To appear in: ***Morphology. A Handbook on Inflection and Word-Formation (Handbücher zur Sprach- und Kommunikationswissenschaft),*** **ed. C. Lehmann, et al. Mouton de Gruyter Publishers, Vol. II (2004).**

153. Morphological Reconstruction

CONTENTS

1. Introduction: Basic Questions and Methods

Morphology is above all concerned with the forms (morphemes and words) of a language. These forms include roots and affixes, as the basic building blocks of words, as well as the patterns of combination — both derivational and inflectional — by which words are built up with these elements, and the phonological adjustments that apply within

words as part of the overall word-formation process. Moreover, of interest too are the categories that these forms express and mark, e.g. gender, number, person, animacy, tense, mood, etc. All of these aspects of morphology can be reconstructed for earlier stages of a language and/or proto-language, once one has appropriate data and methods to work with and some guiding principles.

The basic methods are those that serve reconstruction in other domains (especially phonological). That is, they include primarily the Comparative Method and Internal Reconstruction, as well as whatever other means one can employ to reach a reasonable set of assumptions about the paths of development from reconstructed elements to attested outcomes (e.g., observation of attested changes and models derived from them).

These methods are illustrated, and expanded upon with other useful principles, in the sections that follow.

2. Reconstruction: The First Steps

For the most part, the reconstruction of the formal side of morphological units can be fairly straightforward, drawing on the successes linguists have had with the reconstruction of

sounds and sound systems. That is, once a set of valid sound correspondences holding among (related) languages have been worked out, and phonological units for a proto-language for those languages have been reconstructed based on those correspondences, one can exploit those reconstructions in the reconstruction of the form of morphemes. Essentially, all one has to do is observe the sound correspondences in a comparison of actual morphemes — as opposed to simply matching up sounds that are, so to speak, disembodied — and string together the reconstructible segments to give whole reconstructed forms (a practice Anttila 1972:351 has called doing morphological reconstruction by "applied phonology"). This is a reasonable step to take, since the correspondences themselves emerge out of a comparison of morphemes in the first place.

For example, the comparison sets *p*/*p*/*p*/*f*, *i*/*a*/*a*/*a*, and *t*/*t*/*t*/*d*, etc. across Skt., Gk., Lat., and Gothic, respectively, in *pitar-*/*patér-*/*pater*/*fadar*, and the recognition that each set is a regular, i.e. valid and well-established, sound correspondence that licenses the reconstruction of a Proto-Indo-European (PIE) phonological unit, together allow one to put the reconstructed segments together sequentially to yield a PIE form *pətər-. Similarly, putting *a*/*e*/*e*/*i* and *s*/*s*/*s*/*s* together across these same

languages, along with *t*/*t*/*t*/*t* (in the environment *s*__) and *i*/*i*/Ø/Ø (in the environment __#), in *asti*/*ésti*/*est*/*ist* warrants a reconstruction for PIE of word *esti (actually, more properly *H_1esti in laryngealistic terms). With a less obvious but no less compelling set of correspondences, the equation of Skt. *ch*, Hittite *sk*, Lat. *sc*, and Alb. *h* in verbal forms permits one to reconstruct a PIE morpheme *-sk'-, given that the sound correspondences in evidence here are impeccable (cf. Skt. *cha:ya-* and Alb. *hije* 'shadow', where Gk. *skía* gives a clear indication of *sk' as the starting point for this set).

Within the individual languages, several of the forms that these reconstructions are based on fall into paradigmatic sets with other forms that are equatable across the languages, e.g. along with *asti* in Skt. there is *asmi*, along with Gk. *esti*, there is *emmi* (in the Aeolic dialect), and so on, and these together license a reconstruction *esmi (actually *H_1esmi). To the extent that these reconstructed forms *esmi/*esti constitute members of the same paradigm — as their outcomes in the various languages do — segmentation of the forms into constitutent morphemes *es- and *-ti/*mi is possible, and thus inferences about the order of morphemes in the reconstructed words become possible.

What has not been addressed here is the meaning to be assigned to these reconstructed forms, yet that is a crucial step in morphological reconstruction, to the extent that morphemes are seen to be pairings of both form and meaning. In the simplest case, involving lexical items with a clear meaning, as with *pitar*-, etc. above, all of which mean 'father', or *as*-, etc., all of which mean 'be', the decision is straightforward; thus a form-meaning nexus *pə̆ter- 'father' or *es- 'be' can safely be assumed for PIE. Similar considerations also hold for non-root elements, so that the observation that the reflexes of *-ti all mark third person singular and those of *-mi all mark first person singular forms in their respective languages licenses one to assign those values to the proto-morphemes in question. It is also therefore safe to assume that the person markings were suffixes, since they are suffixal in all of the languages represented.

In somewhat trickier cases, the meanings of the elements in the comparison set do not match perfectly, and thus more difficult decisions need to be made. To some extent, reconstruction in such cases becomes a matter of historical semantics and semantic reconstruction, not morphological reconstruction per se, but the ramifications for reconstructed proto-morphemes should be clear. Thus, Skt. *ma:tar*-/Gk.

ma:ter-/Lat. *ma:ter*, etc. mean 'mother' but reconstructing that meaning for the PIE word must take the Alb. cognate *motër*, meaning 'sister' into account; conceivably the original meaning of the word was broader (e.g. 'female in nuclear family') or else Alb. alone innovated (perhaps through the involvement of an original dvandva (coordinative) compound 'sister-and-mother' as a merism (defined by Watkins 1995:9 as "a two-part figure which makes reference to the totality of a single higher concept") for female kin). At the grammatical level, semantic/functional mismatches for cognate markers can pose problems for reconstruction, especially in the absence of clear bases for understanding how category markings can themselves mutate and be altered. Thus, in the *-sk'- case mentioned above, the Hittite morpheme marks iterative action, the Lat. morpheme marks inchoative (beginning) action, the Skt. morpheme is simply one of several markers for present tense system stems, and the Alb. morpheme marks non-active voice (with passive, reflexive, reciprocal, and stative functions) but only in the present tense system. Thus there is an element of presentiality in most of the reflexes of *-sk'-, but one needs to ask whether the original function of this morpheme was just to mark present, or instead was more specific, with one of the attested values as the starting point in PIE. Any answer here

runs the risk of seeming to be arbitrary, but without an answer, the reconstruction for this morpheme is incomplete; reconstructing the form alone without the meaning or function is satisfying only half of the burden.

The above cases draw on the Comparative Method, the mainstay of historical reconstruction for nearly two centuries, but other methods can be employed as well in reconstructing morphology. In particular, Internal Reconstruction can be used, where, in essence, alternations within a single synchronic stage of a language are "undone" as it were, and an earlier state without the alternation is reconstructed. Morphological reconstruction in such a case consists in the positing of a unique form for a given alternation found in a later stage, generally taking the later allomorphy to be the result of conditioned sound changes. Thus the allomorphy seen in the Modern English plural marker *–s/-z/-əz* allows for a hypothesis of an earlier marker with a single undifferentiated form, e.g. *-əz*, with the variants having arisen, as is so often the case, via sound changes, e.g. syncope and voicing assimilation. But even this method can break down; there is no way to reconcile the plural marker *–en* (restricted in Modern English just to *oxen* and *brethren*, though *children* contains it too) with the *–s/-z/-əz* forms, and that is as it should be, given that its

distribution is lexically idiosyncratic and not rule-governed and not tied to phonological conditioning in any way. But even cases involving clear phonological conditioning of the allomorphy can be problematic for reconstruction; for instance, the Korean nominative markers are in complementary distribution: *-i* occurring after a consonant-final noun, *-ka* after a vowel-final stem, but the phonetic distance between the two forms makes a reconstruction of a single nominative marker that gave rise to both of these alternants most unlikely. The third method, or rather a principle, mentioned above (cf. 1), namely ensuring that there is a reasonable path of development from the earlier reconstructed stage to the attested ones, must be invoked here, so that one does not take *–i* and *–ka* back to the same proto-form, internally reconstructing by brute force, as it were. Presumably they each have an independent origin and have come, by various developments, to stand as phonologically distributed functionally equivalent alternants.

3. Further Guiding Principles

More can be said about these methods and principles. Critical to positing a reconstruction that requires only a reasonable set of changes for the development of any given element is an

understanding of what the expected processes of language change are. What is thus especially important here, besides recognizing the regular nature of sound change (which gives the regular sound correspondences drawn on in doing morphological reconstruction as "applied phonology", as above, cf. 2), is allowing for the workings of other processes of change, such as analogy. For instance, the consonantal matchings between the root nouns (where the root with nothing added constitutes the stem) seen in Gk. *pod-* and Lat. *ped-*, both meaning 'foot', are perfectly regular, but the vowels do not match up as expected, as Gk. *o* usually corresponds to Lat. *o* (reflecting PIE *o) and Lat. *e* usually corresponds to Gk. *e* (reflecting PIE *e). Armed with the knowledge that there are Indo-European languages in which reflexes of *e and reflexes of *o alternate in grammatically determined environments (e.g. English present tense *sing* from *seng^{w}h-, versus past tense *sang* from *song^{w}h-), rather than reconstructing a different vowel for the *o* ~ *e* correspondence, a step which would ultimately lead one to reconstruct a different vowel for every such correspondence (and there are others, especially when languages other than Gk. and Lat. are included), one can instead reconstruct a process by which *o and *e alternate within the grammatical forms of 'foot' in the proto-language.

One can then treat the Gk. and Lat. forms with fixed vowels as the result of analogical change within the paradigm as the individual languages took shape. The exact conditions for the *e/*o alternation may be unclear, but the method of allowing for the "undoing" of the potential effects of analogy leads one to the reconstruction of a morphological process (of vowel gradation) for the proto-language, that is, an aspect of a PIE word-formation process. It can be noted as well that the inferences about morpheme order derivable from the comparison of *esti and *esmi similarly reflect reconstructive assumptions made about word-formation processes of the proto-language, i.e. suffixing (at least for person/number marking).

Moreover, another useful principle can be invoked to illuminate the nature of the reconstructed *e/*o alternation. In particular, it is known that the relics of processes or earlier states that were once productive can be found embedded in compounds; for instance, English *with-* retains its original adversive meaning ('against') only in composite forms such as *withstand* (= "stand against", not "stand alongside of' with the synchronically regular comitative meaning of *with*). Therefore, one can look to compounds for some insights into aspects of earlier formations. The occurrence of *e in the

genitive case of an old root noun found in the Gk. composite form *des-póte:s* 'master, owner', from *dems-pote:s, literally the "dwelling's (*dem-*) master (*potes-*)" as opposed to the consistent *o found in derivatives related to the root nouns, such as Lat. *domus* 'house', invites the suggestion that *e was proper, originally, to the oblique cases (e.g. genitive) whereas *o was found in the direct cases (e.g. nominative) of root nouns in the proto-language. Therefore, one can infer an original paradigm with nominative stem *pod- and genitive (oblique) stem *ped-.

A corollary of the use of relic forms for guidance in reconstructing earlier states is what can be called the "Meillet Principle" (after Antoine Meillet, based on work discussed in Arlotto 1981: 144-145), namely to reconstruct from synchronic irregularities and isolated forms, that is any sort of unproductive material present in a language at a given time. The rationale is that synchronically unproductive material is exactly what demands a historical explanation, whereas synchronically productive forms could in principle have been created at any time by means of the productive and regular processes, and thus are not an indicator of the presence of some element in the proto-language. For example, when confronted with Skt. *bhr≥-ta-* and Old Irish *breth* as past passive

participles of *bher- 'carry', one can mechanically reconstruct *bhr≥-to- as a PIE pre-form underlying them historically, even though the *-to- participial formation is clearly the productive one for PIE and on into the offspring languages, to judge from its widespread occurrence in all of the branches of the family. Adding in Lat. *la:-tus* and Gk. *ois-tós* as the participles to the outcomes of *bher- in these languages (present tense *fero:*/*phero:*, respectively) complicates the picture somewhat, though. The Lat. form seems clearly to be an import from the regular participle of a semantically related verb *tollo:* 'pick up; take on' (*la:tus* from *tla:tus, formed regularly with *-to- from a root *telH$_2$-) and thus presumably simply shows the substitution of a participial form from a different paradigm; such a substitution is understandable, as *tollo:* supplied the form *tuli:* that functions as the perfect tense associated with *fero:*. However, which form did *la:tus* replace, an outcome of *bhr≥-to- or something else again? Gk. *oistós* provides the answer, since it is synchronically isolated within Gk. (the root *ois-* occurs elsewhere only in a suppletive future of *phéro:*) and thus can only be explained historically, as a relic of an earlier irregular (i.e., suppletive) state. We thus reconstruct the PIE participle as *oistos and explain the Skt. and Irish forms as

simply creations within those branches using the productive pattern.

Meillet's Principle thus licenses the reconstruction of suppletion, that is, a proto-language morphological irregularity, and in that sense can lead to a different solution than one might reach with internal reconstruction where irregularities are taken back to earlier regular states. Similarly, allowing for the reconstruction of suppletion means that not every paradigm that one might want to reconstruct has to be fully articulated, with all members intact (i.e., all cells filled), in keeping with the observations from attested language states that there can be defective paradigms and that paradigms can be built up piece-meal (see Watkins 1962 for discussion of how paradigms can be reconstituted).

The overriding factor here, as always, is to give the best account of the facts. In a sense, then, Meillet's Principle is akin to the Accountability Principle, known from sociolinguistics (cf. Labov 1982: 30, Winford 1990: 227), in that it requires that all the forms in a set of comparanda be accounted for — in this case it is not enough to account for just the Skt. and Old Irish forms with a PIE form that does not allow for a straightforward account of the Gk. suppletion. That is, an explanation is possible for why *bhr≥ta-* is to be found in

Skt. or Celtic: at any time due to productivity of *-to- participial formation, such a form could have come into being; the Gk. *ois*-, on the other hand, demands a different kind of account, e.g. continuing an inherited suppletive paradigm, as suggested above.

4. Going Beyond Simple Reconstruction of Forms

What else can be reconstructed of the morphology in a proto-language? Clearly, once one makes assumptions about the functions of particular reconstructed items, one has also made inferences about the relevant grammatical categories for the proto-language, e.g. number, person, tense, etc., based on *-ti and *-mi in PIE. However, it is also possible to go beyond these categories, and engage in some internal reconstruction on the value of categories reconstructed for a proto-language.

For example, the suppletion that was reconstructed in the paradigm of *bher-, as discussed above (cf. 3), is found with other verbs in PIE, most often focusing on present versus past tense forms (as with Lat. *fero:/tuli:,* noted above, or Alb. *ha* 'eats', from a root *(H_1)ed- vs. *(hën)gra* 'ate', from the root *g^wro:- (with prefixes *Ho-en-). One way to make sense of the relatively widespread occurrence of apparent present/past

suppletion is to asssume that it has to do with some inherent properties of the roots in question, derivable from their respective basic meanings. However, present versus past seems to be an unlikely distinction to be associated with the lexical meanings of particular roots (after all, how could an action be inherently associated with past time). Thus, it is more likely that the original distinction was aspectual in nature, e.g. imperfective (durative) versus perfective (completive), since that is a lexically encodable property (what is sometimes referred to as "Aktionsart"). Presumably, these aspectual distinctions were rearranged and altered somewhat on the way to the attested languages. Internal reconstruction, therefore, attempts to rationalize the irregularity of reconstructed suppletion by projecting it back to a previous state where it is not unmotivated but rather follows from some other property, in this case, aspect as determined by the semantics of a given root. This exercise thus leads to some reasonable inferences about the pre-PIE state of affairs regarding verbal categories.

Internal reconstruction can also be carried out on the reconstructed morphemes themselves. For instance, the relation between the reconstructed third person singular present ending *-ti for PIE (as above, 2) and the comparable past ending *-t (cf. Skt. *-t#* = Lat. *-d#*), as well as *-mi vs. *-m in

the first person, invites the analysis that the *-i itself marks present time, and that moreover the present ending was diachronically derived from the endings *-t and *-m through the addition of the *-i. It would follow, therefore, that the original value for *t and *-m was just to mark person (and number), not to mark past time reference directly; only once the opposition of *-ti with *-t emerged was a present/nonpresent distinction relevant.

Hypotheses such as these made by internally reconstructing from reconstructed proto-language elements are not subject to confirmation the way that internal reconstruction on an attested state is (e.g. reconstructing /-əz/ for the English plural allomorphy is verified by examining earlier English *s*-plurals). However, they are compelling scenarios to the extent that they are based on what is known about language in general.

In a similar vein, and as noted above, to the extent that one can be sure of various trends in language change in general, they can be employed in reconstruction. Thus, although it is not impossible for affixes to turn into full-fledged words (see Joseph & Janda 1988, Campbell 1991, and Janda 2001 for discussion), it is nonetheless true that the opposite development, in which words develop into affixes, is by far the

more common direction for changes involving words and bound morphemes. That is, examples such as Old English *ha:d* 'condition; state' giving the Modern English derivational suffix *–hood* (cf. *child*/*child-hood*) from original compounds ("the state of being X") or the English negative adverb *not* giving, at least in part via phonological reduction in an unstressed position, the inflectional affix *–n't* (see Zwicky & Pullum 1983), are relatively common, while examples such as the Old English bound genitive case-suffix *–(e)s* giving, via a reanalysis and aligning with the pronoun *his*, the less-bound and somewhat word-like possessive marker *'s* in Modern English (less bound in that it attaches at the end of a phrase, as in *the King of England's hat*) are significantly rarer, though definitely attested.

This observation means that in some cases, we can reasonably locate the history of a given affix in a word, and this holds whether the basic methodology being employed is the Comparative Method or Internal Reconstruction. For instance, an internal comparison of the English adjectival and adverbial suffix *–ly* with the similarly used and phonologically similar free word *like*, as in *quickly*/*quick-like*, *friendly*/*friend-like*, might suggest an historical derivation of the suffix from a reduction of the free word, perhaps under conditions of low

accentual prominence, and a cross-linguistic comparison of the French adverbial suffix *–ment* with the more word-like element in Spanish, *-mente* (word-like in that a single instance is distributable across two conjoined adverbs, e.g. *clara y rapidamente* 'clearly and rapidly'), permits the reasonable inference that the suffix derives from a once less-bound element. In the case of *–ly/like*, earlier English evidence confirms the reconstruction (cf. the use of *–lic* in Old English), as does Lat. evidence in the case of *–ment/mente* (regarding the use of the ablative *mente* of the noun 'mind' in adverbial phrases, e.g. *obstinata: mente* 'with an obstinate mind', i.e. 'obstinately'). Still, such hypotheses are not iron-clad, and synchronic resemblances can be misleading — the *–less* of *friendless*, for instance, has nothing to do with the independent word *less* (from Old English *læ:s(sa)*) deriving instead from the Old English preposition *le:as* 'without'.

5. Pushing the Limits: Reconstructed States as Real Languages

Some of the successes of morphological reconstruction discussed in the previous sections involve simply applying analytic techniques that are well-known in linguistics, and in

some instances, applying them to a reconstructed proto-language, essentially treating it as just another language, a synchronic state that can be analyzed and, among other things, subjected to internal reconstruction. As a final example of morphological reconstruction, an example is presented of how one can reconstruct morphophonemic rules for a proto-language, based on the PIE paradigm for the present of the verb 'be'. This example serves as a suitable conclusion, as it draws on comparative methodology, internal reconstruction, and Meillet's Principle, as well as general principles of morphological analysis.

Besides the forms first and third person forms *esmi and *esti, it is possible to reconstruct a second person form as well in the singular. The comparison can be made of Skt. *asi* 'you are' with Gk. *ei* 'you are', both of which are irregular within their respective languages (e.g., the non-occurring **assi might be expected in Skt., all things being equal). The "applied phonology" methodology (cf. 2) leads to a reconstruction for PIE of *esi, and the shared irregular status of the Skt. and Gk. forms ensure that this form is to be posited for the proto-language; that is, the paradigm was not a defective one in PIE lacking a second person singular form. Moreover, this form itself is an irregularity within the PIE verbal system, since

**essi would be expected, based on the secure reconstrruction of a root *es- (cf. 2)) plus the second person singular ending *-si (seen, for instance, in Skt. *bhara-si* 'you carry', Gothic *bairi-s*, etc.). Given the difference between the form expected on morphological grounds, *essi, and the reconstructed form based on the comparative evidence, *esi, it is reasonable to reconcile the two forms by treating *essi as the (morphologically motivated) underlying form for PIE and *esi as the surface form, also for PIE; this step means that a morphophonemic rule converting an underlying /-ss-/ into a surface [-s-] must be posited for the proto-language. Moreover, it licenses the inference that the underlying form was the actual pre-PIE form, and that a (sibilant) degemination sound change operated between pre-PIE and PIE to give the surface form *esi, and create the reconstructed irregular singular paradigm with *esmi/esi/esti.

6. Conclusion

In a sense, then, morphological reconstruction is not significantly different in its goals, methods and guiding principles from phonological reconstruction. It is thus not surprising that the topic (often under the rubric of

"grammatical reconstruction") is not discussed at great length in any standard textbooks, except insofar as internal reconstruction leads to results that have consequences of a morphological (most usually morphophonological) nature; still, interested readers should consult Anttila 1972/1989, Fox 1995, Hock 1991, and Trask 1996 for additional general discussion and examples.

7. References

Anttila, Raimo (1972), *An Introduction to Historical and Comparative Linguistics*. New York: MacMillan.

Anttila, Raimo (1989), *Historical and Comparative Linguistics*. Amsterdam: John Benjamins Publishing Co. [revised edition of Anttila 1972].

Arlotto, Anthony ([2]1981), *Introduction to Historical Linguistics*. Washington, D.C.: University Press of America [[1]1972, Houghton Mifflin Co., Boston]

Campbell, Lyle (1991), "Some Grammaticalization Changes in Estonian and Their Implications". In: Traugott, Elizabeth & Heine, Bernd (eds.), *Approaches to Grammaticalization*. Amsterdam: John Benjamins Publishers, Vol. I, 285-299.

Fox, Anthony (1995), *Linguistic Reconstruction: An Introduction to Theory and Method*. Oxford: Oxford University Press.

Hock, Hans Henrich ([2]1991), *Principles of Historical Linguistics*. Berlin: Mouton de Gruyter [[1]1986].

Janda, Richard D. (2001), "Beyond 'Pathways' and 'Unidirectionality': On the Discontinuity of Language Transmission and the Counterability of Grammaticalization". *Language Sciences* (Special Issue — *Grammaticalization: A Critical Assessment*, L. Campbell (ed.)) 23,265-340.

Joseph, Brian D. & Janda, Richard D. (1988), "The How and Why of Diachronic Morphologization and Demorphologization". In: Hammond, Michael & Noonan, Michael (eds.), *Theoretical Morphology: Approaches in Modern Linguistics*. San Diego: Academic Press, 193-210.

Labov, William (1982), "Building on Empirical Foundations". In: Lehmann, Winifred & Malkiel, Yakov (eds.), *Perspectives on Historical Linguistics*, Amsterdam: John Benjamins Publishing Co., 17-92.

Trask, R. L[awrence] (1996), *Historical Linguistics*. London: Arnold.

Watkins, Calvert (1962), *Indo-European Origins of the Celtic Verb*. Dublin: Institute for Advanced Studies.

Watkins, Calvert (1995), *How to Kill a Dragon: Aspects of Indo-European Poetics*. Oxford: Oxford University Press.

Winford, Donald. 1990. "Copula variability, accountability, and the concept of "Polylectal" grammars". *Journal of Pidgin and Creole Languages*, 5,223-252.

Zwicky, Arnold M. & Pullum, Geoffrey K. (1983), "Cliticization vs. inflection: English *n't*". *Language* 59, 502-513.

Brian D. Joseph, Columbus, Ohio (United States of America)