

# Identity Avoidance in Phonology and Morphology

Moira Yip

University of California, Irvine

Many languages avoid sequences of homophonous elements, be they phonemes or morphemes. It is argued that a single principle underlies all such cases of avoidance, and that this principle can interact with the rest of the grammar resulting in the omission of one morpheme, or forcing a choice between different syntactic outputs. This paper is formulated within Optimality Theory, and makes three main points.<sup>1</sup> First, at least some inputs to the Optimality Grammar must be abstract morphological specifications like PLURAL. They are phonologically incomplete outputs of the morpho-syntax. Second, morpheme realization results from an attempt to meet output targets in the form of constraints: REPEAT,  $\sigma = a$ ; PL=s, and so on. Such morphemes do not have underlying forms in the familiar sense (cf Hammond 1995, Russell 1995). Third, the target constraints may be out-ranked by phonological constraints of various kinds, particularly constraints against the repetition of elements, here called the OCP. The elements may be phonological (feature, segment) or morphological (affix, stem). These findings support the view of Pierrehumbert (1993a) that identity has broad cognitive roots. Section 1 gives some background on the handling of morphological data in OT. Section 2 discusses identity avoidance in morphology and sets out the basic proposal. Section 3 discusses cases of adjacent homophonous morphemes in Mandarin, English, and Classical Greek. Section 4 looks at homophonous morphemes on adjacent words (but which are not themselves string-adjacent) in English and Hindi. Section 5 looks at reduplication in Javanese, and argues that echo-words result from the tension between a requirement that *penalizes* a sequence of two identical stems, OCP(Stem), and one that *requires* two identical stems, REPEAT(Stem).

This work took place over some time, and resulted in two papers: the present work, and a detailed exploration of reduplication and identity-avoidance in Javanese, Yip (forthcoming). The early sections form the necessary background to both papers, and are largely identical; the data and analyses in each paper are different, although complementary.

## 1. Morphology in Optimality Theory:

Optimality Theory (Prince and Smolensky 1993, McCarthy and Prince 1993a, and a host of others) as currently conceived is a theory of not only phonology but also many aspects of

---

<sup>1</sup> This paper has benefited greatly from comments from audiences at this conference, and at the Conference on Features in Optimality Theory at the University of Arizona in April 1995, from the participants in the seminar on Constraints in OT at UC Irvine, and the workshop on Theoretical East Asian Linguistics, UC Irvine, and from a number of individuals, including Diane Brentari, Andrew Carstairs-McCarthy, Larry Hyman, Sharon Inkelas, John McCarthy, Orhan Orgun, David Perlmutter and Jerry Sadock. Special thanks to Steven LaPointe, Diane Brentari, Patrick Farrell and the rest of the Davis morphology community for making the conference and this volume possible.

morphology. It includes mechanisms for controlling the size and content of reduplicative morphemes, constraints responsible for the precise placement of affixes, constraints that explain the choice among allomorphs, and constraints that pick the right member of a suppletive set. The essence of Optimality Theory is that it is an output-based grammar in which all possible outputs for some input are assessed by a universal set of ranked and violable output constraints. Only the ranking is language specific. The optimal candidate is evaluated as follows. All outputs which violate the highest ranked constraint are thrown out, and those remaining are evaluated by the next highest ranked constraint. This procedure continues until only a single candidate survives. In the event of a tie at any point in the procedure, the tying candidates are passed on down to the next constraint, which decides matters. Let us see how a selection of morphological phenomena is handled in Optimality Theory.

### 1.1 Reduplication:

McCarthy and Prince 1993a, 1994 lay out an approach to reduplication in which an abstract input morpheme, RED, passes through GEN and results in a set of output candidates in which RED is realized as a full or partial copy of the base. The choice among these candidates is governed by a set of constraints that determine the size of the reduplicant (such as RED=), and its content, controlled by a set of constraints that enforce identity between base and RED, and prefer maximal copying. If RED= is ranked above MAX(imality), the reduplicant will be monosyllabic (e.g. Ilokano *bas-basa*, *da-da.it*). If MAX is the higher-ranked of the two, reduplication will be total. (e.g. Yoruba *agba-agba*, *oru-oru*) These constraints can also interact with syllable structure constraints. For example, if NOCODA dominates MAX, the reduplicant will be coda-less (e.g. Balangao: *tagta-tagtag*, *tayna-taynan*).

### 1.2 Affix placement:

A family of Alignment constraints (McCarthy and Prince 1993b) aligns the edges of prosodic and morphological categories with themselves and with each other. A purely phonological alignment phenomenon would be the placement of feet at the ends of prosodic words: ALIGN-LEFT: (PrWd, Foot) (e.g. English (*Táta*)*ma(góuchi)*, \**Ta(táma)(góuchi)*). A purely morphological example would be the placement of an affix at the beginning of a stem: ALIGN-LEFT(Affix, Stem) (Tagalog prefix *ag-*). A morphology/phonology interface example would place a foot at the end of a root, ALIGN-RIGHT (Root, Foot) as in Indonesian *bi(cará)-kan* (Cohn and McCarthy 1994)

Particularly striking results come in the treatment of affixes that vacillate between prefixation and infixation as a consequence of the interaction between these alignment constraints and syllable structure constraints. For example, in Tagalog the prefix *um-* of *um-aral* is placed in position by a constraint ALIGN-LEFT (*um-*, stem), but this constraint is dominated by NO-CODA. The result is to force infixation of *um-* before C-initial roots, such as *gr-um-adwet*, since the prefixed form \**um-gradwet* would have an extra coda.<sup>2</sup> Zoll (1994) has used a similar approach to explain the behavior of morphemes that surface as floating features at varying positions in the root, like Chaha imperative palatalization; in some cases these moveable affixes may surface as independent segments as well, like Yawelmani suffixal glottalization.

<sup>2</sup> The root codas survive because PARSE dominates NOCODA. Note that ALIGN-ag >> NOCODA >> ALIGN-um, since *ag-* is always prefixed even if it results in NO-CODA violations.

### 1.3 Allomorphy:

Carstairs (1990) discusses several cases in which allomorph choice is phonologically conditioned. This section and the next one summarize the treatment of such facts within OT. Mester (1994) proposes that a prosodic selection process in the lexicon can pick one allomorph from a set of alternatives by looking at which would form the optimal output with respect to a set of constraints. He studies Latin perfect stems, which can be formed by attachment of either *-u-* or *-s-*. The default choice is attachment of *-u*, e.g. *mon-u-i:*, but in stems with final heavy syllables, *-s* is used instead: e.g. *auk-s-i:* \**aug-u-i:*. He suggests that this can be understood as the avoidance of an output in which a single light syllable, *.u.*, cannot be incorporated into a foot because it is 'trapped' between two heavy syllables (one from the root, and one from the final suffix); here I mark foot boundaries with [ ]:

- (1)            \*aug u i:                    cf        auk si:  
                   [ ] [ ]                            [ ] [ ]

The prosodic selection process thus picks *auk-si:* from a set of alternative outputs {*aug-u-i:*, *auk-s-i:*} for the input /auk-PERF-i:/.

### 1.4 Suppletion:

Tranel (1994) examines French determiners, where the feminine 1-sg-poss *ma* is replaced by the masculine *mon* before vowel-initial feminine nouns, and the masculine *ce* 'this' is replaced by the feminine *cet* before vowel-initial masculine nouns.<sup>3,4</sup> Standard accounts simply stipulate this distribution, but Tranel's insight is that both suppletions supply an onset for the following syllable, and are thus phonologically driven. He suggests that a suppletion set is judged against constraints that require gender agreement, and onset satisfaction, and that no one form is basic. Gender agreement can be over-ridden by the need for an onset: ONSET >> GENDER. The result will be, correctly, that the grammar will pick the C-final candidate before a V-initial noun, irrespective of gender.

()

	ONSET	GENDER
a. mon <sub>M</sub> arme <sub>F</sub>		*
b. ma <sub>F</sub> arme <sub>F</sub>	*!	
c. ce <sub>M</sub> abbé <sub>M</sub>	*!	
d. cet <sub>F</sub> abbé <sub>M</sub>		*

<sup>3</sup> *cet* and *cette* are orthographically distinct, but both are phonetically [s t]. I follow Tranel in assuming that both are feminine, and the orthography is irrelevant.

<sup>4</sup> Perlmutter (this volume) independently reaches very similar conclusions.

These results make it hard to identify a clear dividing line between morphology and phonology. What is more, they go much further to blur the distinction than does the interleaving of phonology and morphology found in lexical phonology. In lexical phonology, each component has its own character: the entities are different, and the rules are different. In Optimality Theory, this is not necessarily the case. Alignment is the most striking example. Alignment appears to play a role in pure morphology, in pure phonology, and at the interface.

In this paper, I want to focus on another area in which phonology and morphology appear to overlap, the area of identity avoidance. It is a commonplace in phonology that sequences of adjacent identical elements are avoided, and this is enshrined as the Obligatory Contour Principle, or OCP (Leben 1973, McCarthy 1986, Yip 1988, Odden 1988, Myers 1993, Pierrehumbert 1993a, and others). What has received less attention in OT are superficially similar cases in morphology (but see Golston 1994, 1995), although the generative literature includes many such cases. See for example Sadock 1972, Menn and MacWhinney 1984, Hyman and Mchombo 1992, and particularly Stemberger 1981,

## 2. Identity Avoidance in Morphology

Avoidance of identity in morphology takes several forms. I will divide them into four categories.

- (i) a. The same morpheme cannot appear twice in the same word
- b. Different but homophonous morphemes cannot appear adjacent in the same word, or otherwise adjacent in the sentence
- c. Homophonous morphemes cannot appear on adjacent words
- d. The output of reduplication cannot be total identity

The first type is rare, perhaps non-existent, but it is not clear that the morpho-phonology underlies this: in most cases it seems likely that syntactic and morpho-syntactic principles will achieve this end without identity avoidance being involved at all.<sup>5</sup>

The second type is quite common; the references cited above include numerous examples. A familiar and typical example is the English possessive plural: *\*cats's, cats'*. Further examples discussed in this paper include Mandarin perfective *le* and Currently Relevant State *le* (Chao 1968, Li and Thompson 1981), Classical Greek determiners (Golston 1994) and Mandarin third person pronoun *ta* (Yeh 1994). A common response in these cases is blocking, as in Chichewa. Another strategy, the one chosen in the examples discussed in this paper, is omission of one morpheme, with the remaining one carrying the semantics of both. This phenomenon is called haplology, and will be discussed in section 3.

The third type involves identical morphemes attached to adjacent words, but where the morphemes themselves are not string adjacent. Since the presence of a morpheme on one word does not satisfy the requirements of the second word, omission of a morpheme is rarely the preferred strategy for resolving the situation; instead we are more likely to see syntactic movement, replacement by an alternative morpheme, or simple blocking. Cases of this kind

---

<sup>5</sup> But see Hyman and Mchombo (1992) on Chichewa, where two instances of the applicative morpheme with different semantic roles are acceptable within the same word, so long as they are not adjacent. I assume that the syntax allows this because one morpheme can be used for several semantic roles.

include English *-ing* (see Ross 1972, Milsark 1988, and, for a different view, Pullum and Zwicky (1991)), and Hindi *-ko* Dative and Accusative markers (Mohan 1992). In the Hindi example, sequences of two NP's, each marked by the suffix *-ko*, are avoided. These cases are discussed in section 4.

The fourth type are usually called "echo words": reduplication accompanied by a small change such that the two halves are not quite identical. English *table-shmable* is an example of an echo-word, and they are found in many languages including Turkish, Bengali, Chinese, and a host of others. See Yip 1992 for many examples. I will propose that echo-words result from a tension between two constraints, one requiring repetition (reduplication) and one banning repetition (the OCP). The primary case discussed here is Javanese, in section 5. A much fuller discussion of echo-words can be found in Yip (forthcoming).

## 2.1 A Summary of the Proposal:

The central theme of this paper is the avoidance of complete identity. In phonology the OCP has been the usual way of addressing such issues, and in morphology Menn and MacWhinney (1984) propose the Repeated Morph Constraint (RMC). Both principle rule out sequences of phonologically identical elements; even the RMC does not rule out sequences of phonologically distinct morphemes. The clear similarity between these two constraints suggests that both phonology and morphology are subject to a single general principle that avoids repetition. This principle may even have quite general cognitive roots, as proposed by Pierrehumbert (1993a). I shall continue to use the term OCP for this principle without intending some sort of sovereignty for phonology over a morphological domain. In Yip (forthcoming) I adopt a different term, \*REPEAT, which has its origins in the RMC, and in a suggestion of Diane Brentari (this volume). In the rest of the paper I will continue to use constraints of the OCP family, as defined sweepingly below.<sup>6</sup>

- ( ) OCP: Output must not contain two identical elements

Like many constraints, this is subject to adjacency effects, and it may also be judged gradiently at a featural level. The consequence is that violations will be more serious the nearer two things are, and the more similar they are (Pierrehumbert 1993a). In most of this paper these subtleties will play no role.

The model I am proposing has two main parts. I outline the proposal below; further details will become clear during the body of the paper. First, there is a set of UG constraints, including:

- ( ) OCP: Output must not contain two identical elements
- MORPHDIS: "Distinct instances of morphemes have distinct contents, tokenwise"
- REPEAT: Output must contain two identical elements

---

<sup>6</sup> OCP was the term used in the first draft of the paper, and the oral version. I am thus retaining it here in line with the editors' stated policy. In any case the difference between OCP and \*REPEAT is terminological, not substantive.

The OCP constraint blocks complete repetition.<sup>7</sup> The second constraint, MORPHDIS: "Distinct instances of morphemes have distinct contents, tokenwise", is drawn from McCarthy and Prince (1995:67). They find the need for a constraint that is violated any time a segment does double duty to fulfil more than one morphological role. If this constraint dominates the OCP, sequences of homophonous morphemes cannot be avoided by haplology. If the ranking is reversed, however, we will observe haplology: to avoid repetition, one set of segments is recruited to do the work of two morphemes:

- ( )            OCP >> MORPHDIS    haplology  
                  MORPHDIS >> OCP    no haplology

The REPEAT constraint forces reduplication by self-compounding. Instead of supposing that there is an affix, RED, which must be filled, it assumes that the input has only a morphological annotation such as "PLURAL", and the grammar includes a constraint REPEAT<sub>PLURAL</sub> which must be satisfied for all plural inputs. This can be combined if necessary with constraints governing the size of the reduplicant: I will have nothing to say about this latter point.

If OCP >> REPEAT, we derive the echo-word pattern: reduplication that falls just short of complete identity. This proposal explain why echo-words seem to be most common in the case of word reduplication, where the reduplication would otherwise be total: in the case of partial reduplication, the OCP is satisfied anyway by the failure to copy the entire base. The ranking REPEAT >> OCP will mask the effects of the OCP completely, giving total reduplication.

Summarizing, the tension between the first and last constraints gives rise to the following partial typology:

- ( )            REPEAT >> OCP                    true reduplication  
                  OCP >> REPEAT    echo words: change in one half

REPEAT bears obvious similarities to various constraints proposed in the OT literature on reduplication, particularly McCarthy and Prince (1993, 1994, 1995). It does much of the work of two constraints in their 1995 paper, IDENT-BR, and MAX-BR. For the purposes of this paper, it is sufficient to merge these two into the single REPEAT.

The second part of the proposal is given below:

- ( )    a. Inputs consist of morphologically annotated roots, rather than roots with phonologically specified affixes: /kæt<sub>PL</sub> /, not /kæt-s/, and /udan<sub>HAB-REP</sub>/ not /udan-RED/.
- b. These are realized in order to satisfy specific output constraints

Within OT, this is very similar to proposals of Hammond (1995) and Russell (1995). It is also

---

<sup>7</sup> I will assume that \*REPEAT assesses complete morphological entities, such as stems, but an alternative is to assess all identity as the aggregate of individual identities between pairs of segments.

what seems to be assumed by Mester (1994). Philosophically, it has much in common with the proposals of Anderson (1992). The primary advantage of this proposal in the present context is that it allows for the absence of an affix precisely when some other affix or the root itself is able to satisfy the output constraint in question. The discussion of English in the next section will illustrate this point, and it will play an important role in Javanese.

### 3. Avoidance of adjacent homophonous morphemes

#### 3.1 English 's:

The best known case of haplology comes from English. The plural /s/ and the possessive /s/ cannot co-occur, although adding possessive /s/ to an irregular plural is fine, and so is adding it to a singular ending in /s/, or even a singular ending in /s s/..

( )	<i>Singular</i>	<i>Plural</i>	<i>Possessive Sg.</i>	<i>Possessive Pl.</i>
	child	children	child's	children's
	mouse	mice	mouse's	?mice's
	cat	cats	cat's	cats'      *cats's
	Katz	Katzes	Katz's	Katzes'      *Katzes's
	coreopsis		coreopsis's	

Compare especially *Katz's* vs *\*cats's*; *coreopsis's* vs. *\*Katzes's*. I must emphasize that like all the cases examined in this paper the illicit sequences are ruled out by some principle that disallows *phonological* identity; the sequence of a plural morpheme followed by a possessive morpheme is fine if the two are phonologically distinct, as in *oxen's*.

Two strategies are used to avoid /s-s/. One strategy is haplology: the omission of a morpheme, as in the possessive plural *cats'*. The other is insertion of a buffer vowel, as in the simple possessive *Katz's*, *Kat[s z]* (and between all stridents and suffixal -s).

Optimality Theory, as an output-based grammar, is well-suited to capturing Stemberger's (1981) insight that this and other cases of haplology do not appear to involve deletion so much as a failure to insert a superfluous morpheme if a homophonous morpheme is already in the right position. Thus if the plural /s/ is present, a plural possessive can satisfy the need to end all possessives in /s/ without adding a second /s/. This explanation, though, does not extend to the vowel-insertion between a root /s/ and a suffix /s/, and thus no unified explanation is possible. Within Optimality Theory, we can provide a single straightforward account.

I will now offer an explicit Optimality Theory analysis of the core aspects of identity avoidance, using this as my first example. Suppose, following Myers (1993), that the OCP is a constraint that can be ranked with respect to the other constraints of the grammar. Further suppose that the OCP is a sort of meta-constraint (Pierrehumbert 1993b) which can be instantiated with different arguments, and includes at least the following family;

( )	OCP (feature)	OCP (segment)
	OCP (affix)	OCP (stem)

Consider a case in which insertion is the preferred remediation strategy. The OCP must then

dominate some sort of constraint against epenthesis which, following Prince and Smolensky 1993, I will call FILL. Also high-ranked will be the output constraints that require some morphological category to be phonologically instantiated in a particular way:

- () English 's:
- PLURAL: Plurals must consist of a stem plus an -s affix.
  - POSS: Possessives must consist of a phrase plus an -s affix.
  - OCP (s): OCP (feature), where feature=[strident]
  - FILL: Don't insert
  - MORPHDIS: "Distinct instances of morphemes have distinct contents, tokenwise"

PLURAL, POSS, OCP (s) >> FILL, MORPHDIS (Epenthesis as last resort)

In the tableau below, the possessive plural *cats'* with only one *s* wins because the candidate with two *s*'s violates OCP (s), and the candidate with epenthesis violates FILL. Crucially, the single *s* satisfies the PLURAL and POSS constraints; the winning candidate violates MORPHDIS, which must thus be low-ranked.

()

cat <sub>PL</sub> POSS	PLURAL=s	POSS=s	OCP (s)	FILL	MORPHDIS
cat <sub>PL</sub> POSS-s-s			*!		
cat <sub>PL</sub> POSS-s					*
cat <sub>PL</sub> POSS-s-I-s				*!	

In the possessive of *Katz's*, the affix must be retained to satisfy POSS. FILL is thus violated in order to satisfy the higher-ranked OCP (s).

()

Katz <sub>POSS</sub>	PLURAL=s	POSS=s	OCP (s)	FILL	MORPHDIS
Katz <sub>POSS</sub> -s			*!		
Katz <sub>POSS</sub>		*!			
Katz <sub>POSS</sub> -I-s				*	

These tableaux demonstrate that the omission of one affix after the possessive plural of *cat* versus the epenthesis into the simple possessive of *Katz* follow from the dominance of OCP (s), and of the output requirement that the plural and the possessive must end in an 's morpheme. This output requirement blocks deletion of a lone plural or possessive morpheme, and OCP (s) forces use of the fall-back strategy, epenthesis. Two 's affixes will never be optimal, because they will always violate either OCP (s), if adjacent, or FILL, if separated by an epenthetic vowel, and there is always available a candidate with only one affix that violates neither. This analysis thus allows us to link the morphological "haplogy" of the plural and possessive morphemes with the phonological epenthesis of the English Plural Rule by assuming that OCP (s) plays a role in both

"components".<sup>8</sup>

Note that in an interestingly similar case, Hungarian (Carstairs 1990:17) replaces the usual *-(a)sz* suffix of the 2nd singular indefinite present indicative with *-ol* after sibilants and affricates. Here the OCP(s) apparently selects between two allomorphs, choosing *-ol* to avoid a sequence of two stridents.

English demonstrates the advantages of assuming that affixes are not present underlyingly, but are a response to satisfying an output constraint. In the next section we will see a similar case involving a verbal affix and a sentential affix which are presumably not juxtaposed until the syntax, and where the same avoidance of repetition is found.

### 3.2 Mandarin *le*

It is well known that Mandarin Chinese has a process of haplology that reduces expected adjacent instances of *le* to a single *le*. There are two different kinds of *le*. One is a verbal suffix, and marks perfective aspect. This is illustrated below.

#### ( ) Perfective Aspect

- a. Huo mie-le  
fire went-out-PF  
The fire went out.
- b. Wo wang-le ta-de dizhi  
I forget-PF 3sg-GEN address  
I forgot his/her address
- c. Ta jintian mai-le hen duo shu  
3sg today buy-PF very many book  
He/she bought a lot of books today

The other is a sentence-final particle, and conveys a subtle sense of change-of-state, relevance to discourse, and other things. For details see Chao (1968) and Li and Thompson (1981).

#### ( ) Currently Relevant State (Li and Thompson)

- a. Xia-ge yue wo jiu zai Riben le  
next-CL month I then at Japan CRS  
Next month I'll be in Japan.
- b. Nei tian ta chu-qu mai dongxi le  
that day 3sg exit-go but thing CRS

<sup>8</sup> The epenthesis depends crucially on the OCP, but the haplology in fact does not. The haplology could also follow instead from some notion of economy or faithfulness that penalizes insertion of [s], and from viewing the plural and possessive as output constraints that can be jointly satisfied by a single 's. However, the haplologized forms are certainly *consistent* with the OCP ranking necessary for the epenthetic cases.

That day she went out shopping

If the semantics demand it, a sentence can carry both *le*'s at once, as shown below:

() Both - Non-adjacent

- a. Wo he-le san bei kafei le  
I drink-PF three cup coffee CRS  
I drank three cups of coffee
- b. Feiji chu-le maobing le  
Airplane exit PF trouble CRS  
The airplane has developed some trouble

However, if the verb is intransitive and therefore sentence-final, so that both *le*'s would end up adjacent, only one occurs, and the resulting sentence is three ways ambiguous.

() \*Both - Adjacent

- a. \*Huo mei-le le  
Huo mei-le  
The fire went out (PF) (yesterday)  
The fire has gone out (CRS) (already)  
The fire went out, and that's what I'm telling you (PF/CRS)
- a. Bing dou hua le (\*le)  
ice all melted PF/CRS  
The ice all melted.

This is also true if the verb ends up sentence final because the object has been preposed, showing that it is surface adjacency that matters, not underlying adjacency, and also that trace does not interrupt the adjacency of the two instances of *le*.

- () Wo yijing ba [nei san bei kafei]<sub>i</sub> dou he le t<sub>i</sub> (\*le)  
I already BA [those three cup coffee]<sub>i</sub> all drink PF t<sub>i</sub> CRS  
I already drank those three cups of coffee (that you left me).

Other sequences of homophonous *le* are also pretty bad; most informants reject sentences like the following, where the first *le* of the intensifying suffix *-jile* is arguably not the perfective.

- () ?\* Ta yijing shuo "Hao jile" le  
He already said "Good-INT" CRS  
He already said "Wonderful!"

Sequences of other reduced affixes followed by *le* are acceptable. The following examples show

*de le, ge le*, showing that this is not a constraint on two C unstressed syllables.

- ( ) a. Wo he-le    wo-de le  
       I    drink-PF mine CRS  
       I drank mine.
- b. Wo mai-le    nei-ge    le  
       I    buy-PF    that one CRS  
       I bought that one.

Stemberger points out, though, that two instances of *de*, one a nominalizer and one showing modification, are out:

- ( )    hou de            'thick thing'  
       shu de yanse            'the book's color'  
       hou de (\*de) yanse    'the thick thing's color'

Inspection of these data yields the following observations:

- ( )    Phonological identity is needed : *de de* or *le le*. This is confirmed by the fact that in Cantonese, where the perfective and currently-relevant-state affixes are not homophonous, the combination is fine.
- The phonological identity targets the whole morpheme, not its component segments. This distinction could not be discerned in the case of English /s/, where the morpheme consisted of a single segment.
- Morphological identity is not involved (they are different morphemes), and indeed morphological information may not be needed either, given data like (19) which suggests that any two *le* are unacceptable.

A very similar approach to that taken for English -s will work for Mandarin *le*. Again the crucial idea is that the output constraint and the OCP both dominate the MORPHDIS constraint, so that omission is a way to resolve OCP violations, and the remaining *le* satisfies the output-based morphological constraint.

- ( ) Mandarin *le*:
- a. PERF: The Perfect verb must end in *le*.
  - b. CRS: Currently Relevant State utterances must end in *le*.
  - c. OCP (*le*): OCP (affix), where affix = *le*
  - d. MORPHDIS: "Distinct instances of morphemes have distinct contents, tokenwise"
- ( ) PERF, CRS , OCP (*le*) >> MORPHDIS

()

	Perf, CRS	OCP( <i>le</i> )	MORPHDIS
miePF/CRS	*!		
miePF/CRS -le- le		*!	
miePF/CRS-le			*

### 3.3 Mandarin ta haplology:

I now move to a case that involves not affixes, but pronouns, and ones that are not obviously clitics. Adjacent instances of the third person pronoun *ta* are avoided, but only when they are co-referential, as shown by the following data: (Yeh 1994)

() Wo wen ta<sub>i</sub> ta<sub>j</sub> mingtian lai bu lai  
 I ask he<sub>i</sub> he<sub>j</sub> tomorrow come not come  
 I asked him<sub>i</sub> whether he<sub>j</sub> would come tomorrow

() \*Wo wen ta<sub>i</sub> ta<sub>j</sub> mingtian lai bu lai  
 I ask he<sub>i</sub> he<sub>j</sub> tomorrow come not come  
 I asked him<sub>i</sub> whether he<sub>j</sub> would come tomorrow

() Wo wen ta<sub>i</sub> i/\*<sub>j</sub> mingtian lai bu lai  
 I ask he<sub>i</sub> he<sub>i</sub>/\*<sub>j</sub> tomorrow come not come  
 I asked him<sub>i</sub> whether he<sub>i</sub>/\*<sub>j</sub> would come tomorrow

We must distinguish between the obligatory deletion for at least some speakers in the case of adjacent *ta*, and the optional deletion in the case of a full NP antecedent seen below:<sup>9</sup>

() Wo wen Lao Wang<sub>i</sub> ta<sub>i</sub>/ mingtian lai bu lai  
 I ask Lao Wang<sub>i</sub> he<sub>i</sub> tomorrow come not come  
 I asked Lao Wang<sub>i</sub> whether he<sub>i</sub> would come tomorrow

The following observations hold true; for deletion of *ta*:

<sup>9</sup> Yeh's data do not seem to hold in all dialects. Yeh has other cases, all involving unstressed function words. Somewhat suprisingly, although she claims that phonological phrasing is the domain of this process these deletions/omissions are apparently not blocked by pause.

- ( ) Phonological identity is necessary but not sufficient  
 Referential identity is necessary but not sufficient  
 If vacuous application is invoked here, it is of a very unusual kind. Insertion of an item into a syntactic position must depend on the absence of an identical item from another syntactic position, in a different clause.

Golston (1995) proposes an Optimality Theory model of the interaction between syntax, morphology and phonology in which the syntax and/or morphology provide multiple possible outputs, and phonological constraints may decide among them. Syntactic constraints outrank phonological constraints, so the phonology may not force the selection of a syntactically deviant output, but if more than one syntactic output is possible, the phonological constraint may decide the winner. Applying this model to Mandarin *ta*, note that the syntax clearly provides two choices in the case of adjacent co-referential NP's as shown by the two options in (29): either both NP's may surface, or the second may be lost. Now consider the case when both NP's are *ta*; the syntax will produce both *...ta ta...* and *ta t<sub>i</sub>...*. The OCP will then decide matters, picking the single *ta*. It is the referential identity that allows the syntax to produce two candidates in the first place, but it is the phonology that narrows down the options in the case of homophonous *ta* sequences.

### 3.4 Classical Greek Articles <sup>10</sup>

Golston (1994,5) shows that Classical Greek texts never use sequences of two homophonous articles. Such sequences can arise with center-embedded genitive NP's. The genitive articles are shown below:

- ( ) a. Fem Gen Sg. tées  
 b. Masc Gen Sg tóu  
     Neut Gen Sg  
 c. Fem Gen Pl tóon  
     Masc Gen Pl  
     Neut Gen Pl

A typical center-embedded structure in which two articles (non-homophonous) end up adjacent is shown below:

- ( ) [tées [tóon himatí-on] ergasí-as]  
     the<sub>f.g.s</sub> the<sub>n.g.p.</sub> clothing<sub>n.g.p.</sub> production<sub>f.d.</sub>  
     'of the production of clothing'

<sup>10</sup> This section has been added since the conference; it appeared in the appendix to my handout, but not in the draft paper.

The syntax also allows for extraposition of the embedded NP, like the next example:

- () [[h-ee tólm-a] [t-óon leg-ónt-oon]]  
     the<sub>f,n</sub> courage<sub>f,n</sub>. the<sub>m,g,p</sub> speak-ing<sub>m,g,p</sub>.  
     'the courage of those speaking'

If the two articles are homophonous, the equivalent center-embedded phrase is unacceptable, but an extraposed version is fine:

- () a. [[téés arkh-ées] [téés pól-eoos]]  
     the<sub>f,g,s</sub> dominion<sub>f,g</sub>. the<sub>f,g,s</sub> dominion<sub>f,g</sub>.  
     'of the dominion of the city'  
     b. \*[[téés [téés pól-eoos] arkh-ées]]

The prohibition is morpheme specific, since phonologically identical sequences involving a deictic, not an article, are fine:

- () toútou tóu érgou  
     this<sub>n,g</sub>. the<sub>n,g</sub>. job

My analysis draws again on Golston's proposals. The syntax makes available both center-embedded and extraposed versions, and the phonology forces the choice of the extraposed version exactly when the center-embedded candidate would violate the OCP.

The four cases in this section all involve the avoidance of sequences of adjacent phonologically identical morphemes. The situation is resolved in one of two ways. First, we see omission of one morpheme, where the surviving morpheme does the job of both (*'s*, *le*). This is attributed to the OCP outranking MORPHDIS, and thus allowing haplology. Second, we see the choice of an alternative syntactic output which either lacks the offending morpheme (*ta*) or moves it away (*Greek articles*). Here the OCP chooses between two options made freely available by the syntax.

#### 4. Avoidance of non-adjacent homophonous morphemes

##### 4.1 English *ing*

We move now into a different domain, where the banned sequences involve identical morphemes attached to adjacent words, but where the morphemes themselves are not adjacent. For my first case I turn to English. Sequences of words ending in *-ing* are disliked in English, as pointed out by Ross (1972). For a different view, see Pullum and Zwicky (1991). A recent treatment of this well-known phenomenon is found in Milsark (1988), who proposes the following:

- () The Doubl-*ing* Filter  
     At PF, mark as ill-formed any sentence containing contiguous *-ing*-affixed words.

This accounts for the data below:

- ( ) a. \* John was starting reading the book
- b. \* John was keeping reading the book

Sequences of *-ing* that do not involve the same morpheme are fine:

- (        Good choirs sing ringing choruses all day long.

Milsark argues that "*-ing* is a single lexical element free in category", and given this the constraint bans morphological identity. Interestingly, it is apparently blocked by empty elements with Case features, as pointed out by Jaeggli 1980. This explains why Gen subjects block the filter from taking effect, and (30b) is fine.

- ( ) a. John enjoyed Bill's reading the book.
- b. John was enjoying [PRO]<sub>GEN</sub> reading the book.

Summarizing our observations:

- ( ) Morpheme identity, as well as phonological identity, is involved.
  - Certainly needs morphological information (to allow for the acceptability of (39), even if we do not accept Milsark's view that all *-ing* are a single morpheme with multiple interpretations.
  - Needs syntactic information, because empty categories can block.
  - Adjacency defined at word-level, not string adjacent.
  - Avoidance mechanism not clear, but it is not omission/deletion of one element, as shown by the ungrammaticality of (42).<sup>11</sup>
- ( ) \*\*John was starting reading the book
- \*\*John was starting reading the book

Notice that the reason that omission is not an option for this type of case is obvious. There is no sense in which the presence of *-ing* on one word renders unnecessary the addition of *-ing* to the next word.

Within Optimality Theory, the *-ing* affix presents a new problem. It is fairly clear that we need a constraint OCP(ing), and presumably one requiring that PROG=*ing*. What is new here is that resolution is not possible: neither omission of *-ing* or movement of the offending word are possible strategies. Instead, speakers avoid such utterances altogether. It seems then that there is no optimal output at all, and yet in Optimality Theory there is always an optimal output, no matter how bad. P&S handle this situation by assuming that one candidate is the so-called Null Parse, in which the input morphemes are not parsed at all. If this output is picked as optimal by the grammar, then there is no phonetic output. This situation arises in languages if Parse is outranked by other powerful constraints that rule out all candidates with parsed material.

<sup>11</sup> Junko Itô (p.c.) points out that *John was starting to read the book* could be the surface realization of *John was starting reading the book*. I will not pursue this idea here.

With this as background, I offer the following analysis of English *-ing*. A similar approach could be used for other cases of blocking, such as Chichewa. Suppose that OCP(*ing*) and a requirement that the progressive be marked with *ing* are both equally high ranked.

( ) PROG=*ing*: The progressive must surface marked by *ing*

OCP(*ing*): OCP (affix), affix = *ing*

REALIZE-Verb: Verbs must not be deleted (arguably a syntactic constraint).

Then we not only rule out sequences of V-*ing*, by the OCP, but also any attempts to improve things by omitting *ing*, because then we will violate the PROG= *ing* constraint. The best choice will thus be the fourth candidate, the Null Parse; angled brackets indicate unrealized material:

( )

	PROG= <i>ing</i>	OCP ( <i>ing</i> )	REALIZE-Verb
a. <i>Ving Ving</i>		*!	
b. <i>Ving V</i>	*!		
c. <i>V V</i>	**!		
d. <V> <V>			**

We must also assume that if neither verb is parsed, the sentence itself does not surface. The impossibility of improving things by movement presumably results from the dominance of whatever syntactic constraints block all possible movements of V-*ing*.

#### 4.2 Hindi Case Endings:

My next example comes from Hindi, and is another instance of homophonous affixes on adjacent words being unacceptable. The facts are quite complex, and apply to sequences of *-ko* (Dat, Acc) or *-se* (Inst), and some others. I will give data for *-ko* only. (T. Mohanan 1992) : <sup>12</sup> The dative and accusative suffixes are both *-ko*, and sequences of *N-ko N-ko* are considered strange.

( ) ?? raam-ko baccō-ko samhaalna paaa

Ramm-D children-A take care-NF fall-PERF

Ram had to take care of the children

If they are separated by another word, the sentence is acceptable, as shown below. In fact, extraposition of the accusative noun *bacco-ko* in (35) yields an acceptable sentence, and is one strategy for producing a well-formed version.

( ) raam-ko kal baccō-ko samhaalna paaa

Ramm-D yesterday children-A take care-NF fall-PERF

<sup>12</sup> I am indebted to Utpal Lahiri for help with these data.

Ram had to take care of the children yesterday

Another remedial strategy is to leave the accusative unmarked; this leaves it non-specific, but if the context allows this may not matter.

The prohibition on *N-ko N-ko* holds only if both NP's are arguments. Homophonous non-argument case-endings such as *-ko* (at) may co-occur with argument ones:

- ( ) raam-ko raat-ko ravii milaa  
 Ram-D night-at Ravi-Nmeet-PERF  
 Ram met Ravi at night

It is also clear that the relevant identity here is not abstract case identity, but requires that the cases be realized by phonologically identical elements. After pronouns, the Dative has two options for some speakers, *-ko* or *-e*. The *-e* Dative can freely co-occur with a *-ko* Accusative, as the contrast below shows:

- ( ) a. ?? ham-ko bacce-ko samhaalna paaa  
 we-D child-A take case-NF fall-PERF  
 We had to take care of the child
- b. ham-e bacce-ko samhaalna paaa  
 we-D child-A take case-NF fall-PERF  
 We had to take care of the child

Mohanan also points out that the prohibition only holds if the two morphemes are in adjacent phonological words - simply marking the heads of adjacent NP's is not enough to cause a problem.

- ( ) raam-ko apnii bahin-ke baccō-ko samhaalna paaa  
 Ramm-D self's sister-G children-A take care-NF fall-PERF  
 Ram had to take care of his sister's children

Furthermore, they must be in the same phonological phrase: the effect is blocked by the re-phrasing that accompanies an intervening pause.

- ( ) raam-ko [PAUSE] baccō-ko samhaalna paaa  
 Ramm-D children-A take care-NF fall-PERF  
 Ram [pause] had to take care of the children

These facts converge on the conclusion that:

Identity calculation is phonological, not morphological

It appears to need syntactic and morphological information as well.

Resolution is by omission of *-ko* or movement

An Optimality Theory account must deal with the fact that there are three resolution strategies. The first is omission of the accusative suffix. To derive this we need

- () DAT (*ko*): The dative must be marked by *-ko*, or *-e* for pronouns.
- ACC (*ko*): The accusative must be marked by *-ko*.
- OCP(*ko*): OCP(affix): Affix = *ko*<sub>Acc/Dat</sub>

Crucially, the constraint covering the accusative *-ko* that may be omitted is dominated by the OCP: DAT (*ko*), OCP(*ko*) >> ACC-*ko*

()

	DAT ( <i>ko</i> )	OCP ( <i>ko</i> )	ACC ( <i>ko</i> )
N <sub>D</sub> - <i>ko</i> N <sub>A</sub> - <i>ko</i>		*!	
N <sub>D</sub> - <i>ko</i> N <sub>A</sub>			*
N <sub>D</sub> N <sub>A</sub> - <i>ko</i>	*!		
N <sub>D</sub> N <sub>A</sub>	*!		

The second strategy is limited to pronouns, and chooses the alternative suffix *e*. This is achieved by adding a candidate with *e* to the set, Pronoun<sub>D</sub>-*e* N<sub>A</sub>-*ko*. This candidate satisfies all 4 constraints, and is thus optimal.

The third strategy is movement; usually only fairly heavy NP's are extraposed, but a light *-ko* NP may be extraposed to avoid an OCP(*ko*) violation. So we need:

- () OCP (*ko*) >> \*Extrapose Light NP

Obviously a serious account of this latter strategy depends on our understanding of the nature of the syntactic/prosodic constraints limiting extraposition to heavy phrases. The most appealing accounts, such as Zec and Inkelas (1990) and Truckenbrodt (1994), suggest that extraposed phrases are intonational phrases (IP's), and that IP's are subject to a minimum size requirement, Min IP. In that case (48) can be re-stated as (49):

- () OCP (*ko*) >> Min IP

Whether this account holds up to a more detailed investigation of Hindi extraposition and intonation must be left for further research.

The two cases in this section have looked at more distant adjacency effects. (See also Sadock (1972) on Danish *som*.) Not surprisingly, these are less common than effects on adjacent identical sequences, and also seem to be less strong. The offending sequences are often deviant rather than absolutely impossible. In phonology, Pierrehumbert 1993a points out the same

gradient weakening effect of distance on the OCP, and links it to more general cognitive phenomena.

### 5. Echo-word Formation:

The third type of identity avoidance to be discussed here involves reduplication. Since reduplication apparently strives to achieve identity, it is intriguing to find it apparently shunning total identity: instead, the echo-word outputs I will examine show *almost* perfect identity, but with a single flaw.

Many languages have reduplicative processes that replace one portion of the reduplicant with fixed segmental material. English *table-shmable* is an example of such a process: see Yip (1992) for a range of cases. The segmental material is sometimes arguably the default segment of the language, as argued by McCarthy and Prince for Akan, and Yip for Chaoyang (1993). In other cases, however this is not so: no-one has argued that / , m / are the default consonants of English. A striking characteristic of many such word formation processes is that if the input contains segments identical to the fixed replacement ones, so that the expected output would mimic total reduplication, the process either does not apply at all, or a different set of replacement segments is used. For example, the Tengxian dialect of Chinese (Deng 1995) reduplicates adjectives, replacing the rhyme of the first half by [ ]:

- ( )     dun                 d   dun 'short  
           l                 l l             'cold'  
           kou                k   kou 'tall'

This system is very productive, applying to more than 200 adjectives. Systematically, adjectives whose rhyme is [ ] or [a ] fail to undergo this process, instead using one of several alternatives available in the language: /n / does not yield \*n n , but rather [n h t ].

A second example is drawn from Turkish, which reduplicates the first CV of the adjective to form an emphatic form. This CV addition is followed by a coda consonant from the set /p,s,m,r/, subject to the constraint that this consonant cannot be identical or too similar to any consonant of the base. For details, see Dobrovolsky (1987), Demircan (1987).<sup>13</sup>

- ( ) a. kap-kara 'jet black'                                 ap-aik 'wide open'  
           cep-cevre 'very much around'                 sap-sari 'fully yellow'  
   b. sim-siki 'extremely tight'                         bem-beyaz 'snow white'  
           göm-gök 'sky-blue'                             bum-burusuk  
   c. kas-kati 'extremely hard'                         bes-belli 'unmistakably obvious'  
   d. ter-temiz 'spotless'                                sir-siklam 'wet through'  
           tor-top 'fully round'

The precise choice of consonant depends on a number of factors, and there is some degree of freedom, but the avoidance of repetition is a major consideration. Closer consonants, and coda consonants, exert more influence than do more distant ones, in line with the view of identity avoidance put forward in Pierrehumbert (1993a).

This echo-word type of reduplication accompanied by melody replacement shows a clear

<sup>13</sup> Thanks to Orhan Orgun for help with this section.

tension between a desire for repetition, which can be seen as the need to satisfy a constraint REPEAT, and avoidance of repetition, or a satisfaction of the OCP. The OCP is higher ranked, ruling out total reduplication, but REPEAT plays a central role in ensuring that the overall system is still one of reduplication, with only a minimal difference between base and reduplicant. In the next section I will discuss one complex case of this type, Javanese, in more detail, showing how this tension is played out. Javanese avoids repetition of two kinds: repetition of the entire stem, and repetition of the vowel [a]. It also has two output constraints, REPEAT, and a requirement that the second syllable have the vowel [a]. These interact in interesting ways, as we see in the next section.

### 5.1 Javanese:

Javanese has a pattern of reduplication that is usually referred to as Habitual Repetitive, shortened to Hab-rep. It applies to verbs, adjectives, and even nouns. The whole stem is reduplicated, and then the vowel in the last syllable of the first half is replaced by [a]. Most roots are bi-syllabic, so usually the second syllable has the vowel [a]. However, if the stem is longer or shorter it becomes clear that the locus of [a] is consistently the final syllable of the first half. Some typical data is given below; all examples are given in phonemic transcription unless allophonic details become relevant.<sup>14</sup>

( ) **Normal pattern of Habitual-Repetitive (Hab-reps) Reduplication:**

eli	ela -eli	'remember'
tuku	tuka-tuku	'buy'
ele	ela -ele	'bad'
bul	bal-bul	'puff'

In most cases, it is not possible to tell whether reduplication here is prefixing or suffixing in nature, a point made independently by McCarthy and Prince (1995). This suggests that it is not a type of affixation at all, but rather compounding of the stem with itself, with both halves of equal status. The reduplication is accomplished in response to the constraint REPEAT(Stem), which rules out any output without a reduplicated stem. See Yip (forthcoming) for arguments that reduplication is output-based. In addition to the reduplication itself, Hab-reps consistently use /a/ in the second syllable. This introduced [a] would traditionally be analyzed as an affix that forms part of the Hab-rep morphology. I will argue that its appearance is instead the result of an output constraint requiring the vowel of the appropriate syllable to be [a]. I formulate this constraint below:

- ( ) **2=a:** The final syllable of the first half of Hab-reps must have an [a] nucleus

With these basics in hand, we can return to the OCP. The interest of the Javanese Hab-reps lies in their diverse mechanisms for avoiding identity of various kinds. First, the output may never have both halves completely identical to each other. The constraint in (58) achieves this immediately if the input ends in any vowel other than /a/, but what if it ends in /a/? The data are given below:

( )	udan	udan-uden	'rain'	*udan-udan
	kumat	kumat-kumet	'have a relapse'	*kumat-kumat
	edan	edan-eden	'crazy'	*edan-edan
	tak	tak-tek	'tap'	*tak-tak

<sup>14</sup> Javanese has six vowel phonemes, /i,u,e,o,a,/. For a full treatment of the phonology of Javanese vowels see Dudas (1968), Yallup (1982). The data here is drawn from Dudas (1968), Kenstowicz (1986), and Horne (1964). For more detailed discussion of Hab-rep reduplication, see Yip (forthcoming).

Simple satisfaction of  $2=a$  would result in perfect total reduplication. Instead, the vowel of the second half dissimilates to [e]. The following constraint embodies the avoidance of total identity typical of Hab-reps; I should emphasize that other forms of reduplication in the language do allow complete reduplication, such as *abat-abat* 'century, PL'.

() OCP (Stem): Hab-reps must not consist of two identical stems.

These two constraints are both surface true and undominated. Between them they select *udan-uden* as the optimal candidate as shown below:

()

/udan/	$2=a$	OCP(stem)
a. udan-uden		
b. udan-udan		*!
c. uden-udan	*!	

The simplest argument in favor of treating [a] as the response to an output constraint, rather than as an affix, is based on the fact that identity violations can never be resolved by changing this introduced [a]. The dissimilation site is always the *other* /a/. If /a/ were an affix, it would be necessary to somehow stipulate the choice of target, but the output-based analysis immediately explains the immunity of the introduced /a/ to change.

A different kind of identity avoidance is found if the input has /a/ in the first syllable. From what we have seen so far, we would expect to find outputs in which the first half has /a/ in both syllables, so that /lali/ would have a Hab-rep *lala-lali*, but instead we observe dissimilation of the root /a/ to [o]:

()

lali	lola-lali	'forget'	*lala-lali
adus	odas-adus	'bathe'	*adas-adus
melaku	meloka-melaku	'walk'	*melaka-melaku

Following the same analytical approach used above, I formulate the following constraint:

() \*OCP (a): Sequences of /a/ are not allowed.

Note that the constraint does not apply across stem boundaries, since *lola-lali* is well-formed. The domain of this constraint is the stem, with the introduced [a] analyzed as part of the stem, confirming our earlier claim that it is not an affix. Note also that Kenstowicz (1986) has shown that Javanese roots with two identical vowels must have a single vowel melody occupying two nuclear slots. The argument rests on the fact that allophonic rules show their effects on both root vowels, even if only one of the vowels is in the proper context. Kenstowicz (1986) argues that

this behavior is to be expected if these roots have one melody linked to both nuclei. This representation in turn follows if Javanese obeys the OCP as an MSC on the vowel tier. The OCP(a) constraint can now be seen as a specific instance of a more general constraint OCP(seg) found throughout Javanese vowel phonology.

Following precisely parallel arguments to those we used for a>e, we may understand why it is the root /a/, not the introduced [a], that changes to [o]. The following tableau demonstrates this point: candidate (b), in which the introduced /a/ has changed to [o], violates  $\zeta=a$ , and is thus eliminated.

()

/lali/	$\zeta=a$	OCP(a)
a. lola-lali		
b. lalo-lali	*!	
c. lala-lali		*!

The ranking of OCP(a) >> FAITHFULNESS is validated by the following tableau:

()

/lali-lali/	OCP(a)	FAITHFULNESS
a. lola-lali		*
b. lala-lali	*!	

We have now seen two types of dissimilation that conspire to remove identity violations. I leave for further research the question of why one chooses [e] while the other chooses [o] as their output vowel.

In the previous section I proposed that echo-word reduplication is a response to a constraint REPEAT. This means that there is a tension between a requirement that *penalizes* a sequence of two identical stems, OCP(Stem), and one that *requires* two identical stems, REPEAT(Stem). The two constraints are given here for comparison.

- () REPEAT (Stem) : Hab-Reps must consist of two identical stems.
- () OCP (Stem): Hab-reps must not consist of two identical stems.

It is the former that produces the effect of total reduplication, but this is then minimally destroyed by the latter. The constraint REPEAT(Stem) plays a crucial role in ensuring that the dissimilations we observe are indeed minimal: a single vowel changes, just enough to satisfy OCP(Stem).<sup>15</sup> The following tableau demonstrates how this works; candidate (b), with two

<sup>15</sup> One unexplained issue is why in the second half only the vowel /a/ ever dissimilates. /udan/ could surface as *udan-idan* and satisfy OCP(Stem), and yet such changes are never found.

vowel changes, has two violations of REPEAT(Stem), and loses to candidate (a), which has only one violation.

()

	OCP(Stem)	REPEAT(Stem)
a. udan-uden		*
b. udan-iden		***!
c. udan-udan	*!	

For a discussion of how to handle language-specific and morpheme-specific constraints in OT, see Yip (forthcoming).

I have argued that two aspects of Hab-Reps are best handled by output constraints. Firstly, the introduced [a] is present in response to the constraint  $\text{2=a}$ . Secondly, the reduplication itself, and the dissimilations that accompany it, are the result of a set of contradictory output constraints OCP(Stem), OCP(a), vs. REPEAT(Stem). The Javanese data argue for two conclusions. First, at least some morphology must be handled by output constraints, and the inputs (if any: see Russell 1995) have morphological features, such as the information that something is a Hab-Rep, but not necessarily concrete morphemes, like /a/, or even abstract ones, like RED. Second, conditions that enforce identity, and avoid identity, can make reference to phonological objects, like the vowel /a/, or morphological entities, like Stem.

## 6. Conclusions

I have argued that avoidance of identity is found in many areas of morphology, not just in phonology, and that these phenomena should be given a unified treatment in terms of a family of OCP constraints. This constraint would appear to be part of a more general cognitive pattern, as argued by Pierrehumbert (1993a), suggesting that the nomenclature OCP is inappropriate, and should be replaced by something more general like \*REPEAT. See Yip (forthcoming) for such an approach, prompted in part by the comments of Brentari and others at this conference.

The theme of this conference is the relationship of morphology to other components of the grammar, and I will end by tackling this issue with respect to phonology and morphology. The most important generalization to emerge from the facts examined here is almost trivially obvious:

() The identity avoidance is fundamentally phonological in nature

That is, the starting point for all the cases discussed here is an abhorrence of phonological identity of some kind. This may not be a sufficient reason for avoiding identity, but it is always a necessary one. I conclude that we are dealing here with a phonological phenomenon, but it may or may not *also* be a morphological phenomenon. In other words, we must ask if we can distinguish clearly whether a sharp line can be drawn between the two components as far as their treatment of identity avoidance is concerned?

There are, as we have seen, many kinds of non-phonological information needed for the identity calculation. If we try to draw a firm boundary between phonology and morphology, we do not necessarily deny the possibility that morphological information can be used by phonology, or that morphology can use information from phonology. Such a division does however have two implications:

- ( ) a. Information transfer between the two components should be constrained in certain ways
- b. Each component should have its own grammar, principles of well-formedness, etc.

If neither of these is true, then calling either phonology or morphology "components" is meaningless, merely a descriptive convenience. If only one is true, this is, in my view, sufficient to consider the two components distinct.

The first criterion, restricted information flow, does not block information transfer, but requires that it be somehow limited to certain channels or directions. Morphological information is well-known to be available to phonology in multitudinous ways. Cyclic phenomena, lexically restricted rules, and the whole edifice of lexical phonology, make this point quite clear. It is also true that phonological information plays a role in morphology, as shown by the work of Carstairs, and the recent work of Mester and Tranel discussed above. Information flow, then, must be two-way. It might still be possible to constrain the *type* of information flow, but it is not clear to me at present that this is the case. For example, the phonological information needed by morphology must include syllabic information (French, Turkana *-isi/-u*), metrical information (Latin, Italian *fin-/finisc-*), and segmental information (Hungarian *-ol/- (a)sz*; English *-ion, -ition, -ion*). The type of morphological information needed by phonology must include structural information (root, prefix, suffix, headedness of compounds (Malayalam: Mohanan 1982)), construction-specific information (Habitual Repetitives in Javanese), vocabulary-type (Germanic vs. Graeco-Roman in English, Yamato vs. Sino-Japanese, etc), and categorical information (English extrametricality in noun stress vs. verb stress).

The second criterion, separate grammars, is also unclear. In Optimality Theory, Alignment theory supposes a single set of constraints whose arguments can be morphological or phonological (especially prosodic) entities. And in the phenomena discussed in this paper the OCP appears to take as its arguments phonological, syllabic, morphological or word-sized units. At the very least, then, there is an overlap between the two components in the form of a shared set of rules, principles or constraints.

It looks, then, as though neither of the criteria in (61) can be satisfied, and we may have to conclude that phonology and morphology blur into each other too much for us to consider them as discrete formal components. This does not of course change the fact that both sets of phenomena form worthy objects of study, nor that the terms remain useful as rough labels for domains of study. Rather it reinforces what linguists have always known: phonologists and morphologists have a lot to learn from each other's domains, and the boundary line between them is not sharply defined.

## References

- Anderson, S. 1992 *A-morphous Morphology*. Cambridge University Press.
- Carstairs, A. (1990) Phonologically Conditioned Suppletion. In W. Dressler, H. Luschützky, O. Pfeiffer, and J. Rennison, eds., *Contemporary Morphology*. Mouton de Gruyter, Berlin, New York. pp 35-40.
- Chao, Y-R (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Cohn, A. and J. McCarthy 1994 Alignment and Parallelism in Indonesian Phonology. Ms, Cornell and UMass.
- Demircan, Ömer 1987 Emphatic reduplications in Turkish. H. Boeschoten and L. Verhoeven, ed., *Studies in Modern Turkish: Proceedings of the 3rd Conference on Turkish*. Tilburg University Press. pp 24-41
- Deng, Yurong (1995) The reduplication of monosyllabic adjectives in Tengxian dialect. (In Chinese). *Fangyan* pp33-46.
- Dobrovolsky, M. 1987 Why CVC in Turkish Reduplication. In P. Liliw and M. Saari, eds., *The Nordic Languages and Modern Linguistics 6* Helsinki: 131-146.
- Dudas, K. M. 1968 The Phonology and Morphology of Modern Javanese Ph D Dissertation, U. of Illinois.
- Golston, C. 1994 The Syntax-Phonology Interface in Optimality Theory. Paper presented at WCCFL XIII, Berkeley
- Golston, C. 1995 The Syntax-Phonology Interface in Optimality Theory. Ms, Universität Düsseldorf.
- Hammond, M. 1995 There is no lexicon. Ms, U of Arizona, Tucson.
- Hayes, B. 1993 Metrics as an optimization problem: the case of English folk songs. Handout of talk presented at ROW-1, Rutgers, NJ.
- Horne, E. 1964 *Javanese-English Dictionary*. Yale University Press.
- Hyman, L. and S. Mchombo 1992 Morphotactic constraints in the Chichewa verb stem. In L. Buszard-Welcher, L. Wee and W. Weigel, eds., *Proceedings of the 18th Annual Meeting of the Berkeley Linguistics Society* : 350-364 Berkeley Linguistic Society, Berkeley.
- Kenstowicz, Michael 1986 Multiple Linking in Javanese. *NELS 16*
- Kiparsky, P. 1993 Variable rules. Handout of talk presented at ROW-1, Rutgers, NJ.
- Leben, Will (1973). *Suprasegmental phonology*. Doctoral dissertation, MIT.
- Li, Sandra and C. Thompson 1981 *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: U. of California Press.
- McCarthy, John (1986b). OCP Effects: Gemination and antigemination. *LI 17*: 207\_263.
- McCarthy, John & Alan Prince. 1993a. Prosodic Morphology I: Constraint Interaction and Satisfaction. To appear, MIT Press. Technical Report #3, Rutgers University Center for Cognitive Science. Pp. 184.
- McCarthy, John & Alan Prince. 1993b. Generalized Alignment. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1993*. Dordrecht: Kluwer. 79-153. Technical Report #7, Rutgers University Center for Cognitive Science, 1993. ROA-7.
- McCarthy, John & Alan Prince. 1994. The Emergence of the Unmarked: Optimality in

- Prosodic Morphology. In Merce` Gonza`lez (ed.), Proceedings of the North-East Linguistics Society 24. Amherst, MA: Graduate Linguistic Student Association. 333-379. ROA-13.
- McCarthy, J. and A. Prince 1995 Faithfulness and Reduplicative Identity. Ms., U Mass Amherst and Rutgers University.
- Menn, L. and B. MacWhinney 1984 The Repeated morph constraint. *Language*: 60: 519-541
- Mester, R-A 1994 The Quantitative Trochee in Latin. *Natural Language and Linguistic Theory*.12.1: 1-62
- Milsark, G. 1988 Singl-ing. *LI* 19.4: 611-634
- Mohanan, Tara 1992 Case OCP: A Constraint on Word Order in Hindi. Ms, National University of Singapore.
- Myers, Scott. 1993. OCP Effects in Optimality Theory. Ms., University of Texas, Austin. Pp. 31. ROA-6.
- Odden, David (1988). AntiAntigemination and the OCP. *LI* 19: 451\_75.
- Pater, J. 1994 Optionality and optimality: syllable weight effects in English secondary stress. Ms. McGill.
- Pierrehumbert, Janet 1993a Dissimilarity in the Arabic Verbal Roots. In Amy Schafer, ed., *Proceedings of NELS 23* Amherst, MA: Graduate Linguistics Student Association. pp 367-381
- Pierrehumbert, Janet 1993b Alignment and Prosodic Heads. ESCOL 1993.
- Perlmutter, D. 1995 A grammar can explain allomorphy. This volume.
- Prince, Alan and Paul Smolensky. 1993. Optimality theory: constraint interaction in generative grammar. To appear, MIT Press. TR-2, Rutgers University Cognitive Science Center. Pp. 234.
- Pullum. G. and A. Zwicky 1991 A misconceived approach to morphology. *Proceedings of WCCFL 10*.
- Ross, J. R. 1972 Doubl-ing . *LI* 3: 61-86
- Russell, K. 1995 Morphemes and Candidates in Optimality Theory. Ms, U. of Manitoba.
- Sadock, J. 1972 A conspiracy in Danish relative clause formation. In P. Peranteau, J. Levi, G. Phares eds., *The Chicago Which Hunt: Papers from the Relative Clause Festival*. A Paravolume to Papers from the Eighth Regional Meeting of the Chicago Linguistic Society: 59-62.
- Stemberger, J. 1981 Morphological Haplology. *Language*. 57.4:791-817.
- Tranel, Bernard 1994 French Liaison and Elision Revisited: a unified account with Optimality Theory. Ms, UC Irvine.
- Yallup, C. 1982 Phonology of Javanese Vowels. In A. Halim, L. Carrington, and S. Wurm, eds., *Papers from the Third International Conference on Austronesian Linguistics*. Canberra, Australian National University.
- Yeh, Ling-hsia 1994 Haplology in Mandarin. Ms., UMass Amherst
- Yip, M. 1988 The Obligatory Contour Principle and phonological rules: A loss of identity. *Linguistic Inquiry* 19, 65-100.
- Yip, M. 1992. Reduplication with fixed melodic material. In K. Broderick, ed., *Proceedings of NELS 22* Amherst, Ma: Graduate Linguistics Association.

- Yip, M. (forthcoming). Repetition and its avoidance: the case of Javanese. To appear in K. Suzuki and D. Elzinga, eds., *Arizona Phonology Conference 4: Features in Optimality Theory*. U. of Arizona, Coyote Papers. Tucson, 1995.
- Zoll, Cheryl. 1994. Anchors away: a unified treatment of latent segments and floating features. Ms., University of California, Berkeley. Pp. 55.