TETU and Non-reduplicative correspondence

(1) Today's agenda

- Finish up discussion of reduplication with a brief discussion of fixed segmentism
- Move on to correspondence relations beyond those found in morphological reduplication

TETU and fixed segmentism

(2) Overview of what we have seen so far

- Reduplication is specified in the input, as a RED morpheme
- The source, amount, and placement of copied material can be determined by constraints on the RED morpheme (ALIGN, LOCALITY, etc.)
- The exact segmental makeup is determined by the interplay of faithfulness and markedness
    - Here, $\mathcal{F}$-IR/$\mathcal{F}$-BR) and $\mathcal{M}$
- When $\mathcal{F}_{IB} \gg \mathcal{M} \gg \mathcal{F}_{\{I,B\}R}$ we get a pattern where marked structures are preserved in the base ($\mathcal{F}_{IB} \gg \mathcal{M}$) but eliminated in the reduplicant ($\mathcal{M} \gg \mathcal{F}_{\{I,B\}R}$)
    - The Emergence of the Unmarked (TETU)

(3) Example: reduplication of loanwords Tagalog (Zuraw 1996)

Pattern for at least some speakers:

Base	Redup	Gloss
θeŋkju	mag-te-θeŋkju	'say thank you'
fambʊl	mag-pa-fambʊl	'to fumble'
ʃapiŋ	mag-sa-ʃapiŋ	'to shop'

- IDENT-IO[±contin],[±anterior],etc. $\gg$ *θ, *f, *ʃ $\gg$ IDENT-BR[±contin],[±anterior],etc.

(4) This raises the possibility that *all* IDENT$_{BR}$ constraints can be outranked; the segment is copied, but can't retain any of its feature specifications (unless they happen to be completely unmarked)

Hypothetical example, assuming [ʔ] = least marked C, [ə] = least marked V, CV = least marked σ

Base	Redup
$b_1l_2u_3p_4i_5$	$ʔ_1ə_2$-$b_1l_2u_3p_4i_5$
$m_1o_2f_3o_4$	$ʔ_1ə_2$-$m_1o_2f_3o_4$
$g_1a_2b_3u_4$	$ʔ_1ə_2$-$g_1a_2b_3u_4$

- Technically, the theory provides for a reduplication analysis of this pattern. Should we be worried?

(5) Yoruba (Marantz 1982, Pulleyblank, Alderete et al, etc.)

Verb	Noun	Gloss
lọ	lílọ	'go/act of going'
dùn	dídùn	'be tasty/tastiness'
gbóná	gbígbóná	'be warm/warmth'
jɛ	jíjɛ	'eat/act of eating'
rí	rírí	'see/act of seeing'

- Reduplicant always has [í], no matter what the input vowel
- What are three possible analyses of the [í]?

(6) Alderete et al: phonological fixed segmentism by defaults

- [i] is default: *non-i $\gg$ *i

/d/			NUC	DEP	*NON-I	*i
	a.	d	*!			
	b.	da, de, do, du		*	*!	
☞	c.	di		*		*

- Prevented from copying non-[i] vowels (TETU configuration)
  - IDENT$_{IB}$(vowel features) $\gg$ *non-i $\gg$ MAX$_{BR}$, *[i]
- Result: epenthetic [i] rather than copied vowel

/RED-da/			IDENT$_{IB}$	*NON-I	MAX$_{BR}$	*i
	a.	$\underline{d_1a_2d_1a_2}$		**!		
	b.	$\underline{d_1i_2d_1i_2}$	*!			**
☞	c.	$\underline{d_1id_1a_2}$		*	*	*

- Alternatively: imperfectly copied [a] rather than perfectly copied [a]

/RED-da/			IDENT$_{IB}$	*NON-I	IDENT$_{BR}$	*i
	a.	$\underline{d_1a_2d_1a_2}$		**!		
	b.	$\underline{d_1i_2d_1i_2}$	*!			**
☞	c.	$\underline{d_1i_2d_1a_2}$		*	*	*

Either way, claim is that the quality of the [i] is predictable.

- A desirable type of result: in reduplication, you either get a perfect copy of the original segment, or one that deviates from the original in a predictable way (by being less marked)

(7) Unfortunately, not true of all reduplication cases

- English *schm-* reduplication (*table, schmable*); [ʃm] is almost certainly not epenthetic or less marked than [t]
- Still need "morphological fixed segmentism": /RED+*schm*+*table*/

(8) Claimed distinctions

- Phonological fixed segmentism: segment is a "default" (least marked)
- Morphological fixed segmentism: segments must display properties of an affix
  - Appear near edge of a word (prefix/suffix)
  - Other "morpheme"-like properties (irregularity, contextually conditioned allomorphy, etc.

(9) How can we tell the difference? E.g., Igbo

a. High vowels copied exactly

    ti-ti   'cracking'
    nu-nu   'pushing'
    ji-ji   'snapping'
    ju-ju   'being full'
    mɪ-mɪ   'drying'
    mu-mu   'learning'

b. Non-high: i before alveopalatal, u before labial (with ATR harmony)

    cɪ-cɔ   'seeking'
    ɲi-ɲo   'shadow'
    ɣɪ-ɣɔ   'begging'
    bu-be   'cutting'
    gbu-gbe   'crawling'
    ku-kwe   'agreeing'

    c. Otherwise i, harmonizing in ATR and rounding with stem vowel

> ki-ke   'sharing'
> nɪ-na   'going home'
> kʊ-kɔ   'telling'
> nu-no   'swallowing'

- Alderete & al claim that the vowel is a copy in (a), epenthetic in (b, c)

- But what precludes an underlying /i/? or [+high] feature? (A prefix, forced to infix into redu-plicant to satisfy ONSET)

(10) The upshot

- The idea of phonologically-determined fixed segmentism has the potential to yield a more restrictive theory of possible fixed segments; unfortunately, it's too restrictive (we still need the more powerful mechanism of overwriting in reduplication)

- The choice seems arbitrary in many cases

  - In fact, it's not even clear that a purely phonological analysis is ever truly necessary

- This is not to say it's harmful (the theory predicts it, and it does not generate monsters)

Non-reduplicative correspondence

(11) Tagalog vowel alternations and pseudoreduplication

    [see solution on separate handout]

(12) What creates a correspondence relation between simlar syllables?

- Possibility 1: at some point, when the similarity between consecutive syllables (or feet, etc.) is too great to be a coindence, it is given reduplicative structure in the lexicon

- Possibility 2: rankable constraint remanding that *all* words have reduplicative structure

(13) The REDUP constraint (Zuraw 2002)

- Words must have reduplicative structure (i.e., substrings that stand in correspondence with one another)

(14) Interaction of REDUP with IDENT constraints

- New dimension of correspondence: base to pseudo-reduplicant (I'll call this BP, to distin-guish from BR; Zuraw calls it $\mathcal{F}_{\kappa\kappa}$ (κ = "coupled")

/dodo/			IDENT-BP[$\pm$low]	REDUP
☞	a.	$[d_1 o_2][d_1 o_2]$		
	b.	$[d_1 o_2 d_3 o_4]$		*!

/dado/			IDENT-BP[$\pm$low]	REDUP
	a.	$[d_1 a_2][d_1 o_2]$	*!	
☞	b.	$[d_1 a_2 d_3 o_4]$		*

(15) Allowing BP correspondence in spite of mismatches: REDUP ≫ IDENT-BP

/dado/			IDENT-BP[$\pm$low]	REDUP	IDENT-BP[$\pm$mid]
☞	a.	$[d_1 a_2][d_1 o_2]$	*!		
	b.	$[d_1 a_2 d_3 o_4]$		*	

/dedo/			IDENT-BP[$\pm$low]	REDUP	IDENT-BP[$\pm$mid]
☞	a.	$[d_1 e_2][d_1 o_2]$			*
	b.	$[d_1 e_2 d_3 o_4]$		*!	

- But a problem: what candidate would really win here?

(16) Preventing total agreement with IDENT-IO

/$d_1e_2d_3o_4$/		IDENT-IO[±bk]	IDENT-BP[±low]	REDUP	IDENT-BP[±mid]
☞	a. $[d_1{}^1e_2{}^2][d_3{}^1o_4{}^2]$				*
	b. $[d_1{}^1e_2{}^2d_3{}^3o_4{}^4]$			*!	
	c. $[d_1{}^1o_2{}^2][d_3{}^1o_4{}^2]$	*!			*
	d. $[d_1{}^1e_2{}^2][d_3{}^1e_4{}^2]$	*!			*

- Subscript = IO relations, superscription = surface correspondence (BP) relations
- Crucially, IO relation holds over entire string (a la Inkelas & Zoll)
- The basic scheme: pseudo-reduplicative correspondence is established only when the mismatches are small enough to be tolerable (that is, violate lower-ranked IDENT-BP constraints)
- If IDENT-IO[f] ≫ IDENT-BP[f], then pseudo-reduplicative correspondence will be undetective (doesn't enforce any additional identity beyond what is necessary to establish correspondence in the first place)

(17) Using BP correspondence to block raising of nonfinal mids

/todo-in/		ID-IO[voi]	ID-BP[hi], ID-BP[lo]	ID-IO[hi] /non-stem-final	REDUP	*NONFINAL MID	ID-BP[voi]
☞	a. [to][do]hin					*	*
	b. [to][du]hin		*!				*
	c. [todu]hin				*!		*
	d. [tu][du]hin			*!			*
	e. [to][to]hin	*!				*	
	f. [do][do]hin	*!				*	

/gaston-an/		ID-IO[voi]	ID-BP[hi], ID-BP[lo]	ID-IO[hi] /non-stem-final	REDUP	*NONFINAL MID	ID-BP[voi]
	a. [gas][tos]an		*!			*	*
	b. [gas][tus]an		*!				*
	c. [gastos]an				*	*!	*
☞	d. [gastus]an				*		*

- When IDENT-BP outranks some $\mathcal{M}$, regular phonological process can be blocked (but only when conditions for BP correspondence are met)

(18) An extremely important point

- The blocking of raising when the word looks pseudo-reduplicated is only a statistical tendency in Tagalog, not a hard and fast rule
  - Really REDUP-ish words tend not to undergo (*yoyo, todo*; also many pseudoreduplicated native words, like *dede* and *toto?o*)
  - Moderately REDUP-ish words go one way or the other, or vary
  - Other words tend to raise (though extremely recent or nonce loans may also resist it)
- Lots of exceptions (esp. pseudo-reduplicated words that do raise); we'll talk more about the analysis of exceptions later in the semester

(19) What are we to make of this REDUP constraint?

- Why would words want to have otherwise unsupported reduplicative structure?
- Plausible alternative: don't care about having reduplicative structure per se, but when syllables are similar enough, speakers are driven to conclude that similarity is too great to be due to chance, and must involve correspondence

(20) • How much does it take to conclude the similarity is "too great to be accidental"?

- Tagalog provides abundant evidence for what reduplication should really look like. These aren't right at all
    - E.g., *tonto*: would have to be *toton* or *tonton*
    - Similarly *piloto*: repeated syllable at wrong end of word, has otherwise unlicensed segmental alternation
    - Words are also mostly too short to be reduplicated (native roots are minimally disyllabic)

(21)   English words "fixed" to enforce surface identity (Zuraw 2000)

Standard (etymological)	Google hits	"Fixed"	Google hits
sherbet	62,900	sherbert	12,000
pompon	17,700	pompom	15,500
orangutan	55,600	orangutang	6,130
hara kiri	11,100	hari kari	8,430
smorgasbord	71,500	smorgasborg	1,740
sacrosanct	39,500	sancrosanct	201
Inuktitut	23,500	Inuktituk	751
Abu Dhabi	135,000	Abu Dhabu/Abi Dhabi	126/67
asterisk	613,300	asterist/askerisk	57/100

- OED etymology of *orangutang*: "In many of these languages, the second element *utan* 'woods' has been corrupted to jingle with the first"

(22)   Some other historical changes:

- hotchpot > hotchpotch
- maamajumbo (or something like this) > mumbojumbo
- catapillar > caterpillar > callerpillar

And lots and lots (and lots) of errors, of varying frequencies

frustrum	sumblime	axphyxiate
farmiliar	frangrance (55,400 hits!!!)	Mitshubishi/Mitshubitshi
aprostrophe	rhondodendron	andpersand/ampersamp
perservere	parenthensis/parenthenses	andaconda
proprogate	gongonzola	arkvark/ardvard
porpcorn	gonblins	snapshop
cornucorpia	denfense/denfend	tollboth/toolbooth
crocrodile (also in French)	fan(s)tan(s)tic	Monongahola
heroric	nunchunks	Tuscaloosca
Figuerora	pumpkim	Abbis Ababa

- Ralph on The Simpsons: "I'm bembarrassed for you!"
- Seemingly purely orthography examples:
    - mnemnomic, rhubharb, whirlwhind, Philiphines/Philliphines

(23)   Reduplication as a strategy to "fill up time" when you don't really know the word very well

- Kyracyra [kirəsirə]
    - http://pdaphonehome.com/forums/showthread/t-9761.html
      I actually removed the Palm and Kyracera software twice. Still can't get a conntection.
- gazpazo, okinomiyaki,
- Yasfat Arafat/Yasfak Arafak (by a guy trying to remember Yasser Arafat's name); Afrafat and Arafrat more common

(24)   What we see from English

- Low-level tendency to enforce identity across remote parts of the word

- Tendency is strongest in case the two halves of the word are very similar (orangutan, smorgasbord)
- Also shows up when word not fully known/accurately retrieved

The issue:

- If speakers simply assign (pseudo)-reduplicative structure when they have observed similarity that's too great to be coincidental, why would we get the same effect with these "partially known words"?

(25) Some conjectures

- Zuraw's REDUP constraint is part of a more general bias that speakers have to establish correspondences between elements—even if they are not extremely similar
- Calculating these correspondence relations permits learning morphology, phonological alternations, etc.
  - Start by assuming that A and B are the same item, and see whether the relation between them is a general one that holds of other pairs, too
- The bias to assume a REDUP relation may be stronger than others; reduplicated forms are often ambiguous (they could be sequences of accidentally similar syllables), and the learner needs to be predisposed to try to analyze them as reduplication, or else nothing would ever be learned as reduplication
- REDUP also acts as a form of data compression:
  - Learner wants to reduce unpredictability as much as possible; for each word, attempts to represent word as economically as possible
  - Of course, the segmental makeup of words is by and large unpredictable (it just has to be listed). But two things are predictable:
    * Epenthetic segments
    * Copies of segments that are listed elsewhere
  - Epenthesis can only get you so far in a vocabulary of realistic size, but REDUP could greatly reduce the amount you need to list:
    * [tono] = $/t_1o_2[+nas]_1X_2/$ (schematic; not meant to be a concrete proposal)
    * But how far does this extend? Is *aardvark* underlyingly $/X_2X_3\begin{bmatrix}+\text{cor}\\-\text{dors}\\+\text{voi}\end{bmatrix}_4v_1a_2r_3k_4/?$

☞ It is an open question whether these factors need to be encoded as a grammatical constraint that is rankable (as Zuraw does), or whether they could exert a force as learning biases, that determine how words are represented lexically.

Agreement by correspondence

Rose and Walker (2004) A typology of consonant agreement as correspondence. Language 80, 475-531

(26) Aari (data cited by Rose & Walker, p. 482)

Verb	Causative	Gloss
giʔ	giʔ-sis	'hit'
duuk	duuk-sis	'bury'
sug	sug-zis	'push'
naʃ	naʃ-ʃiʃ	'like'
tʃ'aaq	tʃ'aaq-ʃiʃ	'curse'
ʃaan	ʃaan-ʃiʃ	'urinate'
ʒaag	ʒaag-ʒiʃ	'sew'

- Standard analysis: spreading
- The problem: how do segments affect one another at a distance?

(27)   The observation: consonant harmony typically targets a few particular classes of segments

- liquids/nasals
- sibilants
- anterior segments

These classes of segments seem to involve sets of very **similar** consonants.

(28)   Building in the intuition that similarity drives harmony:

- CORR-C↔C: consonants should stand in a correspondence relation to one another
- Perhaps a universal ranking based on similarity :

  CORR-T↔T (identical stops) ≫ CORR-T↔D (same place) ≫ CORR-T↔K (same voicing), etc. . .

(29)   This idea is very similar to suggestion in (19) above: if you're already similar, you get put into a correspondence relation (which might make you get even more similar)

(30)   A parallel between harmony and reduplication: locality

- Reduplicants tend to take material from neighboring part of the base
- Consonant harmony is stronger when the relevant consonants are in adjacent syllables

"It's good for adjacent things to be identical" (***if they are in a correspondence relation)

(31)   How can we interpret this correspondence relation? (compare possible interpretations of REDUP above.) Can it explain the English "error" data?