

SPLITTING VS. LUMPING IN MORPHOLOGICAL ANALYSIS
EVIDENCE FROM GREEK*

1. Introduction

A standing question in morphological analysis is whether forms should be split up into the smallest units for which any kind of case can be made, at the expense of more complex rules governing how these units can be combined ('splitting'), or whether they should be divided into larger, fewer, and possibly easier to manage units ('lumping'). Here I will adduce some evidence from the history of Greek verb inflection that suggests that fairly aggressive splitting is the correct approach. The two phenomena I will consider are the history of an /s/ formative that arguably meant '2nd person singular' for at least two millennia (and probably just means '2nd person' now), and a /t~st^h/ contrast that emerged to code a distinction between active and medio-passive voice in certain forms in Greek verbal paradigms. Since these are both small formatives, for which the evidence was at times not overwhelming, and which appeared in forms that would be quite complex in analyses that recognized them, evidence that they really existed as independent units is evidence that splitting should be favored over lumping.

The issue of splitting versus lumping holds across a wide variety of contemporary morphological theories. In paradigm-based approaches such as Anderson (1992) or Stump (2001), it corresponds to using more rules making smaller changes as opposed to fewer making larger ones. In Distributive Morphology (Noyer (1997) and much additional literature), it would correspond to more vocabulary items each spelling out fewer features, as opposed to fewer spelling out more.

An issue I won't consider here is exactly what kind of theory of morphology learning would favor splitting over lumping, or vice-versa. Working out explicit methods for determining segmentation is hard (Manning 1998), and depends partly on the logically prior question of what kinds of segmentation we should accept as correct in the first place. It seems to me that the rather fine splittings argued for below are sufficiently noteworthy to deserve corroboration from as many directions as possible.

* I am indebted to two anonymous referees for comments, to Peter Svenonius for discussion and encouragement, and to Harold for book loans and reading suggestions. All errors remain my own.

2. /s/ means ‘you’

2.1 *The Appeal of Lumping*

The primary and secondary medio-passive verb desinences in Homeric Greek were¹

	primary	secondary
1sg	-mai	-mE:n
2sg	-(s)ai	(s)o
3sg	-tai	-to
2dl	-stho	-stho
3dl	-stho	-sthE:n
1pl	-metha	-metha
2pl	-sthe	-sthe
3pl	-ntai	-nto

Table 1

In the 2nd person singular row we see an optional /s/, which was always present after consonants (in the perfective indicative, for example, where there were no mood-markers or thematic vowels), but usually absent after vowels, always absent after the thematic vowel, or optative or subjective mood markers, and sometimes absent and sometimes present after the stem-vowels of athematic verbs including perfects (for example /dunasai/ ‘you are able’, but /dizdE:ai/, not */dizdE:sai/ ‘you seek’; /memnE:(s)ai/ ‘you remember’).

A splitter would recognize an /s/ formative meaning ‘2sg’ in these paradigm, but a lumper might challenge this on various grounds. First, there isn’t any segmentation of the desinences that holds for all the rows. In 1-3sg and 3pl we

¹ The primary endings were used in subjunctive and present-tense indicative forms, the secondary in optative and past-tense indicative forms. The desinences appear after the stem, mood-markers and various other material. In the form-citations, I use ‘E:’ and ‘O:’ to represent the original long vowels rendered as eta and omega in standard Greek orthography.

seem to have a person-number complex followed by a primary/secondary marker (which could be treated as irregularly spelled out as /E:n/ in the secondary 1sg), which can plausibly be regarded as ‘tense’, while in the other forms, this breaks down. The nature and history of the active singular forms is too unclear for them to reveal much, but in the dual and plural we have

	primary	secondary
2dl	-ton	-ton
3dl	-ton	-tE:n
1pl	-men	-men
2pl	-te	-te
3pl	:-si	-(sa/e)n

Table 2

The 1st and 3rd plural forms are also hard to segment in a simple and useful way, but in the other forms we can recognize /t/ vs. /st^h/ as a voice-marker, with the rest of the ending as a combined person-number-tense marker. The net effect is that if we want to recognize the /s/ as a formative, we’re going to need some reasonably complex morphotactic rules to get the expressions of the various relevant features into the correct positions in the paradigm.

The other problem is that /s/ with this meaning is not all that widespread in the medio-passive. Most verb forms have a thematic vowel and/or mood-marker, and with these the /s/ is always missing. In some pre-Homeric earlier stage of the language, the /s/ would have been missing more often, since it disappeared due to a highly regular sequence of sound changes converting prevocalic /s/ to /h/, and then deleting the /h/ intervocalically.

The other place where /s/ appears meaning /2sg/ is in active paradigms. The secondary singular desinences are /n/, /s/, \emptyset , clearly original; in the primary we have \emptyset /mi, s/stha/ \emptyset , \emptyset /si (alternates given in order of frequency), with /s/ by far predominant but not original (it should have been deleted by the sound changes, and is indeed absent with some verbs, such as ‘be’ and ‘go’, but has been restored with all other forms. Although /s/ meaning ‘2Sg’ is quite widespread in the active, a lumpner could plausibly contend that this is irrelevant to the mediopassive, because there isn’t any common scheme of segmentation that works cleanly across both kinds of paradigms.

Therefore, somebody using an explicit morphological theory might well find it quite plausible to simply have a rule adding each desinence to the appropriate stem, rather than trying to have individual rules placing /m/, /s/, /ai/, etc. in the appropriate places.

2.2 *Why it would be wrong*

In spite of its appeal, the historical development of the language seems to indicate that a lumping-style analysis would be wrong. We've already seen some indication of the problem: the apparent restoration of the /s/ in certain forms such as /dunasai/, /memnE:(s)ai/ in Homeric Greek. Although Homeric Greek doesn't appear to be a real spoken language, but rather an artificial *Kunstsprache* combining the features of a number of dialects, it seems plausible to take its treatment of the /s/ in 2^{sg} mediopassive forms as characteristic of the state of the language for a period after the loss of intervocalic /s/.

We can't be entirely sure that the /s/ wasn't retained in some of these positions, but in later stages of the language, we find /s/ successively into all of the positions from which the original /s/-deletion sound-change erased it (cf. Siehler 1995:476). So in Classical Attic, the /s/ is restored in all athematic forms (which are imperfective and perfective (aorist) stems of a small number of common irregular verbs, and perfect mediopassive forms, which aren't much used), but remains absent from forms with a thematic vowel or mood-marker. Then in New Testament (NT) Greek, it spread to the 'contract verbs' with stem-vowel /a/, and later into all verbs, where it is still found in Modern Greek. Here are some representative paradigms (duals omitted due to marginality in Homeric and Attic, and absence from Modern):²

	Homeric	Attic/NT	Modern
1sg	erkhomai	erkhomai	erxome
2sg	erkheai	erkhei/erkhE:i	erxese
3sg	erkhetai	erkhetai	erxete
1pl	erkhometha	erkhometha	erxomaste
2pl	erkhesthe	erkhesthe	erxeste

² In Attic, the result of /s/-deletion undergoes vowel-contraction, with /E:i/ the expected outcome, /ei/ (whose phonetic implementation would have been a closer vowel than /E:i/, details unclear) peculiar to this particular form.

3pl	erkhontai	erkhontai	erxunte
			‘go’

Table 3

Summing up, we see that an /s/ formative which at one time appeared quite widely in various active forms, but only in few places in the mediopassives, due to a sound-change which had erased it, restored itself into those positions. This has always been attributed to the operation of ‘analogy’; the point here is that the analogically driven spread is only explicable if a reasonable number of speakers are analysing the subsequence in question as a formative meaning ‘2sg’ in the positions where it occurs. That is, ‘spitting’ these forms is justified even when the paradigms don’t get a uniform structure, and the postulated formatives fail to appear in many places where they might be expected. It appears to be justified even for dialects similar to Homeric; even more so for Attic and NT, in which the /s/ is appearing in more forms.

An objection (raised by an anonymous reviewer) is that perhaps a lumping analysis could explain the facts anyway, on the basis that after the loss of intervocalic /s/, there would be two different formatives meaning 2sg, /ai/ and /sai/. Then, for some reason, the latter begins to spread. The problem with this is the lack of any plausible reason. The spread of /s/ from a rather marginal corner of the morphology back into its full range of original positions is a steady trend holding over more than a millenium. It is possible that it ‘just happened’, but it would seem worthwhile to try to find possible reasons for it, and a preference for splitting over lumping would appear to be a good candidate.

2.3 Modern Greek

Although /s/ has been preserved in Modern Greek, its meaning has arguably altered somewhat, with the singular property dropping out. In Ancient Greek, the 1st and 2nd person plural pronouns were /hE:meis/ and /hu:meis/ while in Modern they are /emis/ and /esis/ (all nominative case). The accusative/genitive plural forms are furthermore:

	1	2
GenSg	mu	su
AccSg	(e)me(na)	(e)se(na)
Acc/GenPl	(e)mas	(e)sas

Table 4

It seems justified to identify /m/ as a first person marker and /s/ as a second person marker, both indifferent to number.

3. /t/ vs /sth/

Our second case is already evident from the contrast between the dual and plural forms of Table 1 and Table 2. Many of the active forms of Table 2 have ‘t’ where the corresponding medio-passive ones of Table 1 have /st^h/. This is an innovative rather than original feature of the paradigms. The reconstructed PIE forms for the active dual and 2pl have the /t/ (Siehler 1995:455-6; some minor simplifications and orthographic changes here):

	Active	
	primary	secondary
1dl	-wos	-we
2dl	-tH ₁ es	-tom
3dl	-tes	-ta:m
1pl	-mos	-me
2pl	-te	-te
3pl	-nti	-nt/(e:)r

Table 5

Many of the mediopassive forms are ‘more than ordinarily’ conjectural (as marked by parentheses), but no trace of an t~ts^h alternation is present, and in fact the origin of the 2pl st^h is unclear (Siehler:277ff):

	Mediopassive	
	primary	secondary
1dl	(-wosdhH ₂)	(-wedhH ₂)
2dl	(-HtoH ₁)	(-teH ₁)
3dl	(-Hte)	(-te)
1pl	-mosdhH ₂	-medhH ₂
2pl	-dhwo	-dhwo
3pl	-(e:)ror/ntor	-(e:)ro/nto

Table 6

It is taken as obvious that the dual mediopassive forms are formed by analogy with the active ones; the most plausible historical sequence being that it accidentally becomes possible to segment the 2pl forms as /t~st^h/+/e/, with the /t~st^h/ contrast then spreading to the 2/3du forms. And indeed further, to the third person, dual and plural imperatives, where /st^h/ displaces original /t/. The relevant Greek forms are (2sg excluded):

	active	mediopassive
3sg	-to:	-st ^h o:
2dl	-ton	-st ^h on
3dl	-to:n	-st ^h o:n
2pl	-te	-st ^h e
3pl	-nto:n	-st ^h o:n

Table 7

The available PIE forms on the other hand show no trace of a the /t~st^h/ alternation (Siehler:601):

	active	mediopassive
3sg	-tu	-to
2pl	-te	-dhwo
3pl	-(e)ntu	-nto

Table 8

So what we have here is a further case of the spread of a rather ‘small’ position in the morphological form of the verb, containing only two possible occupants.

4. Conclusion:

Historical linguists have of course known about examples of analogy such as these for a very long time. But drawing the consequences for morphological analysis

does not seem to have been a prominent concern of the the generative literature. For example I find no discussion of how refined a morphological analysis ought to be in the recent, comprehensive morphological framework of Stump (2001), or earlier proposals such as Anderson (1992) or Lieber (1992). One of the conceptual difficulties of the theory of inflectional morphology is that, because of the finite nature of the systems, the standard assumption that the members of a speech community all acquire the same grammar is even weaker than it usually is. We cannot therefore assume that these or any other analogical change phenomena show that all learners of a language with rich morphology are avid splitters. But we can conclude that enough of them are to have a significant effect on the historical development of such languages.

References

- Anderson, Stephen R. (1992) *A-Morphous Morphology*, Cambridge University Press.
- Lejeune, Michel (1972) *Phonétique Historique du Mycénien et du Grec Ancien*, Editions Klincksieck, Paris.
- Lieber, Rochelle (1992) *Deconstructing Morphology*, Chicago University Press.
- Manning, Christopher (1998) 'The segmentation problem in morphology learning', in D.M.W. Powers, NeMLaP3/CoNLL98 Workshop on Paradigms and Grounding in Language Learning, ACL, pp 299-305.
- Noyer, Rolf. 1997. *Features, Positions and Affixes in Autonomous Morphological Structure*. Garland Publishing, New York. Revised version of 1992 MIT Doctoral Dissertation.
- Siehlner, Andrew (1995) *New Comparative Grammar of Greek and Latin*, Oxford University Press
- Stump, Gregory (2001) *Inflectional Morphology: A Theory of Paradigm Structure*, Cambridge University Press.