Class 1: Introduction and course overview

To do for next time
- Read K&K ch.2 and turn in study questions (see my web page) on Tuesday
- Check out course web page, especially **feature** links
- Do warm-up problem (ungraded) if you didn't do it already and leave it in my mailbox by end of week
- First assignment (Malagasy—see my web page) is due Friday, Oct. 1

---

1.   What's phonology?

Definition I: Phonology = study of the sound patterns of human language.

For any given language, this includes answering the questions (Goldsmith 1995[1])…
- What are the legal words of the language?
  - phone inventory (set of legal sounds)
  - phonotactics (set of legal sound sequences)
- What alternations occur?
- Which phonetic differences are contrastive? (we will not look at this much)

Definition II: Phonology = study of <u>humans' knowledge of</u> linguistic sound patterns

2.   What is our job as phonologists? There are various answers...
- To answer the questions in Definition I above (i.e., to describe phonologies).
- To develop a theory of "what tools we need in order to provide adequate descriptions of individual languages"[2] (i.e., to develop a descriptive theory).
- To explain why phonologies are the way they are by constructing…
  - a theory of what people's knowledge of linguistic sound patterns is and how they learn, store, and use that knowledge
  - plus a theory of how linguistic sound patterns change over time

This will be our focus

3.   How do we know if an explanation is a good one?
- SPE proposed that if more than one grammar can generate the observed linguistic data, the learner must have some ***evaluation metric*** for choosing one.

- The evaluation metric tentatively proposed in SPE is brevity: learner chooses the grammar with the fewest symbols. (What about ties?)

- If that's right, and if we've got the notation right too, then you can tell which grammar, out of some set of candidate grammars, the learner would choose.

- More plausibly, we want to find independent evidence as to which grammar is right, and make sure our theory explains how/why the learner chose that one—this is a lot harder!

---

[1] Goldsmith, John (1995). Phonological theory. In John Goldsmith (ed.) *Handbook of Phonological Theory.* Cambridge, MA & Oxford: Blackwell. Pp. 1-23.

[2] Dryer, Matthew S. (2003). Descriptive theories, explanatory theories, and basic linguistic theory. Ms., U. of Buffalo.

4. *Example:* **Malay/Indonesian**

Consonant inventory for native words:

p	t	c (or t͡ʃ)	k	
b	d	ɟ (or d͡ʒ)	g	
	s			h
m	n	ɲ	ŋ	
	l, r			
w		j		(Goddard 2005[3])

Nasal substitution (Moeliono & Grimes 1995[4], Blust 2004[5]—see Delilkan 2002[6] for more)

verb	actor-focus form		mC	nC		ngC	subst.	% match
putus	mə-**m**utus	'break'	0	0		0	**1,770**	100.0%
tabrak	mə-**n**abrak	'hit'	0	0		0	**89**	100.0%
cari	mə**ɲ**-**c**ari (see Nomoto 2009)[7]	'look for'	1,130	**909,000**		0	124	99.9%
karaŋ	mə-**ŋ**araŋ	'compose'	0	1		5	**28,600**	100.0%
bakar	mə**m**-**b**akar	'bake, roast'	**79,500**	56		0	3	99.9%
dapat	mə**n**-**d**apat	'get, take'	8,730	**1,150,000**		3	208	99.2%
ɟalan	mə**ɲ**-**ɟ**alan-kan	'walk'	104	**1,480,000**		8	3	100.0%
gabuŋ	mə**ŋ**-**g**abuŋ	'connect'	0	*unspellable*	**1,270**		260	83.0%
səbar	me-**ɲ**əbar-kan	'spread'	0	0		0	**20,200**	100.0%
hituŋ	mə**ŋ**-**h**ituŋ	'count'	0	8	**9,990**		?	99.9%
marah	mə-**m**arah-i	'get angry at'						
nama	mə-**n**ama-kan	'name'						
ɲala	mə-**ɲ**ala-kan	'light'						
ŋaŋa	mə-**ŋ**aŋa	'agape'						
lapor	mə-**l**apor-kan	'report'						
rajap	mə-**r**ajap	'crawl'						
wakil	mə-**w**akil-i	'represent'						
jakin	mə-**j**akin-kan	'convince'						
ikat	mə**ŋ**-ikat	'tie'						

o Formulate rules to account for the different behaviors of the prefix. Let's see which grammar (fragment) is shortest. Ignore the numbers above for now.

Convention for handouts in this course: open bullets indicate a question for you.

o What does each grammar predict for loans that begin with [f, ʃ, x, v, z]?

[3] Goddard, Cliff (2005). *The languages of East and Southeast Asia*. Oxford University Press.

[4] Moeliono, Anton M. & Charles E. Grimes (1995). Indonesian. In *Comparative Austronesian dictionary, part 1, fascicle 1*, ed. by Darrell T. Tryon. Mouton de Gruyter.

[5] Blust, Robert (2004). Austronesian nasal substitution: a survey. *Oceanic Linguistics* 43: 73-148.

[6] Delilkan, Ann (2002). *Fusion and other segmental processes in Malay: the crucial role of prosody*. NYU diss.

[7] Nomoto, Hiroki. 2009. Distantly and prosodically conditional nasal substitution in Austronesian languages. Paper presented at the annual meeting of the Austronesian Formal Linguistics Association. In , University of California, Santa Cruz.

Used Mohd Zamri Murah's database at http://sun1.ftsm.ukm.my/src/zamri/jawiDB.txt to find candidate words. For those where listed form had at least 50 Google results I searched three additional spellings. (Restricted to `.my` domain; quotation marks around target spelling; Sept. 14 2009; numbers on prev. page from same method, for comparison.)

	mC	*nC*	*ŋC*	*nasal substitution*
f	mem-fitnah-kan (369)	men-fitnah-kan (33)	meng-fitnah-kan (33)	me-mitnah-kan (0)
	mem-fitnah (69,000)	3,530	346	5
	mem-fail-kan (29,000)	5,510	134	0
	mem-faks-kan (57)	0	5	0
	mem-fardu-kan (295)	1	0	0
	mem-fatwa-kan (1,840)	443	9	0
	mem-filem-kan (157)	5	1	6
	mem-fithan-kan (369)	25	0	0
	mem-fokus-kan (15,900)	1440	393	61
	mem-formal-kan (70)	9	0	0
	mem-format-kan (133)	7	4	0
	mem-fotostat (144)	0	2	1
	mem-fikir-kan (5)	0	0	94,100
	pem-faktoran (2,380)	8	8	3
	pem-fitnah (832)	237	6	3
	pem-fokus-an (875)	4	3	5
avg.:	*85.8%*	*6.5%*	*1.5%*	*6.2%*
ʃ <sy>	mem-syaratkan (0)	men-syarat-kan (9000)	meng-syarat-kan (10)	me-nyarat-kan (7)
	0	men-syariat-kan (1570)	2	0
	0	men-syirik-kan (1,980)	7	3
	0	men-syukuri (4,480)	39	35
	7	pen-syarah (312,500)	5	789
	0	pen-syaratan (248)	1	9
avg.:	*0.0%*	*98.9%*	*0.3%*	*0.8%*
x <kh>	mem-khabar-kan (0)	men-khabar-kan (5)	meng-khabar-kan (4,710)	me-ngabar-kan (265)
	0	2	meng-khatam-kan (850)	2
	0	1	meng-khatan-kan (610)	1
	0	0	meng-khayal (101)	0
	0	2	meng-khianat (982)	0
	0	6	meng-khianat-i (14,200)	0
	0	0	meng-khidmat-i (92)	0
	0	0	meng-khuatir-i (316)	1
	0	0	meng-khuatir-kan (374)	5
	0	0	meng-khusus (18,200)	0
	0	6	meng-khusus-kan (2,720)	1
	0	0	meng-khusyuk-kan (91)	0
	0	0	peng-khayal (57)	0
	0	9	peng-khianat (19,100)	1

Ling 200A, Phonological Theory I. Fall 2010, Zuraw

	mem-	men-	meng-	me-
	0	3	peng-khianat-an (3,720)	0
	0	8	peng-khusus-an (92,900)	60
	0	0	peng-khutbah (99)	0
avg.:	*0.0%*	*0.1%*	***99.5%***	*0.4%*
v	mem-vakum (29)	men-vakum (0)	meng-vakum (0)	me-makum (0)
	pem-vaksin-an (275)	8	4	0
avg.:	***97.9%***	*1.4%*	*0.7%*	*0.0%*
z	mem-zahir-kan (0)	men-zahir-kan (11,500)	meng-zahir-kan (136)	me-nahir-kan (1) me-nyahir-kan (0)
	0	men-zakat-kan (91)	0	me-nakat-kan (0) me-nyakat-kan (5)
	0	men-zalimi (4,120)	3	0,0
	0	men-zalim-kan (128)	0	0,0
	1	men-ziarah-i (16,700)	10	me-niarah-i (0) me-nyiarah-i (1)
	0	pen-zahir-an (4,740)	114	0,0
avg.:	*0.0%*	***98.5%***	*0.6%*	*0.9%*

Non-majority data are probably rare enough to be dismissed (typos, other languages/dialects, neutralization with other initial consonants [for the last column], speakers who spell the loan consonant the standard way but nativize it in pronunciation...), except maybe the *n-f* forms and the root *fikir*.

o So which grammar, if any, made good predictions? Your thoughts on what conclusions we can draw from this case?

5. *Excursus***: does the learner really have/need an evaluation metric?**

Idea of evaluation metric suggests that learner constructs multiple grammars and chooses one.
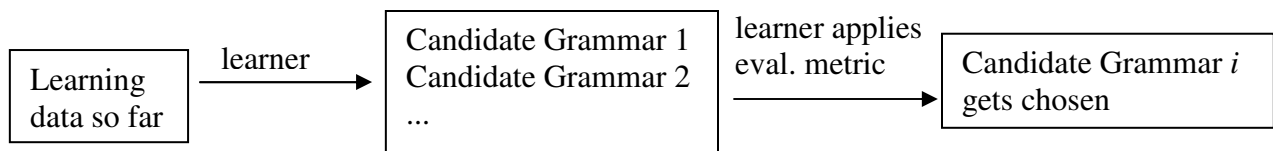
Or: learner follows an algorithm that develops a single grammar, never considering alternatives
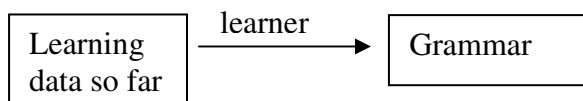
▪ ...and we can state an evaluation metric such that the grammar arrived at always scores at least as well as any alternatives
▪ ...or, we are unable to state such an evaluation metric (except the one that just runs the algorithm and then assigns a winning score to the result).

Or, learner constructs preliminary grammar; considers making a minimal change; accepts change under certain circumstances; considers making a change to result, etc. (i.e., "hill-climbing").

6. *Example:* **French elision/liaison** (SPE p. 353 ff.)

By the logic above, a theoretical innovation is held, in SPE, to be a good one if it allows more <u>concise</u> descriptions of attested/common phenomena than of unattested/uncommon phenomena.

		obstruent- or nasal-initial	liquid-initial	vowel-initial	glide-initial
		/**g**arson/ 'boy'	/**l**ivr/ 'book'	/**ɛ**nfant/ 'child'	/**w**azo/ 'bird'
obstruent- or nasal-final	/pəti**t**/ 'small'	pəti_ garsõ	pəti_ livr	pətit ãfã	pətit wazo
liquid-final	/ʃɛ**r**/ 'dear'	ʃɛr garsõ	ʃɛr livr	ʃɛr ãfã	ʃɛr wazo
vowel-final	/l**ə**/ 'the'	lə garsõ	lə livr	l_ ãfã	l_ wazo
glide-final	/pare**j**/ 'similar'	parej garsõ	parej livr	parej ãfã	parej wazo

For the sake of reconstructing the argument, use the archaic feature [vocalic] and the still-current feature [consonantal]:

	vocalic	*consonantal*
obstruents	–	+
nasals	–	+
liquids	+	+
glides	–	–
vowels	+	–

- Propose rules to account for the C- and V- deletions, without using Greek-letter variables.

- Revise the rules, using Greek-letter variables

- Do Greek-letter variables allow us to compress these two rules:

$$\begin{bmatrix} +\text{voc} \\ -\text{back} \end{bmatrix} \rightarrow \varnothing \text{ / \_\_\_ \# [–cons]} \quad \text{"nonback vowels and liquids delete before vowels and glides"}$$

$$\begin{bmatrix} -\text{high} \\ +\text{cons} \end{bmatrix} \rightarrow \varnothing \text{ / \_\_\_ \# [+nasal]} \quad \text{"nonhigh consonants and glides delete before nasals"}$$

- According to SPE's logic, how should the typology guide us in deciding whether to allow the same Greek-letter variable to apply to different features within a rule?

7.   Reasoning above relies on assumptions about linguistic typology:

- Assume a rule is cross-linguistically common only if it's favored by learners—i.e., learners tend to mislearn, in the direction of a more-favored grammar.
- Assume that learners favor short/simple/whatever rules.
- Therefore, rules that are cross-linguistically common should tend to be short.
- Therefore, our theory of rules, which determines what type of notation length is calculated on, should make common rules shorter than uncommon ones.
- Therefore, a theoretical innovation is good if it makes common rules shorter than uncommon ones.

=> We're not really using "short" (or "simple") in any fixed sense. Rather, we're tailoring the notation to make the rules that we think learners favor appear short. [And of course, that first assumption is questionable...]

This leads us into slippery territory in deciding whether shortness is the right criterion:
- Are learners innately endowed with a certain notation, which they use to calculate grammar length? (i.e., shortness really is the evaluation criterion)
- Or is it the case that learners employ some other evaluation metric entirely, but we've created a system of notation that makes goodness according to the real evaluation metric translate into shortness in our notation?

Something for you to think about, though no answers will be forthcoming: We've seen how to evaluate a particular description or even a theoretical innovation, given a framework like SPE.

o   But how do you evaluate the framework itself—in particular, how can we evaluate a principle such as "if more than one grammar can generate the observed linguistic data, the learner chooses the grammar with the fewest symbols"?

8.   What to expect in this course

Especially if you have a fair amount of background already, the first couple of classes (and readings, from the textbook) should feel like review. And most of the topics on the syllabus should be familiar to you.

But, we will be addressing these topics at an advanced level, relying on the research literature rather than textbook material where possible, and doing problem sets that are less sanitized than in an undergraduate course (though still somewhat sanitized).

9.   Administrative matter to discuss

- I polled last year's students, and they though it would be a good idea for us to have optional discussion sections to go over the homework.
- They'd be one hour long, on Mondays, just in weeks where there was a homework due the previous Friday.
- Let's try it out—we have to figure out what is the best time of day for you. Before 12 or between 2 and 4 work best for me.

Next time: *Notation review and extrinsic rule ordering*