Three Classes of "+" Boundaries

Kenneth W. Church
AT&T Bell Laboratories
Murray Hill, NJ, USA

It is well-known that English morphology has two classes of affixes: "+" morphemes such as *in+*, *ad+*, *ab+*, *+al*, *+ity* and "#" morphemes such as *un#*, *#ness*, *#ly*. The two classes differ in a number of respects, including: (1) Etymology: "+" morphemes are (often) historically correlated with Latin; "#" with German and Greek, (2) Stress Assignment (e.g., *parént+al* vs. *párent#hood*), and (3) Word Formation: + morphemes can attach to bound morphemes (e.g., *crimin-* as in *criminal*); # cannot (**criminhood*). This paper will extend this reasoning in dividing the first class into three parts, Ia, Ib and Ic (see table).

Class Ib contains what we generally think of as "typical" + boundary forms (e.g., *parént+al*, *divin+ity*), both with respect to stress assignment and word formation. It will be argued here that Class Ia obeys a different set of word formation rules and that Class Ic obeys a different set of stress assignment rules.

The notion of compositionality provides a unifying theme across classes. Just as it is often observed that "#" forms have compositional semantics and stress assignment (e.g., *divine#ness* means "the state of" composed with "divine"; the stress of the whole is the concatenation of the stress of the parts) unlike "+" forms (e.g., *divin+ity* has religious implications that cannot be attributed to its parts; the stress of the whole is not the concatenation of the parts because of stress retraction), we would want to say that Class Ia is less compositional than Ib which is less than Ic which is less than II.

1. Word Formation Rules (WFR)

Aronoff proposed two distinct types of word formation rules in his thesis [Aronoff]: stem based wfr and word based wfr.

- Stem Based WFR: *subsume/subsumption, consume/consumption, resume/resumption, expense/expensive, conduce/conductive*
- Word Based WFR: *nominate/nominee, nominate/nominal, feminine/feminism.*

Stem based wfr rules relate pairs of words sharing one of a short (100-1000) list of latinate stems, e.g., *fer, mit, sume, duce, scribe*, whereas word based wfr apply to a large (possibly open) class of forms, often ending with *-ate* or some other archaic affixes such as: *-ine, -uli, -us, -um* that may be stripped off or "truncated" as part of the word formation process. Aronoff distinguished the two types of word formation rules in order to account for the fact that some generalizations, especially productivity and allomorphy, are clearly associated with stems, whereas other generalizations are associated with words.

This paper will use Aronoff's distinction in order to separate Class Ia from other "+" boundary forms. First, though, it may be worthwhile to review Aronoff's reasons for hypothesizing two types of word formation rules.

1.1 Productivity

The contrast in productivity between stem based and word based wfr is very striking. Note that there are very few gaps in stem paradigms:

	0	0 (pp)	-ion	-ive
duce	adduce	adduct	adduction	
	deduce	deduct	deduction	deductive
	conduce	conduct	conduction	conductive
	educe	educt	eduction	eductive
	induce	induct	induction	inductive
	introduce		introduction	
	produce	product	production	productive
	reduce	reduct	reduction	
	seduce		seduction	seductive
	transduce		transduction	
scribe		conscript	conscription	
	describe	nondescript	description	descriptive
	prescribe	prescript	prescription	prescriptive
	subscribe	subscript	subscription	subscriptive
ceive	conceive	concept	conception	conceptive
	deceive		deception	deceptive
	perceive	percept	perception	perceptive
	receive	recept	reception	receptive
here	adhere		adhesion	adhesive
	cohere		cohesion	cohesive
	inhere		inhesion	inhesive

	+ Boundary			# Boundary
	Class Ia	Class Ib	Class Ic	Class II
Examples	ion, ive, ent, or, ory	ity, ic al, ian	ize, ee, itis,ism, ist istic, ment, mental	ness, wise hood, ship
Etymology	Productive in Latin	Norman French	Scientific Literature and Enlightenment	Anglo-Saxon
Stress Retraction	+	+	−	−
Attaches to	stems	bound/free	bound/free	free

In contrast, word based alternations are full of gaps. For example, the word based *-ate/-ee* alternation (e.g., *nominate/nominee, designate/designee*) is limited to just a few cases; the vast majority of words ending with *-ate* do not have variants ending with *-ee*.

1.2 Allomorphy

Stem based word formation rules attempt to capture both productivity and allomorphy generalizations. In stem based forms, allomorphy (e.g., *scribe* vs. *script*) is purely a function of the stem and the suffix, and is independent of derivational history (cyclicity), prefix, part of speech, semantics, phonology, etymology, dialectical variation, etc. In contrast, allomorphy may have more complicated sources in word based forms. Consider, for example, the word *education* which does not follow the stem based pattern found in *adduction, deduction, conduction, eduction, induction, introduction, production, reduction, seduction* and *transduction*, because *education* is derived from the word *educate*, not from the stem *duce*. This example illustrates that derivational history can play an important role in explaining allomorphy, but only in word based derivations, and not in stem based derivations.

Mark Aronoff noticed that stem based allomorphy depended only on the stem and the suffix and attributed this fact to (the mythical) Ben Moshe.

> "The form of the suffix is never determined by a specific word. It is never the case that one verb in a given root will allow one variant, and other verb in the same root a different variant. The form of the suffix is root governed, that is, morphologically governed. There are no exceptions to this. It is the first law of the root originally discovered by the great Semitic grammarian ben-Moshe (ms) [sic] and called Ben-Moshe's First Law.
>
> We will illustrate ben-Moshe's first law in (28) with the root *sume*. The variant of *ion* which appears after *sume* is *+tion*:" [Aronoff, p. 102]

(28)	subsume	subsumption	*subsumation
	consume	consumption	*consumation
	resume	resumption	*resumation
	presume	presumption	*presumation
	consume	consumption	*consumation [sic]
	assume	assumption	*assumation

Aronoff uses Ben Moshe's Law to cover both cases like *sume/sumption* above where the allomorphy alternation is extremely clear as well as cases like *vert/version* and *sert/sertion* where the allomorphy is somewhat more subtle. Note that the orthographic "t" in *invertion* is realized as /zh/ whereas the corresponding "s" in *insertion* is realized as /sh/. Aronoff attributes this distinction to the allophorphy of the stems *-vert* and *-sert*, and then observed that Ben Moshe's Law correctly predicts that this voicing contrast is maintained in related forms such as *diversion, conversion, perversion* which contain /zh/ as in *inversion*, and *desertion, exsertion, assertion* which contain /sh/ as in *insertion*.

Ben Moshe's Law can also be used to cover quantity changing allomorphy as in *confide/confidence*. The "Confidence Puzzle" is intriguing because *-fide* is heavy in *confide* (as evidenced by the long vowel) but light in *confidence* (as evidenced by the stress retraction before the weak retractor suffix *-ence*). Other stems also use allomorphy in order to change quantity (see table). Consider *-side* and *-pel*. Both change their underlying quantity before the suffix *-ent*. *-side* is underlyingly heavy, but acts light in *resident*, whereas *-pel* is underlyingly light, but acts heavy in *repellent*. Note that Ben Moshe's Law correctly predicts that the choice of allomorphy is independent of prefix. The same light *-side* found in *resident* also appears in *president* and *dissident*; the same heavy *-pel* found in *repellent* also appears in *expellent* and *propellent*.

	Acts Light	Acts Heavy
Tense	-fide, -side, -spire, -tain, -stain, -cide, -pare	-hale, -grade, -plain, -flame, -vade, -praise, -rade, -suade, -place, -claim, -rive, -vive, -dign, -mise, -scribe, -quire, -vise, -prise, -fice, -pugn, -clude, -prove, -sume, -lude, -trude, -fuse, -plode, -close, -mote, -pose, -void, -join, -plore
Lax	-fer, -cel	-pel, -mit, -gress, -press, -cess, -cuss

1.3 (Almost) No Exceptions to Ben Moshe's Law

Ben Moshe's Law, according to Aronoff, is exceptionless. After some computer assisted investigation, it appears that the rule is, in fact, nearly exceptionless, if not completely so.[1] Many apparent counter-examples can be dispensed with by attributing the counter-examples to word based wfr, as opposed to stem based wfr, as we did in order to account for *education* which is problematic since most combinations of *duct* and *-ion* yield *duction*, not *ducation*. Aronoff himself uses the word based escape hatch in order to dispense with *consummatation*, which would ordinarily be a problem for Ben Moshe's Law, since *sume* plus *-ion* normally produces *sumption*, not *summation*.

> "Note that the form *consummation*, as in Shakespeare, is not an exception. Rather it is derived from the base *consummate*, by truncation." [Aronoff, p. 102]

Compensative is very much like *consummation*; *compensative* is formed from *compensate* via truncation, as opposed to *expensive* which is stem based and obeys Ben-Moshe's Law. *Friction/frication* also demonstrates the contrast between stem based and word based wfr. *Preventive/preventative* and *interpretive/interpretative* illustrate another class of (apparent) counter-examples to the law. Again, these apparent counter-examples can be accounted for by showing that one of the forms has an alternative source. In this case, *preventative* is from the latin frequentative; the frequenative *-ative* should not be confused with *-ive*.

In general, forms obeying Ben-Moshe's Law show up with a large number of latinate prefixes, as opposed to form like *compensative, expectation, education* and *preventative*, which violate the Law. Thus, for example, *conducive*, another exception to Ben-Moshe's Law (cf., *conductive, deductive, inductive, productive*), is not found with very many other prefixes (e.g., **educive, *deducive, *producive*). Exceptions are unlikely to show up with very many prefixes because prefixes are only productive on stems and these exceptions are word based.

1.4 Class Ia and Stem Based WFR

This paper provides additional evidence in favor of Aronoff's two types of word formation rules by proposing that some affixes (namely, Class Ia affixes) are (generally) associated with stem base wfr and that other affixes (namely, Class Ib and Ic) are associated with word based wfr. Note that Class Ia affixes (e.g., *-ion, -ive, -ent, -or*) are often found after latinate stems (e.g., *permission, permissive, confident, conductor*) but not generally after truncated morphemes (e.g., **nominion, *nominive, *nominent, *nominor*). Similarly, Class Ib and Ic affixes (e.g., *-al, -ee*) are often found after truncated morphemes (e.g., *nominal, nominee*), but not generally after latinate stems **subsumal, *subsumptal, *subsumee, *subsumptee*.

- *The Distributional Claim*: Class Ia affixes (e.g., *-ion, -ive, -ent, -or*) attach to latinate stems (e.g., *fer, mit, sume, duce, scribe*) whereas Class Ib and Ic affixes (e.g., *-al, -ity, -ic, -ee, -ism, -ist*) attach to words (possibly via truncation).

One of the consequences of this claim is that *feral, feric, ferity, ferrous* and *ducal* cannot be related to the latinate stems *fer* and *duce* because Class Ib affixes such as *-al, -ic, -ity* and *-ous* do not attach to latinate stems. This observation may be important for practical computer applications of morphological analysis to unknown words, especially for speech synthesis.

In addition, this distributional claim forces a form of level ordering [Kiparsky], [Mohanan]. Note that Class Ia affixes affixes can be found inside Class Ib affixes (e.g., *festivity, conventional*) but not the other way around (e.g., **fest+ity+ive, *convent+al+ion*), because Class Ia affixes (e.g., *-ive, -ion*) must be attached to latinate stems and therefore, they cannot follow Class Ib affixes.

1.5 Multiple Class Membership

The distributional claim is somewhat weakened, unfortunately, by the fact that some affixes such as *-able* share membership in more than more class. Just as others (e.g., [Aronoff, section 6.2] have assumed that *-able* belongs to both "+" and "#", it will be assumed here that *-able* belongs to all three classes: Ia, Ib and Ic. The difficulty is that *-able* may or may not feed allomorphy, truncation and stress retraction:

- Allomorphy: (with) *circumscriptible, extensible, defensible, perceptible, divisible, derisible* (without) *circumscribable, extendable, defendable, perceivable, dividable, deridable*
- Truncation: (with) *educable, irrigable, navigable, regulable, demonstrable, operable, separable* (without) *educatable, irrigatable, navigatable, regulatable, demonstratable, operatable, separatable*
- Stress Retraction: (with) *cómparable, réparable, préferable*[3]

  (without) *compárable, repárable, preférable*

Aronoff assumed that forms which feed allomorphy, stress retraction and/or truncation contain a "+" boundary and that forms which block these processes contain a "#" boundary. The present proposal would assign *divisible* to Class Ia in order to account for the observed allomorphy, *demónstrable* and *coómparable* to Class Ib in order to account for the observed stress retraction, and *compárable* to class Ib in order to account for the observed lack of stress retraction.

2. Class Ic

The introduction suggested that Class Ib contains what we generally think of as "typical" + boundary forms (e.g., *parént+al, divin+ity*), both with respect to stress assignment and wfr. Section 1 argued that Class Ia obeys a different set of stem based wfr. This section will argue that Class Ic obeys a different set of stress assignment rules.

Within words, one expects to find stress clashes resolved by a rule which forces stressed syllables to alternate. Thus, for example, *degráde* plus *-ation* yields *dègradátion* with alternating stressed syllables, not *degràdátion* with the two adjacent clashing stresses. This prohibition against stress clashes applies to most "+" boundary forms (Classes Ia and Ib), but not to Class Ic. Note, for example, that *depàrtmént*al and *emplòyée* do not become **dèpartméntal* and **èmployée*, as would be predicted if these stress clashes had to be resolved.

Class Ic forms are also exceptions to most so-called "+" boundary rules. Note, for instance, the contrast between *concain+ism* and *profan+ity*. Tri-syllabic laxing, a typical "+" boundary rule, forces the tense vowel in *profane* to become lax in the Class Ib *profanity*, but tri-sylabic laxing does not apply in Class Ic and therefore the tense vowel in *concain* does not become lax in the Class Ic form *concainism*.

It will be assumed here that Class Ic forms are stressed much like compounds. *Assignee*, for example, is formed by combining the two pieces *assign* and *ee* with a right dominant foot [W S] so that the main stress falls on *ee*. Other Class Ic forms such as *cocainism* are combined with a left dominant foot so that the main stress falls on *cocain*.[4] In both cases, the internal metrical structure of the left piece is kept intact. Note that the

1. A dictionary search for orthographic sequences taking both *-ation* and *-ion* produced: *legation (legion), domination (dominion), oration (orion), duration, conversation, cessation, dilatation, natation, labefactation, retractation, affectation, dictation, volitation, indentation, notation,* and *potation*. Of these, *legation, domination, oration, duration, cessation, potation, natation* and *notation* are spurious. *Conversation* is from *converse*, not *convert*. *Indention* is an archaic form of *indentation*. *Dictation* is truncated from *dictate*. *Labefactation, retractation* and *volitation* are extremely rare forms, whose status is dubious. This leaves only *dilatation* and *affectation* as possible problems for Ben Moshe.[2]

3. By reasoning employed above to account for the Confidence Puzzle, *cómparable, réparable, prerable* may be considered examples of allomorphy along side *divisible*.

4. Just as with compounds, it is extremely difficult to decide when to use a left dominate foot and when to use a right dominant foot. We will not attempt to address this question here.

stress on *sign* in *assign* is preserved in *assignee* and the stress on *cain* in *cocain* is preseved in *cocainism*; *assignee* does not become **àssignée*,[5] *cocainism* does not become **cócainism*, *employee* does not become **èmployée*, and so on. Similarly, the internal structure of the left piece is kept intact in *generalize*, *mineralize* and *federalize*, which do not become **genéralize*, **minéralize* and **fedéralize*, respectively.

The following table is presented as further evidence for the claim that Class Ic boundaries do not destroy metrical structure. The table lists a number of words ending in *-ist*, *-ism* and *-ize*. Notice that the stress pattern of the left piece is fixed across all three forms; for example, *romantic* has 010 stress in *romanticist* (010-0), *romanticism* (010-20) and *romanticize* (010-2).

-ist	-ism	-ize	Stress
romanticist	romanticism	romanticize	010
exorcist	exorcism	exorcize	10
humanist	humanism	humanize	10
antagonist	antagonism	antagonize	010
unionist	unionism	unionize	10
communist	communism	communize	10
militarist	militarism	militarize	100
terrorist	terrorism	terrorize	10
systematist	systematism	systematize	100
stigmatist	stigmatism	stigmatize	10
dogmatist	dogmatism	dogmatize	10
hypnotist	hypnotism	hypnotize	10

In this respect, Class Ic affixes differ from most other "+" boundary affixes which induce stress retraction. Strong retractors (e.g., *-ate*, *-ation*) often mung metrical structure: *design* (01) / *designate* (102). Even weak retractors (e.g., *-ent*, *-ant*, *-ence*, *-able*, *ance*, *al*, *ous*, *ary*) can modify metrical structure: *confide* (01) / *confident* (100). Class Ic affixes are unusual, because they do not induce either mode of stress retraction.[6]

Many so-called cyclicity arguments can be used as further evidence that Class Ic boundaries do not destroy metrical structure. Consider *capitalistic* and *militaristic*, where it has been noted [Withgott] that the */t/* can flap in *capitalistic* but not in *militaristic*, presumably because *capitalistic* comes from *capital* where the */t/* flaps, whereas *militaristic* comes from *military* where the */t/* does not flap. These facts are completely consistent with the observation that *-istic* is a Class Ic affix and that Class Ic affixes do not destroy metrical structure. The same flapping facts hold across a wide number of Class Ic affixes; *capitalist*, *capitalism*, *capitalistic*, *capitalize*, *capitalization*, *capitalitis* and *capitalite* all flap, unlike *militarist*, *militarism*, *militaristic*, *militarize*, *militarization*, *militaritis* and *militarite*.

In conclusion, this section has argued that Class Ic cannot be stressed the same way as other "+" boundary forms and therefore they should be assigned a separate class. The previous section argued that Class Ib requires its own word formation rules and therefore, it, too, should be assigned its own class.

5. *Designee* might be considered a counter-example to the claim that Class Ic boundaries do not destroy metrical structure. The contrast between *dèsignée* and *assignée* is accounted for by noting that *designee* is truncated from *designate* (and keeps that structure), whereas *assignee* is formed from *assign* (and keeps that structure).

6. Admittedly there are a few forms ending in *-ist*, *-ism* and *-ize*, where the affix does not appear to be stress neutral (e.g., *immunize*). These forms are extremely problematic for our proposal since they appear to display classic "+" boundary stress alternations (e.g., *immune/immunize*).

References

Aronoff, M., *Word Formation in Generative Grammar*, MIT Press, Cambridge, MA., 1976.

Chomsky, N., and Halle, M., *The Sound Pattern of English*, Harper and Row, 1968.

Kiparsky, P., *Lexical Morphology and Phonology*, unpublished ms., MIT, 1982.

Liberman, M., and Prince, A., *On Stress and Linguistic Rhythm*, Linguistic Inquiry 8, pp. 249-336, 1977.

Mohanan, K., *Lexical Phonology*, MIT Doctoral Dissertation, avaliable from the Indiana University Linguistics Club, 1982.

Withgott, M., *Segmental Evidence for Phonological Constituents*, unpublished Dissertation, University of Texas, Austin, 1982.

Appendix: Lexicon of Stems and Affixes

- Archaic Affixes (Victims of Trunctation): ate, us, um uli, ii, ae, ine, ar, ure
- Class Ia: ion, ation, ive, ative, ent, ence, ency, ant, ance, ancy, or, ory, atory, able, ible
- Class Ib: ity, al, ality, ation, ative, ator, atory, ic, ian, able, ous, osity
- Class Ic: ist, ism, istic, itis, oid, ine (scientific), ate (scientific), ite (scientific) ite (non scientific), ish, able, ability, ee, eer, ette, ify, ize, ization, ification, ment, mental, mentary, mentarian, mentation, er, ery, ectomy, ology, olysis, ometer, imeter, ographer, oscopy, esce, ique, ess
- "#" boundary: wise, less, ness, hood, ship, way, land, ful, most, ly, man, ward, ling, like, dom
- Latinate Stems: act, bate, carp, cast, cave, cede, ceed, ceive, cel, cent, cept, cern, cess, cess, cide, cinct, cise, cite, claim, clam, cline, clive, close, clude, cluse, coct, crease, create, crete, cult, cumb, cur, cure, curse, cuse, cuss, dic, dict, dite, duce, duct, dure, empt, ept, face, fact, fame, fect, fend, fense, fer, fess, fest, fice, fide, firm, fit, fix, flame, flate, flect, flex, flict, flu, flux, form, fort, found, fract, front, funct, fuse, fute, gest, grade, gress, hale, here, hes, hibit, hort, hume, ject, join, joint, junct, lapse, late, lease, lect, lege, licit, lide, lige, line, lise, loc, lude, lume, luse, mand, mend, mense, merge, merse, miss, mit, mote, mount, mune, mute, nate, note, nounce, opt, pact, pand, panse, pare, part, peal, pel, pend, pense, place, plain, plan, plant, plaud, plause, plead, plete, plex, plic, plode, plore, plose, ply, pone, port, pose, posit, pote, pound, press, prise, prize, prove, puga, pulse, punct, punge, puse, quest, quire, quisit, quit, rase, rect, rode, rog, rose, rupt, scend, sciss, scribe, script, sect, sense, sent, sert, serve, sess, sever, side, sign, sist, sole, solve, sorb, sorpt, spect, spense, sper, spire, spond, sponse, stance, stant, strain, straint, strate, strict, stroy, struct, strue, suade, suase, sult, sume, sumpt, sure, surge, tact, tail, tain, tect, tempt, tend tense, tent, test, text, tin, tinct, tire, tone, tort, tract, train, treat, trice, trite, trorse, troverse, trovert, trude, truse, turb, twine, vade, vail, vase, vene, venge, vent, verge, verse, vert, vest vice, vide, vince, vise, vive, voc, voke, volve, vulse

Se 24.3.4