The phonology and phonetics of coalescence

Michael Key

September 10, 2008

1 Introduction

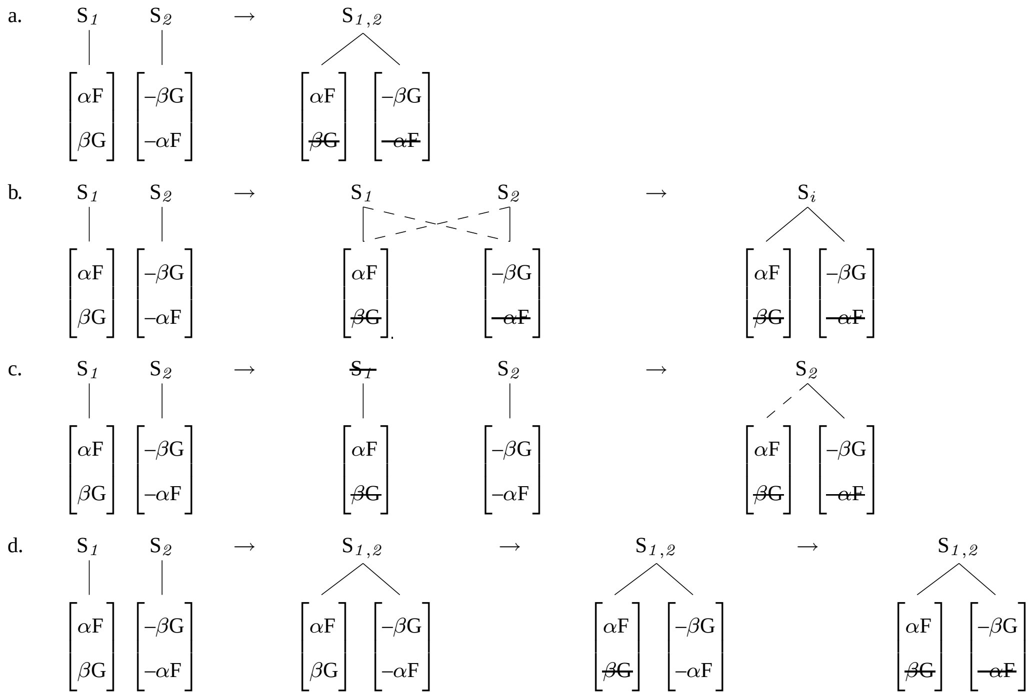
In this dissertation, I would like to address a number of questions regarding coalescence, which I define as a process by which two input segments are merged into a single output segment that is distinct in at least one feature from the segments to which it corresponds. Coalescence is a particularly fascinating phenomenon because it exhibits properties found in other phonological processes, such as assimilation and deletion, as well as in phonetic coarticulation. Thus, coalescence is a case-study in the investigation of the interface between phonology and phonetics. In this regard, I would like to adduce evidence from phonological analysis of coalescence in Optimality Theory (Prince & Smolensky 1993/2004), as well as from experimental work on the acoustics and perception of coalescence. The primary phonological question I would like to pursue in this thesis is whether coalescence best explained in a derivation with a unique fusion operation, or whether it is reducible to a derivation with a combination of other independently-known processes. More specifically, should coalescence be reduced to a special case of assimilation, deletion, or feature stability phenomena? To make this theoretical work general, I would like to construct and compare OT analyses of coalescence as a kind of assimilation, a kind of deletion, as an instance of feature stability, and as fusion under two different architectural assumptions about the grammar: (i) that the candidate set is generated in parallel; (ii) that the candidate set is generated serially. This distinction, explained below, has recently been shown to make important differences in the empirical coverage of OT, and thus it seems a good understanding of the phonological status of coalescence must identify any dependencies upon these architectural assumptions about the grammar.

The phonetic research program to coalescence seeks to explore possible acoustic cues to coalesced segments and to then see if these properties are perceptible. This line of attack is likely to shed light on universal generalizations about the occurence and output of coalescence and to clairfy the relationship between coalescence and coarticulation. Thus, it is envisioned that the combination of results from these two programs will motivate an understanding of coalescence as partly phonological and partly phonetic in nature. In addition, it is hoped that the results of perception experiments using coalescence will help to further articulate the architecture of OT models of production and comprehension, with possible longer-term goals such as a better understanding of how phonologization-as-misperception can be expressed in such models.

2 The phonology of coalescence

Coalescence may be due to a unique fusion operation, or it may be due operations that produce assimilation, deletion, or feature stability phenomena. Derivations representing these four types of analyses are shown in (1) to allow for some general discussion of the theoretical and empirical issues surrounding the analysis of coalescence.

(1) Four possible coalescence derivations



The derivation shown in (1a) represents the classic fusion analysis of coalescence. The two skeletal units S_1 and S_2 fuse to form a single skeletal unit $S_{1,2}$. The feature bundles associated with S_1 and S_2 in the input are preserved and thereby become associated with $S_{1,2}$. The deletion os the features that are not preserved as explained by a subsequent step or steps. This question will be discussed in further detail below. On the theoretical side, coalescence as fusion is a less desirable type of analysis because the operation responsible for it is fundamentally different from the operations that drive other phonological processes. On the empirical side, coalescence environments overlap with those that produce assimilation and/or deletion. Empirical arguments speaking in favor of fusional analyses of coalescence have taken one of two forms: (i) that the case of coalescence in question is not reducible to a combination of independently attested processes, as in Pater's (1999) analysis of

Indonesian; (ii) that an analysis which reduces it to a combination of other process makes incorrect predictions for transparent environments in the same language, as in de Lacy's (2002) analysis of Chipewyan. Both of these arguments and cases will be discussed in detail below.

The remaining three derivations differ in that they attempt to reduce coalescence to a combination of other processes. In (1b), coalescence is derived by a sequence of assimilation(s) (as feature spreading) followed by segmental deletion. When one feature spreads onto another segment bearing an opposite value of the same feature, the oppositely-valued feature must be deleted as is the nature of assimilation. Although I've shown spreading and opposite-value feature deletion in a single derivational step, it is possible that they should be treated as separate steps. Many cases of coalescence appear to be the result of a simple interaction between independent processes of assimilation and deletion. For example, in Bedouin Arabic, palatalization of velars before [i] and syncope occur as independent processes, and they sometimes interact opaquely such that input /ki/ sequences coalesce to [k^j] (Al-Mozainy 1981).

The derivation in (1c) relates coalescence to yet another indepedently observed phonological process, sometimes termed *feature stability*. In this type of derivation, segmental deletion occurs but it leaves behind a 'stable' feature. To produce coalescence, this feature docks onto a neighboring segmental host, which causes any opposite value of the same feature on that segment to be lost, as in assimilation. Stability was first discussed by Goldsmith (1976a, 1976b) in the context of cases in which a vowel deletes but leaves behind its tone, which then either remains 'floating', as evidenced by downstep effects, or reassociates to a nearby vowel, creating either a contour tone or replacing that vowel's original tonal value. In either of the cases in which the tone reassociates, the output is coalescence.

In (1d), coalescence is seen as the fusion of skeletal units, as in (1a), which is followed by the gradual deletion of some of the features associated with each of the input skeletal units. Again, the question of exactly how many feature-deleting steps are actually involved in the derivation of coalescence is an open one. In comparing this analysis with the fusion analysis, we see that the difference lies in whether the skeletal fusion operation and feature-deleting operation(s) occur in the same derivational step or in separate steps. The promise for an analysis of coalescence as gradual feature deletion comes from recent work by McCarthy (2008), who argues that deletion and reduction phenomena can both be explained assuming feature deletion as the basic operation in a serial OT model. On this view, deletion and reduction differ in extent of feature deletion. If this analysis can be extended, then coalescence will be seen as some pattern of feature deletion in which at least one but not all features from each of the input segments are deleted.

2.1 Coalescence in parallel and serial OT

I would like to examine each of these four approaches to coalescence under two different assumptions about the relationship between the candidate-generating function Gen and the ranked hierarchy of constraints Eval: harmonic parallelism (HP) versus harmonic serialism (HS). HP is the classical structure in OT; it is characterized by a definition of Gen that supplies an infinite candidate set, which contains as a subset all the ways in which an output might differ from the input (i.e. the unfaithful candidates). A subset of these unfaithful candidates differ from the input by the result of more than one instance or type of unfaithful mapping. The diversity of candidates produced by Gen in HP allows all the phonology to happen in just one pass through the constraint hierarchy.

In HS, Gen produces only a small number of candidates – the faithful candidate(s) plus all the candidates that differ from the input by just one instance or type of unfaithful mapping. Of course, HS requires a theory of what 'counts' as one instance or type of unfaithful mapping (i.e. a theory of gradualness); see McCarthy (2006) for some discussion of the possibilities. Under any reasonable theory of operations, it is surely the case that one pass through Eval cannot possibly suffice to explain all phonological derivations. Therefore, HS assumes that Gen and Eval are looped such that the output of some evaluation n is the input from which Gen constructs a new candidate set on evaluation $n + 1$. The looping terminates when the derivation converges, i.e. when the output of evaluation k is identical to the output of evaulation $k - 1$.

2.2 Coalescence as fusion

In the phonological literature, both before and since OT, coalescence phenomena have frequently been analyzed as due to a unique *fusion* operation (Chomsky & Halle 1968; Schane 1984; McCarthy & Prince 1995; Pater 1999). Fusion analyses of coalescence posit a relation in which two elements of the input stand correspond to a single element of the output, like the one shown in (1a). For this reason, coalescence as fusion bears no special relation to other phonological processes. In several implementations of this theory in OT, fusion is proposed to violate a special faithfulness constraint that other operations do not: UNIFORMITY (McCarthy & Prince 1995 et seq.) or *Seg-Int (Casali 1996, 1997).[1]

[1] Keer (1999), who proposes that there is no faithfulness constraint that fusion violates, is an important exception.

Nearly all analyses of coalescence in OT treat it as fusion (Lamontagne & Rice 1995; McCarthy & Prince 1995; Casali 1996, 1997, Keer 1999; Pater 1999; Wheeler 2005) and assume the parallel architecture. Fusional analyses posit a unique correspondence relation (like the one shown in 1a) and a special faithfulness constraint against that relation (e.g. UNIFORMITY). In addition, these analyses require crucial rankings between multiple 'basic faithfulness constraints, either IDENT(F) or MAX(F) constraints. Pater's (1999) analysis of Indonesian nasal substitution, described in (2), is representative of fusion analyses in HP. The essentials of Pater's analysis are provided in the tableau in (3).[2]

(2) Indonesian nasal substitution (Halle & Clements 1983: 125)

a.	/məŋ-**p**ilih/	→	məmilih	'to choose'
b.	/məŋ-**t**ulis/	→	mənulis	'to write'
c.	/məŋ-**k**asih/	→	məŋasih	'to give'

The generalization is that prefix-final nasals are substituted for root-initial voiceless obstruents, though the place feature(s) of the latter are retained. In Pater's analysis, the [nasal] feature of the prefix-final consonant and the place feature(s) of the root-initial (voiceless) consonant fuse into a single output segment to avoid violation of a markedness constraint that bans a sequence of a nasal followed by a voiceless obstruent, *NC̥.

(3) Indonesian nasal substitution as fusion (Pater 1999)

/məm$_1$-p$_2$ilih/	ID[+nas]	MAX	ID[voi]	*NC̥	UNIF	ID[-nas]
a. → məm$_{1,2}$ilih					1	1
b. məm$_1$p$_2$ilih				1 W	L	L
c. məm$_1$b$_2$ilih			1 W		L	L
d. məp$_2$ilih		1 W			L	L
e. məp$_{1,2}$ilih	1 W				1	L

As the tableau in (3) illustrates, the optimality of the fusion candidate (3a) requires not only that *NC̥ dominate UNIFORMITY and IDENT[-nas] (motivated by comparison with (3b)), but also that several other faithfulness constraints dominate UNIFORMITY and IDENT[-nas]. In particular, rank-

[2] I've adapted Pater's analysis somewhat in using UNIFORMITY as the faithfulness constraint violated by coalescence. He assumes that fusion violates the same constraint that metathesis does, LINEARITY. In addition, I've assumed a equipollent [nasal] feature and separate IDENT constraints that penalize nasalization and denasalization.

ings like || IDENT[+nas] $\gg$ UNIFORMITY, IDENT[-nas] || (3a vs. 3e) are a hallmark of fusional analyses in HP: a high-ranking faithfulness constraint like IDENT[+nas] can contribute to the explanation of why coalescence is chosen as the solution to the markedness constraint, as in other phonological conspiracies, but also determines *how that solution is realized* – i.e. as nasal-preserving coalescence (/$m_1 p_2$/ $\rightarrow$ [$m_{1,2}$]) rather than oral-preserving coalescence (/$m_1 p_2$/ $\rightarrow$ [$p_{1,2}$]).

This particular consequence of F $\gg$ F ranking is related to an argument against the constraint UNIFORMITY made by Keer (1999) in the context of deriving the lexical OCP. Assuming there are markedness constraints banning pair melody geminates, the ranking || UNIFORMITY $\gg$ $M_{pair-gem}$ || predicts a language that contrasts pair melody geminates with singleton segments. Keer identifies two possible strategies to rule this out: fixed-ranking of $M_{pair-gem}$ above UNIFORMITY and denial of UNIFORMITY. He notes that the former strategy is no guarantee that input pair geminates will coalesce to a single-melody geminate as desired; for example, there may be other faithfulness constraints that dislike coalescence, and if so, they too must be ranked below $M_{pair-gem}$ in the putative fixed ranking, a solution Keer judges to be ad-hoc. By denying the existence of UNIFORMITY, the theory of coalescence as fusion is left to the ranking between faithfulness constraints on featural identity.

In HS, the challenge for the fusion approach comes from the theory of faithfulness in a somewhat different form. Since coalescence involves both unfaithfulness to the featural content of multiple input segments, as well as violation of a special faithfulness constraint such as UNIFORMITY, the theory of gradualness may become rather difficult to state. This is the case if the theory of gradualness is formulated so as to allow violations of more than one 'basic' faithfulness constraint (e.g. IDENT[F]) in a single pass from Gen to Eval. McCarthy (2007: 89) points to the considerable difficulty in formulating a theory of gradualness that would permit violations of multiple basic faithfulness constraints *only* in cases of coalescence, but not in other phonological derivations. For instance, if the theory of gradualness allows /$æ_1 i_2$/ $\rightarrow$ [$e_{1,2}$] in a single pass that violates IDENT[low] and IDENT[high] to explain coalescence, how can it be blocked from also allowing /æ/ $\rightarrow$ [i] in a single pass, since the latter derivation violates the same basic faithfulness constraints? We know that the latter case must be blocked if the theory is to also explain vowel-height chain shifts such as æ $\rightarrow$ e $\rightarrow$ i. If coalescence as fusion is the only argument for multiple basic faithfulness constraint violations in a single pass, it suggests that the evidence for fusion must be carefully examined, which I intend to do in the course of this thesis.

While the prospects for a reasonable theory of gradualness that allows multiple violations of the basic faithfulness constraints are not particularly promising, there are avenues for research that are worth pursuing. In the example just discussed, we might note that the two basic faithfulness constraints implicated are violated at different *loci* in the case of coalescence, while they are violated at the same locus in the height-raising derivation. While this observation alone surely cannot distinguish between coalescence and non-coalescence derivations, it does suggest that properties such as locus of violation must be considered in order to determine whether a such a theory of gradualness is coherent.

To summarize, the general challenge for fusion is that it treats coalescence as *sui generis*; it ignores possible correlations between coalescence and other phonological processes, and has nothing to say about the constraints that determine its precise realization nor restrictions (e.g. locality) on its occurrence. While these problems motivate an attempt to reduce coalescence to one of the other approaches outlined above, it remains a distinct possibility that not all cases of coalescence can be plausibly reduced to one of the other approaches. Since a major part of this dissertation will be to examine the evidence regarding the phonological status of coalescence vis-à-vis assimilation plus deletion, feature stability, and feature deletion, I also plan to investigate the prospects for a theory of gradualness that restricts multiple basic faithfulness constraint violation in a single pass to coalescence.

2.2.1 Determining the result of coalescence: Markedness vs. marked faithfulness

Unfaithful mappings in OT are the result of ranking markedness constraints above faithfulness constraints, however other constraints may decide exactly which unfaithful candidate is optimal. For example, ranking DEP above MAX ensures that deletion is chosen over epenthesis as the solution for some markedness constraint; ranking MAX above FINAL-C will ensure that the locus of deletion will not be in word-final position. In general then, either faithfulness or markedness constraints can determine the choice among competing unfaithful candidates. The question I would like to pursue in the present context is whether certain faithfulness constraints *or* markedness constraints alone are sufficient to explain the realization of coalescence, or whether both types of constraints are needed.

In particular, I'm interested in whether 'marked faithfulness' constraints (de Lacy 2002), which demand faithfulness to certain marked segments or feature values, are necessary to derive the output of coalescence. Marked faithfulness constraints can decide the output in coalescence by allowing marked feature values to be preserved. In de Lacy's analysis of Attic Greek and Dakota vowel coalescence, IDENT[+round] and IDENT[+nasal] respectively outrank markedness constraints against these feature values. However, in other cases, such as the feature [voice] in Swedish coalescence, it is the unmarked value that survives, which motivates ranking markedness above marked faithfulness. It is possible however that marked faithfulness is not necessary to explain the outputs of coalescence. Cases in which the marked value is preserved, as in Attic Greek and Dakota, could be analyzed as the ranking of faithfulness to a feature (irrespective of its value) above the markedness constraint against the relevant value of that feature. The broader implications of this question concern the explanation of typological observations in OT: are universal or near-universal generalizations the reflection of markedness, or does marked faithfulness play a role. In a theory that relies on both types of constraints, it is hard to know whether any given phonological generalization is the result of markedness or marked faithfulness, whereas a theory that attributes the same patterns to markedness alone makes a strong, universal claim.

I've done some preliminary analysis of a generalization about the output of vowel coalescence reported by Causley (1997) using only markedness constraints to explain which unfaithful candidate is selected. The generalization Causley points to, citing Korean and Chaha as examples, is that in languages that show a three-way contrast for vowel backness, an input sequence of a front unrounded vowel plus a back rounded vowel coalesces to a front rounded vowel, but apparently never a central vowel. In languages like Korean, this is surprising given that the language has a high central vowel, but the result is a non-structure preserving front rounded vowel.

In the pair of tableaux below, I show only the markedness constraints, so we can see which candidates for the output of coalescence can improve markedness relative to the input *under some ranking* of the markedness constraints, and whether any of these candidates lead toward the desired front rounded vowel. Thus, the tableaux are designed to show harmonic bounding rather than ranking. As far as markedness constraints are concerned, I assume a constraint against non-low central vowels (*[-front,-back,-low]), a constraint against front rounded vowels (*[+front,+round]), a constraint against back unrounded vowels (*[+back,-round]), and a constraint against non-front vowels (*[-front]). The first three of these are typologically well-motivated – non-low central vowels, front rounded vowels, and back unrounded vowels are rare and are the source of well known implicational universals (e.g. if a language has front rounded vowels, then it must also have front unrounded vowels). The motivation for positing the constraint *[-front] is that the faithful candidate must violate some markedness constraint, else no unfaithful candidate could ever be more harmonic than it.[3]

(4) Front + back coalescence as fusion

/i$_{1}$,u$_{2}$/	*[-front,-back,-lo]	*[+front,+round]	*[+bk,-rnd]	*[-front]
a. *faith* i$_{1}$u$_{2}$				*
b. y$_{1,2}$		*		
c. ☠ ɨ$_{1,2}$	*			*
d. ☠ ʉ$_{1,2}$	*			*
e. ☠ ɯ$_{1,2}$			*	*

An analysis of Causley's generalization using these markedness constraints alone is successful; all bona fide fusion candidates are harmonically bounded by the faithful candidate except the desired [y] candidate (4b).

An interesting condition for this result to hold is that certain candidates with vacuous fusion – in which an output segment bears correspondence to two input segments but is also identical to one of them – must be banned, else they forever would doom the desired unfaithful candidate. In the analysis in (4), that candidate is [i$_{1,2}$], which would violate none of the markedness constraints, nor could the bounding be broken by adding a further markedness constraint (e.g. *[+front]).[4]

2.3 Coalescence as assimilation plus deletion

The second theory of coalescence I want to pursue is the assimilation-plus-deletion approach. It attempts to explain coalescence phenomena from theories of assimilation and deletion, and a theory that allows them to interact opaquely. It is the opacity requirement that makes the assimilation plus deletion approach unviable in most conceptions of HP; since deletion occurs in the coalescing candidate as well as in a candidate that simply deletes the trigger for assimilation, the latter will harmonically bound the former because it vacuously satisfies the faithfulness constraint against assimilation. The only hope for the assimilation plus deletion approach in HP is to revise the representational assumptions about the candidate set such that the trigger of assimilation is somehow

[3]The symbol ☠ indicates a harmonically bounded candidate. The comma in the input form indicates that the order of the input segments does not matter.

[4]The other vacuously-fusing candidate [u$_{1,2}$] would not bound [y], but would instead compete with it. While Causley's generalization suggests that /i,u/ always maps to [y], she notes that [u] is also attested. It is an open question whether /i,u/ → [u] should be analyzed as fusion.

present but not actually realized, as in Containment Theory and Turbidity Theory. I'll have nothing further to say about these possibilities here.

On the other hand, the assimilation plus deletion theory is natural in HS, given the considerations for a responsible theory of gradualness. In McCarthy (2007), the coalescent mapping /ħaːkim-iːn/ → [ħaːkʲmiːn] 'ruling m. pl.' (Al-Mozainy 1981: 49f., 73ff.) in Bedouin Arabic is analyzed as palatalization of /k/ before [i], which counterbleeds syncope of /i/. McCarthy's analysis, abstracting away from some of the details of OT with candidate chains (OT-CC), is reproduced in the tableau in (5).

(5) Coalescence as assimilation plus deletion in HS

/ħaːkim-iːn/	**i*CV	PAL	MAX	PREC (ID[back],MAX)	ID[back]
a. → ħaːkʲmiːn			1		1
b. ħaːkimiːn	1 W	1 W	L		L
c. ħaːkʲimiːn	1 W		L		1
d. ħaːkmiːn			1	1 W	L

The constraint PAL motivates palatalization of /k/ before the vowel [i], which violates the constraint ID[back]. The constraint *iCV drives the syncope of high vowels in open syllables, a mapping that violates MAX. In order to allow both processes to apply in a counterbleeding fashion, the constraint PREC(ID[back],MAX) demands that deletion be preceded in the derivation by changing of the feature [back]; it assigns a mark to (5d) because it has a MAX violation without a (prior) violation of IDENT[back].

Because both palatalization of velar stops and syncope of high vowels in open syllables are independently evident in the transparent phonology of Bedouin Arabic, the assimilation plus deletion approach makes sense in this case. As such, the assimilation plus deletion theory appropriately reduces coalescence to independently known processes, an instance in which a fusion approach would miss an important generalization. In fact, a fusional analysis of coalescence cases like Bedouin Arabic *only* makes sense in an attempt to analyze (some cases of) counterbleeding opacity in HP. It is also worth noting at this point that the assimilation and deletion processes in coalescence need not be paralleled by transparent assimilation and deletion in the same language. Instead, this theory's success depends on whether the assimilation and deletion processes are known in *some language*.

As McCarthy (2007) points out, this is the case for French V+N coalescence; spreading of [nasal] and coda nasal deletion are not observed in transparent environments in French, but are in fact common processes cross-linguistically.

However, there have been some claims in the literature about coalescence cases that require typologically unknown independent processes. One case is the Indonesian nasal substitution facts discussed above. If the mapping /ŋp/ → [m] is analyzable as an opaque interaction between place assimilation (/ŋp/ → [mp]) and deletion (/p/ → ∅), we should expect to see each of these processes attested independently. In this case, its not the assimilation step that is the worry, since place assimilation is obviously amply attested, but rather the deletion process. Pater (1999) argues that deletion of an obstruent after a nasal is only observed cross-linguistically when the nasal has already assimilated in place. Pater's analysis avoids this particular worry since the assimilation and deletion are both embodied in a fusional analysis. In response, McCarthy (2007) has shown that the concern for the plausbility of analyzing this case as assimilation plus deletion is alleviated when an interaction between alternative processes is considered: nasality assimilation (/məŋ-pilih/ → məŋ**m**ilih) precedes coda nasal deletion (/məŋ**m**ilih/ → məmilih). The nasality assimilation process occurs independently from coda nasal deletion in Konjo (Pater 1999: 323), while coda nasal deletion is also known independently.

Secondly, while it is true that the assimilation and deletion processes implicated in coalescence need not be transparent in the language, they should not cause problems in transparent environments. A potentially problematic case of this kind is the pattern of consonant coalescence observed in Chipewyan, an Athabaskan language. In this language, coalescence occurs between the first two Cs of an underlying tri-consonantal cluster, which avoids illicit complex syllable margins. Data are provided in (6).

(6) Chipewyan consonant coalescence (Li 1946)

a.	i.	/ðɛ-θ-l-ɣĩ/	→	ðɛɬɣĩ	'he is exhausted'
	ii.	/hɛs-l-zɛ/	→	hɛszɛ	'start to hunt'
	iii.	/uh-l-ze/	→	huɬzɛ	'start to hunt 2 dual'
b.	i.	/-ðir-ɬ/	→	ðɛɬ	'to be, act, do'
	ii.	/-tsaʒ-ɬ/	→	tsaɬ	'be handling grain-like object'
	iii.	/-tel-ɬ/	→	tɛɬ	'several persons go'

If either C_1 or C_2 is [-voice], the coalesced segment will also be [-voice]. In the assimilation plus deletion analysis, bidirectional spreading of [-voice] must be posited, which should be followed by deletion of the consonant that sponsored [-voice]. This is shown in the tableau in (7).

(7) Chipewyan coalescence in HS: Neutralization to [-voi]

/ðɛ-θ₁-l₂-ɣĩ/	*COMP	*CODA/[+voi]	MAX	PREC	ID[voi]
a. → ðɛɬ₂.ɣĩ ⟨ID[voi]@2, MAX-C@1⟩			1		1
b. ðɛθ₁.ɣĩ ⟨MAX-C@2⟩			1	1 W	L
c. ðɛθ₁ɬ₂.ɣĩ ⟨ID[voi]@2⟩	1 W		L		1
d. ðɛθ₁l₂.ɣĩ ⟨∅⟩	1 W	1 W	L		L

In the analysis in (7), the constraint banning complex margins *COMPLEX is ranked above MAX, causing deletion, while *CODA[+voi] dominates IDENT[voi], ensuring devoicing. As in the Bedouin Arabic analysis above, the constraint PREC(IDENT[voi],MAX) champions the opaque candidate in which devoicing precedes deletion. Both processes are certainly cross-linguistically well-known. Bidirectional spreading of [-voice] is found in Swedish coalescence, while deletion that relieves complex margins is attested in many languages. However, the problem is that the bidirectional [-voice] spreading processes needed to explain coalescence is at odds with the voicing assimilation pattern observed in transparent environments in Chipewyan. In underlying bi-consonantal clusters, *both* values of [voice] spread left-to-right and no deletion occurs, e.g. /hɛ-ɣwa-l-**s**-ɛ/ → [hul**z**ɛ] 'he started to hunt'. The assimilation plus deletion theory is thus in a position of explaining why bidirectional [-voice] spreading only occurs when it counterbleeds deletion, a question also raised by de Lacy (2002).

These observations suggest related research questions that I would like to investigate regarding the assimilation plus deletion approach: are the assimilation and deletion processes implicated in a particular case of coalescence independently known, and, what (if any) are the problems for transparent environments in the same language introduced by these processes?

Another way to evaluate the assimilation plus deletion approach to coalescence is to see if learners will apply each of the processes maximally, resulting in a counterbleeding interaction between assimilation and deletion that produces coalescence. I propose to test this using an artificial language learning design in which different groups of participants are exposed to paradigms that show assimilation, deletion, or both processes (crucially independently of one another) in the training phase. In the test phase, learners will have the opportunity to generalize the process(es) on which they have been trained in new paradigms, thus testing whether the process/processes have been learned, as well as to new paradigms that present the conditions for both processes. If coalescence develops from an opaque interaction between assimilation and deletion, learners that were exposed to both processes in training should produce coalesced outputs more often in paradigms in which the conditions for both processes are presents than learners who were only exposed to assimilation or deletion alone in the training phase.

This could be measured by a forced-choice task in which learners must select a form to complete a new paradigm, as exemplified in the following table by a language in which nasals assimilate to non-coronal stops and voiceless stops delete post-consonantally. In the test items presenting conditions for both processes (e.g. zan + /-pi/ below), learners would have to choose between multiple alternatives for a novel member of a paradigm depending on their exposure in training: learners who were exposed to assimilation or deletion only would choose whether or not to apply the process to which they had been exposed, with the application of the process they had not been exposed to present in both alternatives, in order to see if the presence of an unrelated change would inhibit application of the rule; learners who were exposed to both processes in training would have a three-way choice between alternatives in which both processes apply (i.e. in a counterbleeding fashion) to produce coalescence, those in which deletion bleeds assimilation to produce deletion, or those in which only assimilation applies.

(8) Learning coalescence from assimilation and deletion

Group	Training phase	Testing phase	Alternatives
Assimilation only	dan ∼ da**m**-bi, sha ∼ sha-pi	span + /-bi/ zan + /-pi/	spa**m**-bi *vs.* span-bi za**m**-i *vs.* zan-i
Deletion only	huza-bi ∼ huza-pi, sural-bi ∼ sural-i	turag + /-pi/ zan + /-pi/	turag-i *vs.* turag-pi za**m**-i *vs.* zam-pi
Both processes	dan ∼ da**m**-bi, sural-bi ∼ sural-i, sha ∼ sha-pi	span + /-bi/ turag + /-pi/ zan + /-pi/	spa**m**-bi *vs.* span-bi turag-i *vs.* turag-pi za**m**-i *vs.* za**m**-pi *vs.* zan-i

If coalescence emerges from exposure to independent processes of assimilation and deletion, we predict that the 'both processes' group will choose counterbleeding forms (e.g. za**m**-i) over bleeding forms (e.g. zan-i) more often than the 'assimilation only' group since the latter will be uncertain about whether to apply assimilation when a previously-unseen deletion process also occurs. Similarly, the ‘both processes’ group should choose the counterbleeding alternative over that in which only assimilation occurs (e.g. za**m-**pi) more often than ‘deletion only’ learners since the latter will be unsure about applying deletion when an unforeseen assimilation process also occurs.

2.3.1 Markedness(M) vs. marked faithfulness(mF) revisited

The question of whether markedness constraints alone can explain the output of coalescence is taken up again assuming an assimilation plus deletion account, in which assimilation violates IDENT(F) constraints for the features [front], [back], and [round] and the markedness constraints are the same as those deployed in the fusion analysis of Causley’s generalization. In the pair of tableaux below, I again show only the markedness constraints, using the ☠ symbol to indicate a candidate that is harmonically bounded.

(9) Front + back coalescence as assimilation plus deletion: Spreading of [front], [back], or [round]

/iu/	*[-front,-back,-lo]	*[+front,+round]	*[+bk,-rnd]	*[-front]
a. *faith* iu				*
b. ☠ yu		*		*
c. ☠ ɨu	*			**
d. ☠ iʉ	*			*
e. ☠ iɯ			*	*

In contrast to the fusional approach, none of the unfaithful candidates better the faithful candidate; they are all harmonically bounded. The candidates that we desire under the present assumptions about gradualness are (9b,d) because they maintain the features of the input that we know must be present in the desired surface form [y] – namely [+front] and [+round]. The problem is that they create violations of *[+front,+round] and *[-front,-back,-low], respectively, while failing to get rid of the *[-front] violation shared with the faithful candidate. So even though they are on the way to success, they get stuck in a local minimum which curtails further harmonic improvement (McCarthy 2006).

One solution to this problem is to revise assumptions made about the nature of gradualness. Suppose instead that Gen is defined such that an unfaithful candidate may differ from the input in the value on one segment of (i) any subset of a feature class Color (Odden 1991; Padgett 1995), which I assume to be [front] and [back], or (ii) [round] (but not (i) and (ii) in a single evaluation). This theory of gradualness expands the candidate set a bit, and allows candidates that are ‘on the way’ to desired [y] to be non-harmonically bounded, as shown in (10).

(10) Front unrounded + back rounded: Spreading of {Color} or [round]

/iu/	*[-front,-back,-lo]	*[+front,+round]	*[+bk,-rnd]	*[-front]
a. *faith* iu				*
b. ☠ iʉ	*			*
c. → iy		*		
d. ☠ iɯ	*		*	*
e. ☠ ɨu	*			**
f. ☠ ɯu	*		*	**
g. ☠ yu		*	*	*

Only faithful [iu] and [iy] are non-harmonically bounded in (10); the latter candidate trades the *[-front] violation incurred by the faithful candidate for a violation of *[+front,+round]. This means that the assimilation step, spreading of {Color}, will occur when *[-front] ≫ *[+front,*round]. The deletion step /iy/ → [y] is not shown here, but is similar to the analysis of Bedouin Arabic discussed above.

Markedness constraints alone were able to explain Causley’s generalization in the assimilation plus deletion approach, though the analysis raises two important questions. First, what kind of theory of markedness is necessary to maintain the view that markedness constraints and not marked faithfulness constraints determine the result of coalescence? A comparison between the markedness constraint used here and those used by de Lacy is instructive: the present analysis gets by with markedness only because it assumes a more direct encoding of markedness via prinicipally feature co-occurrence constraints, whereas de Lacy’s theory posits constraints against single feature values. There appears to be a tension between the complexity of the markedness constraints assumed and the necessity of marked faithfulness constraints.

2.4 Coalescence as feature stability

Unlike the first two approaches to coalescence, the feature stability analysis discussed above has not to my knowledge been discussed in the context of the proper account of any case of coalescence. However, consideration of the original feature stability analysis in autosegmental phonology (Goldsmith 1976a,1976b) shows that this is probably an artifact of a more restrictive definition of coalescence phenomena; in cases of tonal stability under vowel deletion, if the tone reassociates to another vowel it either creates a contour tone on that vowel or replaces that vowel's underlying tone. In either case, this is effectively coalescence because the surface segment owes its segmental features to its input correspondent, but inherits its tonal quality from a different input vowel. The feature stability approach certainly must be considered to analyze the tonal stability phenomena that Goldsmith was concerned with (because there is clear evidence of the deletion step), but this analysis might be recruited to explain cases of coalescence that are not usually thought to involve (segmental) deletion.

In HP, the key to the feature stability approach to coalescence is to have some markedness constraint driving deletion dominate MAX, and a markedness constraint motivating reassociation to be ranked above the faithfulness constraint that protects the featural content of the segment to which the stranded feature is reassociating.[5] This analysis is applied to the common /ai/ → [e] case in the following tableau.

(11) Coalescence as feature stability in HP

/ $a_1 i_2$ [-hi] [+hi] /	*VV	*FLOAT	MAX	IDENT[hi]
a. → e_2 [-hi] ~~[+hi]~~			1	1
b. i_2 [-hi] [+hi]		1 W	1	L
c. $a_1 i_2$ [-hi] [+hi]	1 W		L	L

[5]The necessity of the latter ranking is subject to assumptions regarding which operations incur a faithfulness cost that may depend on the kind of feature that is reassociating. For instance, in the case of H tone stability under deletion, it is commonly assumed that the H may redock (thus replacing the would-be L or other default tone) for free. In contrast, redocking of 'segmental' features would typically be assumed to bear a faithfulness cost.

In (11), ranking *FLOAT above IDENT[hi] motivates the redocking of [-hi] under deletion of /a/, ruling out (11b). *VV must not only dominate MAX (to motivate deletion of /a/) but also must outrank IDENT[hi] in order to favor (11a) over faithful (11c). This is an instance of a odd property of some complex mappings in HP: the analysis requires that a markedness constraint (*VV) outrank two faithfulness constraints (MAX, IDENT[hi]), suggesting that violating both is necessary to avoid violation of the markedness constraint, while in fact violation of just one of them is sufficient. This tension between markedness constraints and the faithfulness constraints they can be directly ranked w.r.t. is at the heart of the potential differences in predicted typology between HP and HS.

In HS, the feature stability analysis is essentially the same, but presents an interesting issue of constraint ranking discussed by McCarthy (2007). In the HP analysis of coalescence as feature stability in (11), the constraints *VV and *FLOAT cannot be ranked with respect to each other since the candidates that could provide evidence for their ranking are both suboptimal. However, as shown in (12), in the HS analysis we know that *VV must dominate *FLOAT in order to produce the first derivational step (deletion) in /ai/ → [e]. The tableau in (13) shows the docking step.

(12) Coalescence as feature stability in HS: First pass

/ $a_1 i_2$ [-hi] [+hi] /	*VV	*FLOAT	MAX	IDENT[hi]
a. → i_2 [-hi] [+hi]		1	1	
b. $a_1 i_2$ [-hi] [+hi]	1 W	L	L	

(13) Coalescence as feature stability in HS: Second pass

/ $a_1 i_2$ [-hi] [+hi] /	*VV	*FLOAT	MAX	IDENT[hi]
a. → e_2 [-hi] ~~[+hi]~~				1
b. i_2 [-hi] [+hi]		1 W		L

The biggest questions raised in considering the feature stability approach to coalescence are (i) typological considerations for HP and HS introduced by the differences in required constraint ranking discussed above; (ii) the plausibility of candidates (like (11b), (12b)) that contain floating features. The former issue is at the heart of much current research in OT and this approach to coalescence serves to contribute to it. The latter issue of floating features arises because the candidates in question are non-harmonically-bounded and are thus possible outputs in some language, we must ask whether there is evidence that the feature in question (e.g. [high]) ever floats. The floating status of other features such as tone has been argued for from cases of downstep or upstep (e.g. Clements & Ford 1979), in which the floating tone is a representational explanation for the observed change in F_0. It remains to be worked out whether similar arguments can be marshalled in favor of the floating possibility of other features.

2.4.1 M vs. mF

As in the fusion and assimilation plus deletion approaches, I plan to investigate using markedness constraints alone to explain the precise realization of coalescence in a feature stability account, pending progress on the more basic question of floating features that seems necessary to developing it further.

2.5 Coalescence as feature deletion

The final approach, one which has also not been pursued before, links coalescence to deletion. On this view, coalescence and deletion both obtain by the same process – deletion of features. What distinguishes them is the locus and extent of feature deletion: in coalescence, at least one feature from each input segment is deleted and at least one feature from each input segment is preserved; in deletion, all the features of just one of the input segments deletes. However, when taken together with the independently-proposed idea that segmental deletion is the result of multiple feature deletions (McCarthy to appear, 2008), we can imagine that coalescence might be treated as the result of deletion of *some* of the features of each of the input segments, while segmental deletion would be deletion of *all* of the features of just one of the input segments. Of course, the biggest question for this approach will be whether each feature-deleting step in the derivation of a coalesced form can be motivated.

The empirical observation that seems uniquely applicable to this approach is that both coalescence and deletion occur in the same environments cross-linguistically, and in similar environments within a language, as exemplified in (14). These context/conspiracy facts alone do not necessarily suggest that coalescence and deletion should somehow be formally related. However, these observations do facilitate the analysis of coalescence as gradual feature deletion because, by being intrinisically local (reflexive) in its application, feature deletion vacuously avoids the concern facing the other approaches over whether coalescence and the proposed process that drives it (assimilation or feature floating/docking) share the same locality restrictions cross-linguistically.

(14) Vowel coalescence and deletion in identical/similar hiatus contexts

a.	Bemba	/ba-ka-is-a/	→	bak**eː**sa	'they will come'
b.	Lungu	/a-ma-ino/	→	am**iː**no	'teeth'
c.	Attic Greek	/tiːma-o-men/	→	tiːm**ɔː**men	'we honor'
d.		/tiːma-eːn/	→	tiːm**aː**n	'to honor'

In Bemba and Lungu (14a-b), two Bantu languages, the same input sequence /a-i/ is resolved (mapped unfaithfully), but with different solutions. In Bemba (14a), the solution is coalescence to [eː]. Assuming full feature specification with standard *SPE* features, the coalescence in (14a) involves deletion of [+low] and [+back] (from /a/) and [+high] (from /i/), and preservation of [-high] (from /a/) and [-back] (from /i/). In Lungu (14b), using the terms of the claim above, all of the features of /a/ have deleted while all of the features of /i/ have been preserved. The observation that coalescence and deletion occur in identical contexts in related languages is support for the idea that they are somehow the same process. In Attic Greek (14c-d), coalescence and deletion both relieve hiatus, and the choice between the two is determined by the vowels involved. Coalescence (14c) shows that [+low] and [-round] (from /a/) and [+ATR] (from /o/) are deleted, and [-ATR] (from /a/) and [+round] (from /o/) are preserved. In (14d), all of /e/'s features delete. What determines the choice is the presence of the feature [+round] in the input: either pressure to preserve [+round] or pressure to lose [-round] results in at least one feature of the suffixal vowel being preserved, resulting in coalescence. This at the expense of the preservation of [+low] that otherwise occurs. Deletion of /a/ does not occur in (14c) because of the pressure to preserve [-ATR] or due to the markedness of [+ATR].

To formally relate vowel coalescence and deletion, I follow Casali (1996, 1997) in proposing that they violate the same family of constraints against feature deletion – MAX(F). In both Casali's approach and the one advocated here, permutation of the MAX(F) constraints captures different patterns of deletion and/or coalescence.[6] Each feature-deleting operation is motivated by ranking a feature co-occurence markedness constraint above the corresponding MAX(F) constraint. In the version of this analysis shown in the tableau below, the anti-hiatus constraint *VV is ranked above

[6]However, unlike Casali, I obviously do not assume a separate faithfulness constraint that only coalescence violates.

MAX(RtNode), which causes the adjacent melodies to fuse under a single root node. The feature co-occurence constraints are active on candidates that violate MAX(RtNode) since they each contain conflicting specifications for some feature. This is illustrated in the following tableau.

(15) Coalescence as feature deletion

/$a_1 i_2$/	*VV	*[-lo +lo]	*[-bk +bk]	*[-hi +hi]	MAX (RtNode)	MAX (F)
a. → $ee_{1,2}$ [-hi] [-bk]					1	3
b. $ei_{1,2}$ [-hi] [+hi]				1 W	1	2 L
c. $əi_{1,2}$ [+bk] [-bk]			1 W	1 W	1	1 L
d. $ai_{1,2}$ [+lo] [-lo]		1 W	1 W	1 W	1	L
e. $a_1 i_2$	1 W	L	L	L	L	L

In the tableau above, MAX(F) represents a ranked set of individual MAX constraints for each feature involved. Notice that it is the features [+low] and [+back] from /a/ and the feature [+high] from /i/ that are deleted. Thus, the tableau should be read from bottom-to-top, where each successive candidate represents a waypoint on one possible route to the surface form (15a). To save space, I have only shown the part of the united feature matrix in each candidate that is penalized by each of the markedness constraints as they are written from left to right. An intriguing possible extension of positing fusion of adjacent vocalic root nodes is that it offers a representation to distinguish the so-called 'short' diphthongs reported for Finnish and Rotuman.

What I propose to work out in this dissertation is the nature of the constraints that motivate each feature deletion and that determine whether the feature deletions come from just one of the input segments (as in vowel deletion), or from both input segments (as in coalescence). Yet another issue for this approach is whether coalescence, like segmental deletion, involves the loss of a root node and if so, whether root node deletion incurs a faithfulness cost. It is envisioned that the chief typological challenge for this approach will come from defining a theory of markedness and/or marked faithfulness that generates the set of coalescence and vowel deletion patterns observed.

Although I have limited my exposition of this approach to HP, it is worth noting that a comparison with the HS version of this analysis raises the same puzzle that the feature stability analysis did. Comparison of the faithful candidate with the optimal candidate in (15) requires that *VV dominate both MAX(RtNode) and MAX(F), even though violation of MAX(RtNode) alone would be expected to satisfy *VV, while the HS version (as can be judged by treating (15d) as the optimum of the first pass and comparing it with the faithful candidate) only requires *VV $\gg$ MAX(RtNode).

2.5.1 M vs. mF

As in the previous approaches, I plan to pursue whether markedness constraints alone can account for the realization of coalescence when treated as feature deletion.

3 The phonetic and perceptual properties of coalescence

The second part of this dissertation will examine whether coalescence is fundamentally phonological in nature, or whether at least some cases better submit to explanation in models of phonology-phonetics and/or perception. Answers to specific questions of this type not only tell us something about the disposition of (certain kinds of) coalescence within the realm of phonetics, but also may provide a new source of evidence in ongoing debates regarding levels of representation modularity in OT models of speech production and comprehension, as well as questions regarding how phonologization-as-misperception should be formalized in OT, as discussed in greater detail below.

3.1 Coalescence and completeness of neutralization

The question of whether a process is phonological or phonetic has sometimes been framed in terms of whether it *completely neutralizes* a phonemic contrast in some context. A completely neutralizing process means that the realization of some input category |X| *in the relevant context* (for the process) *c* is identical to the realization of some input category |Y|, thus phonological, while an incompletely neutralizing process results in an underlying contrast between |X| and |Y| being somehow preserved on the surface, thus leaving the burden of explanation to a phonetic implementation component. Because the question concerns the levels of representation at which an incompletely-neutralized input contrast is preserved and lost, this is an interesting issue to consider in light of current research on the architecture of models of speech production and comprehension. In particular, incompletely-neutralizing processes can be used as an argument for one or more levels of phonetic representation (PR) beyond the traditional underlying (UR) and surface (SR) represenations, as discussed below.

Consider a language in which two categories, |X| and |Y| contrast generally, except in context *c*, in which |X| is phonetically realized as {Z} – a realization intermediate between the realization of |X| ({X}) and |Y| ({Y}).[7] The incomplete neutralization of |X| and |Y| in *c* may be modelled in one of the three following ways:

(16) Three models of incomplete neutralization

UR	a.	$\vert X\vert$	$\vert X\vert_c$	$\vert Y\vert_{(c)}$	b.	$\vert X\vert$	$\vert X\vert_c$	$\vert Y\vert_{(c)}$	c.	$\vert X\vert$	$\vert X\vert_c$	$\vert Y\vert_{(c)}$
SR		/X/	/X/	/Y/		/X/	/Y/	/Y/		/X/	/Z/	/Y/
PR		{X}	{Z}	{Y}		{X}	{Z}	{Y}		{X}	{Z}	{Y}

First, |X| and $|X|_c$ may both map to /X/ at the surface level, but /X/s in *c* may map unfaithfully to a subsequent phonetic level as {Z}, while /X/s not in *c* are realized as {X}. |Y| is mapped faithfully on both passes, resulting in phonetic {Y}. This possibility is sketched in (16a). Second, $|X|_c$ may map unfaithfully to a surface representation /Y/ which is identical to the (faithful) surface realization of |Y|, but /Y/s from $|X|_c$ are mapped unfaithfully to {Z} while /Y/s from |Y| are mapped to {Y} (16b). Finally, $|X|_c$ may map to a third category /Z/ that is intermediate between /X/ and /Y/, whereafter /X/, /Z/, and /Y/ are all realized faithfully (16c). This intermediate phonological category is in the same spirit as Trubetzkoy (1969) 'archiphoneme' notion, who used it similarly to explain positional neutralization.

Models (16a) and (16b) account for incomplete neutralization by the use of a third level of (phonetic) representation which supplies the in-between realization. As a result, both of these models capture the phenomenon while maintaining a homogeneous set over which underlying and surface representations are spelled out. In contrast, (16c) sees incomplete neutralization as being produced by the 'phonology' proper, in the traditional generative sense of underlying-to-surface mapping. The tradeoff is that this model must admit an element to the alphabet of surface representations that is not present in the alphabet of underlying representations.

One way of evaluating these models is to take stock of the testable empirical predictions they make via psychophysically well-understood perceptual tasks. For instance, Key (2008) has shown

[7] Below I follow Boersma & Hamann (2007) in using pipes to represent underlying forms and forward slashes to represent surface phonological forms. Curly braces enclose phonetic representations.

that contrasting formats of discrimination tasks can produce responses that correspond to the phonological and phonetic levels of encoding of the same acoustic stimuli. In particular, the AX format of discrimination induces responses on the basis of phonological categories while the 4IAX format elicits a pattern of responses based on a lower-level phonetic encoding, a result also observed by Gerrits & Schouten (2004). All three models predict that |X|, $|X|_c$, and |Y| will be discriminable in the 4IAX task since it accesses a level of representation at which the these underlying categories are encoded distinctively. However, only models (16a,b) predict that $|X|_c$ will be confusable with either |X| (16a) or |Y| (16b) in the AX task since it taps phonological encoding and the surface realization of $|X|_c$ is identical to that of |X| or |Y|; model (16c) predicts similar performance in both tasks since |X|, $|X|_c$, and |Y| are distinct from one another at both levels of representation.[8]

Since the issue of incomplete neutralization has not been investigated with respect to coalescence to my knowledge, I'll begin with some discussion of how it has been examined with respect to word-final devoicing processes in languages like German, Polish, and Catalan. Some researchers have argued that 'devoicing' in word-final contexts results in segments that have a different phonetic realization (i.e. incompletely devoiced) from voiceless segments in other contexts. Acoustic and/or perceptual studies of word-final devoicing by Charles-Luce (1985, 1993), Dinnsen & Charles-Luce (1984), Fourakis & Iverson (1984), Jassem & Richter (1989), Mascaró (1987), Port & Crawford (1989), Port *et al.* (1981), Port & O'Dell (1985), and Slowiaczek & Dinnsen (1985) show mixed results; some suggest that word-final devoicing is incompletely neutralizing while others support complete neutralization.

A study by Port, Mitleb & O'Dell (1981) serves as an instructive example of the kind of methodology that would likely be useful in investigating when and if coalescence is completely neutralizing. First, a spectrographic study that examined the production of minimal pairs for voicing in German speakers uncovered the following observations regarding word-final consonants: (i) vowels before underlyingly voiced consonants were about 10% longer than those preceding underlying voiceless consonants; (ii) glottal pulses persisted into the consonant closure for the underlyingly voiced consonants, but not in the underlyingly voiceless ones; (iii) however, consonant closure durations, a typical cue to voicing distinctions, did not differ. Thus, on acoustic grounds, German word-final devoicing is incompletely neutralizing. In addition, Port, Mitleb, and O'Dell conducted a perceptual study to see if these cues would be sufficient for German listeners to correctly identify

[8] It is also possible that performance should be worse in the AX task than the 4IAX task according to model (16c) due to the fact that Z is a category that intervenes in the contrast between |X| and |Y|.

the underlying voicing of consonants in word-final position in minimal pairs for voicing; subjects had to identify which word of the minimal pair was heard. The results show that German listeners identified the word with the expected underlying voicing value on about 70% of trials.

Assuming that this task accesses the phonological level of representation, these results best support a model like the one in (16c), where |+voice| in word-final position is mapped to a category in between /+voice/ and /–voice/. This model could explain why category-based performance was at a level roughly midway between near-perfection, as model (16a) would predict (because word-final |+voice| and |–voice| are as distinct as /+voice/ and /–voice/), and chance, as model (16b) would predict (because word-final |+voice| and |–voice| map to the same surface category).

Evidence for any of the models in (16) suggests broader implications for the architecture of models of production and comprehension. First, if coalescence patterns like German devoicing in favoring a model like (16c), then they together either show that the sets of underlying and surface phonological categories may differ in that the latter contains a superset of the former, or alternatively, and in line with Richness of the Base, that underlying forms may contrast in a greater number of categories than has previously been assumed; for example, /+voice, –voice, ∼voice/ must now be considered.

Second, evidence from coalescence for models (16a) or (16b), in which incomplete neutralization is effected on the mapping to a level of phonetic representation, would make relevant questions about whether the relation between levels in multi-level models of production or comprehension is *serial* or *parallel*. As discussed by Boersma & Hamann (2007), this distinction concerns whether the constraints that produce mappings between the underlying and surface (US) levels interact (in the same constraint hierarchy) with those that produce mappings between the surface and phonetic (SP) levels, assuming the mappings between all levels are constraint-driven, as I do here. In a serial model, the sets of constraints that control mappings between different pairs of levels do not interact, while in a parallel model they do. The serial vs. parallel distinction corresponds to an ongoing debate in the psycholinguistic literature between proponents of *autonomous* models, which are serial in the sense that they posit modularity, and those that support *interactive* models, which are parallel in that they deny modularity. Evidence for models like (16a,b) argues for a serial and thus autonomous model because the surface and phonetic levels differ in whether a particular pair of underlying categories are represented distinctively.

3.2 Coalescence and coarticulation

Another avenue for studying whether some kinds of coalescence have a phonetic basis is to compare it to coarticulation. Research on coarticulation assumes that listeners have evidence for the *presence* of two segments, and thus focuses on how listeners are able to perceptually 'compensate' for the non-canonical realization of each of the segments. For example, [ʃ] coarticulates with [t], which causes the pronunciation of [ʃ] to become more anterior (in the direction of [s]) and the pronunciation of [t] to be more posterior (in the direction of [k]). Mann & Repp (1981) found that English listeners identify members of a [t-k] continuum as "t" more often when it is preceded by [ʃ], than when it is preceded by [s], when they instead respond "k" more often. Kingston *et al.* (2008) show that the effect works in the other direction as well: English listeners label more of a [s-ʃ] continuum as "ʃ" more often than "s" when it precedes members from the [t]-half of a [t-k] continuum, while the identification bias is the reverse in front of members from the [k]-half of the continuum. To explain this effect, Mann and Repp suggest a universal perceptual mechanism known as compensation for coarticulation, which is essentially a process of parsing bits of acoustic information in coarticulated strings and perceptually attributing them to the intended segments.

Many coalescence cases can be thought of in the same way; for instance, coalescence of |ai| to {e} could be analyzed as coarticulation of /a/ and /i/, causing the pronunciation of /a/ to raise and fronten and the pronunication of /i/ to lower. If coalescence is like bona fide coarticulation, and thus subject to perceptual compensation, listeners should identify tokens of {eː} as |ai|. Another reason to compare any such cases of coalescence with coarticulation is the fact that both show locality effects. Unlike many other phonological processes, coalescence is not known to occur between segments that are not stringwise-adjacent. Similarly, the effects of coarticulation are particularly strong between strictly adjacent segments, although several studies report some degree of coarticulation at a distance.

As in the case of incomplete neutralization, I'll begin by discussing another process that has been examined as a possible case of coarticulation, since no study to my knowledge has looked at the question with respect to coalescence. Between-word glide insertion in Dutch is a process that is often regarded as phonological, but has been argued by van Heuven & Hoos (1991) to be a case of coarticulation. They adduce evidence from two experiments to support this claim. First, they ran an acoustic study that showed that the transitional sound that arises in the context of two abutting vowels across a word juncture is different from non-epenthetic Dutch glides (/j/ or /w/). Second, their perception experiment showed that Dutch listeners can distinguish between two vowels joined by the epenthetic glide and two vowels separated by an non-epenthetic glide. As in the German devoicing studies, Van Heuven and Hoos' work supports a model like (16c) that can explain how acoustic cues can influence the perception of an incompletely-neutralized contrast. As Dutch glide insertion bears a closer resemblence to coalescence than German devoicing due to the involvement of two segments under strict locality, the evidence from this study may be more relevant to the study of coalescence.

The first task I would like to do in this regard is to describe the phonetics of various coalescence cases, something which has not been done before to my knowledge, with the aim of determining whether coalescence is ever incompletely-neutralizing. One procedure that might be serve as a good diagnostic for incompletely-neutralizing coalescence would be to measure the duration of segments derived by coalescence and those mapped faithfully (e.g. |ai| → /e/ vs. |e| → /e/) to see if the former category is reliably longer than the latter since it derives from two underlying segments. A second strategy would be to measure the formant trajectories or transitions, depending on whether vowels or consonants are being tested, of coalesced segments and non-coalesced segments of the same category. Evidence of incomplete neutralization in this case would be a formant trajectory/onset transition characteristic of the first underlying segment followed by a formant trajectory/offset transition characteristic of the second underlying segment. If any cases of coalescence are found to be incompletely neutralizing by either of these methods (or others), then I would like to see whether these acoustic differences are perceptible, as measured principally by contrasting discrimination formats or other methods.

Relatedly, Ohala (1994) has looked at how vowel harmony may be the result of the phonologization of a prior process of between-vowel coarticulation. If listeners occasionally misparse distinct phonetic events *a*, *b*, that properly belong to the same phonological unit *A*, as belonging to distinct phonological units *A*, *B*, this parsing may be then adopted in production and then spread through the speech community over time. The mechanism for phonologization is thus errors of *disassociation* (from phonological to phonetic representations). Finding evidence for a close relation between coalescence and coarticulation would suggest that phonologization can also occur via errors of *misassociation*, whereby phonetic events belonging to seperate phonological units are misapprehended as belonging to the same unit. Although disassociation and misassociation may well be mechanisms behind phonologization, a question that has received little or no attention is how these mechanisms should be formalized within a model of comprehension. Proposals exist for explaining variation and optionality in the underlying-to-surface mapping in OT, such as partial ordering of the constraint set (Anttila 1997), stochastic ranking (Boersma 1998 *et seq.*), access to non-optimal candidates (Coetzee 2004), and weighted constraints in Harmonic Grammar (Coetzee & Pater to appear), and I would like to explore whether misparsing mechanisms driving phonologization bear on the choice between these models or motivate new ones.

3.3 Long range goals

In studying the phonetic and perceptual properties of coalescence, I would like to investigate the questions regarding completeness of neutralization, coarticulation, and possibly other generalizations about coalescence using similar experimental methods as described for the cited studies. As hinted at above, a bigger goal of my dissertation work is to refine OT models of production and comprehension that integrate the phonological grammar within a model that relates additional levels of representation, such as the bidirectional models discussed by Boersma & Hamann (2007). In addition, a longe-range goal of the proposed modelling work is to investigate how coalescence might be the phonologization of an earlier coarticulation process.

References

Al-Mozainy, Hamza Q. (1981). *Vowel alternations in a Bedouin Hijazi Arabic dialect: Abstractness and stress*. Ph.D. thesis, University of Texas, Austin.

Anttila, Arto (1997). Deriving variation from grammar. In Frans Hinskens, Roeland van Hout & Leo Wetzels (eds.), *Variation, Change, and Phonological Theory*, 35–68. Amsterdam: John Benjamins.

Boersma, Paul (1998). *Functional Phonology: Formalizing the interactions between articulatory and perceptual drives*. PhD dissertation, University of Amsterdam. The Hague: Holland Academic Graphics.

Boersma, Paul & Silke Hamann (2007). Introduction to *Phonology in perception*.

Casali, Roderic (1996). *Resolving Hiatus*. Ph.D. dissertation, UCLA.

Casali, Roderic (1997). Vowel elision in hiatus contexts: Which vowel goes? *Language* **73**, 493–529.

Causley, Trisha (1997). Faithfulness and contrast: The problem of coalescence. In *WCCFL*, volume 18. CSLI.

Charles-Luce, J. (1985). Word-final devoicing in German: Effects of phonetic and sentential contexts. *Journal of Phonetics* **13**, 309–324.

Charles-Luce, J. (1993). The effects of semantic context on voicing neutralization. *Phonetica* **50**, 28–43.

Chomsky, Noam & Morris Halle (1968). *The Sound Pattern of English*. Cambridge MA: MIT Press.

Clements, G.N. & Kevin C. Ford (1979). Kikuyu tone shift and its synchronic consequences. *Linguistic Inquiry* **10**, 95–108.

Coetzee, Andries (2004). *What it means to be a loser: Non-optimal candidates in Optimality Theory*. Ph.D. dissertation, University of Massachusetts, Amherst.

Coetzee, Andries & Joe Pater (to appear). The place of variation in phonological theory. In John Goldsmith, Jason Riggle & Alan Yu (eds.), *The Handbook of Phonological Theory*. Blackwell, 2nd edition.

de Lacy, Paul (2002). *The Formal Expression of Markedness*. Ph.D. dissertation, University of Massachusetts, Amherst.

Dinnsen, D. & J. Charles-Luce (1984). Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics* **12**, 49–60.

Fourakis, M. & G. Iverson (1984). On the 'incomplete neutralization' of German final obstruents. *Phonetica* **41**, 140–149.

Gerrits, E. & M.E.H. Schouten (2004). Categorical perception depends on the discrimination task. *Perception and Psychophysics* **66**, 363–376.

Goldsmith, John (1976a). *Autosegmental Phonology*. Ph.D. dissertation, MIT. Distributed by Indiana University Linguistics Club.

Goldsmith, John (1976b). An Overview of Autosegmental Phonology. *Linguistic Analysis* **2**, 23–68.

Halle, Morris & G.N. Clements (1983). *Problem Book in Phonology*. Cambridge, MA: MIT Press.

Jassem, W. & L. Richter (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics* **17**, 317–325.

Keer, Edward (1999). *Geminates, The OCP, and the Nature of CON*. Unpublished Ph.D. dissertation, Rutgers University, New Brunswick, NJ.

Key, Michael (2008). Interactive and autonomous modes of speech perception: Phonological knowledge and discrimination in English and French listeners. Paper presented at Labphon 11, Victoria University of Wellington.

Kingston, John, Shigeto Kawahara, Della Chambless, Michael Key & Sarah Watsky (2008). The independence of auditory and categorical effects on speech perception. Poster presented at LabPhon11, Victoria University of Wellington.

Lamontagne, Greg & Keren Rice (1995). A correspondence account of coalescence. In Laura Walsh Dickey Jill Beckman & Suzanne Urbanczyk (eds.), *University of Massachusetts Occasional Papers*, volume 18, 211–224. GLSA.

Li, Fang-Kuei (1946). Chipewyan. In C. Osgood & H. Hoijer (eds.), *Linguistic Structures of Native America*, 398–423. The Viking Fund.

Mann, Virginia A. & Bruno H. Repp (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America* **69**, 548–558.

Mascaró, Joan (1987). Underlying voicing recoverability of finally devoiced obstruents in Catalan. *Journal of Phonetics* **15**, 183–186.

McCarthy, John J. (2006). Restraint of analysis. In Eric Baković (et al.) (ed.), *Wondering at the Natural Fecundity of Things: Essays in Honor of Alan Prince*. Santa Cruz, CA: Linguistics Research Center. [Published electronically at http://repositories.cdlib.org/lrc/prince/].

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

McCarthy, John J. (2008). Deletion is reduction. Unpublished manuscript, University of Massachusetts, Amherst.

McCarthy, John J. (to appear). The gradual path to cluster simplification. *Phonology* .

McCarthy, John J. & Alan Prince (1995). Faithfulness and Reduplicative Identity. In Jill N. Beckman, Laura Walsh & Suzanne Urbanczyk (eds.), *Papers in Optimality Theory*, University of Massachusetts Occasional Papers 18, 249–384. University of Massachusetts, Amherst: UMOP.

Odden, David (1991). Vowel geometry. *Phonology* **8**, 261–289.

Ohala, John (1994). Towards a universal, phonetically-based, theory of vowel harmony. In *Proceedings of ICSLP 3*.

Padgett, Jaye (1995). Feature classes. In Dickey Laura Walsh Beckman, Jill N. & Suzanne Urbanczyk (eds.), *UMOP*, volume 18, 385–420. GLSA.

Pater, Joe (1999). Austronesian nasal substitution and other *NC effects. In René Kager, Harry van der Hulst & Wim Zonneveld (eds.), *The Prosody–Morphology Interface*. Cambridge: Cambridge University Press. ROA 160-1196.

Port, R. & P. Crawford (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics* **17**, 257–282.

Port, R., F. Mitleb & M. O'Dell (1981). Neutralization of obstruent voicing in German is incomplete. *Journal of the Acoustical Society of America* **70**, S10.

Port, R. & M. O'Dell (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics* **13**, 455–471.

Prince, Alan & Paul Smolensky (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell. Previously distributed as Rutgers University RuCCS Technical Report 2.

Schane, Sanford (1984). The fundamentals of particle phonology. *Phonology* **1**, 129–155.

Slowiaczek, L. & D. Dinnsen (1985). On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics* **13**, 325–341.

Trubetzkoy, N.S. (1969). *Principles of Phonology*. Berkeley: University of California Press. Translated from *Grundzüge der Phonologie* (1939).

van Heuven, Vincent J. & Annalies Hoos (1991). Hiatus Deletion, Phonological Rule or Phonetic Coarticulation? In Frank Drijkoningen & Ans van Kemenade (eds.), *Linguistics in the Netherlands 1991*, 61–70. Amsterdam: John Benjamins.

Wheeler, Max (2005). Cluster reduction: Deletion or coalescence? *Catalan Journal of Linguistics (Special issue on morphology in phonology)* **4**, 57–82.