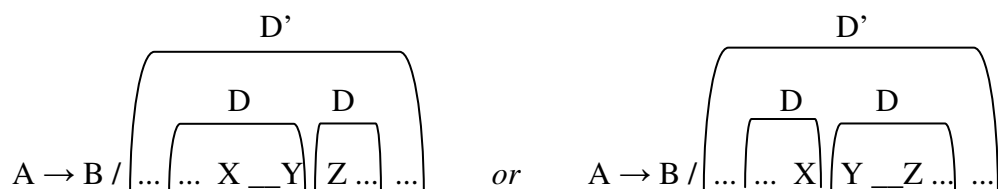
- Rule that applies to **single words**: Italian primary stress

Each (lexical) word has exactly one primary stress.[3]

Only primary-stressed syllables can contain [ɛ], [ɔ] (when secondary-stressed or unstressed, these neutralize with /e/, /o/):

(t[ɔ́]sta)	'toast'
(s[ɔ́]lito)	'usual'
(t[ò]sta-tóre)	'toaster'
(s[ò]lita-ménte)	'usually'
(s[ɔ́]lita) (ménte)	'usual mind'

(7) "Prosodic" approaches to domains

U: utterance
|
I: intonational phrase
|
φ: phonological phrase
|
ω: p-word (aka phonological word, prosodic word)
|
F: foot
|
σ: syllable
|
segment

(many variants exist, of course!)

Example loosely adapted from Nespor & Vogel 1986, henceforth N&V:

[Why the scare quotes on "prosodic"? Because these units don't always have identifiable signatures in pitch or duration.]

---

[3] Where the members of a compound each count as separate words.

(8)   Selkirk's 3 types of rules
- <u>domain span rules</u>: the structural description must be contained within a certain domain

$$A \rightarrow B \: / \quad \overparen{...X\_\_Y...}^{\;D}$$
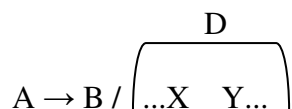
- <u>domain juncture rules</u>: the structural description spans the boundary between two domains D, and is contained within a domain D' (D' is higher than D, but not necessarily the immediately dominating level)

$$A \rightarrow B \: / \quad \Big[... \big[... \: X \_\_Y \big] \big[Z ... \big] ... \Big]^{\;D'} \qquad or \qquad A \rightarrow B \: / \quad \Big[... \big[... \: X \big] \big[Y \_\_Z ... \big] ... \Big]^{\;D'}$$

- <u>domain limit rules</u>: the structural description is at the edge of a domain D

$$A \rightarrow B \: / \quad \big[...X\_\_Y \big]^{\;D} \qquad\qquad or \qquad A \rightarrow B \: / \quad \big[X\_\_Y... \big]^{\;D}$$

(9)   Relation to examples we've seen so far

English *in-* vs. *un-* could be represented as *(un)*$_\omega$*(balanced)*$_\omega$ vs. *(imbalance)*$_\omega$, with strong assimilation being a ω-span rule.

Do prosodic domains help with the Finnish case at all?

(10) Distributional approaches to domains—e.g., Bybee (we will see other such approaches later)

The word, phrase, etc. is not *a priori* a privileged unit.
Rather, sequences that are frequent become lexicalized (with lexicalization being a gradient notion)—these sequences can be bigger or smaller than words.

In this type of approach, we'd want to say that *imbalance* is more lexicalized than *unbalanced* (though seeing whether it is predicted to get more lexicalized requires fully spelling out the model).

2   Case study—Northern Italian intervocalic s-voicing

(11) Basic pattern and prosodic approach (Nespor & Vogel 1986, Peperkamp 1997, also some data from Baroni 2001)

p-word-span rule: /s/ becomes [z] iff it's contained in a V__V span uninterrupted by a p-word boundary.

$$s \rightarrow z \ / \ \overbrace{\left[...V\_\_V... \right.}^{\text{p-word}}$$

[s] and [z] both correspond to letter <s>. Letter <z> represents an affricate.

(12) Within a *stem* or *stem+suffix*: rule applies
I.e., stem and suffix form a single p-word

(stem)	(cá[z]a)	'house'
	(a[z]ílo)	'asylum'
	cf. ([s]ánto)	'saint'
	cf. (bu[s])	'bus'
(stem+suffix)	(ca[z]-ína)	'house-*dim*'
	(cau[z]-áva)	'caused'
	(famo[z]-íssimo)	'very famous'
	(va[z]-áio)	'potter'

(13) Across word boundary, or word/clitic boundary, rule doesn't apply
I.e., these don't form a single p-word

(stem) (stem)	(bella) ([s]era)	'beautiful evening'
(clitic (stem))	(la ([s]iréna))	'the siren'
	(lo ([s]apévo))	'I knew it)
((stem) clitic)	((telefonáti) [s]i)	'having phoned each other'
((stem) clitic clitic)	((andando) ci [s]i insieme)	'(our) going there together'

(14) Two parts of compound act as two separate words

(stem)+(stem)	(pórta)-([s]igarétte)	'cigarette holder'
	(tocca)-([s]ana)	'cure-all'

(15) Prefixes—at least when transparent—stand aloof from their stems...
[But stress data—see below—indicates that monosyllabic prefixes and disyllabic ones have different prosodic relationship to stem.]

(prefix+(stem))	(a-([s]ociále))	'asocial'
	(ri-([s]aláre))	're-salt'
(prefix)+(stem) *or maybe* (stem)+(stem)	(fílo)([s]oviético)	'pro-soviet'

(16) ...except when the /s/ is prefix-final

	(di.[z]-uguale)	'unequal'
(prefix+stem)	(di.[z]-armo)	'disarmament'
	(di.[z]-innescare)	'defuse'
	(ci.[z]-alpino)	'located on this side of the Alps'
	(bi.[z]-avo)	'great-grandfather'

See Peperkamp for full OT analysis, and van Oostendorp (1999) for a somewhat different one.

(17) What about *s*-final free stems?
Kenstowicz 1996 reports that they don't voice (*bu[s]-íno* 'bus-*dim*'), supporting a base-to-derived correspondence analysis.

Peperkamp's consultants disagree, producing occasional [s] but much more often [z] or [s:]:

(lapi[z~s:]-áccio)	'pencil-*aug*'
(lapi[z~s:]-íno)	'pencil-*dim*'

Peperkamp notes that other C-final free stems can (variably) geminate before suffixes too:
> *autosto[p:]-ísta*   'hitchhiker'
> *vermou[t:]-íno*   'vermout-*dim*'

Assuming that the geminate fricative is exempt from voicing (*[z:]), this could explain why *s*-voicing is only optional in these cases.

(18) Prosodic constituents need not match morphological constitutents

ri-[s]uddivi[z]-ione   'resubdivision'   ri-[s]ocial-izzare   'to resocialize'
(ri-(suddivis-ione))   (ri-(social-izzare))
[[ri[suddivis]$_V$]$_V$ ione]$_N$ [ri[[socia]$_A$ izzare]$_V$]$_V$

(19) Other evidence on p-word boundaries: primary stress (Nespor & Vogel 1986, Peperkamp 1997)
Italian primary stress is partly unpredictable, but must fall on one of the last three syllables.

Recall that primary stress can be diagnosed by vowel raising: [ε], [ɔ] are allowed only in primary-stressed syllables:

(stem)	(t[ɔ́]sta)	'toast'
	(s[ɔ́]lito)	'usual'
(stem+suffix)	(t[ò]sta-tóre)	'toaster'
	(s[ò]lita-ménte)	'usually'
(stem) (stem)	(s[ɔ́]lita) (ménte)	'usual mind'
(stem)+(stem)	(t[ɔ́]sta)-(páne)	'bread toaster' (toast bread)
	(t[έ]rra)-(c[ɔ́]tta)	'terra cotta'
	(p[ɔ́]rta)-(ombr[έ]lli)	'umbrella-stand'
	(par[ɔ́]la) (mod[έ]llo)	'model word'
(?)+(stem)	([έ]kstra)-(coniugále)	'extramarital'

unclear whether prefix or just bound stem	([έ]x)-(presidénte)	'ex-president'
	([έ]uro)-(parlamentáre)	'Europarliamentarian'
	(ps[έ]udo)-(concétto)	'pseudoconcept'
	(pr[ɔ́]to)-(notariáto)	'protonotaryship'[4]

(Peperkamp takes the fact that *ex* can stand as a word on its own to indicate that *ex-presidente* is a compound, not a prefixed word.)

Primary stress is also said to be diagnosed by vowel lengthening: vowels in primary-stressed, open, nonfinal syllables lengthen (but see, e.g., D'Imperio & Rosenthall 1999).

(stem)$_\omega$	p[á:]pero	'duck'	
(stem+suffix)	(abbai[á:]-va)	'(it) was barking'	
(fossilized.prefix+stem)	(p[à]ramétrico)	'parametric'	See discussion of Baroni, below!
(prefix?)+(stem)	(c[á]po)-(p[ó:]polo)	'chief'	
	(p[á]ra)-(milit[á:]re)	'paramilitary'	
	(s[ú:]per)-(vel[ó:]ce)	'superfast'	
(prefix+(stem))	([à]-(soci[á:]le))	'asocial'[5]	

(20) Indirect evidence on p-word boundaries: raddoppiamento sintattico (Central & Southern; N&V)
Lengthen a consonant after a p-word-final stressed vowel.
Supports giving *tre Greci* and *pre-greci* different prosody, at least in these dialects

(stem) (stem)	(metá) ([l:]íbro)	'half a book'
	(caffé)-([l:]átte)	'white coffee'
	(tré) ([g:]réci)	'three Greeks'
((word) clitic)	((dá) [m:]i)	'give me'
(prefix (words))	(pre ([g]réci))	'pre-Greek-*pl*'

(21) Problematic case: "familiar" (i.e., frequent) compounds (Peperkamp)
Familiar compounds have contradictory behavior for V-raising and *s*-voicing, and Peperkamp proposes a (stem+(stem)) prosodic structure (see her pp. 127-128 for analysis):

| (c[ò]pri-(létto)) | 'bedspread' | cf. c[ɔ́]pri 'cover' |
| (r[è]ggi-([s]éno)) | 'bra' | cf. r[έ]ggi 'hold' |

(22) Invisibile prefixes (N&V pp. 124-134)

(re[z]istenza) 'resistance'
(pre[z]entire) 'to have a presentiment'

---

[4] I'm assuming the [ɔ́] should also be long, but this datum is from a section of Nespor 1999 that's not concerned with length
[5] I'm extrapolating from Peperkamp's discussion.

cf.   (pre)([s]entire) 'to hear in advance'

Why do some prefixed words behave as though they're not? Are they represented the same way as monomorphemic words? Baroni investigated...

(23) An experiment on *s*-voicing: Baroni 2001
12 (usable) speakers of Standard Northern Italian read 102 target sentences 5 times—plus 10 controls where only one option should be possible, e.g. monomorphemic words—and then rated the semantic transparency of all the items.

3 (usable) speakers judged whether the same words, plus 40 additional controls, were acceptable with [s] only, with [z] only, or with either.

E.g. *Il giallo è un colore che <u>risalta</u> ovunque* 'Yellow is a colour that <u>stands out</u> everywhere'

(24) Results: productivity
<u>The phenomenon is productive</u> (good news! otherwise it would be less clear that it should be represented in the grammar), despite the potential for interference from other dialects.

- In a another study with 58 participants (Baroni 1996), subjects read nonsense words in the expected way: [s]amo, pa[z]a.
- Loans are adapted in the expected way: [s]arland < German *[z]aarland*, me[z]a < Spanish or English *me[s]a*.
- 10 Italian speakers were asked to form *ri-X-izzare* ('re-X-ize') for various proper names *X* and behaved as expected: ri-[s]andr-izzare 're-Sandro-ize', but di[z]-Ald-izzare 'de-Aldo-ize'.

(25) Results: variation
<u>Variation is both between and within items</u>
- To a large extent, subjects agree: the three judges' ratings were highly correlated, as is the composite judges' score with the number of [s] realizations (for the 10 highly-correlated readers).
- But for some items, there's rampant variation: *coseno* 'cosine':
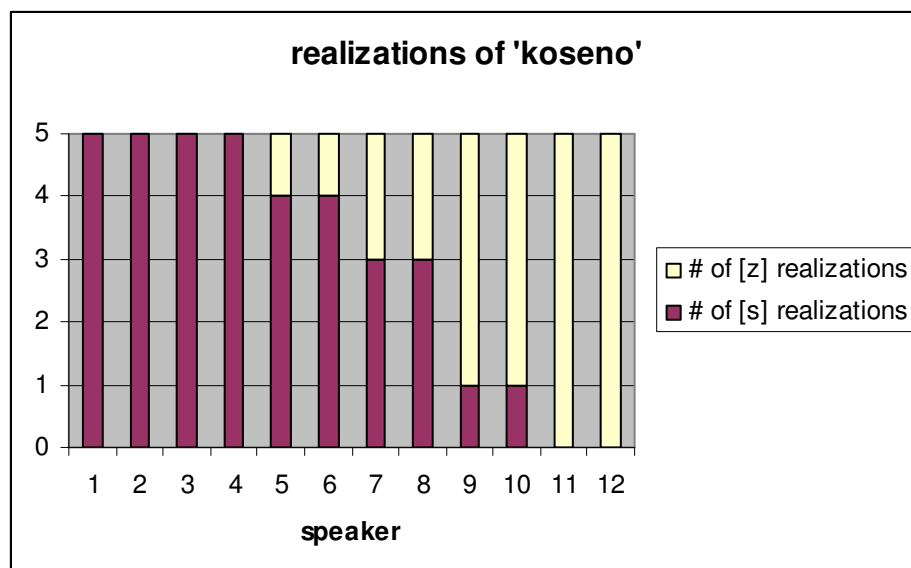
- For other items, there's variation but with a strong trend (a & c):

(17)   N of [s] realizations
  a.   resuscitare   'to resurrect'   55/60
        risolvere   'to solve'   53/60
        riservato   'reserved'   52/60

  b.   coseno   'cosine'   36/60
        residuato   '(war) surplus'   31/60
        bisettrice   'bisecting (line)'   28/60

  c.   risalto   'prominence'   16/60
        risarcire   'to refund'   12/60
        desinenza   '(morphological) ending'   2/60   (p. 22 of ms. version)

(26) Results: representation of opaque forms
Total semantic transparency is not required for a prefixed parse
- *ri-[s]iede* 'resides', *ri-[s]olutive* 'resolutive', and *ri-[s]aputo* 'well known' were consistently produced with [s], despite the lack of iterative meaning (the productive meaning of *ri-*).

(27) Results: factors that influence a word's behavior
Baroni looked at...
- length of word, stem, prefix, and root (i.e., stem minus suffixes),in segments and in syllables
- type and token frequencies of word, prefix, and root
- ratio of prefix frequency to pseudo-prefix frequency (i.e., how many words begin with the prefix string, but don't actually contain that prefix, plus other criteria that would make it possible to parse out the prefix)
- root frequency for productively related forms only
- ratio of root frequency to pseudo-root frequency
- semantic transparency of prefix and stem, as rated by subjects

- stem autonomy (does it occur unprefixed?)—relevant for theories of paradigm uniformity
- frequency of the stem when autonomous
- ratio of autonomous stem frequency to word frequency (cf. Hay 2003)
- whether the root occurs word-initial
- frequency of root when word-initial
- stress position

phew!

The strong predictors turn out to be **stem transparency** and **prefix length** (in syllables).

Baroni points out that the predictive factors account for about 42% of the variation in # of [s] realizations. This is a lot, but means there's still more to the story.

Prefix length is an interesting one. Baroni points out that (from the learner's point of view) a longer string is less likely to have occurred by accident, and thus more likely to be a true prefix.

Nice example: *parasanghe* 'parasangs (ancient Persian measurement unit)' was produced with [s] 37 out of 50 times, even though there's no semantic reason to think it's prefixed.

At the same time, prefix length in syllables could be phonologically significant. If the minimal word in Italian is disyllabic (unclear if it really is—there are few, but then how many do we expect?), then only disyllabic prefixes are able to stand on their own as p-wds.

(28) Interpretation
Baroni's proposal is that speakers maintain two production representations for potentially complex words: a unitary representation (*coseno*) and a decomposed representation (*co+seno*).

The activation threshold of each depends on "the degree of confidence that the speaker has in the fact that the form is complex, plus some degree of (random?) transient fluctuation." (p. 25 of ms.) These fluctuations will be less significant—resulting in uniform productions—to the extent that the activation threshold is extreme (high or low).

(29) Prosodic structure?
Baroni is careful to say that he's looking only at prefixed words here, which is where we see two variants. What about the places where variation is not allowed?

o For semantics and distributional facts alone to account for all the *s*-voicing facts, what would have to be true of *s*-final prefixes? suffixes? compounds? clitics?

3   *Prosodic constituency vs. processing units*

Bybee proposes that "words that are often used together become processing units" (p. 157) and this leads to "phonological fusion".

[Of course, we need to define 'often used together' (or let the model define it implicitly)—do we just mean that the sequence is frequent, or that it's more frequent than would be expected given the frequencies of the components and some assumptions about how things combine? Or...]

In the Hay-ian view (see classes 5 and 6), words that tend to get accessed as single units tend to be more phonologically integrated than words that tend to get accessed via their component morphemes. Access route depends on relative frequency of whole word and subparts.

As we'll see in later classes, there's certainly evidence for such frequency effects. Can models using frequency do all the work, or do we still need prosodic units (or something like them)?

(30) Grammar-dependence vs. distribution/usage dependence : possible scenarios

(i) Units determined entirely by the grammar
E.g., ALIGN(LxWd,L,PWd,L), or, in non-OT terms, the beginning of a lexical word initiates a prosodic word
$\Rightarrow$ compounds 2 words, prefixes and proclitics left out, suffixes and enclitics folded in
p-word then acts as rule/constraint domain

(ii) Units determined entirely by processing
Sequences stored as units (or, accessed in unit stored form, even if decomposed alternatives exist) display phonological fusion internally, propensity to alternate at edges.

In general, (i) should predict a cleaner pattern than (ii), with fewer frequency effects on individual items. (i) also predicts tidy interaction with (presumably?) non-processing considerations such as prosodic minimality.

(iib) Processing masquerading as grammar
Say that processing privileges left edges in such a way that prefixes and proclitics are, in general, more likely than suffixes and enclitics to get left out of the processing unit. If the tendency is strong enough, it could look like the ALIGN constraint above, perhaps with some lexical exceptions.

Similarly, effects of affix length, and differences between compounding and affixation (a given morpheme presumably participates in a wider variety of compounds than it does affixed forms) could come out of a processing model. If strong enough, they could look like grammar (plus exceptions).

(iii) Grammar with processing-grounded constraints
There's a cross-linguistic tendency for prefixes and proclitics to be less integrated than suffixes and enclitics (see Peperkamp 1997).

We often appeal to phonetic motivations for constraint rankings—why not appeal to a processing motivation for the tendency ALIGN(LxWd,L,PWd,L) >> ALIGN(LxWd,R,PWd,R)?

We need a processing version of something like Hayes's (1999) "inductive grounding" (which deals with how messy and fine-grained phonetic patterns could get phonologized into coarse phonological constraints).

(iv) Grammar that can refer to processing
This is something I've tried—constraints like ALIGN(AccessedUnit,L,PWd,L). The idea is to be able to generate a cleaner pattern than the pure processing story would predict, by letting the grammar run things, in the usual way, with limited opportunities for processing to have its say.

But of course, an argument for (iv) should process in tandem with attempts to see if your processing model can generate the clean-looking pattern on its own, à la (iib).

(31) Hierarchical structure
One of the main ideas in Nespor & Vogel and early Selkirk was that different rules would refer to different domains. E.g. primary stress (and associated lack of vowel reduction) at the p-word level, stress retraction at the p-phrase level, spirantization at the intonational phrase level.

What could be the equivalent of prosodic levels in the processing approach? Perhaps looser processing units.

This should predict that rule applicability is actually gradient and not tied to well-defined domains. E.g. stress retraction is a weaker rule, that applies only to tightly cohering units, and spirantization is stronger, applying even to more loosely cohering units.

o   What do you think about primary stress assignment in a framework like this (take the Italian case, where there can be a clear difference between primary and secondary stress, because of vowel reduction)—how do we make it obligatory that every stem or (say) disyllabic prefix has to get a primary stress?

(32) Large units
Once we get up to units like the intonational phrase and utterance, it's implausible that we're dealing with stored units very often. Many intonational phrases will never have been heard or used by the speaker before. But phonology is nonetheless sensitive to those units (or so it's claimed).

o   How could a processing theory deal with large units? Does it make any different predictions than the prosodic theory, and are they plausible?

References
Allen, Margaret (1978). Morphological Investigations. University of Connecticut Ph.D. dissertation.
Marco Baroni (1996). The allophonic distinctions between [s] and [z] in Northern Italian, and what speakers know about it. Manuscript, UCLA.
Baroni, Marco (2001). The representation of prefixed forms in the Italian lexicon: Evidence from the distribution of intervocalic [s] and [z] in northern Italian. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1999*, Dordrecht: Springer. 121-152.
D'Imperio, Mariapaola & Sam Rosenthall (1999). Phonetics and phonology of main stress in Italian. *Phonology* 16: 1-28.
Hay, Jennifer (2003). *Causes and Consequences of Word Structure.* New York & London: Routledge.

Hayes, Bruce (1999). Phonetically-Driven Phonology: The Role of Optimality Theory and Inductive Grounding. In Michael Darnell, Edith Moravcsik, Michael Noonan, Frederick Newmeyer, and Kathleen Wheatly (eds.), *Functionalism and Formalism in Linguistics, Volume I: General Papers*. Amsterdam: John Benjamins. Pp. 243-285.

Kenstowicz, Michael (1996). Base-identity and uniform exponence: alternatives to cyclicity. In Jacques Durand & Bernard Laks (eds.) *Current Trends in Phonology: Models and Methods. Volume One*. Salford: ESRI, 363-393.

Kiparsky, Paul (1982). Lexical phonology and morphology. In I. S. Yang (ed.), *Linguistics in the Morning Calm.* Seoul: Hanshin. Pp. 3-91.

Nespor, Marina (1999). Stress domains. In Harry van der Hulst (ed.) Word Prosodic Systems in the Languages of Europe. Berlin & New York: Mouton de Gruyter.

Nespor, Marina & Irene Vogel (1986). *Prosodic Phonology*. Dordrecht: Foris

van Oostendorp, Marc (1999). Italian s-voicing and the structure of the phonological word. In S.J. Hannahs & Mike Davenport (eds.) Issues in Phonological Structure. Benjamins. Pp. 197-214.

Peperkamp, Sharon (1997). *Prosodic Words*. Den Haag: Holland Academic Graphics.

Siegel, Dorothy (1974). *Topics in English Morphology*. Massachusetts Institute of Technology Ph.D. dissertation.