An OT-CC Analysis of Opacity in Somali

Deborah Cunningham, Graduate Institute of Applied Linguistics student

Abstract: Opaque interactions in Somali are difficult, if not impossible, to explain within Classic Optimality Theory. Underapplication of lenition (e.g. /ba**d**-**t**a/ 'sea' → [ba.**d**a], *[ba.**ð**a]) and overapplication of epenthesis (e.g. /gabḍ-ta/ 'girl' → [ga.β**a**.ḍa] *[gab.ḍa]), necessitate appealing to intermediate stages between the input and the output. I propose McCarthy's (2007) Optimality Theory with Candidate Chains (OT-CC) as a framework capable of explaining these surface forms. OT-CC utilizes precedence (PREC) constraints, which specify that faithfulness constraints must be violated in a particular order. For example, PREC(IDENT(CONT), MAX) specifies that if a surface form violates MAX, it must first violate IDENT(CONT). By ranking the precedence constraint above the leniting constraint, forms like [ba.da], which fail to undergo lenition, are correctly predicted to surface. Similarly, forms like [ga.βa.ḍa], which seem to unnecessarily epenthesize, are correctly evaluated as optimal because IDENT(CONT) is not violated before MAX.

1. Introduction

In Somali, opaque surface forms like [ba.da] 'sea, SG.DEF' resist analysis under the Classic OT[1] framework. In transparent cases, intervocalic obstruents must be fricatives, e.g. /ba**d**-o/ → [ba.**ð**o] 'sea, pl'. However, this process is blocked when the would-be-fricativized obstruent is a coronal adjacent to another coronal in the input, e.g. /ba**d**-**t**a/ → [ba.**d**a], *[ba.**ð**a]. Stating the exception in this way points to the problem: in Classic OT, constraints can only be violated by surface forms, yet forms like [ba.da] seem to require that EVAL appeal to something other than the input or the output.

McCarthy (2007) offers a solution to this problem in the form of Optimality Theory with Candidate Chains (OT-CC). Rather than looking solely at underlying and surface forms, OT-CC recognizes intermediate steps that indicate the particular path of unfaithfulness that an input form takes to become a more harmonic output. These ordered unfaithful mappings either satisfy or violate Precedence (PREC) constraints, which specify which patterns of unfaithfulness are preferred in a particular language. Because of the innovation of PREC constraints, OT-CC is able to maintain a single, consistent constraint hierarchy throughout. In this way OT-CC provides a framework that is reasonably restricted, and yet robust enough to handle the interplay of multiple opaque rankings, such as those found in Somali.

In Section 2, I give a brief overview of OT-CC and then apply the theory to transparent Somali data in Section 3. In Section 4, I look at opaque forms in the classic framework and then in OT-CC. I show that the opaque data can be successfully explained in OT-CC by ranking the constraint PREC(ID(CONT), MAX) above the markedness constraint LENITION.

[1] "Classic OT" refers to Optimality Theory as originally framed in Prince and Smolensky (1993/2004), and modified by McCarthy and Prince's (1995) Correspondence Theory.

2. A Brief Overview of OT-CC[2]

2.1 Candidate Chains

In OT-CC, the concept of “candidate” is augmented to include more than just the output. Specifically, each candidate chain consists of an input, an output, and the gradual intermediate steps required to map from one to the other. For a chain to be considered valid, it must adhere to three restrictions: 1) fully faithful input, 2) gradual unfaithfulness, and 3) harmonic improvement.

2.1.1 Fully faithful input

Every valid chain must begin with the most harmonic, fully faithful parse of a given underlying form. “Fully faithful” here means that the input must not violate any faithfulness constraints. “Most harmonic” is evaluated based on the markedness constraints of the language and primarily refers to syllabification, which can be altered without affecting faithfulness. For example, given the underlying form /bad-ta/, the most harmonic, fully faithful input in Somali is <bad.ta>. This is more harmonic than <ba.dta> because Somali prefers simple codas to complex onsets.[3] However, in a language with the reverse ranking, the latter parsing might be preferred. One implication of the fully faithful input requirement is that every valid candidate chain will begin with the same input.

2.1.2 Gradual unfaithfulness

For a chain to be valid, each intermediate step between the input and the output must violate only one basic faithfulness constraint[4] in a specific location. These violations are called localized unfaithful mappings (LUMs) and are accumulated with each successive member of the chain. For example, while <bad.ta, ba.da, ba.ða> could be a valid chain since only one LUM is acquired at a time (i.e. a violation of MAX followed by a violation of ID(CONT)), <bad.ta, ba.ða> could not, because more than one basic faithfulness constraint must be violated to get from the first member of the chain to the second (i.e. MAX and ID(CONT) must be violated simultaneously).

2.1.3 Harmonic improvement

Finally, each member of a chain must be more harmonic than its predecessor. Harmony is achieved by satisfying a markedness constraint that outranks whichever faithfulness constraint is being violated. For example, the chain <ba.do, ba.ðo> is improving harmonically *iff* a markedness constraint requiring intervocalic obstruents to be fricatives dominates a faithfulness constraint such as IDENT(CONT). One way to think about this is to imagine each member of a

---

[2]The following discussion of OT-CC is, by necessity, somewhat simplified. Although I omit some details of the theory, I do not think I contradict or deviate from the spirit of the framework in any way. See McCarthy (2007:59-99) for a much more thorough explanation of these concepts.

[3]Note that <ba.dta> would also violate the Sonority Sequencing Principle in addition to *COMPLEX.

[4]McCarthy (2007) defines the “basic faithfulness constraints” as MAX, DEP, and IDENT(*f*), allowing for the possibility of a few others. The primary requirement is that no member of the set of basic constraints should overlap with another.

given chain as competing candidates in a Classic OT tableau. Given the chain <bad.ta$_1$, ba.da$_2$, ba.ða$_3$>, 2 must be more harmonic than 1, and 3 must be more harmonic than both 1 and 2. If this is not the case, then the chain is invalid.

2.2 LUM Sequence

In addition to candidate chains, another important piece of information submitted to Eval is the LUM sequence (LUMseq). This is a set containing each of the LUMs of a given chain in the order they were accumulated. The LUMseq for the chain <b$_1$a$_2$d$_3$.t$_4$a$_5$, b$_1$a$_2$.d$_3$a$_4$, b$_1$a$_2$.ð$_3$a$_4$> is <Max@4, Id(cont)@3>.[5]

2.3 Prec Constraints

While output forms violate markedness constraints, and unfaithful mappings from input to output violate faithfulness constraints, precedence (Prec) constraints are violated by dispreferred LUM sequences. A Prec constraint is a set containing two crucially ordered faithfulness constraints, e.g. Prec(Max, Dep). This constraint states that Max must be violated before Dep. Prec(Max, Dep) can be violated in one of two ways. First, if Dep is violated and Max is never violated, then Prec(Max, Dep) has been violated once. Second, if Dep is violated and then Max is subsequently violated, then Prec(Max, Dep) has been violated twice.

3. Application to Transparent Data

3.1 Preliminary Constraints

Most of the data[6] to be analyzed can be accounted for by the six constraints listed below in Figure 1. The first three are markedness constraints and the following three are faithfulness constraints.

Figure 1: Preliminary Constraints Defined

Name:	Abbreviation[7]:	Application:	Source:
*Complex	*Comp	assign one violation mark for every syllable position node (whether onset, nucleus, or coda) associated with more than one segment; i.e. *CCV, *VV[8], *VCC	Prince & Smolensky 2004

[5]The @index notation may be omitted if it is not needed to distinguish between two similar LUM sequences. I will not use it in the tableaux that follow.
[6]All Somali data is taken from Kenstowicz and Kisseberth (1979).
[7]These abbreviations were not necessarily posited by the original authors of these constraints; I am simply employing them for the purpose of economy.
[8]Saeed (1982:8) notes that long vowels count as single vowels in Somali. That is, a word like [daar] ‘house’ would be interpreted CVC and would not violate *Comp, but a hypothetical word [dair] would be interpreted as CVVC and would violate this constraint.

LENITION	LEN	assign one violation mark for every intervocalic obstruent that is [-cont]; i.e. *VC[-son][-cont]V	Kennedy 2008[9]
OCP$_{COR[-SON][\alpha CONT]}$	OCP	assign one violation mark for every coronal obstruent adjacent[10] to another coronal obstruent with the same value for the feature [±cont]; e.g. *dt	Coetzee & Pater 2005
DEP		assign one violation mark for every segment in the output without a corresponding segment in the input	McCarthy and Prince 1995
IDENT(CONT)	ID(CONT)	assign one violation mark for every segment in the output that has a different value for the feature [±cont] than its corresponding segment in the input	Kenstowicz & Banksira 1999
MAX		assign one violation mark for every segment in the input without a corresponding segment in the output	McCarthy and Prince 1995

3.2 Preliminary rankings

Before addressing the opaque data, we can begin to form the basic constraint hierarchy based on the transparent data. The following combination tableau[11] gives evidence that *COMP and MAX both dominate DEP. Both *COMP and MAX favor the winning candidate, so they must be ranked above DEP which favors both losers. *COMP and MAX cannot be ranked with respect to each other since they do not come into conflict.

Tableau 1: *COMP, MAX » DEP

	Input: /nirg/ 'baby female camel'	*COMP	MAX	DEP
a.	→ni.rig			*
b.	nirg	*! W		L
c.	nir		*! W	L

[9]Kirchner's (2000) constraint LAZY could also be used here, as it serves a similar function and was proposed earlier. However, since it is only fully satisfied when articulatory gestures are reduced to Ø, it would unnecessarily complicate this analysis, requiring many more high-ranking faithfulness constraints to ensure the correct output. Hence, LENITION is preferable in this case since it specifically targets intervocalic obstruents that are [-cont].

[10]When this constraint was originally proposed (Coetzee & Pater 2005), it was in the context of triconsonantal roots in Arabic, so the definition of adjacency crucially relied on ignoring intervening vowels. I adopt the same definition for this analysis, so that hypothetical candidates [bad-ta] and [badata] would violate this constraint equally.

[11]"Classic OT ranking arguments…are a legitimate shortcut in OT-CC when they deal with the effect of only one process at a time and they include the most harmonic faithful candidate" (McCarthy 2007:110).

Tableau 2 establishes that ID(CONT) must be dominated by both MAX and LENITION since the former favors the losers and the latter two favor the winner. In other words, in Somali it is better for an obstruent to be unfaithful to its input value [-cont] than to remain [-cont] intervocalically.

Tableau 2: MAX, LENITION » IDENT(CONT)

	Input: /bad-o/ 'sea-PL'	MAX	LEN	ID(CONT)
a.	→ ba.ðo			*
b.	ba.do		*! W	L
c.	bad	*! W		L

4. Application to Opaque Data

4.1 The Opacity Problem in Classic OT

In trying to rank OCP and MAX, we get our first glimpse of the problem opacity poses for the Classic model. Since the [t] in /bad-ta/ does not surface, it seems intuitive to rank OCP above MAX. However, this ranking cannot be established because the paradoxical tableau below does not actually allow any candidate to be evaluated as optimal. With regard to the three constraints that have not been previously ranked (i.e. OCP, MAX, and LEN) candidate (c) harmonically bounds candidate (a), so the real output cannot win no matter how these constraints are ranked. If the intended winner has no hope of winning, then the necessary ingredients for establishing a ranking between OCP and MAX are not present (i.e. a direct conflict between the winner and a loser). The only alternative would be to reverse the ranking of LEN and ID(CONT), which has already been established on the basis of transparent [ba.ðo].

Tableau 3: Rankings between OCP, MAX, and LEN cannot be established

	Input: /bad-ta/ 'sea-SG.DEF'	OCP	MAX	LEN	ID(CONT)
a.	(real output) ba.da		*!	*!	
b.	*bad.ta	*!			
c.	*ba.ða		*!		*

4.2 Opacity Explained in OT-CC

Tableau 4 expands the candidate set from Tableau 3 to include candidate chains and their respective LUMseq. Furthermore, the constraint PREC(ID(CONT), MAX) has been ranked above LEN (i.e. the real output's fatal violation[12]) to insure that the proper candidate is successful.

Tableau 4: OCP » MAX » PREC(ID(CONT), MAX) » LEN » ID(CONT)

	Input: /bad-ta/	OCP	MAX	PREC(ID(CONT), MAX)	LEN	ID(CONT)
a.	→ <bad.ta, ba.da> <MAX>		*	*	*	
b.	<bad.ta, ba.da, ba.ða> <MAX, ID(CONT)>		*	**! W	L	*
c.	<bad.ta> < >	*! W	L			

The transparent candidate (b) violates the PREC constraint twice since it violates MAX before ID(CONT). Candidate (a) only violates it once and so is permitted to win. Since this tableau produces the correct winner, it may be used to establish the ranking OCP » MAX. The PREC constraint is necessarily dominated by MAX due to a theory-internal meta-constraint which states that B » PREC(A,B) (see McCarthy 2007:98-99). To summarize, the following direct rankings have been established: OCP » MAX » PREC(ID(CONT), MAX) » LEN » ID(CONT). In addition, *COMP remains undominated, and *COMP, MAX » DEP, though DEP's ranking with regard to the other constraints has not yet been determined due to lack of situations of direct conflict. This will be partially remedied below in Tableau 5.

The following summary tableau incorporates the undominated constraint *COMP, as well as ordering DEP below the PREC constraint, since these two conflict between candidates (a) and (b). Although PREC must dominate DEP, LEN, and ID(CONT), DEP's relationship with LEN and ID(CONT) is still uncertain. I have placed it on par with LEN, dominating ID(CONT), but this could be verified or countered by further data.

[12] Although the paradoxical tableau shows the intended winner incurring a fatal violation at the hands of either MAX or LEN, it is clear from comparing candidate (a) to its harmonically bounding enemy (c) that LEN is the real problem.

Tableau 5: Summary Tableau

	Input: /gabḍ-ta/ 'girl-SG. DEF.'	Ocp	*Comp	Max	Prec(Id(cont), Max)	Dep	Len	Id(cont)
a.	→ <gabḍ.ta, ga.baḍ.ta, ga.βaḍ.ta, ga.βa.ḍa> <Dep, Id(cont), Max>			*		*	*	*
b.	<gabḍ.ta, gab.ḍa> <Max>			*	*! W	L		
c.	<gabḍ.ta, ga.baḍ.ta, ga.ba.ḍa, ga.βa.ḍa, ga.βa.ða> <Dep, Max, Id(cont)>			*	*!*	*		**
d.	<gabḍ.ta>[13] < >	*!	*!					

It is worth noting that the additional constraint Prec(Dep, Max) could more accurately be invoked to explain overapplication of epenthesis and eliminate candidate (b). This constraint would establish a precedence for the counterbleeding order of epenthesis before deletion. Since candidate (b) allows deletion to occur first, bleeding epenthesis, it would fatally violate Prec(Dep, Max). However, this additional constraint is not necessary to account for the data since candidate (b) equally violates Prec(Id(cont), Max).

5. Conclusion

OT-CC provides a framework that is able to account for the opaque interactions found in Somali. Within this framework, underapplication of lenition is analyzed as a consequence of the requirement that, in any given candidate chain, violation of Id(cont) must precede violation of Max. Overapplication of epenthesis is also explained by this precedence requirement, since it crucially states that Max cannot be violated first, preventing deletion from bleeding epenthesis. Some might argue that candidate chains unnecessarily complicate the analysis of transparent data. However, as Ashley et al. (2010) assert, "the ultimate test for evaluating a linguistic model is not what it looks like on paper or how much ink it takes to write up its formalisms. Rather, the ultimate test to evaluate a linguistic theory is what you can and cannot do with it." Since OT-CC's requirements for gradual unfaithfulness and harmonic improvement protect it from making bizarre, unattested predictions, its ability to provide an intuitive analysis of complicated opaque interactions is, in my opinion, well worth the extra ink.

[13]Note that the candidate <gab.ḍta> would receive the same evaluation as candidate (d) since *Comp prohibits complex onsets as well as complex codas.

References

Ashley, Karen, et al. 2010. How many constraints are there? A preliminary inventory of OT phonological constraints. Graduate Institute of Applied Linguistics Occasional Papers in Applied Linguistics No. 9. <http://www.gial.edu/opal.index.htm>.

Coetzee, Andries, and Joe Pater. 2005. Gradient phonotactics in Muna and Optimality Theory. University of Michigan & University of Massachusetts Amherst.

Kennedy, Robert. 2008. Evidence for morphoprosodic alignment in reduplication. *Linguistic Inquiry* 39:589-614.

Kenstowicz, Michael, and Degif Petros Banksira. 1999. Reduplicative identity in Chaha. *Linguistic Inquiry* 30:573-585.

Kenstowicz, Michael and Charles Kisseberth. 1979. *Generative Phonology: Description and Theory*. New York: Academic Press.

Kirchner, Robert. 2000. Geminate inalterability and lenition. *Language* 76:509-545.

McCarthy, John J. 2007. *Hidden Generalizations: Phonological Opacity in Optimality Theory*. London: Equinox.

McCarthy, John J. and Alan Prince. 1995. Faithfulness and Reduplicative Identity. In Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, (eds.). *University of Massachusetts Occasional Papers in Linguistics* 18:249-384. Amherst, MA: GLSA Publications. [Available on Rutgers Optimality Archive, ROA-103].

Prince, Alan and Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA & Oxford, UK: Blackwell. [Revision of 1993 technical report, Rutgers University Center for Cognitive Science. Available on Rutgers Optimality Archive, ROA-537.]

Saeed, John J. 1982. Central Somali: A Grammatical Outline. *Afroasiatic Linguistics* 8:1-43.