Optimal Morphology

Jochen Trommer

Institut fuer Linguistik/Allgemeine Sprachwissenschaft
Universitaet Potsdam Postfach 601553 D-14415 Potsdam

Abstract

Optimal morphology (OM) is a finite state formalism that unifies concepts from Optimality Theory (OT, Prince & Smolensky, 1993) and Declarative Phonology (DP, Scobbie, Coleman & Bird, 1996) to describe morphophonological alternations in inflectional morphology. Candidate sets are formalized by inviolable lexical constraints which map abstract morpheme signatures to allomorphs. Phonology is implemented as violable rankable constraints selecting optimal candidates from these. Both types of constraints are realized by finite state transducers. Using phonological data from Albanian it is shown that given a finite state lexicalization of candidate outputs for word forms OM allows more natural analyses than unviolable finite state constraints do. Two possible evaluation strategies for OM grammars are considered: the global evaluation procedure from Ellison (1994) and a simple strategy of local constraint evaluation. While the OM-specific lexicalization of candidate sets allows straightforward generation and a simple method of morphological parsing even under global evaluation, local constraint evaluation is shown to be preferable empirically and to be formally more restrictive. The first point is illustrated by an account of directionality effects in some classical Mende data. A procedure is given that generates a finite state transducer simulating the effects of local constraint evaluation. Thus local as opposed to global evaluation (Frank & Satta, 1998) seems to guarantee the finite-stateness of the input-output-mapping.

1 C/Ø-Alternations in Albanian Verbs

In many cases C/Ø-alternations in the Albanian verb paradigm are triggered by the tendency to avoid sequences of vowels (VV, hiatus) or consonants (CC), e.g.

(1)

(a)	(b)	(c)	(d)
pije	**pie*	*hape*	**hapje*
CVCV	CVV	CVCV	CVCCV
'drink it!'		'open it!'	

(b) and (d) seem to be out because they contain illicit sequences of VV and CC respectively[1]. For illustration of a declarative account I implement the morphological candidate set for each form and the phonological constraint as regular expressions The corrrect forms are then obtained by intersecting the two expressions:

(2)

Morphology:[2](pi|hap)(j)e

Phonology:

(CV)*, for C = (p|h|j) and V = (i|a|e)

[1] Alternatively these alternations can be viewed as a strategy for arriving at "perfect" syllable structures, i.e. syllables with onsets and without codas. This probably more adequate view can be formalized using the same methods, but for expositional ease I will ignore syllable structure and assume the simpler analysis above.

[2] A more adequate morphological characterization follows in (2')

Intersection(Morphology, Phonology):

(pije|hape)

Intriguing as this simple model is; it runs into difficulties with occuring CC and VV sequences in Albanian verbs:

(3)

(a)	(b)
bie	*hapje*
CVV	CVCCV
'he falls'	'you opened'

Since *hapje* is ungrammatical as V+OBJ but grammatical as V+IMF:2SG there seems no purely phonological way to predict the observed contrasts.By lexicalizing the alternations though, say by stating that OBJ is *-je* after vowels and *-e* after consonants we lose the insight that the process is motivated by the perfectly natural and crosslinguistically observable avoidance of of CC and VV (cf. Prince & Smolensky, 1993). The solution I propose to this dilemma is the following:

(4)

a. Replace the CV-constraint by a soft constraint marking CC and VV as constraint violations.
b. Annotate the candidate set with independently needed morphosemantic interpretation and choose as the correct form for certain morphological features (e.g.PI+OBJ) the phonologically optimal form annotated with it. [3]

More concretely (2) is replaced by (2'):

[3] The idea to use optimality to decide between allomorphs with respect to an morphological characterization can also be found in Kager (1995) See Russel (1995) for some discussion of morpheme signatures.

(2')

Morphology:

$$\begin{pmatrix} PI \mid HAP \\ pi \mid hap \end{pmatrix} \begin{pmatrix} OBJ \mid OBJ \mid IMF \\ e \mid je \mid je \end{pmatrix}$$

Phonology:

$$\begin{pmatrix} C\ C\ * \\ 0\ 1 \end{pmatrix}^{?} \left(\begin{pmatrix} V\ V\ * \\ 0\ 1 \end{pmatrix} \begin{pmatrix} C\ C\ * \\ 0\ 1 \end{pmatrix} \right)^{*} \begin{pmatrix} V\ V\ * \\ 0\ 1 \end{pmatrix}^{?}$$

The phonological constraint following Ellison (1994) has the form of a regular relation mapping phonological strings into sequences of constraint violation marks ('1') and '0's which stand for no violation. The effect of the constraint can best be seen in tableau form familiar from OT:

(5)

PI+OBJ	*CC/VV
pije	
pie	*!
HAP+OBJ	
hape	
hapje	*!
PI+IMF	
pije	
HAP+IMF	
hapje	*

Differently from OT optimal forms are not computed with respect to an underlying phonological form, but directly referring to a morphological characterization of the generated word form.

2 Formalism

An OM grammar is a quintuple {MA, PA,M,P,{0,1}} where MA and PA are sets of symbols, the morphological and the phonological alphabets. M is a finite set of regular relations, each mapping MA into PA, while P is a finite sequence of regular relations mapping MP into {0, 1}.

3 Generation

Specific word forms are characterized as strings of abstract morphems, e.g. PI+OBJ[4]. Specific candidate sets are obtained by the crossover product of word forms[5] with the candidate relation. For ease of exposition I give the algorithm for finite state automata and transducers which are formally equivalent to regular expressions and binary regular relations respectively. For a transducer T and an automaton A the crossover product AxT is generated in the following way. I_T and I_A are the initial, F_T and F_A the final states:

(6) **Crossover product(A,T)**[6]

make (I_A,I_T) initial in AxT
make (F_A,FT) final in AxT
for each arc from z to t in T labeled M_2/P
  for each arc from x to y in A labeled M_1
    if $M_1 = M_2$
      then add to AxT an arc
      from (x,z) to (y,t) labeled P.

Obviously, the resulting automaton contains all and only the phonological output candidates for the given morphological input. (7) shows the application of the algorithm to PI+OBJ and the candidate set from (2'):

(7)

Input

Initial State:	0
Final State:	2
Transitions:	(0, 1, PI), (1, 2, OBJ)

Morphology

Initial State: A
Final State: C
Transitions:
(A, B, PI/pi), (A, B, HAP/hap),
(B, C, OBJ/je),(B, C, OBJ/e),
(B, C, IMF/je)

Resulting Candidates

Initial State: (0, A)
Final State: (2, C)
Transitions:
((0, A), (1, B) pi), ((1, B), (2, C), e)
((1, B), (2, C), je)

Since the candidate set and the constraint in (2') are regular Ellisons (1994) algorithms for getting an automaton containing only optimal candidates, as long as candidate set and evaluation transducer are regular, can be used. For details I refer the interested reader to Ellisons paper.

4 Parsing

The candidate transducer constitutes a kind of backbone for parsing phonological strings into strings of abstract morphemes. For example the candidate transducer in (2') will allow correct parsing of *hape* into HAP+OBJ. A complication arises when the transducer maps phonological forms to more than one morphological form, possibly including incorrect ones. E.g. the example transducer will map *hapje* onto (correct) HAP+IMF and (incorrect) HAP+OBJ. Then given the generation procedure above for every obtained morphological string it can be checked if it generates the input phonological form. A special case of this are not existing word forms. For example we will get PI+OBJ as the only possible parse for *pie*. But since the optimal output for PI+OBJ isn't *pie* but *pije* there is no actual parse.

[4] For sake of readability concatenation of morphological symbols is represented by '+'.

[5] Strings are a trivial case of regular expressions.

[6] Crossover product(A, T) is equivalent to the image of A under T (Kaplan & Kay, 1994:340-42), defined as the range of the composition Id(A) o R, where Id(A) is the identity relation that carries every member of A into itself. See Frank & Satta (1998:5-6) for the same concepts under different terminology.

5 Comparison with other approaches

Optimal Morphology clearly is a hybrid. Candidate sets are given by lexicalized monotonic constraints as in DP. Phonological constraints on the other side are violable and ranked as in OT[7].

Looking at constraint based formalisms as tentatives to restrict the rather rich theoretical inventory of generative SPE theory (Chomsky & Halle, 1968) it becomes obvious that OM indeed fills a conceptual gap:

(8)

	SPE	OT	DP	OM
arbitrary rule/ constraint order	yes	yes	no	yes
language specific rules/constraints	yes	no	yes	yes
underlying representations	yes	yes	no	no

Neglecting here the difference between rules and constraints, OT has chosen to eliminate language specific constraints[8] while maintaining underlying representations. It is not a priori clear that this is favorable to eliminating arbitrary rule (constraint) ordering or underlying representations, and indeed this last choice is what happens in OM, while DP has eliminated two of SPEs theoretical instruments. In this respect it is the most restrictive framework, but we have seen empirical and conceptual problems with its assumptions in the preceeding section. Consider the following: Lexicalized candidate sets in both DP and OM are given by language-specific constraints, thus this point makes no difference. Now, when the toy example above proves to be representative, OM (like OT) allows to maintain universal constraints where DP cannot.

[7]See section 6 for an example of ranking.

[8]Note that the OT-claim to use only language-specific constraints is crucially weakened by the family of alignment-constraints that can be instantiated in a language-specific way.

6 Regular phonological processes

It might seem that OM can handle regular phonological processes only by stipulation of allomorphs with a high degree of redundancy. Thus for the case of German final devoicing we would have to assume two allomorphs for each morpheme showing the alternation, e.g. {Tod, Tot}[9] as in *Tot*, 'death(sig)' and *Tod-e*,'death-plu'. The choice between these allomorphs could be accomplished by the two constraints !DEVOICE that marks voiced obstruents in coda position and !VOICE that marks unvoiced obstruents generally. The ranking[10] !DEVOICE $\gg$!VOICE will make 'Tot'- the optimal candidate in *Tot* and 'Tod-' in *Tode*. *Tot*, 'dead' with an "underlying" t will remain the same in every context, assuming that this is the only allomorph. Though this works technically it seems highly undesirable to have two allomorphs for *Tod* one of which can be predicted.But the regular expression (Tod | Tot) can equally well be written as To(d|t) or even as To[-continuant -sonorant +coronal] since regular languages can be enriched straightforwardly by bundles of finite-valued features (Kaplan & Kay, 1994:349-351). Thus allomorphy in this and comparable cases reduces to underspecification.[11]

7 Locality of Constraint Evaluation

Ellisons (1994) algorithms offer a way of globally finding optimal candidates out of regular candidate sets. This however is not the

[9]German orthography is used except for the phonemic writing of voiceless 't')

[10]Ranking can be implemented along the lines of Ellison (1994).

[11]It might be argued that this move doesn't explain the nonexistence of voiced Coda-obstruents in German, since nonalternating voiced obstruents could be accomplished by fully specifying them as voiced in morphology. But there are languages like Turkish (Inkelas:1994:3), where certain morphemes resist otherwise regular final devoicing and this descriptive possibility thus seems to be well motivated.

only possible method to generate word forms given a particular OM-grammar. In the example above the optimal word form can also be found in a much more local way: by traversing at once the candidate automaton and the evaluation transducer and choosing at each state the '0'-transition, when possible. The reader can check this by constructing finite state automata out of the regular relations in (2'). Though this is a case of extreme locality, probably not representative for phonological phenomena in general, it seems promising to study, how far constraint evaluation can work locally. An argument that constraint evaluation not only can but SHOULD happen locally can be constructed out of the following well known data from Mende[12]:

(9)

H:	*kó*	'war'	*pélé*	'house'
L:	*kpà*	'debt'	*bèlè*	'trousers
HL:	*mbû*	'owl'	*ngílà*	'dog'
LH:	*mbǎ*	'rice'	*fàndé*	'cotton'
LHL:	*mbã*	'companion'	*nyàhâ*	woman'

H:	*háwámá*	'waistline'
L:	*kpàkàlì*	'tripod chair'
HL:	*félàmà*	'junction'
LH:	*ndàvúlá*	'sling'
LHL:	*nìkílì*	'groundnut'

In Mende nouns have one of five specific tone patterns (indicated in the left column). Note that atoms of tone patterns like L in HL can be realized as sequences of high tone syllables as in *félàmà* or as part of contour tones[13] as in *mbû*. Hence for *mbu*, *nyaha* and *felama*, there are the following logical possibilities of tonal realization:

(10)[14]

mbu:	*mbû*
felama:	*fêlàmà, félámà,*
	félàmà, félâmà, félámâ
nyaha:	*nyǎhà, nyàhâ, nyǎhâ*

The problem of selecting out of these the correct for is solved by Goldsmith (1976) through the following mapping procedure:

(11)

Tone mapping

a. Associate the first tone with the first syllable, the second tone with the second syllable and so on.

b. Tones or syllables, not associated as a result of (a) are subject to the wellformedness condition.

Well-Formedness Condition

a. Every tone is associated with some syllable.
b. Every syllable is associated with some tone.
c. Association lines may not cross.

This mapping procedure is (apart from its use in a wealth of other phenomena) stipulation. I will sketch how the same effects can be derived from plausible constraints on phonological wellformedness and locality of constraint evaluation.

Let's suppose that the candidate sets in (10) are given by lexical constraints realized as transducers[15]. It's natural to assume that contour tones are a marked case in human language, violating the constraint *Contour and that there's a ban on adjacent syllables with the same tone,

[12]The ideas presented here owe much to the concept of incremental optimization developed in Walther (1997). Locality in Constraint evaluation is also invoked in Tesar (1995), but there it isn't conceived as empirically different from global evaluation. For an interesting hybrid of directional rule application/constraint evaluation see Kager (1993)

[13]Contour tones are the falling tone (*mbû*) the rising tone(*mbǎ*) and the falling rising tone (*mbã*). Acute stands for H, grave for L tone. As Leben, I analyze contours as sequences of Hs and Ls.

[14]In Goldsmiths system these possibilities would correspond to the possible noncrossing exhaustive associations of the stems with their respective tone pattern. For details see Trommer (1998).

[15]For an implementation see Trommer (1998).

an instance of the Obligatory Contour Principle (!OCP(Tone)).

(12)

***Contour**[16]

$$\begin{pmatrix} C \mid S \\ 1 \mid 0 \end{pmatrix}^{*}$$

,where C stands for contour tones and S for simple tones.

!OCP(Tone)

$$\begin{pmatrix} L\ L^{*} \\ 0\ 1 \end{pmatrix}^{?} \left(\begin{pmatrix} H\ H^{*} \\ 0\ 1 \end{pmatrix} \begin{pmatrix} L\ L^{*} \\ 0\ 1 \end{pmatrix} \right)^{*} \begin{pmatrix} H\ H^{*} \\ 0\ 1 \end{pmatrix}^{?}$$

, where L = Low Tone, H = High Tone

mbû violates *Contour, but being the only candidate for *mbu* its optimal. For *felama* *fêlàmà* and *félámâ* will be excluded since they violate both constraints and the other candidates only one. But[17] *félámà*, *félàmà* and *félâmà* for that reason are all on a par. *Nyǎhâ* is out because it violates two times *Contour, but there's no way to decide between *nyǎhà* and *nyàhâ* that violate it only once.

However once locality of constraint evaluation is taken seriously only the correct candidates remain: Suppose we are following the candidate automaton for felama, for '*e*' '*é*' will be chosen since the word has to start with H and a contour tone would violate *Contour. Next, '*à*' will be selected since '*á*' violates !OCP(Tone) and '*â*' again *Contour. For the same reason '*á*' would be best for the next '*a*'. But now according to the unviolable lexical constraint that requires the tone pattern for this noun, only '*â*' is possible, even if it violates !OCP(Tone). More or less the same story can be told about *nyaha*. Hence the left to right-assymmetry in the mapping convention emerges out of locality or, put in a more placative way, the tendency to choose marked options as late as possible. Note that the use of violable constraints is necessary in this account to get contour tones and tone plateaus at all, but only when necessary.For example an unviolable constraint prohibiting contours would exclude correct forms as *mbû* and *nyàhâ*.

[16] Obviously H and L in (12) have to be spelled out segmentally, i.a. in their shape as contour tones. In Trommer (1998) I present a procedure to convert constraints on tones in constraints on corresponding segments. A more autosegmental treatment following the lines of Eisner (1997) is also considered there and shown to be compatible with the results presented here. Consonants are ommitted for ease of exposition.

[17] For simplicity I ignore possible results out of ranking the constraints and treat them like one complex constraint

8 Locality and Directionality

In Mende tone patterns marked (tonal) structure appears as far to the right as possible. Although this seems to be the unmarked case in tone languages the mirror image too seems to exist (e.g. Kannakuru, Odden, 1995:460). To describe such cases it's sufficient to use candidate automata that are interpreted from right to left.

Directionality effects like the ones found in tone mapping are also typical in many other phonological domains, e.g. foot construction (Hayes, 1985; Kager, 1995), root and pattern morphology (McCarthy, 1979), and syllabification (Îto 1988). Even if a locality-based reanalysis isn't possible in all of these cases traditionally handled by derivational mapping and parsing devices, this is a promising area for further research on locality and markedness.

9 Locality and Generative Power

Frank & Satta (1998) show that input-output mapping in OT isn't regular in the general case even if GEN and the constraints are implemented as finite state transducers. This result carries over immediately to OM under global constraint evaluation. However if local evaluation is used there seems to be an algorithm to construct an equivalent finite state transducer. Since the complete algorithm requires a more elaborate representation of morpheme structure than I can give here I illustrate only the basic idea.[18] Let's first consider a slightly modified version of (2') and the corresponding transducers[19]:

(2")

Morphology:

$\begin{pmatrix} 11 \mid 222 \\ pi \mid hap \end{pmatrix} \begin{pmatrix} 3 \mid 33 \\ e \mid je \end{pmatrix}$

As Transducer:

Initial State: 0
Final State: 6
Transitions:
$(0,1,h_1)$, $(1,3,a_1)$, $(3,4,p_1)$, $(0,2,p_2)$, $(2,4,i_2)$,$(4,5, j_3)$,$(5,6, e_3)$, $(4,6,e_3)$

Phonology:

Initial State: A
Final States: C,V
Transitions:
(A, C, C/0), (V, C, C/0),(C, C, C/0)
(A, V, V/0), (C, V, V/0),(V, V, V/0)

For convenience abstract morphemes are represented by numbers that are repeated for each segment realizing the respective morpheme. The following algorithm yields from a candidate transducer and an evaluation transducer a transducer with three tapes integrating candidate generation and their mapping to markedness values. The first tape of this new transducer ($Kx_{23}C$) corresponds to the candidate set, the second one to the morphological signatures and the third one to the evaluation under the constraint:

[18]Problems for the described method arise with recursivity since '33' mapped to 'aa' is ambiguous between two a-morpemes and one aa-morpheme. Further states that can be reached and left by transitions of the same morphological index would need a seperate treatment in the procedure described on p. 9.

[19]the regular expression for the phonological part is exactly as in (2').

(13) **23-product(A,O)**[20]

make (I_A,I_O) initial in $Ax_{23}O$
make (F_A,F_C) final in $Ax_{23}O$
for each arc from z to t in A labeled M/P1
  for each arc from x to y in O labeled P2/N
    if P1 = P2
      then add to $Ax_{23}O$ an arc
      from (x,z) to (y,t) labeled M/P1/N.

Applied to the transducers in (2") we get:

(14) **23-product(Morphology,Phonology)**

Initial State: A_0
Final State: V_6
Transitions:
$(A_0,C_1,h_1/0)$, $(A_0,C_2,p_2/0)$, $(C_1,V_3,a_1/0)$, $(C_2,V_4,i_2/0)$, $(V_3, C_4, p_1/0)$, $(C_4, C_5,j_3/1)$, $(C_4, V_6, e_3/0)$, $(V_4,C_5,j_3/0)$, $(V_4,V_6, e_3/1)$, $(C_5,V_6, e_3/0)$

Again we can find the optimal output for a certain morphological input in a local way. E.g. traversing (14) and the automaton corresponding to '1+3+' (HAP+OBJ) at once we will arrive without choice at , C_4 ('111'). There we choose locally to go to V_6 over $e3_0$ since this is less marked than the transition over $j_3/1$ to C_5. We get 1113/hape.

Alternatively we can also dispense with the evaluation tape altogether: We delete each transition from every state S if there is another transition from S with the same morphemic index and a smaller number on the third tape. For (14) this means that we remove the transition from V_4 to V_6 over $e_3/1$ since there is a "better" transition namely over $j_3/0$ to C_5. Similarly the transition from C_4 to C_5 over $j_3/1$ is eliminated in favour of the transition over $e_3/0$ to V_6. Since for each state and every index there remains only one transition the third tape becomes superfluous and is removed. The result is an input/output transducer mapping '1113' (i.e. HAP+OBJ) to hape and '2233' (i.e. PI+OBJ) to pije:

[20]A is mnemonic for candidate, O for constraint transducer. I_A, I_O and F_A, F_O are the initial and final states of A and O respectively

(15)

Initial State: A_0
Final State: V_6
Transitions:
(A_0,C_1,h_1), (A_0,C_2,p_2), (C_1,V_3,a_1), (C_2,V_4,i_2), (V_3, C_4, p_1), (C_4, V_6, e_3)
(V_4,C_5,j_3), (C_5,V_6,e_3)

10 Final Remarks and Conclusions

It's quite probable that locality as developed in the preceeding three sections is too simple to account for phonological data that require evaluation of segments at a certain distance from each other. But the point was to show that locality of constraint evaluation in the introduced framework even in this form is empirically supported and preferable on theoretical grounds. A promising extension would be to evaluate locally in multiple phonological domains using autosegmental representations along the lines of Eisner (1997), but the technical realization of this still has to be worked out. As the Albanian data the tonal patterns in Mende reveal the advantages of using OM's violable constraints in the context of lexicalized candidate sets. On the other hand this lexicalization allows a simple generation procedure. Parsing phonological output forms onto morphological signatures in OM is relatively straightforward while the question is not even adressed seriously in finite state formalizations of OT. (Ellison, 1994; Eisner, 1997). Both parsing and generation are even simpler if OM is interpreted as a finite state transducer under local constraint evaluation. It remains to be seen, if the specific mixture of OM borrowing from DP AND OT will give rise to further linguistically worthwhile analyses and to efficient computation of morphophonology.

References

CHOMSKY, N. & Halle, M. (1968) The Sound Pattern of English. New York: Harper & Row.

EISNER, J. (1997) Efficient generation in Primitive Optimality Theory. Ms., University of Pensylvania.

ELLISON, M. (1994) Phonological Derivation in OT. In: COLING'94 Vol. II, 1007-1013, Kyoto, Japan.

FRANK, R. & SATTA, G. (1998) Optimality theory and the generative complexity of constraint violability. Computational Linguistics (forthcoming).

GOLDSMITH, J.A. (1976) Autosegmental Phonology. Doctoral Dissertation. Massachussets Institute of Technology.

HAYES,B. (1985) Iambic and trochaic rhythm in stress rules. BLS 11, 429-446.

ÎTO, J. (1988) Syllable Theory in Prosodic Phonology, Garland Publishing, New York.

KAGER, R. (1993) Alternatives to the Iambic-Trochaic Law. Linguistic Inquiry 11, 381-432.

KAGER, R. (1995) On Affix Allomorphy and Syllable Counting. Paper presented at Conference on Interfaces in Phonology, Berlin.

KAPLAN, R. & KAY, M. (1994) Regular models of phonological rule systems. Computational Linguistics, 20(3), 331-378.

LEBEN, W. R. (1978) The Representation of Tone. In: FROMKIN, V.A. (ED.) Tone - A Linguistic Survey, 177- 219. Accademic Press.

ODDEN, D. (1995) Tone: African Languages. In: Goldsmith, J. (ed.) The Handbook of Phonological Theory, 444-475. Oxford: Basil Blackwell.

PRINCE, A. & SMOLENSKY, P. (1993) Optimality Theory: Constraint Interaction in Generative Grammar. RuCCs TR-2, Rutgers University.

RUSSEL, K. (1995) Morphemes and candidates in Optimality Theory. Ms., University of Manitoba.

SCOBBIE, J.M., COLEMAN, J.S. & BIRD, S. (1996) Key Aspects of Declarative Phonology. In: DURAND, J. & LAKS, B. (Eds.) Current Trends in Phonology: Models and Methods (= Proceedings of the Royaumont meeting 1995), vol. II, 685-710. European Studies Research Institute, University of Salford Publications.

TESAR, B. (1995) Computational Optimality Theory. PH. D. thesis, University of Colorado, Boulder, CO.

TROMMER, J. (1998) Mende tone patterns revisited: Two finite state implementations. Ms.,University of Potsdam.

WALTHER, M. (1997) Deklarative Prosodische Morphologie. Doctoral Dissertation, Universität Düsseldorf.