Tobias Scheer
Université de Nice

A representational theory of morphological information in phonology

Morphological information plays an important role in a number of phonological processes. It is usually referred to by non-phonological symbols such as "-", "+" or "#". These are taken to interact with properly phonological categories such as consonants and vowels, while remaining truly morphological in both Signifiant and Signifié.

Lexical Phonology (e.g. Kiparsky 1982,1985, Mohanan 1986) has proposed to encode morphological information through different derivational levels. As a result, morphological boundaries disappear in the statement of rules.

The purpose of the present talk is to explore the possibility of eliminating boundary-symbols in yet a different way. A claim is made to the effect that Morphology is a device of grammar whose Signifiant has a phonological identity. That is, the existence of a morphological boundary is decided by the morphological component alone, but once a boundary has come into being, it is expressed through a unit that enjoys a truly phonological identity.

An immediate benefit thereof are homogeneous phonological representations: phonological processes refer only to phonological objects.

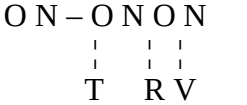
It is well established that prosodic categories such as stress may materialize as phonological objects. For instance, stress in Italian Tonic Lengthening and Raddoppiomento Sintattico (e.g. Chierchia 1986) is best represented as a syllabic constituent.

Taking on this precedence, I explore the possibility of identifying the word-initial boundary "#" as a CV-unit in the sense of Lowenstamm (1999), i.e. as an empty Onset followed by an empty Nucleus.

This kind of representation allows for a non-circular account of the non-existence of word-initial #RT-clusters (where "R" stands for any sonorant, "T" for any obstruent). Indeed, it is commonly held that #RT do not exist because sonority must increase within branching Onsets, and a word may not begin with a Coda-Onset sequence. However, the ONLY reason to posit a constraint of that kind rather than its reverse is the factual observation that word-initial clusters are always of rising sonority. If languages displayed only #RT clusters to the exclusion of #TR, the constraint would simply state the reverse: sonority would be said to always decrease within branching Onsets. This cannot be the ideal move since a trivial requirement for scientific activity is to propose a theory that is able to account for the world observed, AND to rule out all non-existing worlds (typically generative: "all and only the Xs that occur").

If on the other hand "#" identifies as a CV-unit, and under the assumption that syllabic constituency boils down to a strict sequence of non-branching Onsets and non-branching Nuclei (Lowenstamm 1996, Scheer 1998), the following picture obtains.

(1) a. #TRV



b. #RTV

Following Government Phonology, empty Nuclei in the representations under (1) are subject to the phonological Empty Category Principle (Kaye et al. 1990, Kaye 1990). In representations that recur to strict sequences of Onsets and Nuclei, a branching Onset TR is

interpreted as an R-headed domain of consonantal interaction. Under (1a), the ECP applying to the Nucleus that separates T and R is satisfied through the governing relation that R establishes over T (Infrasegmental Government, cf. Scheer 1998). R is entitled to do so because it is licensed to govern by its own Nucleus, which contains the first vowel of the word (Government Licensing, cf. Charette 1991). The Empty Category Principle may also be satisfied by an internuclear relation called Proper Government. Under (1a), the first vowel of the word is able to properly govern the initial empty Nucleus precisely because the empty Nucleus flanked by the T and the R is taken care of by the consonantal relation mentioned. Under (1b), however, the empty Nucleus separating R from T cannot be cicumscribed by a consonantal relation because the head of the putative domain, R, fails to be licensed by its Nucleus, which is empty. It therefore calls for Proper Government from the first vowel of the word, which in turn is unable to reach the initial empty Nucleus. The structure in (1b) is ill-formed because the ECP of the initial empty Nucleus remains unsatisfied.

Note that this approach is not circular because none of the theoretical devices mentioned (Proper Government, Infrasegmental Government, Government Licensing, strict CV constituent structure) are built on the observation that sonority always increases within word-initial clusters.

Now, it is well known that the empirical picture is not quite what the preceding discussion suggests. In addition to languages where #TR, but not #RT is attested, there are languages that exhibit both types of word-initial clusters. Among those count many modern occidental Afro-Asiatic languages such as Berber and Moroccan Arabic, but also most of the modern Slavic languages (e.g. Russian, Czech, Serbo-Croatian). No language, however, shows #RT, while lacking #TR.

Traditional accounts of word-initial clusters (e.g. Clements 1990, Selkirk 1984) have nothing to say about languages like Moroccan Arabic where #RT is just as good as #TR. Languages of that kind are usually viewed as bare exceptions to a set of constraints or markedness conventions. This, of course, is done to the expense of losing a UG-generalisation about branching Onsets: sonority, then, "normally" increases within this type of structure, but may also decrease. Deciding that initial R in #RT-clusters are extrasyllabic does not take any further either: it simply deletes the offending item from the set of data considered. The sonority-sequencing generalisation is save, but no consequences of either empirical or theoretical nature obtain (e.g., unsyllabified consonants would be expected to delete or at least be unstable, which is not the case at all for initial R in #RT-sequences). Namely, extrasyllabicity does not provide any clue to the question in which sense #RT-languages are different from #TR-only languages.

On the other hand, if "#" identifies as a CV-unit, a parameter may be set which does not invalidate any of the theoretical claims that are made when "regular" Indo-European languages are looked at. In fact, the representation of languages where the distribution of word-initial clusters is free follows from the preceding discussion. Recall that clusters of the #RT type were ill-formed because of the existence of the initial CV, whose empty Nucleus requires Proper Government. If there were no initial CV, no restriction would rely on word-initial clusters because the only empty Nucleus calling for Proper Government, i.e. the one separating #R from T in (1b), will always be satisfied by the first vowel of the word.

The crucial feature of this approach is its ability of expressing the difference between the two kinds of languages by a PRIVATIVE opposition: the word-initial context is either marked by the presence of an initial CV-unit ("regular" Indo-European type), or by its absence (Moroccan Arabic type). This option is not available unless a phonological identity is given to

the morphological information "beginning of the word". Under the traditional approach, there is no way of considering that # is absent: it occurs in all word-initial representations because no part of its identity is independent from this particular context. Only when it is thought of as a phonological unit can "#" possess an identity that is partly independent from its position in the string: "word-initial" refers to the Signifié of "#" alone, which is morphological and inalterable. By contrast, its Signifiant enjoys an independent phonological life and hence may be conceived of as either present or absent.

The privative opposition discussed is comparable to the one that is argued for in monovalent segmental representations. While traditional approaches using features claim that a phonological prime may have a positive or a negative value, but is present in any case, Dependency Phonology (Anderson & Ewen 1987), Particle Phonology (Schane 1984), Government Phonology (Kaye et al. 1985) and underspecification theories (e.g. Steriade 1995) argue for primes that may or may not be present in representations. In this sense, the proposal made here enlarges the scope of privative representations from the segmental to the syllabic level.

Finally, I inquire about the different situation of both groups of languages that admit #RT clusters, i.e. occidental Afro-Asiatic and Slavic. In either case, we are lucky enough to have a record of former stages of the languages under concern (although on a comparative basis for Berber only). As a matter of fact, neither Classical Arabic nor Common Slavic (or Old Church Slavonic) contained #RT-sequences. That is, a diachronic generalisation may be established to the effect that ALL cases of known #RT clusters arose through the loss of a vowel that separated both consonants in former times.

This ties in with another observation, that is the number of #RT-words in a language like Moroccan Arabic as compared to a modern Slavic language like Czech. In the former, the distribution of #RT-words is random, and the lexicon contains about as many #TR- as #RT-items. In the latter, however, there are no more than some 40 roots with an #RT structure (a list of Slavic #RT-words that aims at exhaustivity is available at http://www.unice.fr/dsl/rt/slavicRT.htm).

Moreover, the words featuring on this list are the same all through Slavic: e.g.

(2) Czech *lhát* = Slovenian *lgati* = Polish *Lgac'* = Russian *lgat'* "lie (verb)"
Czech *mha* = Upper Sorbian *mhLa* = Polish *mgLa* = Ukrainian *mhLa* =Russian *mgla* "fog"
Czech *rvát* =Serbo-Croatian *rvati se* =Slovenian *rvati* =Polish *rwac'* = Russian *rvat'* "tear"

Therefore, all Slavic #RT-words must share a property amounting to Common Slavic, and which made them become of the #RT-kind. This common feature is well known: all words mentioned bore a yer within today's #R and T in Old Church Slavonic, i.e. *lUgati, rUtI, rUvati* (where "I" is the soft, and "U" the hard yer). Yers fell out and thereby created the incriminated #RT-clusters.

This fact also explains the inordinate rareness of #RT-words in Slavic: Common Slavic had 11 different vowels that may be assumed to have enjoyed an identical distribution in the environment #C__C. Only two of them disappeared, thus creating only a minority of #RT-items.

By contrast, a Classical Arabic verb was of the shape #CaCVC-. In the evolution towards Moroccan Arabic, ALL short vowels were lost in verbal forms regardless of their timbre (except if there were two short vowels in a row, in which case the second one is retained as a schwa), to the effect that #CaCVC- comes out as #CC@C ("@"=schwa), e.g. *katab-a* > kt@b "he has written". Thus, the sonority slope of #CC in modern verbs is free since the distribution of the first two consonants in the ancestor #CaCVC- was free. Just as many #RT-

as #TR-clusters are expected, and this is in line with fact.

Hence, the difference in number of #RT-words in Slavic and Afro-Asiatic turns out to be a simple reflect of particular diachronic events, that is loss of all word-initial vowels in the latter, but only of two vowels out of 11 in the former. The fact of losing initial vowels, in turn, provoked the diachronic change appealed to before: as the first vowel was unable to govern the initial CV (="#"), this initial CV had to disappear.

This view also makes a prediction on non-existing initial clusters in Slavic: namely, these gaps must be accidental. In Moroccan Arabic, any logically possible #CC and its mirror image is attested. In Slavic, on the other hand, only a small subset of those combinations of two consonants that are logically possible do actually occur word-initially. For instance, #ml and #mn exist (e.g. Czech *mlýn, mnich* "mill, monk"), but #lm and #nm do not occur. It has been tried for a long time to discover the principle that defines existing and non-existing initial clusters, e.g. Kurylowicz (1952), Gussmann & Cyran (1998) for Polish. No such thing could be identified.

If on the other hand the observed gaps are not systematic but accidental, #lm and #nm do not occur simply because there happened to be no word in Common Slavic with a structure #lIm, #nIm (where "I" stands for yers). And if, in turn, modern Slavic languages lost the initial CV, actual grammar does not prohibit words with #lm, #nm. They are well-formed and could well enter the system as loans, acronyms or new creations. I have tested this prediction on putative acronyms and nonsense-words with native speakers from Polish, Czech, Slovak and Serbo-Croatian, and it was not falsified.

In sum, this presentation makes one general claim, which is instantiated by an analysis of word-initial clusters. The claim is that it may be interesting to conceive of morphological boundaries as objects with a syllabic materialization.

The ensuing view on initial clusters opens the possibility of setting a parameter (presence versus absence of "#"=CV) that is able not to release any of the devices that are involved in the analysis of regular only-#TR-languages when facing Afro-Asiatic and Slavic.

It also derives properly all and only the configurations allowed for by natural language. That is, either a language possesses the initial "#"=CV, in which case only #TR-clusters occur. Or it lacks the initial CV-unit, releasing thereby any co-occurrence restriction on word-initial consonant clusters, cf. Afro-Asiatic and Slavic. The third logical possibility whereby a language contains only #RT-words is ruled out, as shown in the discussion following (1).

References

Anderson, J.M. & C.J. Ewen 1987. *Principles of Dependency Phonology*. Cambridge: Cambridge University Press.

Charette, Monik 1991. *Conditions on Phonological Government*. Cambridge: CUP.

Chierchia, G. 1986. Length, syllabification and the phonological cycle in Italian. *Journal of Italian Linguistics* **8**.

Clements, G.N. 1990. The role of the sonority cycle in core syllabification. *Papers in Laboratory Phonology, I: Between the Grammar and Physics of Speech*, edited by J. Kingston & M. Beckman, 283–333. Cambridge: CUP.

Gussmann, E.& E. Cyran 1998. Polish consonantal sequences: a phonological testing ground. *Structure and interpretation: studies in Phonology*, edited by E. Cyran, 127-138. Lublin: PASE.

Kaye, J. 1990. 'Coda' licensing. *Phonology Yearbook* **7.2**, 301-330.

Kaye, J., J. Lowenstamm & J-R. Vergnaud 1985. The internal structure of phonological representations: A theory of Charm and Government. *Phonology Yearbook* **2**, 305–328.
Kaye, J., J. Lowenstamm & J-R. Vergnaud 1990. Constituent structure and government in phonology. *Phonology Yearbook* **7**, 193–231.
Kiparsky, P. 1982. From Cyclic Phonology to Lexical Phonology. In: *The structure of phonological representations I*, edited by Harry van der Hulst & Norval Smith, 131-175. Dordrecht: Foris.
Kiparsky, P. 1985. Some Consequences of Lexical Phonology. *Phonology Yearbook* **2**, 85-138.
Kurylowicz, J. 1952. Remarks on Polish consonantal groups. *Biuletyn Polskiego Towarzystwa Jezykoznawczego* **11**.
Lowenstamm, J. 1996. CV as the only syllable type: Full interpretability in phonology. *Current trends in Phonology: Models and Methods*, edited by J. Durand & B. Laks, 419–441. Salford: ESRI.
Lowenstamm, J. 1999. The beginning of the word. In: *Syllables?!*, edited by J. Rennison & K. Kühnhammer, 153-166. The Hague: Holland Academic Graphics.
Mohanan, K. 1986. *The Theory of Lexical Phonology*. Dordrecht: Reidel.
Schane, S.S. 1984. The fundamentals of Particle Phonology. *Phonology Yearbook* **1**, 129–155.
Scheer, T. 1998. A theory of consonantal interaction. *Folia Linguistica* **32**.
Selkirk, E. 1984. On the major class features and syllable theory. *Language Sound Structure*, edited by M. Aronoff & R.T. Oehrle, 107–136. Cambridge, Mass.: MIT Press.
Steriade, D. 1995. Underspecification and markedness. *The handbook of phonological theory*, edited by J. Goldsmith. Oxford: Blackwell.