Loanwords and the Perceptual Map: A Perspective from MaxEnt Learning

by

Erin Olson

B.A., McGill University (2012)

Submitted to the Department of Linguistics and Philosophy
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2020

Author ....
Department of Linguistics and Philosophy
September 9, 2020

Certified by ....
Adam Albright
Professor of Linguistics
Thesis Supervisor

Certified by ....
Michael Kenstowicz
Professor of Linguistics
Thesis Supervisor

Certified by ....
Donca Steriade
Professor of Linguistics
Thesis Supervisor

Accepted by ....
Kai von Fintel
Andrew W. Mellon Professor of Linguistics
Linguistics Section Head

Loanwords and the Perceptual Map: A Perspective from MaxEnt Learning

by

Erin Olson

Submitted to the Department of Linguistics and Philosophy
on September 9, 2020, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This dissertation examines the predictions of two computational models of grammar within the domain of loanword phonology. These models, formulated within a Maximum Entropy (MaxEnt) framework, have been shown to be successful when simulating the effects that a substantive bias such as the Perceptual Map (PMap) hypothesis of Steriade (2001) may have on a phonological learner. While previous studies have focused primarily on modelling data taken from artificial grammar learning experiments (Wilson, 2006; White, 2013), this dissertation will instead model loanword adaptation. Loanword adaptation was chosen as a useful test domain as speakers will often choose to repair phonotactically-illicit loanwords in ways that are not attested in their native grammar. It thus provides a wealth of data about how speakers structure their grammar in the absence of overt phonological evidence. To this end, a case study of English loanword adaptation in Cantonese is undertaken. It will be shown that the patterns of consonant deletion and vowel epenthesis used by speakers of Cantonese to adapt English words are compatible with the PMap, and can be modelled through the MaxEnt learners mentioned above. It will also be shown through a series of computational simulations that Wilson's (2006) learner fails to acquire the grammar necessary to account for the patterns of loanword adaptation, while White's (2013) learner succeeds. This is a result of the way in which the PMap is encoded within these learners. While both encode the PMap as a series of asymmetrical Gaussian distributions on the weights of constraints, Wilson (2006) encodes this asymmetry through the variances, or plasticities, of the distributions, while White (2013) encodes it through the means, or target weights. A grammar which encodes the PMap through asymmetrical plasticities must encounter evidence from the phonology of the language in order to alter the weights of constraints. However, the loanword phonology of Cantonese crucially lacks such phonological evidence, and Wilson's (2006) model cannot make use of it when establishing constraint asymmetries. White's (2013) model, however, allows constraint asymmetries to be maintained in the absence of overt evidence, and results in more accurate grammars of Cantonese loanword adaptation.

Thesis Supervisor: Adam Albright
Title: Professor of Linguistics

Thesis Supervisor: Michael Kenstowicz
Title: Professor of Linguistics

Thesis Supervisor: Donca Steriade
Title: Professor of Linguistics

Acknowledgments

First and foremost, this dissertation would not have existed were it not for the encouragement, insight, and patience of my committee members, Adam Albright, Michael Kenstowicz, and Donca Steriade. Talking with you all was always an immense help, and I walked away from each meeting with a better and broader understanding of the questions at hand and new avenues of research to pursue.

I would also like to thank the other faculty members of the Linguistics Department who helped me feel welcome and supported in my research throughout my time at MIT. Thanks to Edward Flemming, Norvin Richards, and Kai von Fintel for lending me and my peers an enthusiastic and compassionate ear when it came to all things academia. Thanks also to Shigeru Miyagawa and Michel DeGraff for being on my generals committee, and helping me rise to meet your expectations.

Thank you to those who helped me decide to pursue this degree in the first place. Thanks to Martin Hackl, Michael Wagner, Meghan Clayards, Sarah Mackenzie, Jessica Coon, and Ayaka Sugawara for providing me with the opportunities to expand my horizons prior to my degree, and for your enthusiasm in helping me take my first steps towards this dissertation. Thanks also to Heather Goad, whose overwhelming excitement for phonology and phonological theory solidified my decision to study linguistics in the first place.

Thank you to everyone who has been my peer throughout my time as a graduate student. You all helped so much to make me feel welcome, and I am forever grateful to those of you who lent an ear or a hand when I was feeling frustrated or anxious, or when I was feeling particularly excited about a topic of research. I am also grateful to those of you who simply made me feel included, because sometimes it is the smallest things which help the most. Thanks to my officemates throughout the years (Juliet Stanton, Despina Oikonomou, Amanda Swenson, Ishani Guha, Snejana Iovtcheva, Newell Lewey, Chris Baron, Felipe Hisao Kobayashi, Boer Fu, Maša Močnik), to my fellow student organizers (Naomi Francis, Michelle Yuan, Athulya Aravind, Nick Longenbaugh), to my fellow phonologists (Daniel Asherov, Christopher Yang, Michelle Fullwood, Cora Lesure, Rafael Abramovitz, Ezer Rasin, Hyun Ji Yoo, Suyeon Yun, Sam Zukoff), to those of you who always had interesting phonological questions (Elise Newman, Danfeng Wu, Neil Banerjee, Yadav Gowda), and to those of you who were always willing to drop by the office for a casual chat (Verena Hehl, Daniel Margulis, Ömer Demirok, Peter Grishin, Agnes Bi, Sherry Chen).

Thanks also to my friends outside the department who were there for me when things were at their darkest. Thanks to my roommates Abby Johnson, Leah Arnold, and Sunny He, and to my other friends in Boston, Amy Goodnough, Laina, and Mark Leach for making me feel welcome in this city, and for reminding me that sometimes we all need a break from the ivory tower. Thanks to Elise McClay, Yasha Frankel, and John-Erik Hansson for your continued friendship and support from time zones across the globe.

A special thanks to my parents, Barbara and Scott, for being so patient and understanding while I navigated this strange new world, and to my brother Corey for picking his own path in life and inspiring me to continue on with the one I chose.

To Claudia Lo, my partner over the last three years, thank you so much for your support and perspective while I finish this long journey and start a new one with you. The world is not always what we want it to be, but we have each other, and I would not want it any other way. Thanks also to our cat CousCous and her helpful additions to this thesis, such as “po0p[;l”, “,.......luij”, and “tygrt”.

Finally, thank you to Dr. Amy Mitchell, for being the mentor I needed for all six of my years in Boston. Thank you so much for all your time, concern, patience, and perspective while I did the hard work of teasing apart my own Gordian knots. This dissertation would not have been written without you.

Contents

List of Figures

Chapter 1

Introduction

The **Perceptual Map** (PMap) of (Steriade, 2001) has long been recognized as a method of solving the "too many solutions" problem that is inherent to constraint-based frameworks of phonological theory. This theory has also led to fruitful research within experimental paradigms such as artificial grammar learning, where participants are tasked with acquiring a variety of small grammars in order to see how successful they can be acquired or how quickly they can be learned. The PMap has proved useful here in accounting for patterns of phonological generalization (Wilson, 2006) and biases against certain patterns of segment-to-segment correspondence (White, 2013). The experimental data obtained through these experiments have, in addition, been shown to be adequately modelled through use of **Maximum Entropy** (MaxEnt) grammars. These grammars provide a convenient framework for modelling the effects of the PMap as they make incorporation of the PMap into the learning algorithm quite simple. Furthermore, their computational properties are well-understood, both inside and outside of linguistics (Della Pietra et al., 1997; Hayes and Wilson, 2008, *a.o.*).

However, while use of these computational tools has been shown to be promising when examining how learners of a language behave when presented with **overt** evidence of a particular phonological process, they have not been tested on more extreme cases related to phonological learning. The models used in the experiments above make differing predictions about how speakers will behave in the **absence** of overt phonological evidence – either at the beginning of the learning process, or after a learner has acquired the full phonology of their language. As there does not exist any language which makes use of the full range of phonological segments, restrictions, or processes, the final language learned will, by necessity, contain some gaps. What might happen, then, if a speaker is tasked with producing a word that exists in one of these gaps?

One domain of phonology where this exact situation is encountered is **loanword adaptation**.

Speakers who may not encounter words with, for instance, particular combinations of adjacent consonants, must nevertheless attempt to produce these kinds of structures within their own language, and make the word sound as natural as possible. They are known to employ a variety of different phonological repairs – some of which are **unattested** in their native language – in order to accomplish this task.

The adaptation of loanwords has long been hypothesized to be a result of the speaker attempting to best match their **perception** of the loanword onto the native phonology (Kenstowicz, 2005, 2007; Adler, 2006; Shinohara, 2006; Peperkamp et al., 2008; Boersma and Hamann, 2009, *a.o.*). More specific hypothesis assert that this perceptual matching is achieved through the PMap itself (Adler, 2006; Shinohara, 2006; Kenstowicz, 2007). If this is the case, then making use of loanword data should prove to be a fruitful line of inquiry for examining the behaviour of MaxEnt learners which incorporate a PMap bias.

The goal of this dissertation is to begin this line of inquiry by examining a case study in loanword adaptation through the lens of MaxEnt learning. A series of computational experiments were performed in order to assess the performance of the MaxEnt learners detailed in Wilson (2006) and White (2013) when tasked with acquiring the sophisticated patterns of English loanword adaptation present in Cantonese (Silverman, 1992; Yip, 1993, 2006; Kenstowicz, 2012). While the model used in White (2013) was shown to perform better on this task than the model used in Wilson (2006), the experiments outlined here have shown that there is much more to be understood about how these speakers make use of the PMap, and much more to be understood about how to properly implement the PMap in a computational model. As a starting point, I will review how MaxEnt grammars function within phonology, and I will provide a sketch of how the two proposed computational models of the PMap function within this framework.

1.1 Maximum Entropy Grammars

A series of MaxEnt grammars (Hayes and Wilson, 2008, *a.o.*) will be produced and examined in detail throughout the course of this thesis. These grammars, their associated learning algorithms, and how they compare to other constraint-based phonological grammars are discussed in detail in Chapter 2, and any reader who is interested in the exact mechanics of these grammars and their associated learning algorithms is encouraged to consult this chapter. For those who only require a brief overview, a MaxEnt grammar is similar to one constructed in **Harmonic Grammar** (HG) (Legendre et al., 1990) in that it is characterized by a unique set of constraint **weights**. In order to evaluate a set of potential phonological outputs for a given input, the weight of each constraint is **multiplied** by the violations that the output candidate incurs, and each of these values is then **summed** to determine the overall penalty incurred by that output candidate. Where MaxEnt deviates from HG is in how these penalties are treated. In HG, they are simply compared, while in MaxEnt, they are instead converted into output candidate **probabilities**, such that the output candidate with the smallest penalty receives the most probability of being selected as the correct output.

Converting penalties into probabilities in this way allows MaxEnt grammars to be able to accommodate **variable** language data, where multiple outputs are attested for a given input. This feature of the MaxEnt framework also allows researchers to make use of general-purpose probabilistic **learning algorithms**, such as that of Della Pietra et al. (1997). The main goal of such algorithms is to **maximize** the probability of each of the attested output forms, while **minimizing** the probability of the unattested candidates.[1]

In addition to being able to accommodate and learn from probabilistic inputs, MaxEnt learning algorithms also have the advantage that they can quite easily incorporate **additional grammatical requirements**. As the grammar deals directly in probabilities, other probabilistic requirements can simply be **added** to the function which defines how probability is calculated in MaxEnt. The most common requirement is that the resulting weights should fall within some pre-determined **range** of values, centred around some **average** or **target** value. This is, in fact, standard practice for most MaxEnt learners, such as those used in Hayes and Wilson (2008), in order to prevent the resulting grammar from **overfitting** to the data it is acquiring. These requirements are most commonly incorporated as **Gaussian** distributions, or **bell curves**, the properties of which are well-understood within statistics. The average or target value corresponds to some default value the constraint should attain the absence of evidence, and the range is determined by the **variance** or overall width of the bell curve. The flatter the curve, the larger the range considered by the learner throughout the acquisition process.

This process of adding an additional requirement on the permissible **range** of possible weight values is also known as **bias incorporation**. The learner can be considered to be biased in the sense that it prefers grammars which meet this requirement and will rule out those which do not. Of particular interest here is that this particular bias does not have to be **uniform** across all constraints. Rather, it is equally possible for **each constraint** to be assigned its own bias, where the acceptable

[1] The algorithm will always assign as **uniform** a probability as it can, given the data it is attempting to acquire. This preference for uniform probability is also known as **entropy** – hence the name Maximum Entropy grammar.

ranges and target values can be **different** for each. This is, in fact, what is undertaken by Wilson (2006) and White (2013) when modelling how the PMap of Steriade (2001) can be incorporated into a MaxEnt learner.

The PMap and its importance with respect to loanword data will be discussed in some detail in the following section, and more extensively in Chapter 3. It is of interest to researchers who make use of MaxEnt learners because it provides a method by which **principled asymmetries** between constraint weights can be established. Furthermore, as the PMap is considered to arise from the **perceptual system** of speakers, as opposed to the grammatical system, it is a good candidate for being modelled via bias incorporation. As discussed above, the most common way of incorporating a bias into a MaxEnt learner is by establishing a Gaussian distribution over the weight values that a constraint is allowed to take. This Gaussian is defined by **two** parameters – the mean, or target value; and the variance, or size of the acceptable range of weights. There are thus at least two possible ways of establishing an asymmetrical, PMap-like bias over the set of constraints: either the target weights can be assigned asymmetrically, or the ranges can change between constraints.

In fact, both of these approaches have been implemented for MaxEnt learners. The first such bias was implemented by Wilson (2006), who chose to vary the **range**[2] of possible values between constraints which are considered to be governed by the PMap. This meant that some constraints were **more capable** of achieving weights that were substantially different from their target weight than others. Those which had smaller ranges and were less capable of varying from the target weight thus had more **rigid** weighting requirements, and those which had larger ranges had more **plastic** requirements. This notion of the **plasticity** of a constraint will be referenced throughout this thesis, and the bias used by Wilson (2006) will be referred to as a **plasticity** bias or a **plasticity-based** bias.

The alternate approach was undertaken by White (2013), where the asymmetry between the constraints governed by the PMap was established over the **target weights** of each constraint. This meant that some constraints were predicted to have **low weights** "by default", or in the absence of overt phonological evidence, while others were predicted to have **high weights**. As the asymmetry for White's (2013) bias was established directly over the weights of constraints, it will be referred to as a **weight** bias or a **weight-based** bias throughout this thesis.

Both approaches outlined here made use of only **one** method of establishing asymmetrical biases

[2]It should be noted that the variance of a Gaussian distribution centred on some target value is not the same as imposing a requirement on the range of values that a value may take. Constraints are still capable of achieving weights outside the desired range, but the variance will impose a much harsher penalty on those that do. When providing readers with an intuition of how the biases function, however, I have chosen to use the term "range" here to make reasoning about the constraint asymmetries easier.

on constraint weights. In other words, if one parameter varied according to which constraint was being discussed, the other parameter remained **constant** across all constraints. Thus, in Wilson's (2006) plasticity-based bias, the target weight of all constraints was set uniformly at 0.0; and in White's (2013) weight-based bias, the plasticity or range of all constraints was set uniformly as well.

This choice, as will be shown through the results obtained by the MaxEnt learners in Chapter 4, is not without consequence. While both biases are equally capable of predicting how a MaxEnt learner can acquire a grammar with an asymmetrical weighting of constraints in the **presence** of phonological evidence, they make vastly different predictions about how a learner will behave in its **absence**. This is ultimately due to how the MaxEnt learning algorithm of Della Pietra et al. (1997) interacts with the two biases proposed by Wilson (2006) and White (2013). In order for the weight of any constraint to **differ** from the target introduced as part of the bias, the learner must encounter **direct evidence** about the weight of that constraint from the language it is acquiring. Under White's (2013) weight-based bias, this means that the final grammar will contain asymmetrical weights for constraints governed by the PMap, modulo any evidence to the contrary. However, under Wilson's (2006) plasticity-based bias, this means that the final grammar will assign an identical weight of 0.0 to all constraints which are not referenced during the acquisition process.

The studies undertaken in Wilson (2006) and White (2013) are not concerned with this issue, as they aim to model how learners differ in their acquisition of overt phonological patterns. This thesis aims to fill this gap in knowledge by examining how these models perform when tasked with modelling patterns of **loanword adaptation**. As will be discussed in the following sections, some instances of loanword adaptation cannot be accounted for by simple application of phonological processes from the host grammar to the novel word. Rather, the speakers of the language must have access to some **additional source** of knowledge which allows them to posit novel phonological processes that are applicable to loanwords. In other words, these speakers must have arrived at these repairs in the **absence** of overt phonological evidence from their own language. This is the exact state of affairs for which the MaxEnt learners of Wilson (2006) and White (2013) make different predictions.

This thesis will be concerned with **applying** both of these MaxEnt learners to **Cantonese**, a language which is known to have a sophisticated set of epenthesis and deletion repairs to English loanwords (Silverman, 1992; Yip, 1993, 2006; Kenstowicz, 2012). Each learner will be tasked with acquiring the native Cantonese vocabulary, and will then be **tested** on how well each predicts the patterns of adaptation seen in English loans. The results of each of these tests is presented in Chapter 4. This chapter will show that the grammar used by Wilson (2006) fails to predict the correct series of repairs, while the grammar used by White (2013) is substantially more successful. In order to gain a better intuition of how the experimental paradigm was structured, the following section will go into detail about how the study of loanwords can aid in the study of the PMap and other asymmetrical biases. The section after will go into detail about the specific repairs seen in Cantonese, and how these can be used to make inferences about the PMap.

1.2 Loanwords

The process of **loanword adaptation** from one language to another provides a rich source of information about aspects of grammar which might otherwise have been obscured by the language in general. As no two languages have the exact same phonology, words from one language will occasionally have to be **repaired** when they are incorporated into a second language. In some instances, these repairs are predictable from the morpho-phonology of the language. For example, a language which already contains a process of vowel epenthesis to repair consonant clusters which occur as the result of combining a root morpheme with an affix might then choose to repair unattested consonant clusters in loanwords in the same manner. These speakers would have had **direct evidence** from their language about how to repair consonant clusters, and would be able to use that evidence to acquire the constraint weighting necessary to predict it.

However, the kinds of loanword adaptations which are of interest here are those where speakers apply a repair that is used **nowhere else** in the language. Thus, a learner of this language must be using an **alternate source** of knowledge to construct this repair (Uffmann, 2015). These cases are of special interest because they are the ones for which MaxEnt learners equipped with asymmetrical constraint weight restrictions will predict different outcomes. As discussed above, learners equipped with a plasticity-based asymmetrical bias are not predicted to be able to replicate these patterns of loanword adaption, while those equipped with a weight-based asymmetrical bias will.

The source of the biases proposed by Wilson (2006) and White (2013) is the **PMap** of Steriade (2001). The PMap, as originally formulated within classical OT (Prince and Smolensky, 1993), is a way of translating what speakers know about the **perceptual discriminability** of potential repairs into **hierarchical rankings** of Faithfulness constraints. Say, for instance, that a speaker is presented with a loanword which contains a consonant cluster in onset position, but their language does not allow consonant clusters that occur within syllables. A number of potential repairs might be considered here, such as deletion of one member of the cluster, or epenthesis of a vowel between both members of the cluster.

(1) **Potential repairs to an illicit onset cluster** C_1C_2V

a. C_1V

b. C_2V

c. C_1VC_2V

The PMap allows for a **principled way** of selecting which repair should be preferred when **no evidence** is available from the language being studied. The basic mechanism by which this is done is to **compare** how different the repair is to the original structure in terms of perceptibility. Naturally, there will be differences in pronunciation and perception between a word with an onset cluster and a word with a singleton onset. However, there may be a greater difference in pronunciation and perception if the first member of the cluster is deleted than if the second member is deleted. This difference can then be used to postulate a difference in the ranking of Faithfulness constraints, such that the constraint which governs the deletion of the first consonant, MAX-C_1 **outranks** the constraint which governs the second, MAX-C_2. Within a MaxEnt grammar, this would be translated into an asymmetrical **weighting** of constraints, such that the weight of MAX-C_1 **exceeds** the weight of MAX-C_2. Readers who are interested in a more in-depth discussion of the PMap should consult Chapter 3.

The PMap has been proposed as the specific mechanism by which speakers of a language choose to adapt loanwords in a variety of papers (Adler, 2006; Kenstowicz, 2005, 2007; Shinohara, 2006), although many others have proposed that perceptual similarity plays a crucial role regardless (Peperkamp et al., 2008; Boersma and Hamann, 2009). Kenstowicz (2007) examines a series of English loans into Fijian, and notes that many of the phenomena present within the loans can be explained if speakers are taking into account the perceptual **similarity** of illicit phonological structures contained within the loans and the phonologically-licit structures of their own language. For example, as Fijian does not have voiced stops, any English loanwords which contain such segments must be repaired in some way. There are two attested kinds of repairs to these segments in Fijian, exemplified by the data in (2).

(2) **Attested repairs to voiced stops in Fijian** (Kenstowicz, 2007)

English Loan	*Fijian Adaptation*	*Gloss*
dɛsk	**ⁿd**esi	‘desk’
ɹæ**b**aj	ra**ːp**ai	‘rabbi’

This asymmetry in repairs can be explained by a difference in the perceptibility of prenasalized stops in Fijian. This is due to the fact that the nasal portion of the prenasalized stop is quite **weak** in onset position, so that it is very hard to distinguish between the English voiced stop and an onset prenasalized stop. However, in word-medial position, the difference between a voiced stop and a prenasalized stop is much stronger, as there is a preceding vowel which can better host the nasal portion of the Fijian prenasalized stop. As such, a **different** repair must be chosen in order to produce English loanwords with voiced stops. This asymmetry in the perceptibility of prenasalized stops, under the PMap, will lead to an asymmetry in the weights[3] of their governing constraints, such that the constraint which governs a change in nasalization in word-medial position will receive a **higher weight** than the constraint which governs a change in nasalization in word-initial position.

Any language for which the PMap is proposed as a source of loanword repairs will be one that is a good candidate for testing the proposals of Wilson (2006) and White (2013). However, locating such cases is not quite as simple as locating a repair strategy which is otherwise unattested in the language. For instance, vowel epenthesis is often attested as a repair to consonant clusters, even when the language does not show evidence of such a repair in its morpho-phonological system. This has been proposed as a **default** repair to consonant clusters, since a language will attempt to **preserve** as much of the incoming loanword as possible (Paradis and LaCharité, 1997). While this isn't necessarily a bias that can easily be accounted for by the PMap, it is a bias which might be incorporated into a MaxEnt grammar via asymmetric restrictions on constraint weights. However, it has been shown not to be universal, and may instead be due to a difference in the way in which loanwords are first encountered. Loanwords which enter the language via the **orthography** are more likely to show vowel epenthesis as a repair to consonant clusters, while those which enter the language via **speech alone** are more likely to show deletion repairs (Smith, 2006; Dohlus, 2010). Without an understanding of how loanwords are first encountered in the language, establishing such a bias for any given language is unwarranted.

The phonotactics of the language may also **prevent** certain repair options from being considered. As discussed in Pakendorf and Novgorodov (2009); Vasilyeva (2010) and Yun (2016), Yakut shows an asymmetry in the site of vowel epenthesis within loanwords. This is shown by the data in (3), where epenthesis is internal within obstruent-sonorant clusters, but external for fricative-stop clusters.

(3) **Positional asymmetries in vowel epenthesis in Yakut** (Vasilyeva, 2010; Yun, 2016)

a. *Internal epenthesis in obstruent-sonorant clusters*

Russian Loan	*Yakut Adaptation*	*Gloss*
klad	**kɯl**aːt	‘treasure’
platʲə	**bɯl**aːçːa	‘dress’
sluz̧ba	**sul**uːspa	‘service’
kora**blʲ**	xaraː**bɨl**	‘ship’

[3]Assuming that the framework used to construct the grammar is a MaxEnt or HG framework, rather than the classical OT framework used in Steriade (2001).

b. *External epenthesis in fricative-stop clusters*

Russian Loan	*Yakut Adaptation*	*Gloss*
stol	**ost**uol	‘table’
ʃkaf	**ɯsk**aːp	‘closet’
spitsɨ	**isp**iːsːe	‘knitting needles’

It might be tempting to postulate that this positional asymmetry could be due to some asymmetrical perceptual difference between fricative-stop clusters and obstruent-sonorant clusters. However, as noted by Yun (2016), Yakut **disallows** obstruent-sonorant clusters **throughout** the language, even though consonant clusters are otherwise allowed. This is shown by the loanword data in (4), where word-medial obstruent-sonorant clusters from Russian loanwords are still repaired by epenthesis.

(4) **Epethesis in word-medial clusters in Yakut** (Pakendorf and Novgorodov, 2009; Vasilyeva, 2010; Yun, 2016)

Russian Loan	*Yakut Adaptation*	*Gloss*
ku**kl**a	ku**ːkul**a	‘doll’
kre**sl**o	kirie**hil**e	‘armchair’
ka**bl**uk	χo**bul**uk	‘heel’

It is therefore premature to consider that every asymmetry in loanword adaptation is always due to some asymmetry present within the set of constraints governed by the PMap or some other substantive bias. Instead, the phonotactics of the language may be the source of such asymmetries.

However, this is not to say that finding such cases is impossible, merely that caution should be taken in doing so. The following section will present the reader with a selection of loanword data from Cantonese, and show how adaptations in this language arise not through the native grammar of Cantonese, but instead through rankings or weightings which can be inferred directly from the PMap.

1.3 Loanwords in Cantonese

This thesis will primarily be concerned with exploring how loanwords can be modelled within a MaxEnt grammar through the lens of a case study of Cantonese. As has been observed by many researchers before, Cantonese disallows both **consonant clusters** and **sibilants in coda position** (Hashimoto, 1972; Bauer and Benedict, 1997, *a.o.*). Nevertheless, speakers of Cantonese have a sophisticated system of repairs to loanwords which is dependent both upon the **position** of the illicit structure and the **segmental makeup** of the structure (Hashimoto, 1972; Silverman, 1992; Yip, 1993, 2006; Bauer and Benedict, 1997; Kenstowicz, 2012, *a.o.*). Each of these will be discussed in turn.

In general, when a cluster is **word-initial**, either epenthesis or deletion is possible, often within the same lexical item. Within the database assembled from data in Silverman (1992); Yip (1993, 2006) and Kenstowicz (2012), presented together in Appendix A, each repair is roughly equivalently observed. Deletion of the second consonant occurs in 22 items and epenthesis of a vowel between the two consonants occurs in 20 items. Some examples are provided in (5). Note that all English transcriptions match those of RP and not any variety of North American English, as this was assumed to be the variety of English that speakers of Cantonese would have the most exposure to.

(5) **Attested repairs to word- or syllable-initial consonant clusters**

a.	ˈ**pʰɹ**ɒm	→	**pʰ**ɛ̄́ŋ	‘prom’ (Yip, 2006, p.953)
	ˈ**fɹ**iːzə	→	**f**ísā́ː	‘freezer’ (Silverman, 1992, p.290)
	ˈ**kʰl**ɪk	→	**kʰ**ék	‘click’ (Kenstowicz, 2012)
	ˈ**bl**ɛndə	→	**p**ɛ́ntā́	‘blender’ (Silverman, 1992, p.318)
	ˈvɑ**lj**uːm	→	wɔ́ː**l**ɛ̄́m	‘volume’ (Silverman, 1992, p.299)
	ˈ**kʰɹ**iːm	→	**kʰ**íːm	‘cream’ (Yip, 2006, p.953)
b.	ˈ**pʰɹ**ɪnt	→	**pʰìl**ín	‘print’ (Silverman, 1992, p.290)
	ˈ**fɹ**æŋk	→	**fàtl**ɔ́ːŋ	‘frank’ (Kenstowicz, 2012)
	ˈ**kʰl**iːn	→	**kʰìl**ín	‘clean’ (Silverman, 1992, p.324)
	ˈ**bl**ɒnd	→	**pìl**án	‘blonde’ (Silverman, 1992, p.317)
	ˈ**kʰɹ**iːm	→	**kʰèjl**íːm	‘cream’ (Silverman, 1992, p.290)

In order to account for the data above, a MaxEnt grammar must penalize vowel epenthesis and deletion of the second consonant roughly **equally**. One grammar which is capable of making this distinction – or one which is sufficiently close to it – is given in (6). While the constraints used below are not necessarily those used throughout the remainder of this thesis, they give the reader a good idea of the kinds of constraints considered. The constraints make use of shorthand to characterize classes of segments with identical sonority: T stands for non-sibilant obstruents, S for sibilants, R for sonorants, C for all consonants, and V for vowels.

(6) **Accounting for initial cluster repairs**

/ˈkʰɹiːm/	p (%)	*Complex $w = 10$	Max-T/_R 10	Max-R/_V 4	Dep-V/_C 3	Dep-V/C_ 2
a. [kʰlíːm]	0.335	1	0	0	0	0
b. [kʰíːm]	49.665	0	0	1	0	0
c. [líːm]	0.335	0	1	0	0	0
d. [kʰèjlíːm]	49.665	0	0	0	1	1

The tableau above is an example of a MaxEnt tableau. Those readers wishing to gain a better understanding of how such a tableau works should consult Chapter 2. For now, it will be sufficient to understand that the probability p assigned to each candidate is calculated from the sum of weighted constraint violations, where the weight w is presented below the name of each constraint. For the sake of clarity, all constraints will be listed in **descending** order by weight. In order to illustrate the most notable weights, some constraints will be **left out** of the tableau, although they are considered when calculating the total amount of probability assigned to each candidate.

For the tableau in (6), it should be noted that in addition to the highly-weighted constraint *Complex, the constraint Max-T/_R also receives a high weight, one which is much higher than Max-R/_V. This weight difference will account for why the **initial** segment is selected when the structure is repaired via deletion. The **combined** weight of the Dep-V constraints is close to being **equivalent** to the weight of the constraint Max-R/_V, which leads to the grammar's uncertainty about which repair to prefer – epenthesis, or deletion.[4]

When a cluster is **word-final**, the second member of the cluster is almost always **deleted** (except when that segment is a sibilant — see data to follow).

(7) **Attested repairs to word- or syllable-final consonant clusters**

ˈpʰaʊ**nd**	→	pʰɔ́ː**ŋ**	'pound' (Kenstowicz, 2012)
ˈlɛ**ŋθ**	→	lɛ́ː**n**	'length' (Silverman, 1992, p.324)
ˈlɪ**ft**	→	líː**p**	'lift' (Silverman, 1992, p.299)
ˈsʌ**ɪd**ˌbɔːd	→	sá**j**púːt	'sideboard' (Silverman, 1992, p.297)
əˈsʌ**ɪn**mə**nt**	→	āːsá**j**mɔ̃́**n**	'assignment' (Silverman, 1992, p.308)

These data can be accommodated in a MaxEnt grammar by the constraints and weights listed in (8). Note that the accommodation will not be **perfect**, but that the attested repair is instead assigned the majority of the probability available for the tableau. This will be considered to be an adequate measure of **accuracy** throughout the thesis. More discussion of the accuracy measures used is provided in Chapter 4.

(8) **Accounting for final cluster repairs**

/ˈlɛŋθ/	p (%)	*Complex $w = 10$	Dep-V/_C 3	Dep-V/C_ 2	Max-C/V_ 2	Max-C/C_ 1
a. [lɛ́ːnt]	0.007	1	0	0	0	0
b. [lɛ́ːn]	52.923	0	0	0	0	1
c. [lɛ́ːt]	19.469	0	0	0	1	0
d. [lɛ́ːntíː]	19.469	0	0	1	0	0
e. [lɛ́ːníːt]	0.969	0	1	1	0	0
f. [lɛ́ːníː]	7.162	0	0	1	0	1

The asymmetries which help account for why deletion is the attested repair lie in the higher overall weight of the Dep-V constraints when compared with the Max-C constraints. The further asymmetry which accounts for the selection of the second member of the cluster for deletion resides in the higher weight of Max-C/V_ when compared with Max-C/C_.

When it comes to words or syllables that end in a **sibilant** – also banned in native Cantonese words – speakers favour **epenthesis** of a vowel after the sibilant.

[4]The reason they are exactly equivalent is due to the effect of a very low-weighted constraint Max-C/C_, to be discussed below.

(9) **Attested repairs to word- or syllable-final sibilants**

ˈbʌ**s**	→	páː**sī́**	‘bus’ (Silverman, 1992, p.300)
ˈkʰæ**ʃ**	→	kʰɛ́ː**sȳ́ː**	‘cash’ (Yip, 2006, p.961)
ˈt͡ʃiː**z**	→	t͡síː**sī́**	‘cheese’ (Silverman, 1992, p.298)
ˈkʰlʌ**t͡ʃ**	→	kʰìklík**t͡sī́ː**	‘clutch’ (Silverman, 1992, p.318)
ˈkʰæ**ʃ**ˌmɪə	→	kʰɛ́ː**sìː**mɛ́ː	‘cashmere’ (Yip, 1993, p.270)
ˈwɪ**s**ki	→	wɐ́j**sìː**kɛ̄́j	‘whiskey’ (Kenstowicz, 2012)

This is accounted for quite simply by positing that there exists some highly-weighted constraint Max-S, in addition to the constraints discussed previously. A tableau illustrating this analysis is presented below. The constraint CodaCond stands for the coda restrictions present in Cantonese.

(10) **Accounting for syllable-final sibilant repairs**

/ˈkʰæʃ/	*p* (%)	CodaCond $w = 10$	Max-S 10	Dep-V/C_ 2	Max-C/V_ 2
a. [kʰɛ́ːs]	0.034	1	0	0	0
b. [kʰɛ́ː]	0.005	0	1	0	1
c. [kʰɛ́ːsī́ː]	99.962	0	0	1	0

The next question a reader may have is what happens when a sibilant appears within a consonant cluster in a loanword of Cantonese — do speakers treat them as they do any other cluster, as in (5) and (7), or do they treat sibilants as they do word-finally, as in (9)? For the most part, speakers prefer to retain sibilants via **vowel epenthesis** in both cases, in line with how word-final sibilants are repaired, rather than deleting them. In the case of word-final sibilants in consonant clusters, the sibilant is occasionally deleted (5 instances), instead of retained with vowel epenthesis (13 instances).

(11) **Attested repairs to word- or syllable-initial sibilant-consonant clusters**

ˈ**sp**ɛː	→	**sìːp**ɛ́ː	‘spare’ (Yip, 1993, p.267)
ˈ**st**æmp	→	**sìːt**áːm	‘stamp’ (Silverman, 1992, p.301)
ˈ**sk**ɔː	→	**sìːk**ɔ́ː	‘score’ (Kenstowicz, 2012)
ˈ**sm**ɑːt	→	**sìːm**ɐ́k	‘smart’ (Yip, 1993, p.270)
ˈd͡ʒɪnˌ**sl**ɪŋ	→	t͡síː**sìːl**éŋ	‘gin sling’ (Kenstowicz, 2012)
ˈ**sw**ɪt͡ʃ	→	**sìːw**íːtt͡sī́ː	‘switch’ (Yip, 1993, p.270)

(12) **Attested repairs to word- or syllable-final consonant-sibilant clusters**

a.	ˈgʌ**ts**	→	kɐ́**tsī́ː**	‘guts’ (Kenstowicz, 2012)
	ˈmæ**θs**	→	mɛ́ː**tsī́ː**	‘maths’ (Kenstowicz, 2012)
	ˈtʰɪ**ps**	→	tʰíː**psī́ː**	‘tips’ (Silverman, 1992, p.301)
	ˈɒɹɪ**nd͡ʒ**	→	ɔ́ːlø̄**nt͡sī́ː**	‘orange’ (Kenstowicz, 2012)
	ˈɪ**nt͡ʃ**	→	íː**nt͡sī́ː**	‘inch’ (Yip, 1993, p.270)
	ˈfæ**nz**	→	féː**nsī́ː**	‘fans’ (Kenstowicz, 2012)
b.	ˈd͡ʒiː**nz**	→	t͡síː**n**	‘jeans’ (Kenstowicz, 2012)
	ˈlʌɪsə**ns**	→	láːjsɛ̄́**n**	‘license’ (Yip, 1993, p.288)
	ˈspɔː**ts**ˌʃɛːt	→	sìːpúː**t**sø̄́t	‘sports-shirt’ (Kenstowicz, 2012)

This behaviour is largely already what is predicted by the constraints and weights discussed thus far within the sketch of the grammar. The only addition is of a constraint Max-T/_V, which is predicted to have a weight at least as high as MAX-T/_R. A brief illustration of each is provided in the examples below.

(13) **Accounting for syllable-initial sibilant-consonant clusters**

/ˈskɔː/	p (%)	*COMPLEX $w=10$	MAX-S 10	MAX-T/_V 10	DEP-V/_C 3	DEP-V/C_ 2
a. [skɔ́ː]	0.663	1	0	0	0	0
b. [sɔ́ː]	0.244	0	0	1	0	0
c. [kɔ́ː]	0.663	0	1	0	0	0
d. [sìːkɔ́ː]	98.430	0	0	0	1	1

(14) **Accounting for syllable-final consonant-sibilant clusters**

/ˈɪnt͡ʃ/	p (%)	CODACOND $w=10$	*COMPLEX 10	MAX-S 10	DEP-V/C_ 2	MAX-C/V_ 2
a. [íːnt͡s]	0.000	1	1	0	0	0
b. [íːn]	0.011	0	0	1	0	0
c. [íːt͡s]	0.004	1	0	0	0	1
d. [íːt͡síː]	11.918	0	0	0	1	1
e. [íːníː]	0.001	0	0	1	1	0
f. [íːnt͡sɪ̄ː]	88.065	0	0	0	1	0

Cantonese loanword repairs are notable in that, when a sibilant-consonant cluster is word-final, Cantonese speakers choose to both **epenthesize a vowel** and **delete the final consonant**, as in (15).

(15) **Attested repairs to word-final sibilant-consonant clusters**

ˈkʰɑː**st**	→	kʰáː**síː**	‘cast’ (Yip, 1993, p.267)
ˈpʰəʊ**st**	→	pʰów**síː**	‘post’ (Silverman, 1992, p.306)
ˈlɑː**st**	→	láː**síː**	‘last’ (Kenstowicz, 2012)
ˈtʰəʊ**st**	→	tʰɔ́ː**síː**	‘toast’ (Kenstowicz, 2012)
ˈmɑː**sk**	→	máː**síː**	‘mask’ (Kenstowicz, 2012)

Previous literature has ascribed the lack of a final consonant in these cases to a **failure to perceive** the final consonant in clusters (Silverman, 1992; Yip, 1993, 2006). Under such an analysis, the words in (15) should be treated as identical to those in (9). However, it has also been observed that speakers of Cantonese are often **bilingual** in English, and must at some point have learned both how to perceive and produce word-final consonant clusters (Steriade, p.c.). A complete grammatical analysis of these data must be able to account for why word-final consonant clusters prefer deletion of the second consonant, while word-initial consonant clusters are also capable of retaining both members of the cluster via vowel epenthesis.

This behaviour is capable of being modelled by the present Max-Ent grammar, since this grammar makes a **distinction** between the DEP-V constraints. Due to the **combined** weight of these DEP-V constraints, the grammar cannot select a candidate that simply epenthesizes a vowel between the members of the cluster. The only other viable candidate also opts to delete the final member of the cluster. This candidate will not receive too high of a penalty due to the low weight of Max-C/C_, and so will receive the highest probability of the set.

(16) **Accounting for syllable-final sibilant-consonant clusters**

/tʰəʊst/	p (%)	CodaCond $w = 10$	*Complex 10	Max-S 10	Dep-V/_C 3	Dep-V/C_ 2
a. [tʰɔ́ːst]	0.000	1	1	0	0	0
b. [tʰɔ́ːs]	0.030	1	0	0	0	0
c. [tʰɔ́ːt]	0.011	0	0	1	0	0
d. [tʰɔ́ːsíː]	88.043	0	0	0	0	1
e. [tʰɔ́ːtíː]	0.001	0	0	1	0	1
f. [tʰɔ́ːsíːt]	11.915	0	0	0	1	1

Of course, the grammar provided in the examples above is only intended as a **sketch** of what a grammar might look like for Cantonese loanwords. The MaxEnt learners which are tested below will have access to a much more expanded set of constraints, and the resulting grammars are not guaranteed to be identical to the one above. It is, however, useful to examine the sketch to get an intuition about how the Faithfulness constraints are weighted with respect to one another, and how they may be weighted in a grammar derived by a MaxEnt learner. Any such grammar which wishes to account for the behaviour of these words in Cantonese must, at a minimum, account for the series of facts in (17).

(17) **Generalizations about repairs to illicit structures in loanwords in Cantonese**

a. Word- or syllable-initial consonant clusters can be repaired either by **deletion** of the second member of the cluster or by **vowel epenthesis** between the consonants.
b. In contrast, word- or syllable-final consonant clusters are always repaired by **deletion**.
c. Word- or syllable-final sibilants are always repaired by **epenthesis**.
d. When a sibilant is present in a consonant cluster, it biases the attested repairs further towards **vowel epenthesis**, instead of or in tandem with consonant deletion.

The observations in (17a,b) can be thought of as the **default** repair strategies to illicit consonant clusters in Cantonese. Already, the imbalance in repair strategies between word-initial and -final positions is compatible with an analysis based on the PMap of Steriade (2001). In this instance, it can be argued that inserting a vowel into a word-final cluster leads a greater difference between the fully faithful form containing the cluster (*e.g.*, [ˈpʰaʊnd]) and a form containing an epenthetic vowel (*e.g.*, [pʰɔ́ːntī]) than deleting a member of the offending cluster (*e.g.*, [pʰɔ́ːŋ]).[5] Thus, a speaker of Cantonese knows that a Faithfulness constraint that governs vowel epenthesis in this context, such as Dep-V/C_ in the grammar above, should have a higher weight than a constraint governing consonant

[5]There are, of course, many discrepancies between the English input [ˈpʰaʊnd] and the hypothesized Cantonese forms [pʰɔ́ːntī] and [pʰɔ́ːŋ], such as the difference between tone and stress, the quality of the nasal consonant, the voicing of the final coronal consonant, *etc.* These differences in segment quality — while a rich and productive vein of study — will not be pursued in this thesis.

deletion, such as Max-C/C_. Likewise, this speaker also knows that a constraint governing deletion of the first member of the cluster, Max-C/V_ should receive a higher weight than a constraint governing deletion of the second member, such as Max-C/C_V. While there are very few experimental studies that directly test this hypothesis, evidence from the phonetic cue distribution literature hint that the asymmetry in ranking between Max-C/V_ and Max-C/C_ should be as posited above (Steriade, 2001; Wright, 2004). Consonants preceded by a vowel will have access to the transitional cues present at the right edge of the vowel (*e.g.*, formant transitions, differences in F_0), but consonants preceded by another consonant will lack these cues. It is reasonable to assume that consonants with fewer cues are less perceptible, and thus closer to silence and more easily deleted, than their better-cued counterparts (Steriade, 2001).

When it comes to clusters in word-initial position, an inequality in the perceptibility of the initial obstruent and its following sonorant can be inferred, as the initial obstruent is always retained under deletion. Thus, it is likely the case that deleting an initial obstruent is more perceptible than deleting a medial sonorant, and the constraints governing their presence (Max-T/_R and Max-R/_V) must receive asymmetrical weights. However, there is no inequality in the relative number of deletion versus epenthesis repairs, indicating that these kinds of repairs are of roughly equal perceptibility. Thus, it can be surmised that the constraint governing deletion in onset position, Max-R/_V, should receive a weight roughly equal to the constraints governing epenthesis in this same position, Dep-V/C_ and Dep-V/_C. It can thus be stated that, at least with respect to these constraints, the PMap either makes **no predictions** about their weights relative to one another, or it predicts that the perceptual difference between inserting a vowel between a pair of consonants and deleting a sonorant adjacent to a vowel is of roughly the **same magnitude**.

An additional fact about the PMap is hinted at here in the repairs made to words that contain sibilants in illicit positions, outlined in (17c,d). Since sibilants are more often **retained** via vowel epenthesis in all positions, rather than deleted, it can be inferred that constraints that govern the deletion of sibilants, such as Max-S in the grammar above, are ranked higher than the constraints that govern vowel epenthesis, Dep-V/C_ and Dep-V/_C. This general ranking is expected given what is known about the salience of sibilants in the acoustic signal (Wright, 2004), and also from how sibilants are treated in loanwords in other languages. For example, rather than being deleted from clusters in Javanese, sibilants are **metathesized** with an adjacent consonant, so that they can be retained in a licit position within the language. Thus, it is unsurprising that constraints governing their retention are ranked above constraints that favour their deletion — deleting such a salient consonant will lead to a greater discrepancy between the perceived input and the produced output than inserting a vowel, so the PMap will favour a ranking that places the Max-S constraint (or constraints) above the Dep-V constraints.

In sum, the data from Cantonese loanwords are compatible with an analysis that makes use of a ranking of Faithfulness constraints via the PMap such that:

(18) **PMap requirements for Cantonese loanwords**

a. Presence vs. absence of a consonant before a vowel is more distinct than presence vs. absence of a consonant before the end of a word (silence) or another consonant.
b. Presence vs. absence of a consonant before a vowel is roughly equivalent to presence vs. absence of a vowel adjacent to a consonant, or is not governed directly by the PMap.
c. Presence vs. absence of a sibilant is more distinct than presence vs. absence of a vowel in the neighbourhood of that sibilant.

What is remarkable about these observations is not merely that they line up with predictions about the PMap, but that Cantonese speakers **do not receive any grammatical evidence** from their first language that these rankings or weightings should hold. As stated above, Cantonese speakers are never presented with words that contain consonant clusters at syllable edges, and they are never presented with words that contain sibilants in coda position. This is reflected in the sketched grammar by the high weight of the constraints *Complex and CodaCond, which are never violated in Cantonese. Nevertheless, speakers consistently employ the repairs listed in (5–15), despite never having been taught how to repair illicit structures in this way.[6]

Thus, it appears that the study of Cantonese loanwords will prove to be a rich one with respect to the PMap and with respect to modelling loanword adaptation within a MaxEnt framework. The method by which this is achieved will be discussed in full in Chapter 4. A full outline of this thesis is given in the next section.

1.4 Overview

The remainder of this dissertation is structured as follows:

Chapter 2 provides details about the MaxEnt grammar framework used throughout this thesis. It goes into detail about how MaxEnt grammars compare to other constraint-based grammars, and how probabilities are assigned to candidates within this framework. The MaxEnt learning algorithm,

[6] Of course, many of these words are now established lexical items of Cantonese, and learners of the language who are also fluent in English may have noticed a connection and actively arrived at a consistent repair strategy using their knowledge of both grammars. Nevertheless, there must have been, at some point in the past, a population of speakers that had just encountered these nonce words for the first time, and needed to adapt them for use in their own monolingual communities. The fact that this group of speakers arrived at such a sophisticated and nuanced series of repairs is what is of interest in this thesis, although I do provide a sketch of how a computational model may accommodate a multi-generational in Chapter 5. I will leave further questions of what kinds of knowledge bilingual speakers have about these processes up to future research.

originally from Della Pietra et al. (1997) and adapted by for use in phonology by Hayes and Wilson (2008), is discussed in detail. This chapter will also discuss how a Gaussian prior can be incorporated into the learning algorithm. Finally, it will discuss the exact implementation of the algorithm in R (R Core Team, 2017) by use of a modified version of the `solve.R` script of Pater and Staubs (2013).

Chapter 3 discusses the PMap in greater detail. It reviews how the PMap functions in classical OT, as in Steriade (2001), as well as the various models of the PMap undertaken in the MaxEnt grammatical framework by Wilson (2006) and White (2013). The reasoning behind the predicted differences between these two approaches will be sketched in this chapter as well.

Chapter 4 will discuss three separate experiments undertaken to test the hypotheses about the MaxEnt learners of Wilson (2006) and White (2013). Each experiment involves **training** a MaxEnt learner on a subset of the native Cantonese vocabulary drawn from the online dictionary CC-Canto (Pleco Software, 2016) in order to arrive at a grammar which is compatible with the language. Each grammar derived in this way is then applied to a test set consisting of the English **loanwords** discussed in the section above. The first experiment aims to test the null hypothesis that Cantonese contains sufficient evidence from the vocabulary to posit an asymmetrical weighting of PMap-governed constraints. This is shown not to be the case. This experiment also includes discussion of a **best-case** grammar derived from training the MaxEnt learner on the set of loanwords. This grammar was used for purposes of comparison throughout the thesis, and as a basis for establishing later PMap-style Gaussian biases.

The second experiment aims to test the behaviour of a learner equipped with Wilson's (2006) plasticity-based bias. This bias was derived largely from the best-case grammar, as direct evidence of the perceptual similarity of segment presences versus absences was not available. As such, this best-case grammar is examined in detail, to see how well it matches up with cue-based predictions of how the PMap is organized (Steriade, 2001; Wright, 2004; Yun, 2016). While the best-case grammar does, in some ways, match predictions based on the availability of cues, it also shows many discrepancies in this regard. The effect of establishing a plasticity bias was found not to predict the asymmetrical patterns of loanword adaptation outlined above, regardless of whether the prior was as PMap-compliant as possible or whether it contained extraneous information.

The third experiment discussed in Chapter 4 aims to test the behaviour of a learner equipped with White's (2013) weight-based bias. This grammar again made use of the best-case grammar when constructing a bias, and was shown to predict asymmetrical patterns of loanword adaptation, in line with the data discussed above. However, the bias predicted to be closest in nature to the PMap did not perform as well as one which contained additional information from the best-case grammar, indicating that additional research into the perceptibility of whole segment contrast should be undertaken in order to construct more accurate priors.

Chapter 5 discusses some of the assumptions made during the experiments reviewed in Chapter 4, and potential improvements to the experimental protocol used in this thesis. Some assumptions made led to discrepancies in predictions. Some of these discrepancies were deemed avoidable, such as those due to coding errors; and others were deemed to be inherent to the nature of the experiments undertaken, such as the choice of which candidates to generate. Other improvements were either proposed, such as a change to the **bias term** itself; or undertaken directly, such as a change to the **constraint set** used. Those which were undertaken directly were examined in detail through an experiment which attempted to replicate some of the results from the previous chapter.

Finally, Chapter 6 summarizes the findings of this thesis, and provides the reader with future directions that this kind of research could take.

Chapter 2

MaxEnt Grammars and MaxEnt Learning

2.1 Overview of MaxEnt grammars

In the broad spectrum of phonological theories, MaxEnt grammars belong to the family of **constraint-based** grammars, as opposed to rule-based grammars. The constraint-based grammar that most phonologists are familiar with is Optimality Theory (Prince and Smolensky, 1993), hereafter referred to simply as OT. MaxEnt shares much of its structure with OT, so I will begin by providing a very abbreviated review of how OT works, and then I will switch to describing the crucial differences between the two theories.

OT – along with all other constraint-based grammars – is a grammar that makes use of ordered constraints to select among possible phonological outputs. The grammar consists of three main components: Gen, which generates a broad and fairly comprehensive set of possible outputs for a given input form; Con, the ordered set of constraints that assess the possible outputs and assign violations; and Eval, the method by which the correct output form is selected. In OT, the constraints in Con are ordered in a **strict hierarchy**, and Eval proceeds through this hierarchy, eliminating all output candidates that violate the top-ranked constraint more than minimally until only one candidate is left. That candidate is the **optimal** candidate, and is selected as the correct output form for the given input. A basic example is provided in (19).

(19) **A basic OT example**

/ɪnpʊt/	Agree[Pl] *rank* = 1	Max-C 2	Dep-V 3	Ident[Pl]$_T$ 4	Ident[Pl] 5	*Coda 6
a. .ˈɪn.pʊt.	*!					**
☞ b. .ˈɪm.pʊt.					*	**
c. .ˈɪn.tʊt.				*!	*	**
d. .ˈɪ.pʊt.		*!				*
e. .ˈɪ.nə.pʊt.			*!			*

The grammar receives an input, /ɪnpʊt/, and Gen provides the grammar with the five candidate output forms listed in (19a–e). In principle, Gen can supply a **potentially infinite** set of candidates for the input. However, it is common to only supply those candidates that will be relevant for the constraints displayed in the tableau. Eval then proceeds through the constraints in Con, in order. The first constraint, Agree[Pl], assigns violations to all candidates that have a sequence of consonants that disagree for the feature [Place]. The only candidate that has such a sequence is (19a), so it is assigned a violation (*). Since there exist other candidates that do not violate this constraint, the violation is marked as fatal (!) and is eliminated from consideration as an optimal candidate (indicated by greying out of the cells for the remaining constraints). The grammar then proceeds to the next constraint in Con, Max-C, and repeats this process until only candidate (19b) remains. It is selected as the correct output form for the input, also known as the **optimal candidate**, and is marked as such (☞). All other constraints and their violations are no longer considered at this point, even though the optimal candidate may violate them.

All constraint-based grammars have this basic Gen-Con-Eval structure. However, the exact nature of each component varies from theory to theory. For instance, in Harmonic Serialism (hereafter HS) (McCarthy, 2010), Gen is constrained so as to only provide the set of candidates that instantiate one change from the local input. This necessitates a slight change to Eval, such that the output of the first pass through the grammar is selected as the next input for a second pass through the grammar, *et cetera*, until the local input and output are identical. The result of this last pass through the grammar is then selected as the correct output form.

Harmonic Grammar (hereafter HG) (Legendre et al., 1990; Smolensky and Legendre, 2006), retains the same conception of Gen as in OT, but changes the nature of the ordering of constraints in Con. Instead of a strict ranking, each constraint is assigned a **weight**. Higher-ranked constraints receive higher weights, and lower-ranked constraints receive lower weights. The structure of Eval likewise also changes to accommodate this difference, such that the violations assigned by a constraint are **multiplied** by the weight of that constraint, and these are then **summed** to generate a score known as a **harmony value**. The candidate with the **lowest** harmony value is the one that is selected as optimal by the grammar.

MaxEnt, likewise, shares the same basic Gen-Con-Eval structure as the theories sketched above, but the details of how Con and Eval are implemented are changed from that of OT. Like HG, a constraint **weighting** system is employed, such that constraint violations are multiplied by their weights and summed for each candidate. Unlike HG, the value determined by the sum of weighted constraint violations is not used to directly pick a single optimal candidate. Instead, it is converted into a **probability** for each candidate by the equation in (20).

(20) **Equation for determining probabilities in MaxEnt**

$$P(\omega) = \frac{exp(-\sum_i^N w_i \cdot C_i(\omega))}{\sum_{\omega'}^{\Omega} exp(-\sum_i^N w_i \cdot C_i(\omega'))}$$

where:

- $P(x)$ is the probability of x,
- ω is an output candidate,
- Ω is the set of all output candidates,
- C_i is the i^{th} constraint,
- $C_i(x)$ is the number of violations assessed by constraint C_i to x,
- w_i is the weight of the i^{th} constraint,
- N is the total number of constraints, and
- $exp(x)$ is a function that raises the mathematical constant e to the power of x.

This equation takes the harmony value for a candidate ω, as calculated by the sum of weighted constraint violations $\sum_i^N w_i \cdot C_i(\omega)$, and raises e to the negative of this value. This is done to ensure that the resulting probability is always **positive**, as negative probabilities do not exist. Then, this value is then divided by a **normalizing constant** – taken to be the sum of all of the exponentiated harmony values of every candidate ω' in Ω – to ensure that all probabilities sum to 1. In theory, Ω consists of all possible candidates supplied by GEN for any potential input. However, as this is commonly assumed to be an infinite set, calculation of its joint exponentiated harmony value is impossible, and Ω is taken to be just the set of candidates displayed in a given tableau.

(21) **A basic MaxEnt example**

/ɪnpʊt/	AGREE[Pl] $w = 8$	MAX-C 5	DEP-V 4	ID[Pl]$_T$ 3	ID[Pl] 1	*CODA 1	H	$P(\omega)$
a. .ˈɪn.pʊt.	1	0	0	0	0	2	10	0.000587
b. .ˈɪm.pʊt.	0	0	0	0	1	2	3	0.643
c. .ˈɪn.tʊt.	0	0	0	1	1	2	4	0.237
d. .ˈɪ.pʊt.	0	1	0	0	0	1	6	0.0320
e. .ˈɪ.nə.pʊt.	0	0	1	0	0	1	5	0.0871

As in OT, the grammar receives an input, /ɪnpʊt/, and GEN provides the same set of five output forms (21a–e). The constraints in CON are then applied to the outputs to generate the violations for each. So, for instance, candidate (21a) is assessed by the full set of constraints, and receives 1 violation for AGREE[Pl], 2 violations for *CODA, and 0 violations for all other constraints. The same is done for candidates (21b–e). As an intermediate step for the reader, the HG harmony values (in the column headed H) were calculated for each candidate, as these are equivalent to the sums of the weighted constraint violations to be used in the probability calculation. For candidate (21a), this value is 10, which is arrived at as in (22).

(22) **A sample harmony value calculation**

$$\begin{aligned} H(.\text{ˈɪn.pʊt.}) &= 8 \cdot \textsc{Agree}[\text{Pl}](.\text{ˈɪn.pʊt.}) + 5 \cdot \textsc{Max-C}(.\text{ˈɪn.pʊt.}) + 4 \cdot \textsc{Dep-V}(.\text{ˈɪn.pʊt.}) + \\ &\quad 3 \cdot \textsc{Ident}[\text{Pl}]_{\text{OBST}}(.\text{ˈɪn.pʊt.}) + 1 \cdot \textsc{Ident}[\text{Pl}](.\text{ˈɪn.pʊt.}) + 1 \cdot {*}\textsc{Coda}(.\text{ˈɪn.pʊt.}) \\ &= 8 \cdot 1 + 5 \cdot 0 + 4 \cdot 0 + 3 \cdot 0 + 1 \cdot 0 + 1 \cdot 2 \\ &= 8 + 0 + 0 + 0 + 0 + 2 \\ &= 10 \end{aligned}$$

It should be noted that, in contrast to OT, **all** constraint violations are taken into account, rather than only the violations of the highest-ranked constraints. Thus, the violations assigned to candidate (21b) by the constraint *Coda must be taken into account when calculating the harmony value in (22). In order to ensure that they do not influence the overall value too much, the constraint *Coda is assigned a low weight, while the other constraints that assess violations to that candidate, like Agree[Pl], are assigned substantially higher weights. However, say there is a hypothetical candidate (21f) that violates only *Coda four times. In OT, this candidate would win, since it violates none of the high-ranked constraints. When assigning a harmony value, however, this constraint would receive a harmony of 4, equivalent to that of the losing candidate (21c), and (21b) would still be selected as the winning candidate. This is illustrated in (23), below. Note that the probabilities assigned have changed, as there are now more candidates in the tableau, and the probability must be spread equally between them.

(23) **Ganging-up effects in MaxEnt**

/mpʊt/	Agree[Pl] $w = 8$	Max-C 5	Dep-V 4	Id[Pl]$_{\text{T}}$ 3	Id[Pl] 1	*Coda 1	H	$P(\omega)$
a. .ˈɪn.pʊt.	1	0	0	0	0	2	10	0.000474
b. .ˈɪm.pʊt.	0	0	0	0	1	2	3	0.520
c. .ˈɪn.tʊt.	0	0	0	1	1	2	4	0.191
d. .ˈɪ.pʊt.	0	1	0	0	0	1	6	0.0259
e. .ˈɪ.nə.pʊt.	0	0	1	0	0	1	5	0.0704
f. ?????	0	0	0	0	0	4	4	0.191

Whether this **ganging-up** behaviour is desirable has been a subject of much debate in the literature (Smolensky and Legendre, 2006; Jäger and Rosenbach, 2006; Pater, 2009; Potts et al., 2010; Jesney, 2011, 2016; McPherson, 2016; O'Hara, 2016). For the purposes of this thesis, it should simply be noted that there may not always be a straightforward link between an OT ranking and a MaxEnt weighting, and whether one provides a more accurate fit to the data than the other will not be explored.

Once the harmony values for each candidate are calculated, they are negated, $-H$; then exponentiated, $exp(x)$; and summed to provide the normalizing constant for the tableau in (21). Then, each individual exponentiated harmony value is divided by this normalizing constant to derive the probability of that candidate, $P(\omega)$. This is illustrated for (21a) in (24).

(24) **A sample probability calculation**

$$\begin{aligned}
P(.\text{ˈɪn.pʊt.}) &= \frac{exp(-H(.\text{ˈɪn.pʊt.}))}{\begin{array}{c} exp(-H(.\text{ˈɪn.pʊt.})) + exp(-H(.\text{ˈɪm.pʊt.})) + exp(-H(.\text{ˈɪn.tʊt.})) \\ +exp(-H(.\text{ˈɪ.pʊt.})) + exp(-H(.\text{ˈɪ.nə.pʊt.})) \end{array}} \\
&= \frac{exp(-10)}{exp(-10) + exp(-3) + exp(-4) + exp(-6) + exp(-5)} \\
&= \frac{4.539\text{E} - 05}{4.539\text{E} - 05 + 0.0498 + 0.0183 + 0.00248 + 0.00674} \\
&= \frac{4.539\text{E} - 05}{0.0774} \\
&= 0.000587
\end{aligned}$$

In contrast to OT, **no optimal candidate** is selected – rather, each and every candidate is assigned a **non-zero** probability. This means that it is impossible to ensure that any candidate is selected categorically, as there will always be some losing candidate in a tableau that is assigned some portion of the probability space, no matter how small. However, it is possible to assign the losing candidates such a vanishingly small probability that in practice, the winning candidate is assigned a probability incredibly close to 1, and a speaker will almost never produce the losing candidates in their lifetime. The tableau in (21) does not meet this criterion, as the candidate that is selected as the categorial winner in OT, (21b), is only selected as the winner 64.3% of the time. In order to mirror categorical behaviour, this candidate would have to be selected over 99% of the time, which could be achieved with a different set of weights. However, as this is an example made only for the purposes of illustrating the basic theory behind MaxEnt, this will not be pursued here.

The crucial points to take away from this abbreviated discussion of MaxEnt as a constraint-based grammar are: 1) that it is a grammar that converts constraint violations into **probabilities**; and 2) that it does so by making use of constraint **weights**. In the following sections, the advantages of using a probabilistic grammar will be enumerated (section 2.2), and the basic architecture of MaxEnt-based learners will be explored (section 2.3).

2.2 Advantages to using MaxEnt grammars

While this thesis is not meant to serve as an argument for using MaxEnt grammars or other weighted constraint-based grammars over others, there are nevertheless certain advantages to using MaxEnt grammars for this research project that do need to be acknowledged.

The data discussed throughout this thesis come from **loanword adaptations**, which oftentimes can be more **variable** than data drawn from the native vocabulary of a given language, and Cantonese is no exception (Yip, 2006, *a.o.*). As a grammar that deals exclusively in assigning probabilities to output forms, MaxEnt is well-suited to modelling such data.

Furthermore, as will be discussed in section 2.3, below, the learning algorithm makes it straightforward to add an **extra-grammatical bias** into the learning process in the form of **regularization**. This can be done by simply adding a **Gaussian prior** on the weights of each individual constraint, and allowing the algorithm to proceed as usual (Goldwater and Johnson, 2003; Hayes and Wilson, 2008). Comparable success in adding such an extra-grammatical bias has been explored by learning algorithms based within OT such as Low Faithfulness Constraint Demotion (Hayes, 2004) and Biased Constraint Demotion (Prince and Tesar, 2004), but these algorithms are only concerned with establishing a bias of ranking Markedness constraints above Faithfulness constraints wherever possible. As of this writing, a comparable OT-based learning algorithm that is capable of instantiating more sophisticated ranking requirements has not yet been developed.

On the other hand, the most promising incorporation of multiple hierarchical ranking requirements into phonological learning has been done using MaxEnt grammars (Wilson, 2006; White, 2013; Do, 2013). All of the above authors explore methods for incorporating the PMap of Steriade (2001) into a successful learning algorithm. As this thesis aims to investigate the extent of learners' knowledge of the PMap, having recourse to an established method of modelling such a bias is an excellent starting point for investigation. Further discussion of the exact ways in which Wilson (2006) and White (2013) incorporate the PMap into a MaxEnt-based learning algorithm are explored in the following chapter. For now, I will turn to discussing the general algorithm, so that the reader can gain a better understanding of why it is so simple to incorporate a variety of extra-grammatical biases into the learning process itself.

2.3 Learning using MaxEnt

2.3.1 The MaxEnt learning algorithm

The ultimate goal of a learning algorithm for a constraint-based grammar is to discover the **set of constraint weights** (or rankings, within OT) that predict the correct output for any input the grammar encounters. For MaxEnt, which assigns probabilities to outputs rather than selecting an outright winner, the goal is to find the set of constraint weights that **maximizes the probability of observed forms**. In other words, the goal is to assign the majority of the probability space to the candidate or candidates that are produced by a typical adult speaker of the language. This must be done for every potential input-output pairing that the learner encounters. Rather than attempting to find the weights that maximize the probability for every tableau individually, the algorithm instead tries to maximize the probability of all observed outputs at once. The probability of the observed outputs can be found by finding the **product** of the probabilities of the observed forms. The goal is to find the highest possible value for this function, hereafter known as the **objective function**.

(25) **Objective of a MaxEnt grammar**

$$max(P(D)) = max\left(\prod_{\omega}^{D} P(\omega)\right)$$

where:

- D is the observed data,
  ☞ **Note**: This is not equivalent to Ω, used in (20). Ω contains unobserved output candidates as well as observed ones; D only contains the observed candidates.
- ω is an output candidate,
- $P(x)$ is the probability of x as calculated by the equation in (20), and
- $max(f(x))$ is the maximum possible value for the function $f(x)$.

According to Della Pietra et al. (1997), maximizing this value is equivalent to maximizing the entropy (information-to-size ratio) of the data – hence the name "Max(imum) Ent(ropy)." I will not discuss this link further, as it will be sufficient for the reader to understand how MaxEnt works in terms of probabilities.

The reader should also note that for ease of computation, the objective function will not find the probability of the output forms directly, but instead will be computed over the **natural logarithm** of this probability. It has been proven that the weights that maximize $P(D)$ will be identical to the weights that maximize $log(P(D))$ (Della Pietra et al., 1997), so most MaxEnt learning algorithms make use of this shortcut. Hereafter, when discussing probability, it will be assumed that the learner is actually maximizing the log probability, or using the objective function in (26).

(26) **Simplified objective of a MaxEnt grammar**

$$max(\ log(P(D))\ ) = max\left(\ log\left(\prod_{\omega}^{D} P(\omega)\right)\right)$$

As the goal of the grammar is to maximize the probability of the observed forms, this entails that it must **minimize the probability of the unobserved forms**. Since the probabilities for any tableau must sum to 1, assigning the maximum amount possible to the observed candidate(s) will leave as little as possible to be distributed among the unobserved candidates. Thus, it can be guaranteed that maximizing the probability of the observed forms will effectively rule out the unobserved forms.

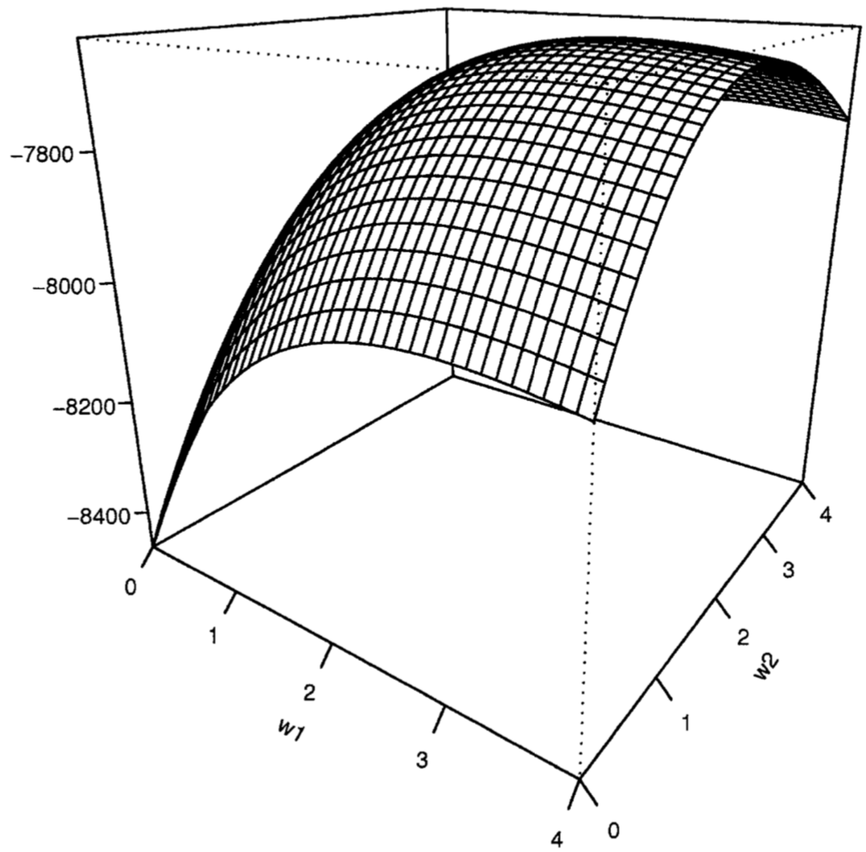


Figure 2-1: The surface defined by the probability of a representative training set for a two-constraint grammar from Hayes and Wilson (2008).

A natural question that may arise at this point is whether there could exist multiple maximum probabilities. If so, a learning algorithm may select the incorrect maximum probability as its target, and arrive at a set of constraint weights that do not match the observed data. Della Pietra et al. (1997) have proved that this is an impossibility. The search space defined by the objective function for any MaxEnt grammar is always **convex**, indicating that there is always some absolute maximum to be arrived at. Thus, an algorithm making use of a MaxEnt objective function can never "get stuck" at some local maximum and fail to learn the appropriate set of constraint weights. A sample convex space for a grammar with two constraints is reproduced in Figure 2-1 from Hayes and Wilson (2008). The two lower axes represent the weights of each individual constraint, C_1 and C_2, and the third axis represents the overall log probability of the grammar that they define. As can be seen in the figure, the search space forms a uniform dome, with only one possible maximum. This remains true no matter how many weights, or dimensions, must be optimized by the learner (although displaying this fact graphically is impossible).

Now that it is understood that the search space as defined by the MaxEnt objective function has a single maximum, it remains to be seen how to arrive at this maximum. The way in which the learning algorithm proceeds is broadly detailed in (27).

(27) **MaxEnt Learning Algorithm**

a. Assign all constraints a starting weight, commonly assumed to be 1 (Hayes and Wilson, 2008; Pater and Staubs, 2013) (although in principle the algorithm could begin at any randomly chosen set of weights).

b. Calculate the joint (log) probability of the observed data with the current constraint weights, according to the equations in (20) and (25).

c. Check whether the maximum has been reached by examining the **partial gradients** of each constraint weight.

d. If the partial gradients are all sufficiently close to or at 0, the maximum has been reached and the algorithm **terminates**.

e. Otherwise, the weights are **adjusted** according to the partial gradients calculated in step (27c), and the algorithm returns to step (27b).

The partial gradients in this process can be thought of as the **upward slope** of the convex search space along each weight dimension. The gradient is partial since the corresponding downward slope is not calculated. If the slope is at or very near 0 for some weight parameter, it indicates that the weight has reached a plateau in its trajectory, and adjusting it any further will not lead to an increase in probability for the observed data. If, however, it is not 0, that weight can be adjusted according to the slope calculated, and learning must continue. There are a few ways in which the partial gradient can be found – more details about how this is done for the algorithm used in this thesis will be provided in the following section.

For the sake of illustration, the same search space from Figure 2-1 is reproduced in Figure 2-2, except that this time, only the weight dimensions are shown, and the probability dimension is indicated by topographical rings. The learning trajectory for one iteration of the algorithm employed by Hayes and Wilson (2008) is indicated by the line that straddles these topographical boundaries.

The algorithm begins at the point in the lower left-hand corner, where both w_1 and w_2 are equivalent to 1. It then calculates the joint log probability of the observed data with these weights, and finds it somewhere between -7850 and -7900. It then calculates the partial gradients for both w_1 and w_2 and finds that they are not close enough to 0 to be considered a plateau. The weights for both constraints are then adjusted according to these partial gradients – the weight of constraint 1 is increased by ~ 1.75, and the weight of constraint 2 is increased by about 1. The process of calculating the joint log probability and the partial gradients is then repeated, and it is found that

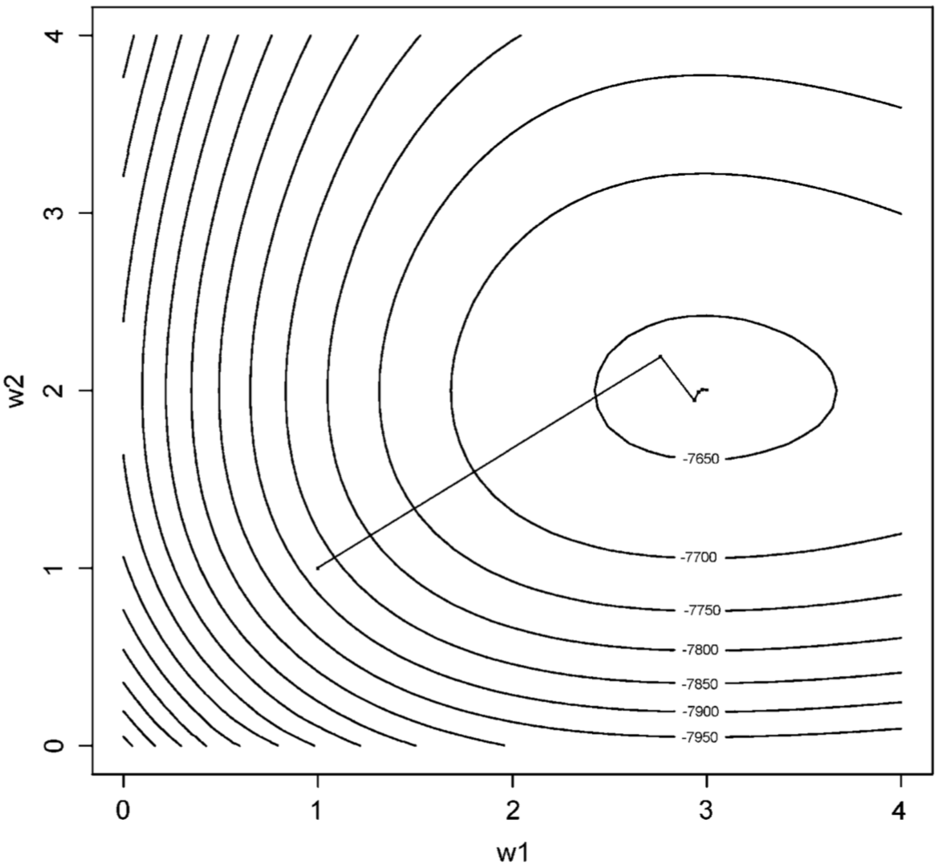


Figure 2-2: Iterative ascent of the surface given in Figure 2-1

the weights must be adjusted again. This time, the weight of constraint 1 is further increased by about 0.25, while the weight of constraint 2 is decreased by approximately the same amount. The process repeats three to four more times before the maximum is found and the algorithm terminates.

Two problems nevertheless arise with an objective that is simply to find the weights that maximize the probability of the observed data. One such problem, acknowledged by Hayes and Wilson (2008), is that there is a risk of **overfitting** to the data. For instance, accidental gaps identified by the phonological analyst may be deemed to be significant by the MaxEnt grammar, and constraints that rule out these gaps may receive an artificially high weight. Another problem that may arise is that the data fed to the MaxEnt grammar are **ambiguous**, and the learning algorithm may assign identical weights to constraints that the analyst has deemed to be distinct. A solution that can address both of these problems simultaneously is to impose some sort of **restriction** on the weights of each constraint, so that there is an additional penalty assigned to weights that deviate too far from the desired range.

The most common way to assign a restriction on constraint weights is to establish a **Gaussian**

prior, also known as a normal distribution or bell curve, around a pre-established average weight. This is known as **regularization**. The equation in (28) defines the Gaussian distribution used to regularize a set of constraint weights.

(28) **Gaussian distribution**

$$\varphi(w) = \prod_{i}^{N} exp\left(-\frac{(\ w_i - \mu_i\ )^2}{2\sigma_i^2} \right)$$

where:

- $\varphi(x)$ is the Gaussian distribution,
- w is the vector consisting of all constraint weights,
- w_i is the weight of the i^{th} constraint,
- μ_i is the mean (target weight) of the i^{th} constraint,
- σ_i^2 is the variance (plasticity) of the i^{th} constraint, and
- N is the total number of constraints.

Hayes and Wilson (2008) employ this method in their paper, and establish a uniform bias on constraint weights such that $\mu = 0$ and $\sigma^2 = 1$. This can help solve the problem of overfitting, since the grammar that uses such a regularization scheme will prefer to assign weights as close as possible to 0, and prevent the weight of any one constraint from deviating too far from this weight by assigning the plasticity the low value of 1. This does not, however, help in solving the problem of ambiguity – instead, it has the potential to exacerbate it, as all constraints ideally have an identical weight of 0. As will be discussed in the following chapter, the PMap is one partial solution to the problem of ambiguity, in that it provides a method by which an **a priori** ranking can be established for the set of Faithfulness constraints. This has been also been modelled using the Gaussian prior by Wilson (2006); White (2013), and Do (2013), the details of which are reserved for Chapter 3.

The regularization function can be straightforwardly incorporated into the objective function by adding it onto the main objective, as is done in (29).

(29) **Objective of a regularized MaxEnt grammar**

$$max(\ log(\ P(D)\ )\ ) = max\left(\ log\left(\prod_{\omega}^{D} P(\omega) + \varphi(w) \right) \right)$$

As the search space for a Gaussian prior is guaranteed to be convex, and as the sum of two convex search spaces is also convex, the same algorithm can be used to search for the maximum log probability in this case as well.

Thus far, three main properties of MaxEnt grammars have been introduced. First, it has been shown how probabilities are derived from constraint weights in a MaxEnt grammar, as in equation

(20). Second, the objective function for a MaxEnt grammar both with and without regularization has been defined, as in equations (29) and (26), respectively. Third, the algorithm for how this objective is achieved is sketched, as in (27). A reader who is interested in learning more about how the PMap has been incorporated using a Gaussian prior is encouraged to skip directly to Chapter 3. Those interested in learning more details of how the algorithm is implemented in this thesis are encouraged to read section 2.3.2, below.

2.3.2 Implementing the algorithm

Recall from the previous subsection that while the goal of a MaxEnt-based learning algorithm is to maximize the log likelihood of the data, the method by which this result is arrived at may differ from algorithm to algorithm. The main point of difference between algorithms is in how the partial gradients needed for determining whether the algorithm must continue or terminate are calculated. In this section, I will outline the general solution to this problem, and give the details of how this is done for the learning algorithm implemented by Pater and Staubs (2013), which is employed for the remainder of this thesis.

As has been proven by Della Pietra et al. (1997), the partial derivative of a log probability function $log(P(D))$ is equal to the difference in **observed** versus **expected** violations of the constraint under consideration C_i.

(30) **Partial derivative for a constraint**

$$\frac{\partial}{\partial w_i} log(P(D)) = O[C_i] - E[C_i]$$

where:

- $\frac{\partial}{\partial w_i}$ is the partial derivative associated with the i^{th} constraint,
- C_i is the i^{th} constraint,
- $O[C_i]$ is the observed violations for the i^{th} constraint, and
- $E[C_i]$ is the expected violations for the i^{th} constraint.

When calculating these values over the entire grammar, the equations in (31) are employed. The only difference between the two is how the value for $P(\omega)$ is calculated – when finding the observed violations, it is derived from the empirical values for each phonological form in the language; when finding the expected violations, it is derived from the constraint weights entertained by the algorithm at that point in time.

(31) **Cumulative violation calculation**

a. *Observed*

$$O[C_i] = \sum_{\omega}^{D} P_o(\omega) \cdot C_i(\omega)$$

where:

- D is the empirically observed data,
- $P_o(x)$ is the probability of output candidate x as determined by empirical observation, and
- All other notational conventions are as above.

b. *Expected*

$$E[C_i] = \sum_{\omega}^{\Omega} P_w(\omega) \cdot C_i(\omega)$$

where:

- Ω is the set of all output candidates,
- $P_w(x)$ is the probability of output candidate x as determined by a set of constraint weights w, and
- All other notational conventions are as above.

As is noted by Hayes and Wilson (2008), and as should be familiar from section 2.1, calculation of the true value of $E[C_i]$ is often impossible, as it requires summing over the infinite set of phonological forms Ω. Various strategies have been employed in order to arrive at an estimate of this value; I will employ the one taken by Pater and Staubs (2013), following Goldwater and Johnson (2003). As is done when calculating the MaxEnt probability of an output form, the value of $E[C_i]$ is estimated by taking only the candidates provided in the tableau and ignoring the others.

The reader may wonder why it is admissible in the calculation of the derivative in (30) to compare sums of constraint violations across two different sets of data – the (approximation of the) infinite set of all candidates Ω and the substantially smaller set of observed forms of a language D. The reason is that if a candidate is unobserved in the language – that is, if it belongs to Ω but not to D – it will by necessity have probability 0. Since the constraint violations introduced by that candidate are multiplied by the probability of the constraint (*i.e.*, zero), any potential contribution that candidate could have introduced into the cumulative violation calculation is nullified. Thus, regardless of which set is used when calculating $O[C_i]$, the result will be the same, and it is simpler to use only the candidates that will produce non-zero results.

Pater and Staubs (2013), following much of the literature on probabilistic models of this type, further note that this method, whereby the learning algorithm is performing a gradual, ascending search over the space defined by the log likelihood of the data, is a kind of **error minimization**. Error is defined as the **Kullback-Leibler divergence** (hereafter KL-divergence) between an expected and an observed probability distribution. The closer this value is to 0, the smaller the error.

The equation for KL-divergence is given in (32) (Kullback and Leibler, 1951).

(32) **Kullback-Leibler divergence**

$$D_{KL}(O||E) = -\sum_{x}^{\chi} O(x) \cdot log\left(\frac{E(x)}{O(x)}\right)$$

where:

- $D_{KL}(X||Y)$ is the KL-divergence between two probability distributions X and Y,
- E is the probability distribution associated with the expected data,
- O is the probability distribution associated with the observed data, and
- $P(x)$ is the probability assigned to x in P.

Intuitively, the smaller the difference between the observed and the expected probability distributions, the closer the two are to being identical. Thus, if the goal is to arrive at an expected probability distribution that matches the observed distribution as closely as possible, it should also be the case that the value for D_{KL} should be as low as possible. A reader who is interested in the exact mathematical relation between this objective and the MaxEnt objective defined in (26) are encouraged to consult Eguchi et al. (2014). For the purposes of this paper, it will be sufficient to note that they are intuitively very similar, and that there is a link between calculation of KL-divergence using the observed and expected probability distributions and calculation of the partial gradients for a more traditional MaxEnt grammar.

In order to make their algorithm more generally applicable, Pater and Staubs (2013) make use of the method error minimization when constructing their algorithm, rather than employing the narrower, more traditional method of maximizing the log probability of the data. Thus, the objective function used in the code they develop is that of minimizing the KL-divergence between the observed and expected data, rather than the objective function outlined in (26). As these two objectives are mathematically equivalent, it will not make a difference as to which is chosen as the objective function.

Another small difference in the implementation of the MaxEnt learning algorithm in Hayes and Wilson (2008) and the one undertaken by Pater and Staubs (2013) is that the entire MaxEnt objective function is **converted** into a minimization problem. In other words, the objective function is no longer to find the weights which maximize the log probability of a set of data, but to find the weights which minimize the negative log probability of the data. This is shown in (33). These two problems are equivalent, and so there is no penalty to thinking of the objective function in this way.

(33) **Inverse objective of a MaxEnt grammar**

$$min(\ -log(\ P(D)\ )\ ) = min\left(\ -log\left(\ \prod_{\omega}^{D} P(\omega) + \varphi(w)\ \right)\ \right)$$

While this difference does not have any consequences for the algorithm used, it does have consequences for how alternate biases must be constructed. This will be discussed further in the context of difference biases, introduced in Chapter 5.

A slightly modified version of the code used in Pater and Staubs (2013) is used to perform MaxEnt learning for the simulations outlined in the following chapters. This code is written to be compatible with the `R` software package (R Core Team, 2017) and the functions defined therein. The code takes as its objective function the minimization of KL-divergence, and in order to perform the gradient search required by the algorithm, it makes use of the `optim()` function provided by `R` run using the `L-BFGS-B` method (Byrd et al., 1995).

A reader familiar with the implementation of a MaxEnt learning algorithm in Hayes and Wilson (2008) may wonder if any of the implementational differences between their algorithm and the algorithm employed in Pater and Staubs (2013) leads to any significant differences in the results obtained. While this will not be discussed in this thesis, a comparison was done in order to asses whether this was the case. It will be sufficient to note for now that while there were consistent differences between the exact values found by either implementation, the results were qualitatively very similar. The implementation used in Pater and Staubs (2013) was chosen over the implementation used in Hayes and Wilson (2008) primarily due to the ease with which the algorithm code could be altered by the author. It is, however, presumed that any future research into modelling the adaptation of loanwords that makes use of the algorithm developed for Hayes and Wilson (2008) will arrive at qualitatively similar results as the ones reported in future chapters.

Chapter 3

The Perceptual Map and MaxEnt

3.1 The Perceptual Map

As laid out in Steriade (2001), the Perceptual Map (PMap) is a solution to the **too many solutions problem** of strict-ranking OT and other constraint-based grammars. This problem arises from the typology defined by constraint-based grammars. If the set of possible constraints is universal, then differences between grammars are reduced to differences in how they rank constraints. This, in turn, means that the number of possible grammars is equivalent to the number of different constraint rankings. However, not all languages that are predicted by the typology are empirically observed. As discussed in Steriade (2001), there are a variety of languages that enforce a process of **word-final devoicing**, as is the case for the German example in (34).

(34) **German word-final devoicing**

a. *Data from Plural formation*

Voiceless				Voiced			
ʁaʊ**p**	↔	ʁaʊ**p**ə	'caterpillar/s'	ʁaʊ**p**	↔	ʁaʊ**b**ə	'robbery/ies'
ʁaː**t**	↔	ʁaː**t**ə	'council/s; counsel/s'	ʁaː**t**	↔	ʁaː**d**ə	'wheel/s'
ʁuː**k**	↔	ʁuː**k**ə	'jolt/s; tug/s'	t͡suː**k**	↔	t͡syː**g**ə	'train/s'

b. *Ranking*

/ʁaːd/	*D#	Dep-V/_#	Ident[nasal]/_#	Ident[voice]/_#
i. ʁaːd	*!			
ii. ʁaːdə		*!		
iii. ʁaːn			*!	
☞ iv. ʁaːt				*

In the above example, the mapping from underlying /ʁaːd/ to surface [ʁaːt] is due to the low ranking of the constraint Ident[voice]/_#. All other relevant constraints are ranked above it, and so all other repairs – such as the nasalization repair, *[ʁaːn], and the vowel epenthesis repair, *[ʁaːdə] – are eliminated by higher-ranked constraints. However, typologically, these four constraints could be **re-ranked**, so that DEP-V/_# is lowest, predicting the language outlined in (35).

(35) **A language with word-final epenthesis**

a. *Data from a hypothetical morphological alternation*

<u>Voiceless</u>				<u>Voiced</u>			
lej**p**	↔	lej**p**na	—	lej**bi**	↔	lej**b**na	—
luː**t**	↔	luː**t**na	—	luː**di**	↔	luː**d**na	—
lɑː**k**	↔	lɑː**k**na	—	lɑː**gi**	↔	lɑː**g**na	—

b. *Ranking*

/lejb/	*D#	IDENT[nasal]/_#	IDENT[voice]/_#	DEP-V/_#
i. lejb	*!			
☞ ii. lejbi				*
iii. lejm		*!		
iv. lejp			*!	

While this is an intuitively simple change from the grammar in (34) – all that is changed is which constraint is lowest-ranked – it is nevertheless an **unattested** pattern of alternation. In fact, repairs to word-final voiced obstruents are heavily biased towards devoicing, rather than any other phonological process (Steriade, 2001). The solution that is enforced by the PMap is to establish a **hierarchical ranking** of Faithfulness constraints, so that the ranking proposed in (35b) is ruled out, and the ranking proposed in (34b) is preferred. Establishing such a ranking will also help with the problem of **ambiguity** introduced in Chapter 2 – if there are fewer possible rankings of constraints, there are fewer possible grammars that can be entertained over the course of learning, and greater certainty that the grammar arrived at is the correct one.

The hierarchical ranking is not arbitrarily selected – it is instead based on the **perceptual similarity** of pairs of output forms. A brief example involving DEP-V/_# and IDENT[voice]/_# will be used to illustrate how this is achieved. IDENT[voice]/_# – a context-specific form of the more general constraint IDENT[voice] – is violated by forms which differ in their specification for the feature [±voice]. It can thus be used to differentiate between two phonetic forms which vary only in the specification of [±voice] – *e.g.*, [ʁaːd] and [ʁaːt]. When these two forms are compared perceptually, they are found to be **similar**, as indicated by a variety of factors, including how **confusable** they are in a laboratory setting (Steriade, 2001).

The same process can be performed for the constraint DEP-V/_#, which is violated by forms which differ in the presence vs. absence of a word-final vowel. It can be used to differentiate between forms which vary only in this way – *e.g.*, [ʁaːd] and [ʁaːdə]. When these two forms are compared perceptually, they are found to be quite **different**, at least in comparison to the similarity of [ʁaːd] and [ʁaːt]. Thus, the following inequality in perceptual similarity can be established:

(36) **A sample ranking of differences along a scale of perceptual similarity**

$$\Delta(\text{ə} - \varnothing/\_\#) > \Delta(\text{t} - \text{d}/\_\#)$$

where:

- $\Delta(x - y/E)$ is the perceptual difference between x and y in the environment E

From this inequality, it can be inferred that any constraint which can differentiate between [ə] and [∅] word-finally must be ranked **above** a constraint which can differentiate between [t] and [d] in the same environment. Thus, the ranking in (37) can be established.

(37) **A sample ranking of constraints, according to the inequality in (36)**

$$\text{Dep-V}/\_\# \gg \text{Ident[voice]}/\_\#$$

The same process can be applied to the other Faithfulness constraints in Con, so that a more complete hierarchy can be arrived at for any given language. For example, the perceptual difference between two forms that differ only in the nasality of the word-final consonant – *e.g.*, [ʁaːd] and [ʁaːn] – can be estimated by a speaker, and found to be more similar to one another than to a pair of forms that differs in the presence vs. absence of a word-final vowel, but less similar than a pair of forms that differs in word-final consonant voicing. The constraint that is capable of differentiating [ʁaːd] and [ʁaːn] – Ident[nasal]/_# – would then need to be ranked between the two constraints governing the other two perceptual differences, Dep-V/_# and Ident[voice]_#. Ranking the constraints in this way results in a refinement of the ranking established for German in (34).

(38) **A potential ranking of Ident[nasal]_#, according to the PMap**

a. *Perceptual similarity hierarchy*

$$\Delta(\text{ə} - \varnothing/\_\#) > \Delta(\text{n} - \text{d}/\_\#) > \Delta(\text{t} - \text{d}/\_\#)$$

b. *Resulting PMap ranking*

$$\text{Dep-V}/\_\# \gg \text{Ident[nasal]}/\_\# \gg \text{Ident[voice]}/\_\#$$

In this way, this mechanism of comparing forms along a scale of perceptual similarity can establish a ranking among Faithfulness constraints that biases a language to prefer certain kinds of repairs over others. A language that contains the ranking outlined in (35), where Ident[voice]/_# outranks Dep-V/_#, cannot be generated under the PMap since the ordering of constraints that it predicts contains a contradictory ranking, namely, Dep-V/_# ≫ Ident[voice]/_#. All languages are hypothesized to have access to this mechanism, which also explains why biases towards certain kinds of repairs are **uniform** cross-linguistically (Steriade, 2001). Thus, a language like the one outlined in (35) is predicted not to exist, as Dep-V/_# will always be ranked too high in the hierarchy to be selected as the default repair.

Of course, it should be noted that Steriade (2001) does not make the claim that all Faithfulness constraints can be ranked in this way. For instance, another constraint that can distinguish a pair of words that varies in the presence vs. absence of a word-final vowel is Max-V/_#. As the same perceptual similarity judgement is used for both Max-V/_# and Dep-V/_#, it might be expected that these two constraints cannot be ranked with respect to one another – at least not without explicit evidence from the grammar or some other source.

However, it might be expected that the rankings established over the set of contextual Max constraints would **mirror** the rankings established over the set of contextual Dep constraints. If, for instance the PMap establishes a ranking such that Dep-V/_T $\gg$ Dep-V/_#, then it is predicted that there should be a ranking such that Max-V/_T $\gg$ Max-V/_#. As both make use of the similarity judgements between the presence vs. absence of a vowel in the same contexts, they should order the similarity judgements in the same way.

When it comes to the research done to investigate perceptual similarity, much of it has been tested using experimental studies into the **confusability** of various sounds in context (Miller and Nicely, 1955; Wang and Bilger, 1973; Guion, 1998, *a.o.*). However, for various reasons, this is a difficult task, as certain contrasts – such as presence vs. absence of a segment – are often so easily discriminable that participants perform at ceiling. Furthermore, not all contrasts in all positions have been investigated, and this has not been done for many languages other than English (Yun, 2016). In this instance, other sources of information are used to hypothesize about what the perceptual similarity of two sounds in context could be. One such source is the acceptability of **imperfect rhymes** – for example, rhymes where the pair differ in the voicing of the final obstruent are observed more often in corpora than rhymes where they differ in nasality (Zwicky, 1976; Steriade, 2001). Another source comes from the phonetic **cues** to a certain segment in that context. For instance, cues to the place of a consonant are often most robust and more numerous before a vowel than before an obstruent or silence (Flemming, 1995, 2001; Steriade, 2001; Wright, 2004).

It should also be noted that the exact phonetic cues associated with a segment may cause differences in the perceptual similarity of that segment in context. This will, in turn, result in **different hierarchical rankings** of Faithfulness constraints cross-linguistically. For example, Pater (1996a) observes that there are a variety of possible repairs to the illicit nasal-voiceless obstruent cluster (*NT) within the Austronesian language family and beyond. In Mandar (Sulawesi, Indonesia; Austronesian language family, $\sim$ 480,000 speakers), a prefix-final nasal is **denasalized** when followed by a voiceless obstruent, indicating that the constraint Ident[nasal] is lowest-ranked among the

Faithfulness constraints in this language.

(39) **Ranking which achieves denasalization in Mandar** (Pater, 1996a)

/maN$_1$ + t$_2$unu/	*NT	MAX-C	LINEARITY	IDENT[nasal]
a. man$_1$t$_2$unu	*!			
b. mat$_2$unu		*!		
c. man$_{1,2}$unu			*!	
☞ d. mat$_1$t$_2$unu				*

In Puyo Pungo Quechua (Ecuador; Quechuan language family, demographics unknown), which has the same phonotactic restriction, the nasal and the obstruent are **fused** into a single segment, indicating that LINEARITY is lowest-ranked.

(40) **Ranking which achieves fusion in Puyo Pungo Quechua** (Pater, 1996a)

/məN$_1$ + p$_2$ilih/	*NT	MAX-C	IDENT[nasal]	LINEARITY
a. məm$_1$p$_2$ilih	*!			
b. məp$_2$ilih		*!		
☞ c. məm$_{1,2}$ilih				*
d. məp$_1$p$_2$ilih			*!	

Finally, in Kelantan Malay (Kelantan, Malaysia; Austronesian language family, ∼ 5 million speakers), the attested repair is **deletion of the nasal**, indicating that MAX-C is lowest ranked.

(41) **Ranking which achieves deletion in Kelantan Malay** (Pater, 1996a)

/məN$_1$ + p$_2$ilih/	*NT	LINEARITY	IDENT[nasal]	MAX-C
a. məm$_1$p$_2$ilih	*!			
☞ b. məp$_2$ilih				*
c. məm$_{1,2}$ilih		*!		
d. məp$_1$p$_2$ilih			*!	

While no phonetic studies have been performed to investigate the exact phonetic realization of nasals and obstruents in this language, it might be surmised that each language has a different conception of how the cluster should be realized **phonetically**. In Mandar, the nasal might be prone to devoicing in similar contexts (*e.g.*, word-finally), making it indistinguishable from a geminate when located before a voiceless obstruent. In Kelantan Malay, the nasal may be coarticulated with the vowel to such an extent that its consonantal portion is not easily separable, making it most similar to a phonetic form without a nasal consonant at all. Much more research should be done into the fine phonetic details of these languages to establish what kinds of phenomena are present that could account for these differences – for now, it is only necessary to understand that the differences in ranking could be attributable to differences in the phonetic realization of the nasal or the obstruent from language to language.

The fact that the PMap can be so sensitive to differences in phonetic cues raises the question of how more **general** constraints such as MAX-C should be ranked. After all, MAX-C collapses across a broad variety of consonants: obstruents, fricatives, nasals, liquids, and potentially glides. Each kind of consonant has a very different set of relevant cues, each of which can vary widely in different segmental contexts. This will, in turn, lead to a wide variety of discriminability judgements that the constraint might use to establish itself in the hierarchy. It remains an open question as to which of these should be used as the default PMap ranking, or whether there should be some way of averaging across them all when ranking such a general constraint. Wilson (2006) hypothesizes that they should be ranked according to the **most discriminable** pair that could be covered by that constraint – in the case of Max-C, this might be equivalent to the constraint Max-T/V_V. In order to avoid this issue, I will be making use of constraints with **fully-specified** contexts for the majority of the simulations conducted here. More details about the exact constraint set used can be found in Chapter 4 and Appendix B. Those interested in what impact the inclusion of more general constraints could have on PMap-assisted learning should consult the experiment discussed in section 5.2 of Chapter 5.

3.1.1 Deletion and insertion and the PMap

As stated before, there is no *a priori* reason to rank Max constraints and Dep constraints with respect to one another, since both constraints must make use of the phonetic comparison between a segment in context and its absence. There are also challenges to establishing a PMap ranking of these kinds of constraints, as it is more likely for participants to perform at ceiling in discrimination or confusability tasks that make use of such drastically different stimuli. There have nevertheless been a few attempts at conducting these kinds of studies, and as the purpose of this thesis is to model how the PMap may help establish a ranking among Max and Dep constraints in Cantonese, I will draw on them when attempting to establish a PMap ranking for use in the experiments discussed in Chapter 4. Of particular note are the series of studies undertaken by Yun (2016), who investigated the nature of the cues relevant to vowel epenthesis in various consonantal contexts. Her findings will be summarized below.

In Yun's (2016) studies, speakers of a variety of languages – such as English, Mandarin, and Korean – were presented with pairs of sounds which varied in the presence or absence of a vowel in a variety of consonantal contexts, both word-initial and word-final. The participants were asked to perform an AX discrimination task using these sounds, which were spoken by a native speaker of either Russian or Ukrainian, as these languages allow for a greater variety of word-initial and word-final cluster combinations than any of the target languages examined. Some example stimuli are provided below. Consonants that are followed by a ⌜ are audibly released, and those followed by a ˺ are unreleased.

(42) **Sample AX stimuli from Yun (2016)**

ˈt˹mate	∼	t˹əˈmate
ˈg˹nute	∼	g˹əˈnute
ˈp˹tite	∼	p˹əˈtite
ˈg˹bute	∼	g˹əˈbute
ˈm˹date	∼	m˹əˈdate
ˈn˺mite	∼	n˺əˈmite
ˈlkate	∼	ləˈkate
ˈn˺kite	∼	n˺əˈkite
	. . .	

Yun (2016) discovered that pairs of consonants which exhibited a **rise in intensity** were more confusable with their vowel epenthesis counterparts, and pairs of consonants which exhibited a **plateau** or a **fall** in intensity were more easily discriminated from the same structures with an epenthetic vowel. This is an intuitive result – as vowels are often the segments which have the highest intensity in a word, they are often accompanied by a rise in intensity from the preceding segment. Thus, inserting a vowel at an intensity rise is less of a change than inserting one at an intensity fall or plateau. This drives Yun to establish a universal ranking of Dep-V constraints, such that those which penalize inserting a vowel at an intensity fall are ranked above those which penalize inserting a vowel at an intensity rise. This is summarized in (43).

(43) **Universal ranking of Dep-V constraints** (Yun, 2016)

$$\textsc{Dep}\text{-V}/\neg\text{IR} \gg \textsc{Dep}\text{-V}/\text{IR}$$

Similarly, Yun (2016) discovered that pairs of consonants which begin with a **voiced consonant** are more confusable with their vowel epenthesis counterparts than those which begin with a voiceless consonant. This is also an intuitive result – as vowels are almost always voiced, inserting a voiced segment after another voiced segment results in a lesser change than inserting a voiced segment after a voiceless one. Yun does not establish this constraint within the hierarchy outlined in (43), as these are orthogonal concerns – a cluster which begins with a voiced consonant may or may not contain a rise in intensity to the next consonant. Both concerns must nevertheless be taken into account when establishing a PMap-compliant hierarchy of Dep and Max constraints.

It should also be noted that the hierarchy established in Yun (2016) is **partial**. It does not make any predictions about vowel epenthesis adjacent to other vowels, although it might be inferred that these fall within the category of the constraint Dep-V/IR, and will be ranked quite low on the overall PMap hierarchy. It also does not make any predictions at all about **consonantal** epenthesis (or deletion) in any context. Other sources will be used to establish a ranking for these constraints.

While Yun (2016) examines a particular cue to segmental contrast in great experimental detail, this is not always true of other cues to segmental contrast. Instead, I will be relying on a survey of cue availability and attentional preference covered in Wright (2004) in order to flesh out predictions for the remainder of the Faithfulness constraints considered here. Wright discusses two broad tendencies within his survey that are immediately applicable to the PMap inequalities predicted to have the most influence on the patterns of Cantonese loanword adaptation studied as part of this dissertation. The first lies in the fact that segments of different sonority also differ in the number and strength of the **internal** cues to their presence. According to Wright (2004), **sibilants** have the strongest internal cues, followed by **sonorants** and then by **obstruents**.

(44) **Weighting schema for Max-C and Dep-C constraints** (Wright, 2004)

$$\begin{array}{ccccc} \textsc{Max-S} & \gg & \textsc{Max-R} & \gg & \textsc{Max-T} \\ \textsc{Dep-S} & \gg & \textsc{Dep-R} & \gg & \textsc{Dep-T} \end{array}$$

Wright also makes a distinction among the cues available to segments in a variety of contexts, as done in Steriade (2001). He uses obstruents to illustrate the effects of segmental context, as obstruents have the weakest internal cues and as a result are almost exclusively cued by the **transitions** present between them and adjacent segments. An obstruent will be best cued when it is between two sonorants, such as in the contexts V_V, R_V, *etc.* This is a result of it receiving strong transitional cues from both the preceding and following sonorant in the form of formant transitions, F_0 differences, release bursts, voice onset time, and the like. It will be next best-cued when there is a sonorant following the obstruent, as there exists an attentional asymmetry in the perceptual system towards phonetic cues which are at the onset of a sonorant, and cues that occur in this position will be perceived as being stronger (Sinex and Geisler, 1983; I. and Sachs, 2983; Delgutte and Kiang, 1984b,a; Sinex, 1995). An obstruent will only be moderately well-cued by a preceding sonorant for this reason as well. Finally, an obstruent will be very poorly cued when located between other obstruents or silence, as it will only have access to its weak internal cues in these contexts. These considerations are used to construct a series of predicted rankings for a set of Max-T constraints, below.

(45) **Weighting schema for constraint contexts, exemplified by the family of Max-T constraints** (Wright, 2004)

$$\left\{\begin{array}{c} \textsc{Max-T/V\_V} \\ \textsc{Max-T/V\_R} \\ \textsc{Max-T/R\_V} \\ \textsc{Max-T/R\_R} \end{array}\right\} \gg \left\{\begin{array}{c} \textsc{Max-T/\#\_V} \\ \textsc{Max-T/\#\_R} \\ \textsc{Max-T/T\_V} \\ \textsc{Max-T/T\_R} \end{array}\right\} \gg \left\{\begin{array}{c} \textsc{Max-T/V\_\#} \\ \textsc{Max-T/V\_T} \\ \textsc{Max-T/R\_\#} \\ \textsc{Max-T/R\_T} \end{array}\right\} \gg \left\{\begin{array}{c} \textsc{Max-T/\#\_\#} \\ \textsc{Max-T/\#\_T} \\ \textsc{Max-T/T\_\#} \\ \textsc{Max-T/T\_T} \end{array}\right\}$$

The findings of Yun (2016) and the predictions of Wright (2004) will be combined in an effort to evaluate a loanword grammar of Cantonese, discussed in detail in section 4.4.2.1 of Chapter 4. This loanword grammar will be examined in order to determine if the weights learned by a MaxEnt model are consistent with the predictions made in the papers outlined above. While the weights learned by the model are not always a perfect match to the predictions outlined here, they are able to account for the asymmetrical patterns of loanword adaptation, indicating that there is either 1) more to be understood about how these various cues are integrated into the PMap as a whole; or 2) an additional source of the bias needed in order to account for the data from Cantonese. Both are possibilities, and are discussed further in Chapter 5. At present, however, it will be sufficient for the reader to understand how phonetic cues contribute to the PMap, and what kinds of research have been done in order to determine how it is structured.

In sum, the PMap is a mechanism that has been proposed to establish an *a priori* partial ranking of Faithfulness constraints, according to whether speakers judge the change introduced by a violation of a Faithfulness constraint to be more perceptually similar than other changes introduced by violations of other Faithfulness constraints. Perceptual similarity is most analogous to the **discriminability** of two acoustic forms – one which violates the Faithfulness constraint in question, and one which does not. However, data about these comparisons are scarce, either due to a lack of phonetic testing, or due to the high perceptibility of the change involved. Therefore, other sources of information, such as **cue distribution**, are relied upon in order to establish a plausible PMap ranking. This thesis will largely rely upon the latter, as it deals with epenthesis and deletion data, which are difficult to test using discrimination tasks. It is also meant to model speakers of Cantonese, and it is not guaranteed that results from studies performed by English-speaking participants will necessarily extend to a speaker of another language. Literature about the acoustic cues to the relevant segments, however, are more readily available, and can be used as a reliable source of inference about the ranking of constraints.

With these caveats in mind, I will now turn to how a PMap-compliant ranking of Faithfulness constraints can be successfully incorporated into a MaxEnt learning algorithm.

3.2 Incorporating the PMap into a MaxEnt grammar

Recall from Chapter 2 that the general method of incorporating the PMap into a MaxEnt learner is to formulate it as a **Gaussian prior** on the weights of constraints, and to add it as a second goal of the objective function used during learning. The full objective function is repeated in (46).

(46) **MaxEnt Objective Function, repeated from (29)**

$$max(\ log(\ P(D)\ )\ ) = max\left(\ log\left(\ \prod_{\omega}^{D} P(\omega) + \varphi(w)\ \right)\ \right)$$

where $P(\omega)$ is defined as:

$$P(\omega) = \frac{exp(-\sum_i^N w_i \cdot C_i(\omega))}{\sum_{\omega'}^{\Omega} exp(-\sum_i^N w_i \cdot C_i(\omega'))}$$

from (21)

and $\varphi(w)$ is defined as:

$$\varphi(w) = \prod_i^N exp\left(\ -\frac{(\ w_i - \mu_i\ )^2}{2\sigma_i^2}\ \right)$$

from (28)

In terms of notation, I will focus on the Gaussian prior only for the remainder of this chapter; readers who are interested in learning more about the calculation of probability in MaxEnt should refer to Chapter 2 for more details.

There are three main variables to be established for the Gaussian prior: the **actual weight** of a given constraint, represented by w_i; the **target weight** of that constraint as established by the prior bias, μ_i; and the **plasticity** of the constraint weight as established by the prior bias, σ_i^2. The target weight refers to the weight that the bias has established independently of the learning process. The higher the target weight, the more influential the constraint is when selecting an output candidate. The plasticity refers to the ability of the constraint weight to deviate from the target weight. In general, the larger this number is, the more it is able to deviate from the target.

Gaussian priors are used not only to establish PMap-compliant rankings of constraints, but have been used to **regularize** the outputs of MaxEnt grammars. A regularized MaxEnt grammar will be less prone to overfitting, and will better match the researcher's expectations about the grammar's ability to generalize to new data. In order to regularize a grammar, a **uniform** Gaussian prior is established, so that the target weight μ and the plasticity σ^2 are set at the same value for all constraints – usually, μ is set to 0, indicating that the constraint is not predicted to have any effect, and σ^2 is set to 1, indicating that the constraint's learned weight should not deviate too much from the target weight. An example of what this might look like for a three-constraint grammar is provided in Figure 3-1.[1]

A Gaussian prior which instantiates the PMap crucially does not make use of a uniform prior

[1] An astute reader may have noted that these Gaussians are not strictly identical – if they were, they would overlap, and only one colour would be visible. I have changed the value for σ^2 slightly for each so as to make all three constraints visible in the plot.

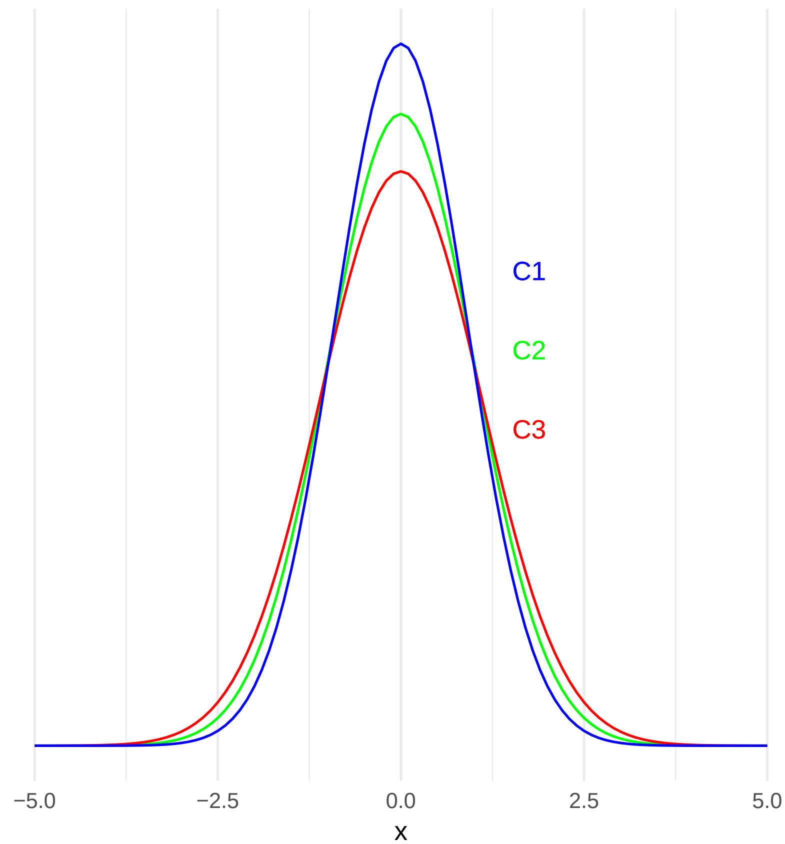
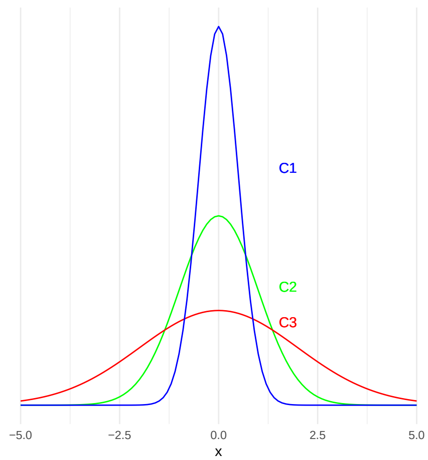


Figure 3-1: Sample Gaussian distributions with uniform regularization

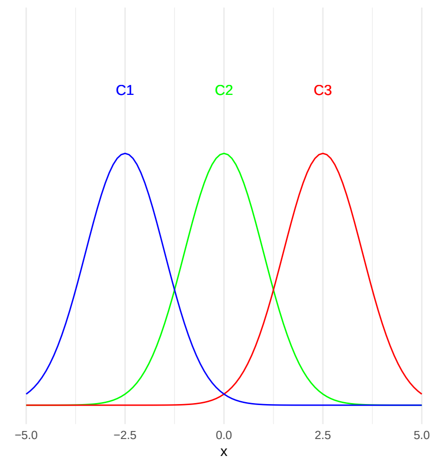
– rather, each constraint receives **its own Gaussian distribution**. There are two main ways of achieving this result, both of which have already been explored in the literature, and both of which will be explored in more detail, below. One way is to establish a **different plasticity** for each constraint, so that some constraints are more able to deviate from the uniform target weight than others; the other way is to establish a **different target weight** for each constraint, so that some constraints are predicted to be more influential than others across the board. These are both shown in Figure 3-2.

It is fairly straightforward to see how the PMap is instantiated when the target weights for each constraint are different, as in Figure 3-2b. If the weight of constraint C_1 must remain beneath the blue curve, it is unlikely to exceed the weight of constraint C_2, as most of its area does not overlap with the area under the green curve. It is even less likely to exceed the weight of C_3, associated with the red curve, as the overlap between these areas is almost non-existent.

It is less intuitive to see how the PMap is instantiated when the plasticities for the constraints are varied, as in Figure 3-2a, but for now, it should be noted that constraint C_3 contains more area



(a) PMap established by varying plasticity



(b) PMap established by varying target weights

Figure 3-2: Sample PMap-compliant Gaussian distributions

on the edges of the scale than the other two constraints, indicating that it is the constraint that is most likely to have a high weight, followed by C_2, and finally C_1. This will be discussed further when each instantiation is described in more detail, beginning with section 3.2.1.

There are other logical ways of instantiating a PMap bias using a Gaussian prior – for example, each constraint could receive both its own plasticity and its own target weight, or there could be requirements that a constraint C_1 should be weighted not only above C_2, but that their distributions should not overlap within a certain threshold of the target weight. The latter is actually closer to the original conception of the PMap as laid out by Steriade (2001), in that it would force the constraint C_1 to nearly always be weighted above C_2 – in other words, it would mirror the principle of **strict ranking** that is inherent in OT. Other potential biases will be discussed in Chapter 5.

Simply establishing a separate Gaussian distribution for two constraints is not sufficient to model a strict ranking within MaxEnt. Say that the grammar establishes a PMap-compliant Gaussian distribution for constraint C_1 and constraint C_2, such that C_1 receives a higher target weight than C_2, as in Figure 3-2b and in Figure 3-3, below. Let us also say that the learning data contain an abundance of information that C_2 (in blue) must have a high weight, but that there is no information about how to assign the weight of C_1. The grammar will then assign a high weight to C_2 and keep C_1 at its target weight, μ_1. Depending on how much data is present in the source, this may mean that C_1 and C_2 are assigned similar weights, or it may mean that C_2 outweighs C_1, as illustrated above – a **reversal** of the expected weights, reflected by the target weight in the PMap prior. The ranking is not inconsistent with the prior, however, as the weight of w_2 falls squarely under the

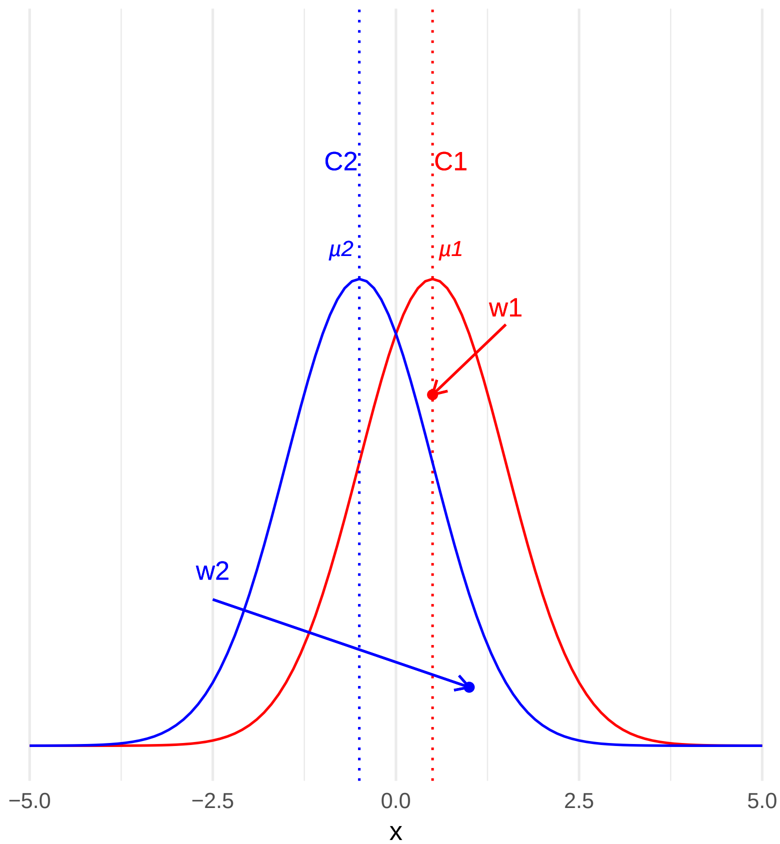


Figure 3-3: Hypothetical reversal of a weighting established by the PMap

blue curve of C_2, and the weight of w_1 falls under the red curve of C_1. In this way, the PMap bias established for MaxEnt models is a **softer bias** than the one established for strict ranking OT in Steriade (2001).

It is an open question as to whether or not this is a desirable outcome. For example, the existence of nasalization of word-final voiced obstruents in Noon (Thiès, Senegal; Niger-Congo language family, 10,000 – 50,000 speakers) would argue that the PMap ranking hypothesized in (38) can be broken, with sufficient data (Merrill, 2015), as does the existence of phonological **saltations**, where one phoneme is mapped to another that is more than minimally different than itself (White, 2013). I will leave the question of the rigidity of the PMap bias for discussion in Chapter 5. Both methods that will be discussed in more detail immediately below make use of the softer bias, and this is the kind of bias that will be explored throughout this thesis.

3.2.1 The PMap as differing constraint plasticities

The first instantiation of a PMap-compliant bias was pursued in Wilson (2006), who showed through a series of artificial grammar learning experiments that adult speakers of English are more biased towards generalizing palatalization rules when they are presented with exemplars of palatalization before mid front vowels than when they are presented with exemplars of palatalization before high front vowels. In modelling the results of his experiments through MaxEnt, he shows that this result is only achievable if we hypothesize that there is a **substantive bias** on the weights of constraints that mirrors the perceptual confusability of velar vs. palatal consonants in a variety of vowel contexts – in other words, that there is a PMap bias on constraint weights. Wilson (2006) establishes this bias over the set of constraints governed by the PMap by setting the **target weight** uniformly at **zero**, and setting each **plasticity variably**, so that some constraints can be more easily re-weighted than others. In other words, Wilson makes use of the constraint schema outlined in Figure 3-2a.

The way in which Wilson models the PMap, however, is in some ways the **reverse** of the PMap outlined in Steriade (2001). Rather than establishing a ranking among a set of contextual Faithfulness constraints, Wilson establishes a ranking among a set of **contextual Markedness** constraints, and sets them in opposition to a much smaller set of general Faithfulness constraints which receive a high weight of 10. In order to link the ordinarily output-oriented Markedness constraints to the notion of perceptual similarity – which must make a comparison between multiple output forms – Wilson posits that the relevant point of comparison is the output form that instantiates the fewest changes in order to satisfy the Markedness constraint. So, for example, a constraint that prohibits the combination of a velar consonant and a high front vowel, *ki, would be evaluated with respect to an output form that the constraint set deems to be minimally different. In Wilson's case, this is [t͡ʃi], as he only considers changes to the consonant, and not the vowel.

In order to derive the plasticities for each of these contextual Markedness constraints, Wilson converts **confusability data** from Guion (1998) into σ values by applying a **generalized context model** of classification (Nosofsky, 1986). Those readers interested in the details of a generalized context model, or GCM, are invited to consult Wilson (2006) for further details. The plasticities he derived via this method are reproduced below, in (47).[2][3] As the data from Guion (1998) lack comparisons for velar and palatal consonants before mid front vowels, Wilson chooses to interpolate

[2] It should be noted that the superscript $^{-2}$ in the table below should be interpreted literally, as the inverse square of that number, rather than as a shorthand E−2, as is sometimes done in the experimental literature.

[3] It should also be noted that the inverse square of a large number, such as 126.93, will result in an incredibly small number. Thus, these values are still consistent with the generalization from section 3.2, that the smaller the number, the less able the target is able to deviate from the target weight.

values for constraints that must make reference to these comparisons by setting their plasticity values to be the midpoint of the values selected for the high and low vowel contexts.

(47) **Varying plasticities as an instantiation of the PMap from Wilson (2006)**

Constraint	μ	σ^2
*ki	0.0	9.23^{-2}
*ke	0.0	12.68^{-2}
*kɑ	0.0	88.72^{-2}
*$\text{kV}_{[-\text{low}]}$	0.0	12.68^{-2}
*$\text{kV}_{[-\text{high}]}$	0.0	88.72^{-2}
*kV	0.0	88.72^{-2}
*gi	0.0	21.13^{-2}
*ge	0.0	40.60^{-2}
*gɑ	0.0	126.93^{-2}
*$\text{gV}_{[-\text{low}]}$	0.0	40.60^{-2}
*$\text{gV}_{[-\text{high}]}$	0.0	126.92^{-2}
*gV	0.0	126.93^{-2}

As discussed previously in section 3.1, Wilson makes use not only of highly specific constraints, such as *ki and *kɑ, but also of more general constraints, such as *$\text{gV}_{[-\text{low}]}$.

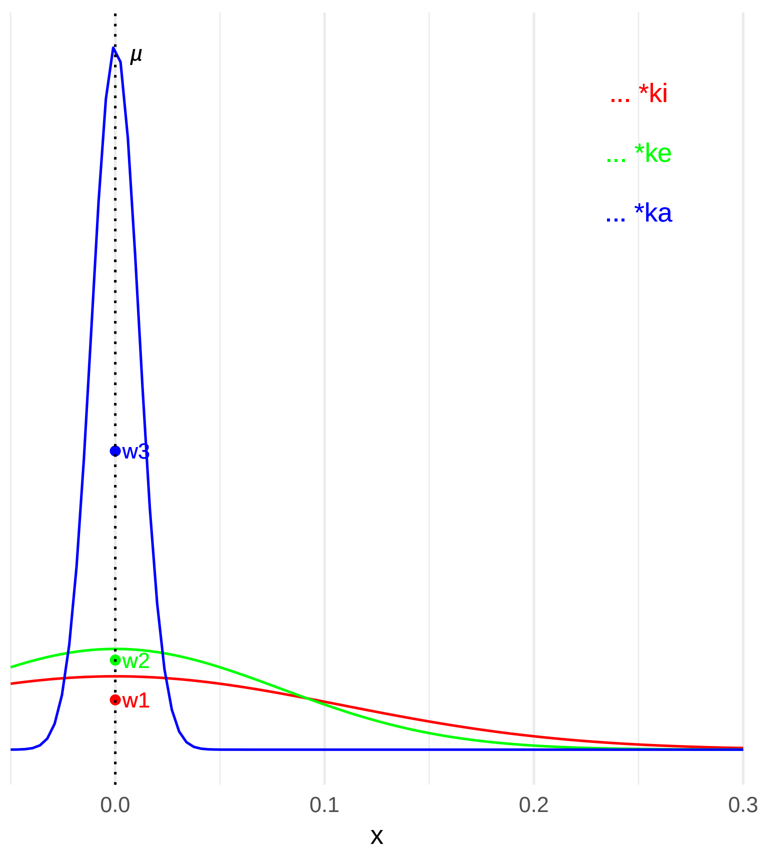


Figure 3-4: Gaussians and weights associated with the *kV constraints from Wilson (2006), prior to learning

In order to resolve the question of where these more general constraints should be ranked within the PMap, Wilson chooses to rank them along with the most discriminable pair that they encompass. So, for example, the constraint $*\mathrm{kV}_{[-\mathrm{low}]}$ encompasses both the constraints *ki and *ke, as a candidate which violates $*\mathrm{kV}_{[-\mathrm{low}]}$ will by necessity also violate either *ki or *ke. Since *ke is ranked according to its association with the comparison $\Delta(\mathrm{ke} - \widehat{\mathrm{t}\int}\mathrm{e})$, and this comparison is deemed to have a greater perceptual discriminability than the comparison $\Delta(\mathrm{ki} - \widehat{\mathrm{t}\int}\mathrm{i})$, $*\mathrm{kV}_{[-\mathrm{low}]}$ will also be ranked according to the more perceptually distinct comparison. Thus, it receives an **identical** σ^2 value to *ke in the table in (47), namely, 12.68^{-2}.

In order to show the reader how a difference in plasticity can lead to a PMap-compliant weighting, a brief example will be worked through using the priors for the constraints *ki, *ke, and *kɑ. Figure 3-4 shows the Gaussian distributions associated with each constraint. The coloured points on the dotted line, representing the target weight for each constraint, represent the weights that the MaxEnt algorithm will entertain in the absence of any phonological evidence.

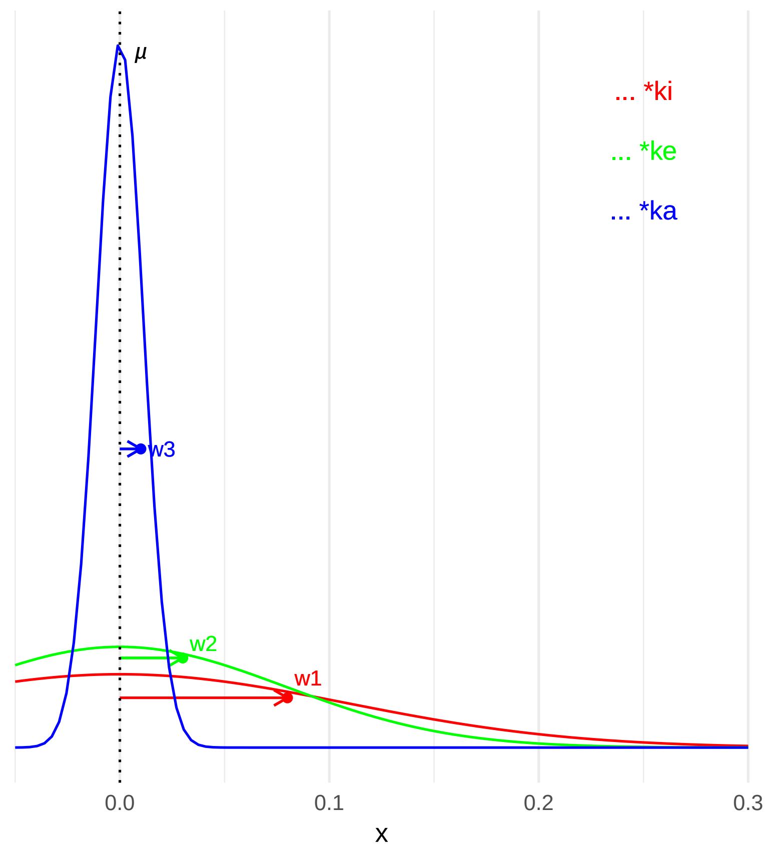
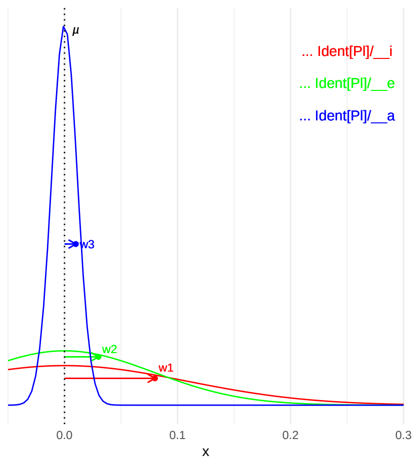


Figure 3-5: Gaussians and weights associated with the *kV constraints from Wilson (2006), after learning

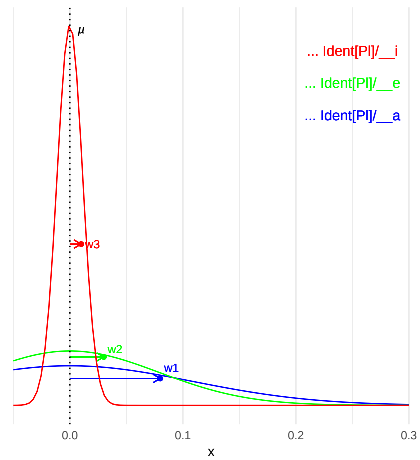
Say that the MaxEnt learner is presented with data that words containing any kV sequence must be palatalized – that is, that all constraints must be active in the grammar, and ranked above 0. When calculating the penalty for deviating from the target weight according to each constraint's Gaussian distribution, the constraint *kɑ will receive a high penalty. As a result, its weight will be adjusted only slightly upwards. In contrast, the constraint *ki will receive a low penalty, and its weight can be adjusted more freely. The constraint *ke will receive a penalty somewhere between the two, and will be adjusted accordingly as well. This is reflected in Figure 3-5. In this way, while the target weights of individual constraints do not reflect the PMap, the PMap can nevertheless **emerge** over the course of learning as a result of each constraint's plasticity.

In order to adapt this concept for use with a more standard, Faithfulness-based instantiation of the PMap, a few adjustments must be made. First of all, naturally, the plasticities of the **Faithfulness** constraints must be varied, rather than those of the Markedness constraints. Second, as the MaxEnt algorithm will be employed to model the behaviour of L1 learners of a language, rather than the behaviour of adult speakers of English learning a portion of an artificial language in a laboratory setting, the assumptions about the target weights of Markedness and Faithfulness constraints must be examined. Wilson assumes that the Markedness constraints are weighted at 0 and the Faithfulness constraints are weighted at 10, as he is modelling adult speakers of English who have learned that palatalization is not a productive process in English. However, in order to model L1 acquisition, it has been commonly assumed that the reverse should be true, so that the Markedness constraints receive a high target weight, and the Faithfulness constraints should receive a low target weight (Smolensky, 1996; Boersma and Levelt, 2000; Gnanadesikan, 2004; Hayes, 2004; Prince and Tesar, 2004, *a.o.*). The latter assumption will be pursued throughout the experiments performed in this thesis.

If the Faithfulness constraints are to be weighted low, then it must also be the case that the way in which plasticity is assigned must also be **reversed**. That is, constraints which govern more discriminable contrasts must be **more plastic** – and more likely to be weighted above 0 – than constraints which govern less discriminable contrasts. This is illustrated in Figure 3-6. The Faithfulness constraint analogs to Wilson's *ki, *ke, and *kɑ can naturally be thought of as IDENT[Pl]/_i, IDENT[Pl]/_e, and IDENT[Pl]/_ɑ, respectively. If a simple substitution were to be made, the result would be as in Figure 3-6a, where it would be predicted that the constraint with the highest weight would be IDENT[Pl]/_i, followed by IDENT[Pl]/_e, and finally IDENT[Pl]/_ɑ. However, this weighting predicts that palatalization should be most likely before **low vowels** – the opposite of what is observed cross-linguistically.



(a) Direct substitution



(b) Reversal of plasticity assignment

Figure 3-6: Gaussians and weights associated with the IDENT[Pl] analogs of Wilson's *kV constraints, after learning

In order to correctly capture the observed data, the plasticities of the Faithfulness constraints must be reversed, so that the constraint which governs the least discriminable acoustic difference is most likely to retain a low weight. This is illustrated in Figure 3-6b. In order to make a more transparent comparison between Wilson's (2006) model and the model employed by White (2013), this will be the method of assigning differing plasticity values for the remainder of this thesis. This does, of course, mean that it will not be possible to make use of the values derived by Wilson's generalized context model when assigning plasticities, as values derived by these methods will correspond to the situation outlined in Figure 3-6a. The exact method used will be discussed further in Chapter 4.

This method of encoding the PMap appears to have the advantage that the PMap will only emerge as the learner encounters evidence from their native language. As discussed previously in this chapter, this is an explicit prediction of the PMap, as speakers across different languages are not guaranteed to treat segments in similar contexts equally (Steriade, 2001). However, this is not the case: while the **weights** of the constraints emerge over the course of learning, the **plasticities** are set beforehand, and remain unchanged. As such, this model is not capable of simulating this particular hypothesis about the origins of the PMap. The construction of explicit models of how this knowledge might be built up over time are discussed briefly in Chapter 5, but are ultimately left to future research.

3.2.2 The PMap as differing target weights for constraints

In contrast to the method pursued in Wilson (2006), White (2013) instead establishes the PMap as a series of Gaussian distributions where the **target weight** of each constraint is **variable** and each **plasticity** is set **uniformly** at 0.6. The target weights were found by training an unbiased MaxEnt learner on the confusability data from Guion (1998) and using the weights learned as the target weights for future MaxEnt models. When modelling the experimental data from Wilson (2006), White made use of the constraints and weights in (48). As will be done throughout this thesis, the PMap-governed constraints are taken to be the Faithfulness constraints – in White's case, the *Map constraints of Zuraw (2007). *Map constraints assign violations to a candidate which maps segment A onto segment B, perhaps in some context or more generally. For the purposes of his comparison to Wilson's model, and for this thesis, it will be helpful to think of them as Ident[Pl] constraints that are specific to the velar stops.

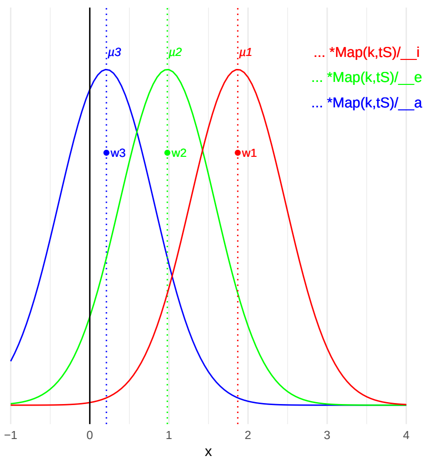
(48) **Varying target weights as an instantiation of the PMap from White (2013)**

Constraint	μ	σ^2
*Map(k, t͡ʃ)/_i	0.21	0.6
*Map(k, t͡ʃ)/_e	0.98	0.6
*Map(k, t͡ʃ)/_ɑ	1.87	0.6
*Map(g, d͡ʒ)/_i	1.22	0.6
*Map(g, d͡ʒ)/_e	1.66	0.6
*Map(g, d͡ʒ)/_ɑ	2.27	0.6

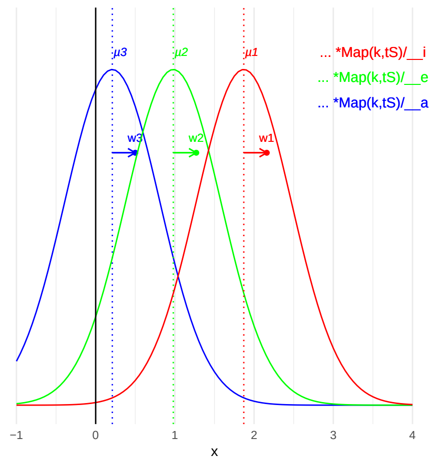
In contrast to the method undertaken by Wilson (2006), White does not make use of constraints which overlap in the contexts they apply to. Thus, there is no need to postulate how to rank more general constraints, such as *Map(k, t͡ʃ)/_$\mathrm{V}_{[-\mathrm{low}]}$.

For the sake of completeness, the Gaussian distributions associated with the *Map(k, t͡ʃ) constraints are shown in Figure 3-7a.

Even without presenting the learner with data that would require re-weighting any of the *Map(k, t͡ʃ) constraints, it would be expected that the model would be more likely to palatalize a velar consonant when located before a high front vowel than before a mid front vowel, and in turn, it would be more likely to palatalize a velar consonant before a mid front vowel than before a low vowel. As each constraint is associated with a Gaussian with the uniform plasticity of 0.6, presenting the learner with identical amounts of palatalization data will lead to identical penalties for deviating from the target weight. Thus, each constraint will be **adjusted identically**, as in Figure 3-7b, and the PMap will be maintained.



(a) Prior to learning



(b) After learning

Figure 3-7: Gaussians and weights associated with the *Map(k, t͡ʃ) constraints from White (2013)

3.2.3 Predictions regarding loanword phonology

All of the studies undertaken in Wilson (2006) and White (2013) were done in the **artificial grammar learning** experimental paradigm, where adult speakers of a language were tasked with learning a portion of an artificial grammar and tested on how well they learned the process under study. The purpose of the MaxEnt models in both studies was to show that inclusion of a substantive, PMap-style bias was a better model of the experimental data than an unbiased model that made use of the same constraints. Both kinds of PMap-style priors improved the performance of the MaxEnt models in each study, and it can be inferred from this that they are both adequate for modelling the **end-state** of learning in adult subjects.

However, the two different kinds of priors make different predictions about the **initial state** of the learner. The prior used by Wilson (2006) – hereafter the **plasticity prior** – makes the prediction that in the absence of phonological evidence, all Faithfulness constraints have an **equal weight**. Thus, if a learner is presented with a novel phonological form that violates some Markedness constraint in the language, there should be no predictions about how that phonological form should be repaired, and it is equally likely that **any repair** could be chosen. In other words, this prior predicts that at the outset of learning, the **too many solutions** problem is still at large.

In contrast, the prior used by White (2013) – hereafter the **weight prior** – makes the prediction that in the absence of phonological evidence, there will nevertheless be a **preference for weighting** certain Faithfulness constraints above others. Thus, if a learner is presented with a novel phonological form that violates a Markedness constraint in the language, it would be expected that the repair that violates the constraint at the bottom of the weighting hierarchy should be chosen. This is exactly what is observed in adult languages, where there is a **preferred repair** for cross-linguistically marked phonological structures (Steriade, 2001).

These predictions only apply to areas of the grammar where there is an **absence of evidence** as to what the desired repair should be. This is just the situation that can be encountered within **adaptation of loanwords** into a new phonological system. It is often the case that a loanword can contain an illicit segment or phonological structure in the language of adoption, such as the voiced palato-alveolar fricative [ʒ] of French [ljɛʒ] into English [liːd͡ʒ], 'liege'; or the [sr] cluster of Sinhala [sriː lam̩kaː] into English [ʃɹiː lɑːŋkə], 'Sri Lanka'. Furthermore, this illicit segment or phonological structure is often one that has never been seen before in the native phonology, meaning that the phonological grammar should have **no hypotheses** about how to repair the structure based on what they may have learned from phonological alternations or regular phonological processes. However, these speakers should still have access to their knowledge of the PMap, and will be able to use that knowledge to select an appropriate and consistent repair or set of repairs to the illicit loanword structure.

The goal of this dissertation is to test how well these two models of the PMap perform when they are incorporated into a MaxEnt model of the native Cantonese vocabulary. Recall from Chapter 1 that Cantonese not only does not have consonant clusters within syllables, but also does not allow sibilants in syllable-final position. Cantonese speakers nevertheless repair English loanwords containing these structures in specific ways: sibilants in illicit positions within a syllable will almost always trigger vowel epenthesis, while consonant clusters can trigger either epenthesis or deletion, depending on their location within a syllable. Based on the discussion above, the preference for only two repairs seems to indicate that the weight prior is most likely to be the correct model of the PMap for use in MaxEnt grammars. However, it may still be the case that the distributional facts about the phonological segments of Cantonese could cause some of the weights of the relevant Faithfulness constraints to be increased, meaning that the plasticity prior may also be a viable option for modelling these data. These models will now be explored in Chapter 4, below.

Chapter 4

Predictions for Loanwords: A Case Study in Cantonese

As stated at the end of the previous chapter, the two methods for modelling the PMap through the establishmet of variable Gaussian priors on constraint weights make different predictions about how a MaxEnt model will behave in the **absence of phonological evidence**. If the Gaussians are established as done in Wilson (2006), by varying the **plasticity** of each constraint, presenting a model with a phonotactically illegal structure may lead to a variety of **equally likely** repairs to that structure. This arises because the target weights of each constraint are equivalent, and if there is no reason to adjust the weights of these constraints, they will remain at the target weight. However, if the Gaussian priors are established as done in White (2013), by varying the **target weights** of each constraint, presenting a model with a phonotactically illegal structure will lead to a preference for a single repair, or at the very least, to **asymmetric probabilities** for repairs.

One domain where researchers can be reasonably sure that there is an absence of overt phonological evidence for repairs to marked structures is that of **loanword phonology**. This makes it an excellent testing ground for the two models of the PMap introduced above. For example, say that a language has banned consonant clusters, but that speakers are tasked with adapting an English word such as [ˈstæmp], 'stamp', into the language. Assuming the language itself has no morphophonological processes that dictate which repair to choose, as outlined in Chapter 1, the repairs that are ultimately chosen can help elucidate how the PMap-governed constraints are weighted in that language.

Recall from Chapter 1 that Cantonese contains two phonotactic constraints which are **never violated** in the native language – *Complex and CodaCond, which ban **tautosyllabic consonant**

clusters and **sibilants in coda**, respectively. Since these constraints are never violated by words from the native vocabulary, it can be reasonably assumed that speakers do not need to repair them, and thus will never learn a repair through such means as morpho-phonological alternation. Speakers nevertheless have come up with a sophisticated series of repair strategies, which are summarized in (49), repeated from (17).

(49) **Generalizations about repairs to illicit structures in loanwords in Cantonese**

a. Word- or syllable-initial consonant clusters can be repaired either by **deletion** of the second member of the cluster or by **vowel epenthesis** between the consonants.

b. In contrast, word- or syllable-final consonant clusters are always repaired by **deletion**.

c. Word- or syllable-final sibilants are always repaired by **epenthesis**.

d. When a sibilant is present in a consonant cluster, it biases the attested repairs further towards **vowel epenthesis**, instead of or in tandem with consonant deletion.

As discussed in Chapters 1 and 3, the above generalizations are **consistent** with a weighting of constraints determined by the PMap. For onset and coda clusters, the **availability of cues** in each environment can help explain the difference in attested repair types. When a consonant is word-final, it lacks cues from a following segment. Thus, the only cues to its presence must come from the preceding segment, and if that preceding segment is also a consonant, there will be fewer cues than if the preceding segment is a vowel. For example, if a word-final stop consonant is preceded by a vowel, the cues to its presence can include formant transitions from the vowel to the stop, a sharp decrease in intensity, a drop in fundamental frequency, the presence of a voice bar in the stop itself, *etc.* (Steriade, 2001; Wright, 2004). However, if it is preceded by another consonant, some of these cues may be less noticeable, such as a smaller decrease in intensity if the preceding consonant is a sonorant; or absent altogether, if the preceding consonant is an obstruent and has no formant structure (Steriade, 2001). Thus, it is unsurprising that a word-final consonant in this environment is more easily confusable with silence, and is overwhelmingly deleted by speakers of Cantonese. When a consonant is located before a vowel, however, it has access to all of the cues present in the following segment, such as formant transitions, a sharp increase in intensity, a rise in the fundamental frequency, presence of a release burst, *etc.* (Steriade, 2001; Wright, 2004). Since consonants in this environment have access to these additional cues, it is expected that this consonant is less confusable with silence, and has a much higher rate of retention via vowel epenthesis in Cantonese.

The special status of sibilants is also consistent with a PMap interpretation. Sibilant obstruents are inherently very perceptible due to their abundance of **high frequency noise** (Wright, 2004). Thus, even though their transitional cues may be attenuated or non-existent in the context of other consonants, their internal cues are sufficiently salient so as to be difficult to confuse with silence.

Thus, it might be expected that speakers of Cantonese would prefer to repair illicit structures that contain sibilants via vowel epenthesis across the board, rather than deleting the sibilant. This pattern is not unique to Cantonese, but is also exhibited by other unrelated languages when adapting loanwords with sibilants, lending credence to this interpretation of the facts.

The above discussion, of course, assumes that any differential weighting of constraints is due to the PMap. There is, however, a simpler explanation for the source of the differential weights necessary to generate the patterns of Cantonese loanword adaptation. Namely, it could be that these weights arise throughout the course of **unbiased phonological learning**, due to differences in the distributions of the different kinds of phonological segments within Cantonese. After all, Cantonese-learning children must learn to preserve stop consonants in coda, which requires promotion of any MAX-C/_# constraints over *CODA; and they must likewise learn that *COMPLEX must have a much higher weight than any constraint that would call for retention of both members of a consonant cluster. Differences in the contextual distribution of consonant types within the native Cantonese vocabulary could very well cause differences in the amount that each kind of Faithfulness constraint is adjusted, which could in turn lead to the asymmetries exhibited when Cantonese speakers are tasked with adapting phonologically-illicit loanwords into their own language. The exact process by which this could occur is discussed more in section 4.3.

Under the assumption that a suitable quantitative translation of the PMap can be found for the generalizations of Cantonese loanword adaptation, the next question to be asked is how well the proposed models predict them. The hypotheses for each are outlined in (50).

(50) **Hypotheses for Experiments 1, 2, & 3**

a. *No prior (Null hypothesis)*: Predictions about Cantonese loanwords may be met, depending upon the distribution of relevant segmental contexts in the native vocabulary.

b. *Plasticity prior*: Predictions about Cantonese loanwords may or may not be met, depending on the distribution of relevant segmental contexts in the native vocabulary. This is because the PMap will only emerge in the presence of evidence. If the model does not accurately predict the correct kinds of repair for each loanword, it is expected that the grammar will select a variety of equally likely repairs.

c. *Weight prior*: Predictions about Cantonese loanwords should always be met, as the PMap should be maintained in the absence of evidence.

The hypothesis about the plasticity prior should not be considered an outright failure to predict PMap-compliant behaviour in loanwords, contrary to what might be expected given the discussion in the previous chapter. While the MaxEnt learner may not encounter any **direct** evidence as to the proper repair to a consonant cluster or to a sibilant in an unlicensed position, it must still rule out extraneous deletions and epentheses when learning the native phonology of Cantonese. In doing so, it will still have reason to adjust the weights of the Faithfulness constraints, and the amount by which these weights are adjusted is governed by the plasticity prior. Thus, the simple fact that the weights of these constraints must be adjusted throughout the course of normal phonological learning could lead to a PMap-like asymmetry, provided the plasticities of each constraint are different. This will be discussed in more detail in section 4.4.

The goal of the following experiments is to test each of these hypotheses by simulating the result of learning only from the **native Cantonese vocabulary** using MaxEnt. The general scheme for the experiments reported on in this chapter is summarized in (51).

(51) **General outline of methodology**

a. **Generate** a set of tableaux for native Cantonese words.
b. **Train** a MaxEnt model on the generated tableaux to arrive at a proposed **weighting** for Cantonese.
c. **Test** the resulting weighting of constraints on a set of both native Cantonese words and English loanwords to see which repairs are selected.
d. **Compare** the results of each test to an idealized model.

In training the model only on native Cantonese words, the result will mirror the knowledge that a **monolingual** Cantonese speaker should have by the time they are an adult. If the experiment calls for it, the PMap – or a similar proxy bias – will be incorporated during learning as an independent condition on the MaxEnt objective function described in (20) in Chapter 2. This is consistent with the idea that the grammar is influenced by independent phonetic knowledge that the speaker has obtained through some other source. The test is then to present the learner with a novel set of English-language data and predict how it will repair illicit structures, without performing an additional learning step to incorporate the new loanwords. If the trained model is accurate in its predictions, we can then say that the prior is successful in replicating the PMap.

The structure of this chapter is as follows. First, the set of **constraints** used across all experiments will be outlined. Then, the Cantonese **data set** that is used to train the models will be introduced, followed by a test of the null hypothesis. Then, a plasticity prior will be constructed by making use of the weights generated by a grammar trained to best model the English-language data presented in the test set. The strengths and weaknesses of this approach will be discussed in detail before testing the behaviour of the plasticity prior. Finally, the behaviour of the weight prior will be tested, and broad results discussed.

4.1 Constraints

The constraints which are of the most interest for the experiments outlined in this chapter are, of course, the Faithfulness constraints which govern the insertion and deletion of segments – *i.e.*, MAX and DEP. As discussed in Chapter 3, these constraint classes will be broken up into a series of contextual Faithfulness constraints which can be **independently ranked** according to the PMap of Steriade (2001). There are two main ways in which the Faithfulness constraints will be divided – first by the **target segment** and then by its **context**. All points of division will be made according to the **manner** of the segments involved, as this is a major point of difference in the confusability of segments with silence (Wright, 2004; Yun, 2016). The main points of division for the target segments is listed in (52), and the main points of division for the contexts is listed in (53).

(52) **Classes of Target Segments for Faithfulness Constraints**

a. Non-sibilant **obstruents**, represented by T
b. **Sibilant** obstruents, represented by S
c. **Sonorant** consonants, represented by R
d. **Vowels**, represented by V

(53) **Classes of Contexts for Faithfulness Constraints**

a. **Word edges**, represented by #
b. **Obstruent** consonants, represented by T
c. **Sonorant** consonants, represented by R
d. **Vowels**, represented by V

For the **target segments**, the divisions made were meant to mirror established patterns of sonority among vowels, sonorants, and obstruents, with the addition of a division between sibilant and non-sibilant obstruents. This additional division was made in order to adequately capture the differential behaviour of obstruents in Cantonese, since sibilants pattern differently with respect to loanword adaptation than other obstruents. These segments are also known for having much stronger internal cues to their identity than other obstruents, and a division is often made between them when it comes to their perceptibility (Wright, 2004, *a.o.*).

For the **contexts**, the division between sibilant and non-sibilant obstruents was not maintained, as it was hypothesized that both would be relatively weak carriers of information as to the presence or absence of an adjacent segment. The division between sonorants and vowels was maintained, however, as they have been known to be stronger carriers of information about the nature of an adjacent segment, with vowels being better at conveying information than sonorants. The addition of the word edge as a possible context was done to ensure that all constraints had both a preceding and following context, and is meant to represent an environment where there are no cues to an adjacent segment's identity.

All combinations of preceding context, target segment, and following context for both Max and Dep constraints were included in the constraint set. This was done to avoid **significant constraint overlap**, which would make assigning weights to PMap-governed constraints difficult. If there were contextual overlap, a decision would have to be made as to which context to use as the default when a constraint spans many contexts. For example, the constraint Max-T/_V covers both the case where an obstruent is preceded by silence, and the case where an obstruent is preceded by a vowel. As discussed previously, these two environments differ widely with respect to the number of cues available for identifying the obstruent – there will be substantially more cues as to the identity of the obstruent if preceded by a vowel than if preceded by silence. How, then, should we establish a PMap-compliant weight of Max-T/_V, when some instances of pre-consonantal obstruents will be better cued than others? Establishing different constraints for each of these environments avoids this problem, and also allows for a more transparent application of the PMap to the constraint set. However, readers who are interested in learning more about the potential effects that an overlapping constraint set could have on a MaxEnt learner should consult section 5.2 of Chapter 5.

Constructing the set of Faithfulness constraints to avoid contextual overlap also allows each constraint to be weighted **independently**. Thus, there should be no candidate with a single epenthesis or deletion which requires the MaxEnt algorithm to re-weight multiple Faithfulness constraints. When this constraint set is used to train a model on the native Cantonese vocabulary, it will aid in making inferences about the distribution of segments within the training set, since the relative proportions of each segment in context will only contribute to the weight of a single constraint. Likewise, when this constraint set is used to train a model on the attested English loanwords, it will allow straightforward inferences about the Cantonese-specific PMap to be made. Since the constraints are independently weighted, only those constraints which contribute to the retention or deletion of particular segments will receive high weights, while those which do not contribute will remain low-weighted. These relative rankings can then be compared against what we know of the PMap to determine if they are consistent with this hypothesis. This will be undertaken in section 4.4.2.1, below.

Constructing a set of independent constraints was done in the following way: Using the classes in (53) as both preceding and following contexts results in $4 \times 4 = 16$ different unique contexts. Each context is then applied to each of the target segments in (52) to give $16 \times 4 = 64$ different segmental trigrams. Finally, each segmental trigram is then used to construct one Max and one Dep constraint, for a total of $64 \times 2 = 128$ Faithfulness constraints. A sample set of constraints is provided in (54). Ident constraints are not considered, since they have no impact on whether a segment is present or absent.

(54) **Schema for Faithfulness Constraints**

$$\left\{ \begin{array}{c} \textsc{Dep} \\ \textsc{Max} \end{array} \right\} - \left\{ \begin{array}{c} \mathrm{T} \\ \mathrm{S} \\ \mathrm{R} \\ \mathrm{V} \end{array} \right\} / \left\{ \begin{array}{c} \# \\ \mathrm{T} \\ \mathrm{R} \\ \mathrm{V} \end{array} \right\} \_ \left\{ \begin{array}{c} \# \\ \mathrm{T} \\ \mathrm{R} \\ \mathrm{V} \end{array} \right\}$$

which generates constraints such as:

- Dep-V/#_T
- Dep-S/R_R
- Max-R/V_T
- Max-T/R_#
- *etc.*

In addition to the Faithfulness constraints considered, a series of **syllabic Markedness constraints** was included in order to correctly model the syllable structure of Cantonese. This set of constraints included a series of *Peak constraints which governed the licit syllable peaks in Cantonese; a constraint enforcing syllable bimoracity, Bimoraic (Kenstowicz, 2012); the familiar Onset and *Coda constraints; and the two constraints which are hypothesized to motivate the repairs attested in Cantonese loanwords, *Complex and CodaCond. These two are defined in more detail in (55).

(55) **Markedness constraints responsible for loanword repairs**

a. *Complex: A syllable onset or coda may not contain more than one segment.

b. CodaCond: A syllable coda may only consist of the segments { j, w, m, n, ŋ, p, t, k }. (Silverman, 1992)

*Complex is used to account for the ban on tautosyllabic consonant clusters, while CodaCond is used to account for the ban on sibilants in coda, among other syllabic restrictions. It is hypothesized that after a model is trained on the native Cantonese vocabulary, these two constraints will have high weights, reflecting their lack of violations in the output forms of Cantonese.

Finally, a set of **phonotactic** constraints consistent with the Cantonese syllabary presented in Bauer and Benedict (1997) were included in order to prevent possible confounds in the data. Since the simulations explored below make use of actual native words of Cantonese, the data that the MaxEnt algorithm will be learning from will contain a variety of gaps in the inventory of possible syllables, such as the lack of sequences of a rounded vowel followed by a labial consonant. In order to avoid having the MaxEnt algorithm account for these gaps by adjusting the weights of the Faithfulness constraints, the phonotactic constraints were included. There were a total of 12 of these constraints. A full list of constraints, both Faithfulness and Markedness, is included in Appendix B.

4.2 Data

4.2.1 Sources

The data used for the experiments outlined in this chapter come from a variety of sources. For the **loanwords**, I primarily worked off of the list of attested English loanwords in Kenstowicz (2012), supplementing them with additional data from Silverman (1992) and Yip (1993, 2006), all of whom cite Bauer and Benedict (1997) as a main source, with the addition of their own elicitations in many cases. It will be assumed throughout that speakers have access to the **full phonetic form** of the English loanwords in question, without significant misperceptions. This is not a position that is widely held in the literature on English loanword adaptation in this language – see especially Silverman (1992) and Yip (2006), who claim that the final consonant in word-final clusters is not perceived. I will nevertheless be adopting the assumption that they are perceived, as all loanword adaptation generalizations outlined in (49) are consistent with an interpretation which attributes all differences in repair frequencies to the PMap, while other interpretations must make use of an additional mechanism to account for at least one of the repair types. With this in mind, the **RP** forms are provided as the underlying or original forms of each loanword, as this is the dialect that most Cantonese speakers will have had the most exposure to. The full list of words and their sources can be found in Appendix A.

For the **native vocabulary**, I made use of **CC-Canto**, an open-source dictionary of approximately $22,350$ Cantonese words and phrases, which accommodates not only the traditional and simplified Chinese spellings of vocabulary items, but also provides both the Mandarin Pinyin and the Cantonese Linguistic Society of Hong Kong, or Jyutping, spellings. The dictionary is based off of another open-source Chinese dictionary called CC-CEDICT, which provides only the Mandarin pronunciations and interpretations of words. The current expansion is the result of twelve native speakers of Cantonese providing transcriptions for each of the vocabulary items listed in CC-CEDICT. Each new transcription was checked by up to three different editors for consistency (Pleco Software, 2016).

I have largely chosen to maintain the Jyutping transcription system throughout. The Jyutping graphemes for each phoneme and tone of Cantonese are provided in (56). As I am not a speaker of Cantonese, converting the Jyutping transcription into the IPA standard for over 22,000 words is beyond my capabilities. For example, converting the Jyutping graphemes into IPA phonemes would require the disambiguation of the graphemes *e* and *o*, which can correspond either to [e] or [ɛː] and [o] or [ɔː], respectively. While this means I am working off of a slightly simpler Cantonese grammar in this respect, it is hoped that collapsing across vowel categories will not impact the analysis of vowel epenthesis and consonant deletion, and that this is an inconsequential simplification.

(56) **Phoneme-to-Grapheme Correspondences of Jyutping Transcription**

Consonants				*Vowels*		*Tones*		
Jyutping	*IPA*	*Jyutping*	*IPA*	*Jyutping*	*IPA*	*Jyutping*	*Diacritic*	*Description*
b	p	g	k	aa	aː	1	é	High level
p	p^h	gw	k^w	a	ɐ	2	e᷄	Mid rising
f	f	k	k^h	e	e	3	ē	Mid level
m	m	kw	k^{wh}		ɛː	4	ê	Low falling
d	t	h	h	o	o	5	ě	Low rising
t	t^h	ng	ŋ		ɔː	6	è	Low level
s	s	w	w	oe	œ			
z	t͡s			eo	ø			
c	$\widehat{ts}^h$			i	iː			
n	n			u	uː			
l	l			yu	yː			
j	j							

When adapting the Jyutping tonal system into IPA, I have chosen to use diacritics for the tonal markers rather than the tone bars which are standard. I feel that this makes the overall IPA transcription easier to read, and allows the reader to focus on the segmental phonemes rather than the tones. As the tones are orthogonal to the loanword adaptation phenomena discussed in this thesis, and as they are removed from the Jyupting transcriptions to be used in the learning process, this is largely a cosmetic change.

I have chosen to modify the Jyutping transcription when it comes to coda consonants somewhat, in order to simplify the code necessary to perform the large-scale simulations detailed below. Rather than using the graphemes *i* and *u* for glides in coda, I have chosen to replace them with the graphemes *j* and *w*, respectively. All other transcriptions remain as they are in the CC-Canto dictionary.

Since CC-Canto is an open source project and relies on crowd-sourced data entry, transcriptions which contain loanwords may be inconsistently transcribed or tagged as loans in the English translation. Furthermore, even if a loanword is tagged as such, its language of origin is often not included. For safety, I have chosen to filter out all entries identified as loanwords within the dictionary, regardless of whether their language of origin is English, Mandarin, or some other language. I have also chosen to filter out any entries which contained fullwidth Latin characters[1] in the Chinese logographic transcription, as it was assumed that the use of the Latin characters represented the inclusion of a loanword into the vocabulary item. Finally, any entries which contained an illicit syllable according to the modified version of the Jyutping transcription listed above was removed. This

[1]These graphemes look the same as the graphemes of the Latin alphabet, with the exception that they include additional whitespace around them in order to be compatible with the fixed-width fonts used to write traditional and simplified Chinese characters.

included transcriptions which contained syllables such as 'p5', which consists only of an obstruent and a tone; and also transcriptions which contained the symbol '/', which indicated the presence of a variable pronunciation for one of the syllables. This eliminated 138 entries, leaving around 22,200 for inclusion in the simulations reported on below.

4.2.2 Tableau Generation

Of course, having access to full word and phrase forms of words of Cantonese is not sufficient for training a MaxEnt learner – it is necessary to have full a **tableau** for each of these entries. Due to the volume of the data considered, I developed a Python script for automatically generating these tableaux from both the CC-Canto data and the Jyutping transcriptions of the English loan-word data. The Python materials necessary for running the script are available via Github at `https://github.com/erin-daphne-olson/canto-tools`.

The script had slightly different procedures for words which were native to Cantonese and for words which were loaned from English. For the native Cantonese words, the script was designed to take a list of Jyutping transcriptions and generate all possible **single-change candidates**, as outlined in (57).

(57) **Candidates Generated for Native Cantonese Words**

- The **fully faithful** candidate, with any potential tonal information removed.
- All **single deletion** candidates, excluding tones.
- All **single epenthesis** candidates, where the epenthetic segment could be:
  - A non-sibilant obstruent, T
  - A sibilant obstruent, S
  - A sonorant, R
  - A vowel, V

It should be noted that in the cases of epenthesis, a **proxy character** was used to represent the epenthetic segment – T for obstruents, S for sibilants, R for sonorants, and V for vowels. This was done in part so that the model could remain agnostic about which segment to choose for the purposes of epenthesis. Epenthetic segments in Cantonese can range from [iː] and [yː] to vowels as low as [o], so having a single character to represent them all allowed them to be unified under a single approach. This orthographic convention was also done to aid in the automatic computation of Dep constraints by making the epenthetic segment distinct from all other segments of Cantonese.

In all these cases, the **fully faithful** candidate was assumed to be the winning candidate, and as such was assigned a probability of 1. All other candidates were assigned a probability of 0. Each candidate was then automatically assessed by the constraints detailed in the previous section, and their violations were tabulated. All candidates and violations were then saved as a single row of a larger training file, along with their underlying form and their probability. This generated tableaux such as in (58).

(58) **Sample entries for *ze1 jyun3*, 'to sigh or whine'**

input	output	p	CodaCond	Peak(Vowel)	···	*Note*
ze1 jyun3	ze jyun	1	0	0	···	Fully-faithful candidate
ze1 jyun3	e jyun	0	0	0	···	Deletion candidates
ze1 jyun3	z jyun	0	0	1	···	↓
ze1 jyun3	ze yun	0	0	0	···	
ze1 jyun3	ze jn	0	0	1	···	
ze1 jyun3	zeT jyun	0	0	0	···	T-epenthesis
ze1 jyun3	ze jTyun	0	0	0	···	
ze1 jyun3	ze jyunS	0	1	0	···	S-epenthesis
ze1 jyun3	zeS jyun	0	1	0	···	
ze1 jyun3	zRe jyun	0	0	0	···	R-epenthesis
ze1 jyun3	ze jyuRn	0	0	0	···	
ze1 jyun3	V ze jyun	0	0	0	···	V-epenthesis
ze1 jyun3	ze V jyun	0	0	0	···	
···	···		···	···		

The candidates generated were limited to one change each for the purposes of simplicity in computation. In addition, since most candidates will be ruled out as possible repairs on the virtue of a single constraint violation, adding more violations to the same constraint is unlikely to change the outcome of learning. For example, a candidate such as *ce1 leon4 zin3*, the Jyutping transcription of [t͡sʰé lɵ̂n t͡sīn], 'a wheel war', could surface with the deletion of the first vowel, *c leon zin*, or with deletion of the first and third vowels, *c leon zn*. Both candidates would violate the constraint Max-V/T_R, except that the latter candidate would violate it twice, while the former would violate it only once. Since the fully faithful candidate is predicted to win in both cases, it should be sufficient to establish the weight of Max-V/T_R high enough to eliminate *c leon zin*, and trust that the additional violation introduced by *c leon zn* will also cause it to be ruled out.

When dealing with English loanwords, a slightly different procedure was followed. First, a Jyutping transcription of the English loan was approximated for each loanword. This was taken to be an approximation of the perceived form of the English word by a monolingual speaker of Cantonese. For example, the word [ˈɒɹɪnd͡ʒ] 'orange' was transcribed as Jyupting *o1 leonc2*, the orthographic equivalent of the attested surface form [ɔ́ː lɵ̄n t͡sí], minus the word-final epenthetic vowel. This was done to make the English forms conform to the orthographic conventions established for the native Cantonese words and to simplify the process of adapting the candidate generation procedure from Cantonese to English.

In addition to the candidates generated by the procedure outlined in (57), a select number of candidates with **multiple** changes to the underlying form were considered. This was due to the fact that some English loanwords with word-final SC clusters exhibit both deletion of the word-final consonant and vowel epenthesis in order to save the sibilant. For example, the English word [ˈmɑːsk] ‘mask’ is adapted in Cantonese as [máː síː], where there is **simultaneously** epenthesis of the vowel [íː] and deletion of the consonant [k]. The procedure in (57) will not be able to generate these candidates, but they must be present in the tableaux in order to be selected as winners by a MaxEnt learner. This change is reflected by the following schema:

(59) **Candidates Generated for English Loanwords**

- The **fully faithful** candidate, with any potential tonal information removed.
- All **single deletion** candidates, excluding tones.
- All **single epenthesis** candidates, where the epenthetic segment could be:
  - A non-sibilant obstruent, T
  - A sibilant obstruent, S
  - A sonorant, R
  - A vowel, V
- All combinations of **one deletion** and **one vowel epenthesis**.

The above procedure will, of course, generate candidates which are also largely transcribed in Jyutping, excluding any numerals associated with tone. As such, the winning candidate will be the transcription that is closest to the **attested** form of the loanword. So for the proposed input for ‘orange’, *o1 leonc2*, the winning candidate will be *o leon cV*, with final vowel epenthesis. These winning candidates were stored next to their proposed underlying forms in a text file to be read in by the generation script. While the majority of the MaxEnt learners that will be tested in the following experiments will not have access to any of the winning candidates stored in this text file, it is marked in the tableau files for two reasons. First, it allows the `solve.R` script of Pater and Staubs (2013) to assess learner performance. Second, it allows for the construction of an accurate loanword grammar within MaxEnt, to replace the sketch provided in Chapter 1.

As with the native Cantonese words, each candidate generated by the procedure in (59) was then assessed by the full set of constraints described in section 4.1 and stored in a test file to be read by the MaxEnt learner. Some sample candidates for the English loanword ‘orange’ are displayed in (60).

(60) **Sample entries for *o1 leonc3*, 'orange'**

input	output	p	CodaCond	*Complex	···	*Note*
o1 leonc2	o leonc	0	1	1	···	Fully-faithful candidate
o1 leonc2	leonc	0	1	1	···	Deletion candidates
o1 leonc2	o eonc	0	1	1	···	↓
o1 leonc2	o lnc	0	1	0	···	
o1 leonc2	To leonc	0	1	1	···	T-epenthesis
o1 leonc2	o Sleonc	0	2	1	···	S-epenthesis
o1 leonc2	oR leonc	0	1	1	···	R-epenthesis
o1 leonc2	o leon cV	1	0	0	···	V-epenthesis; winning candidate
o1 leonc2	o lVnc	0	1	1	···	Deletion and V-epenthesis
···	···		···	···		

For the majority of the experiments outlined below, the procedure for generating candidates outlined in (57) was applied to all of the data used to **train** the MaxEnt learner. The procedure for generating candidates outlined in (59) was used for all of the data used to **test** the end result of each learning process. Thus, whenever mention is made of a "training set", it can be assumed that each vocabulary item in the training set will appear in a tableau consisting only of single-change candidates. Likewise, whenever mention is made of a "test set", it can be assumed that each vocabulary item appears in a tableau with all of the relevant single-change candidates, plus those with one vowel epenthesis and one deletion.

4.3 Experiment 1: Testing the Null Hypothesis

As discussed in the introduction to this chapter, a MaxEnt learner may be able to arrive at a ranking of constraints which accurately predicts loanword repairs through some encoding of the PMap. However, it could be that such a learner arrives at this ranking through the course of relatively **unbiased learning**. The latter solution is the simplest, in that it does not require the differential ranking of constraints, and as such does not require that the learner access a separate source of linguistic knowledge throughout the learning process. The purpose of this experiment is to see whether this simpler model can explain the patterns of Cantonese loanword adaptation. If it is sufficient, then there is no need to incorporate the PMap through an asymmetrically encoded bias, either via plasticity or via target weights, and a different pattern of loanword adaptation must be sought.

Two questions may arise to the reader at this point. First, how can Cantonese – a language which respects *Complex universally and that has a vocabulary which consists mostly of monosyllabic words – ever encounter evidence on how to treat consonant clusters, which are so prevalent in English loanwords? And second, how can an **asymmetric** constraint weighting arise from the native Cantonese lexicon alone?

While it is true that Cantonese words consist largely of monosyllables lacking complex margins, this is not true of all lexical entries. Here, I make the distinction between a **word**, which I take to be essentially equivalent to a prosodic word, and a **lexical entry**, which can include constructions such as compound words or idiomatic readings of entire sentences. Although no word of Cantonese will ever contain a sequence of consonants that appear in English loans, there are many lexical entries of Cantonese which can contain the consonant clusters in question. For example, while there are no Cantonese words which contain the sequence [kl], as in the English word [ˈkʰliːn], 'clean', there are lexical items which do contain this sequence, albeit split across syllable boundaries. A selection of examples are presented in (61).

(61) **Lexical items which contain [kl] sequences**

Lexical Item	*IPA*	*Gloss*
jan4 do1 saw2 goek3 lyun6	[jân tó sɐ̌w kœ̄**k l**ỳːn]	'too many cooks in the kitchen' (Pleco Software, 2016)
ng5 zok3 low2	[ŋ̌ t͡sɔ̄**k l**ɔ̌w]	'post-mortem examiner; funeral parlour worker' (Pleco Software, 2016)
zuk6 laj6	[t͡sʊ̀**k l**ɐ̀j]	'customary rule' (Pleco Software, 2016)
bok3 laam5 lwan4 syu1	[pɔ̄**k l**ǎːm kʷɐ̂n sýː]	'to be well-read' (Pleco Software, 2016)
sej3 sej3 luk6 luk6	[sēj sēj lʊ̀ː**k l**ʊ̀ːk]	'appropriately, justly' (Pleco Software, 2016)

If these lexical entries are taken into account, they can act as a source of consonant-consonant sequences. These sequences, in turn, can act as a source of contexts over which the trigram constraints discussed in section 4.1 can be applied. In this way, a native speaker of Cantonese can have access to data about consonant-consonant sequences that are present in many of the English loanwords they have adapted.[2]

In response to the question of how **asymmetric** constraint weightings can arise within Cantonese, I will walk through a brief example using the [kl] sequences shown above in order to illustrate. One of the attested repairs to such sequences when they appear in loanwords is to delete the second member of the consonant cluster. For instance, English [ˈmʌj**k**ɹəˌfəʊn] 'microphone' is adapted as Cantonese [mɐj **k**ow foŋ] (**?**). In order for this repair to be attested, it must be the case that the weight of MAX-T/V_R exceeds that of that of MAX-R/T_V. This is schematized in (62).

[2]It should, of course, be noted that while the **graphemes** *kl* in the Cantonese and English sequences do coincide, their actual pronunciations do not. In Cantonese, this sequence will always be pronounced as [k̚l], while in English it is pronounced [kʰl]. This discrepancy will be discussed further in Chapter 5.

(62) **Constraints at play when determining which consonant to delete**

/ˈmʌjkɹəˌfəʊn/	*Complex	*Coda	Max-T/V_R	Max-R/T_V
a. mɐj klow foŋ	*!	*		
☞ b. mɐj kow foŋ		*		*
c. mɐj low foŋ		*	*!	

Now assume that a completely naïve learner of Cantonese is presented with the data outlined in (61). The learner – who must learn that Cantonese allows codas, albeit codas of a particular type – may produce an error whereby the coda [k] is deleted.[3] This will incur a violation of Max-T/V_R, which the learner can then use to **increase** the weight of the Max-T constraint and **decrease** the weight of any constraint which bans codas altogether.

(63) **Error produced over the course of learning Cantonese**

/t͡sùk lɛ̀j/	*Complex	*Coda	Max-T/V_R	Max-R/T_V
a. t͡sùk lɛ̀j ∼ t͡sù lɛ̀j		↓	↑	

Crucially, these kinds of errors will **not** cause the learner to promote the Max-R constraint, as the sonorant is kept in both instances. Furthermore, the learner will presumably have no reason to produce an additional error whereby the sonorant is deleted, as it appears in onset position and is a licit onset in Cantonese. As such, this constraint will not be promoted at all, and the result is that the grammar will have a weighting consistent with the ranking schema in (64). In other words, the grammar will contain an asymmetrical weighting of faithfulness constraints.

(64) **Result of learning from errors like those in (63)**

/t͡sùk làj/	*Complex	Max-T/V_R	*Coda	Max-R/T_V
☞ a. t͡sùk lɛ̀j			**	
b. t͡sù lɛ̀j		*!	*	
c. t͡sùk ɛ̀j			**	*!

This **same weighting** of constraints can also be used to derive the correct result for adapting the English loanword [ˈmʌjkɹəˌfəʊn] ‘microphone’, which the reader can check by comparing the crucial violations assessed by the constraints in (62) to those in (64). The only difference between the native Cantonese example and the English loan is that there is a highly-weighted markedness constraint which forces the underlying asymmetrical weighting of faithfulness constraints to have a visible effect in the English loanword.

Of course, the situation is substantially more complex than this brief example would lead the reader to believe. A learner of Cantonese is not limited to positing only deletion errors, and actual

[3] The reasoning behind this example comes from the literature on error-driven learning (Tesar and Smolensky, 2000; **?**; Hayes, 2004; Prince and Tesar, 2004; Magri, 2012). This is not the method by which the learner receives evidence for the MaxEnt models constructed throughout this dissertation, but it is an intuitive way to illustrate the point.

speakers of Cantonese do not always choose to repair consonant-consonant clusters with deletion, as displayed in example (5) of Chapter 1. After all, vowel epenthesis is a viable solution in these scenarios as well, and a learner will also have to entertain epenthesis repairs alongside the deletion repairs. Nevertheless, this example shows that it is **possible** to arrive at an asymmetrical weighting of constraints that is consistent with the PMap without needing to build the PMap in to the learning algorithm.

The purpose of this experiment is to test whether this is a viable way of deriving the PMap by training a learner on a **representative sample** of the Cantonese data outlined in section 4.2.1. A representative sample is used in order to expedite the learning process. The full corpus of words from CC-Canto is around 22,000 entries, and if the entire corpus is used when generating candidates via the method outlined in (57), this will result in a set of nearly one million candidates from which to learn. As the `solve.R` program learns by examining the entire candidate set on each pass through the algorithm, this will greatly slow down the learning process. As such, selecting a sample of the corpus that preserves as much of the distribution of segments and their contexts as possible is necessary in order to run multiple computational experiments. This experiment will be used to test this methodology, and to help establish the ideal **size of the representative sample** of the Cantonese lexicon to be used in the experiments to come.

4.3.1 Data

This experiment and all following experiments discussed in this chapter will consist of a **training phase**, where the MaxEnt algorithm is tasked with acquiring a grammar that is consistent with the native Cantonese vocabulary; and a **test phase**, where the resulting grammar is applied to a set of English loanwords and similar native Cantonese words. The training phase is meant to simulate the learning trajectory of a **monolingual** Cantonese speaker[4], from the point at which they begin speaking into adulthood. The test phase is meant to test whether the grammar acquired by the learner is 1) adequate for native Cantonese words, and 2) sufficient to explain the patterns of loanword adaptation outlined in (49). In general, the training and test data sets will remain **constant** throughout all experiments – the only differences will be between what priors, if any, are used.

[4]It should be noted that this is not the case for most speakers of Cantonese, who may also be fluent in Mandarin and/or English. The consequences of this discrepancy will be discussed further in Chapter 5.

4.3.1.1 Training Data

All of the training data were drawn exclusively from the CC-Canto corpus discussed in section 4.2. As the entire corpus would be too large of a training set to efficiently learn from, it was decided to instead use a **representative sample** taken from this corpus as the training set. A total of **ten** representative samples were used for this experiment. Five of the samples had a limit of 1,000 lexical items (around 5% of the total corpus), and five had a limit of 2,000 lexical items (around 10% of the total corpus). Each sample was assembled according to the method outlined in (65).

(65) **Sampling procedure for CC-Canto**

- Find the **total count** of each unique syllable within the entire 22,000 word corpus.
- **Divide** this value by the total number of entries in the entire corpus to estimate how often that syllable occurs.
- **Multiply** each of these values by the desired sample size to determine roughly how many of each kind of syllable to include in the representative sample. This value is then **rounded down** to arrive at a whole number.
- **Randomly order** the corpus.
- **Check** whether all syllables of the first entry can be included in the representative sample by ensuring that the value for each in the desired total is non-zero.
  - If all syllables have a non-zero value for the desired number of syllables, add the word to the representative sample and subtract each count from the desired total.
  - If not, skip to the next word and repeat until the sample size is met.

No loanwords were selected for inclusion in any of the training data sets used in this experiment, as it was intended to model how a **monolingual** speaker of Cantonese might behave when presented with a series of English loanwords for the first time.

Each set of words selected to be part of a training data set was then passed through the automatic tableau generation procedure outlined in (57) in section 4.2.2. Recall that the candidates generated by this procedure included all candidates which instantiated a **single change** from the Jyutping transcription provided by CC-Canto, and no others. This was done in order to provide the learner with a data set that was not only maximally informative, but as small as possible. This ensured that the learner would infer as much as could be inferred about the insertion and deletion of segments without having to reason over too large a candidate set. The resulting file was then used as the training data for an iteration of the `solve.R` MaxEnt learner. The consequences of using this generation procedure for the test set are discussed in detail in Chapter 5.

4.3.1.2 Test Data

The data used to construct the test data set were selected both from the set of loanwords listed in Appendix A and from CC-Canto. Each loanword was classified according to whether it had any sequences of interest, such as onset or coda clusters, or sibilants in coda. Any loanwords which had **more than one** phonotactically-illicit sequence were excluded from the analysis. This was done in order to more easily assess model performance. The remaining loanwords were categorized according to which illicit sequence they contained, a full list of which is provided in (66). Ten words of each category were selected for inclusion in the training set, where possible. In addition, if there were multiple attested repairs to a particular sequence, the ten exemplars were split between repair types, according to how often each repair was attested.

(66) **Illicit Cantonese Structures to be Tested**

Structure	*Example*	*Repair*	*Gloss*	*N*	*Note*
CCV	ˈ**pʰɹ**ɒm → ˈ**kʰɹ**iːm →	**pʰ**ếŋ **kʰèj l**íːm	‘prom’ ‘cream’	5 5	Split to reflect distribution
VCC	ˈpʰaʊ**nd** →	pʰɔ́ː**ŋ**	‘pound’	10	
VS	ˈt͡ʃiː**z** →	t͡síː **sɨ́**	‘cheese’	10	
SCV	ˈ**sk**ɔː →	**sì k**ɔ́ː	‘score’	10	
VCS	ˈɒɹɪ**nd͡ʒ** → ˈd͡ʒiː**nz** →	ɔ́ː lø̄**n t͡sɨ́ː** t͡síː**n**	‘orange’ ‘jeans’	8 2	Split to reflect distribution
VSC	ˈtʰəʊ**st** →	tʰɔ́ː **sɨ́ː**	‘toast’	5	Not enough data for ten
(C)V(C)	ˈkʰɪŋ →	kʰéŋ	‘king’	40	
[Native]	–	t͡sʰé lø̂n t͡sīn	‘a wheel war’	90	

In addition to each of the phonotactically-illicit structures, 40 additional phonotactically-licit loanwords were included into the standard set. This was done to test whether the model was treating loanwords differently than the native Cantonese vocabulary. The number of additional loanwords selected was meant to mirror the distribution of unrepaired vs. repaired loanwords in the entirety of the corpus in Appendix A. Once this selection process was completed, the training set contained a total of 95 loanwords.

Once all of the loanwords had been selected for inclusion in the test set, an additional 95 entries from CC-Canto were selected to **balance** the set of loanwords. These words were included in order to test whether the learner had succeeded in acquiring the phonotactics of Cantonese. Each word selected was chosen to be as similar as possible to the surface form of the loanwords that were included in the test set. For example, the the Cantonese word [t͡sʰé lø̂n t͡sīn], ‘a wheel war’, was included since it was similar in form to the surface form of the loanword [ɔ́ː lø̄n t͡sɨ́ː], ‘orange’. Such forms were found by searching within CC-Canto for entries which matched in all segments, excluding tone, or which differed only in a few segments. This increased the number of words in the test set from 95 to 190.

This data set was then passed through the automatic tableau generation procedure outlined in (59) in section 4.2.2. This procedure was identical to the one used for the training data sets, with the exception that candidates which showed both one segment deletion and one vowel epenthesis were included. This deviation was necessary in order to capture the behaviour of the loanwords in the VSC category, which are repaired by deleting the final consonant and epenthesizing a vowel after the sibilant. The resulting tab-delimited text file was then passed to the `solve.R` script for purposes of evaluation, which will be discussed in more detail in section 4.3.3, below.

4.3.2 Model parameters

Once the training and test data sets were generated, the data were then imported into the `solve.R` script outlined in Chapter 2. The `solve.R` script contains, among other things, a MaxEnt grammar learner. This learner is, in essence, a well-defined **objective function** that can then be fed to R's native `optim()` function (R Core Team, 2017). Readers who would like to know details of how the objective function was defined should consult Chapter 2, in particular, examples (25) and (26). The `optim()` function can make use of a variety of different methods for iterating over the custom-defined objective function, but `solve.R` requires that the `L-BFGS-B` method (Byrd et al., 1995) is used for this process. The `optim()` function also imposes an iteration limit, in order to ensure that computations have a designated stopping point. By default, this limit is set to 100, but it was increased to $1,000$ for the following experiments, so that the MaxEnt learners could accommodate the large amount of training data.

As is standard, a **Gaussian prior** was added to the MaxEnt objective function. This is done in order to ensure that the grammar the learner acquires is able to **generalize** to novel data – a requirement when the test data includes loanwords which contain sequences not seen in the training data. If there were no prior incorporated into the objective function at all, it would introduce the risk of **overfitting** to the training data, and arriving at a grammar that would only be able to produce the words present in the training set. The Gaussian prior used for testing the null hypothesis is given in (67).

(67) **Regularization parameters for the Null Hypothesis Model**

Constraint Type	μ	σ^2
Markedness	10.0	100
Faithfulness	0.0	100

The value for σ^2 was set to be quite high, so as to allow all constraints to be highly plastic and more capable of being re-weighted. A Markedness $\gg$ Faithfulness bias was encoded by setting the target weight for all Markedness constraints at 10.0 and the target weight for all Faithfulness constraints at 0.0. This was done in order to make sure the learned grammars were sufficiently **restrictive** when presented with novel data (Hayes, 2004; Prince and Tesar, 2004, *a.o.*). If such a bias is not incorporated into a phonotactic learner, it runs the risk of arriving at grammars which erroneously predict that novel forms are never repaired.

There are a variety of ways that the Markedness $\gg$ Faithfulness bias could have been instantiated within the `solve.R` script. For example, Pater and Staubs (2013) make use of an addition to the objective function that aims to maximize the difference between the weights of both kinds of constraints. For the sake of simplicity, I will be making use of the same methods as Wilson (2006) and White (2013) in incorporating this bias as a bias on the target weights of constraints. However, readers who would like more details about this bias should consult Chapter 5.

It is crucial to note that making use of a Gaussian prior is **not** the same as making use of a PMap-like prior. In order for a prior to be consistent with the PMap, it is necessary for the prior to treat the Faithfulness constraints **asymmetrically**, either by varying their plasticities (Wilson, 2006) or their target weights (White, 2013). The prior used for this experiment treats all Faithfulness constrains **identically** in this respect, so it cannot mimic the behaviour of the PMap.

In addition to the training data and optimization function, the `solve.R` script requires a vector of **constraint weights**, which is used to predict which candidates will be selected as output forms. The constraint weight vector was defined in the `solve.R` script itself, and was initialized with all weights being equal to 1.0.

Once the full objective function with Gaussian prior and starting weights were set up, the `solve.R` script was run a total of 10 times, once for each training set outlined in in section 4.3.1.1. This created a total of 10 possible native Cantonese grammars, five of which were trained on a vocabulary of $1,000$ lexical entries, and five of which were trained on a vocabulary of $2,000$ lexical entries.

Each grammar was then applied to the **test set**, to see which candidate was selected as optimal for each of the 190 entries. Crucially, all of the empirical information about which candidate was preferred by human speakers of Cantonese was **withheld** from the `solve.R` script at this stage. In this way, each trained grammar was forced to assign probabilities to candidates using the **trained weights alone**. The predictions of the model were then **categorized** according to which Repair Type had been chosen, and the probabilities of each type of repair were then **aggregated** according to the type of Input Structure the word belonged to.

(68) **List of Repair Types**

None	R-Epenthesis
Deletion	V-Epenthesis
T-Epenthesis	Deletion + V-Epenthesis
S-Epenthesis	

(69) **List of Input Structures**

Native	VS
(C)V(C)	SCV
CCV	VCS
VCC	VSC

Each probability was divided by the total number of input forms present in each input structure category, so as to ensure that all probabilities for each input structure category summed to 1.

Once predictions were obtained for each of the ten grammars, the results were **compared to one another** by calculating each pairwise KL-divergence between training sets of the same size. Recall from Chapter 2 that KL-divergence, or D_{KL}, is a measure of the similarity of two probability distributions. The smaller the value of D_{KL}, the more similar the distributions are to one another. The training sets associated with the sample size which resulted in the **least amount** of divergence were then used as the training sets for the experiments to follow.

Of course, it is not sufficient to determine which training sets were most **internally consistent** – after all, a set of grammars may be highly consistent with one another, but nevertheless extremely far off from the target grammar. While the adequacy of the grammars may be assessed simply by examining how often a grammar **correctly** predicted the attested output form for each of the items in the test set, this was deemed to be too stringent a test for the purposes of this thesis. The grammars examined were trained on data which had very **sparse** probability distributions over candidates, wherein only one candidate received 100% of the probability of being chosen as the output for a particular input. However, the MaxEnt learner used to derive these grammars does not replicate this pattern, since it must assign non-zero probability to all candidates by definition. This means that optimal candidates will receive less than 100% probability, even if they are otherwise perfect examples of Cantonese words. It can therefore not be expected that the probabilities assigned by a MaxEnt grammar will match empirical probabilities.

There are two ways of arriving at a more apt comparison. One way is simply to find the candidate or candidates which have the **highest probability** of being selected as the output form for a particular input form, and to count them as the "optimal" form. These optimal forms can be recorded and then compared to the empirical data to see if the same form was selected in both cases. In this way, we can arrive at a rough estimate of **accuracy**.

In addition to this method, there is a more fine-grained way of determining how well a grammar has performed. This involves assessing how well the grammar's predictions line up with the **best possible outcome** of learning, given the test set. This best possible outcome is derived by allowing the MaxEnt learner to arrive at a grammar by **training it** on the **test set alone**. This will generate a set of weights which will not only be able to account for the Cantonese data present in the test set, but also the English loanwords and their attested repairs. While the best-case grammar cannot arrive at a perfect replica of the empirical data, as discussed above, it will arrive at one which is guaranteed to be able to predict loanword behaviour. This grammar can then be **re-applied** to the test set, and its predictions can be used to assess how well the grammars trained exclusively on the native Cantonese vocabulary have fared. This is done by calculating the KL-divergence between each grammar trained only on native Cantonese lexical items and the best-case loanword grammar. Both methods of assessing grammar performance – accuracy estimates and comparison to a best-case grammar – will be employed in the following sections to determine how successful a particular grammar is at correctly predicting loanword repairs.

To summarize, a detailed sketch of the methodology used for this experiment and future experiments is given in (70).

(70) **Methodology for the Null Hypothesis Experiment**

1. **Sample** from the full list of Cantonese lexical items in CC-Canto.
2. **Assemble** the data to be used in the **test set**.
3. **Generate** tableau files for each sample and the test set.
4. **Define** the objective function for MaxEnt Learning and the Gaussian prior.
5. **Set** all initial constraint weights to be 1.
6. **Run** the `solve.R` script with the objective function, initial constraint weights, and the **training sets** derived from the samples to arrive at a **grammar** for each.
7. **Run** the `solve.R` script with the same objective function, initial constraint weights, and the **test set** to arrive at a **best-case grammar**.
8. **Apply** each grammar to the test set to generate a set of predictions for each.
9. **Compare** the predictions of the sampled grammars to the predictions of the best-case grammar.

The results of this process are discussed in more detail in the next section.

4.3.3 Results

4.3.3.1 1000-item samples

First, each of the grammars derived from the 1000-item training sets will be examined and compared to **one another**. Examining the predictions of these grammars will give a general idea of how a grammar with a simple Markedness ≫ Faithfulness bias trained only on native Cantonese words will behave. These results will be compared in more detail with the predictions of the **best-case grammar** later, in section 4.3.3.3.

The **accuracy** for one of the 1000-item grammars is provided in Figure 4-1. As the results from each of the 1000-item training sets are qualitatively similar, only one graph is presented here. This will also be the case for subsequent experiments.

Figure 4-1 is limited to only those candidates of the test set which are assigned the **highest probability** by the grammar in question. Each bar on the graph displays the **proportion** of repair

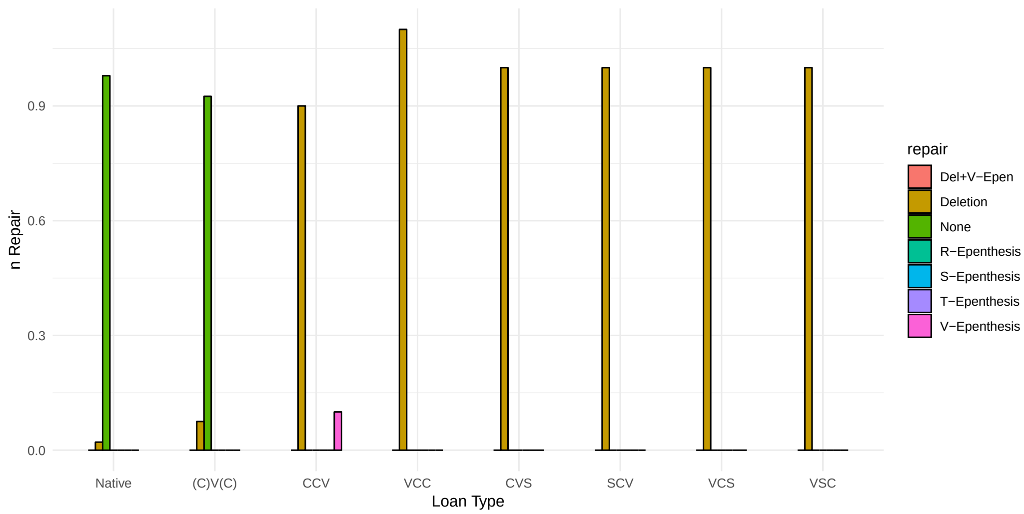


Figure 4-1: Proportion of winning candidates divided by repair and input structure category, representative sample of 1000 Cantonese lexical entries with Markedness ≫ Faithfulness bias

types per word category. Thus, it can be shown that for the ten CCV loanwords present in the test set, 90% of them selected a deletion candidate as the most likely repair, and 10% selected a candidate with vowel epenthesis as the most likely repair. It should also be noted that some proportions are **above** 100%, such as in the case of the VCC-category loanwords. This is due to the fact that there was occasionally **uncertainty** about which member of a particular cluster to delete, so both viable deletion candidates were assigned an **identical** probability. A sample of repairs for each word category are given in (71). The reader should keep in mind that the probability p values displayed are those assigned to the **winning candidate**, not the probability of seeing the repair type overall.

(71) **Sample repairs for a grammar trained on 1000 Cantonese lexical entries with Markedness ≫ Faithfulness bias**

Category	*Repair Type*	*Input*	*Gloss*	*Output*	p
Native	None Deletion	`naap seoj` `o pej`	'tax' 'to fart'	`naap seoj` `pej`	0.99782 0.99624
(C)V(C)	None Deletion	`kaa low lej` `e sej`	'calorie' 'essay'	`kaa low lej` `sej`	0.99879 0.99624
CCV	Deletion V-Epenthesis	`klip` `wo ljam`	'clip' 'volume'	`lip` `wo lV jam`	0.84961 0.78763
VCC	Deletion	`sapt`	'shaft'	`sap` `sat`	0.38123 0.38123
CVS	Deletion	`paas pot`	'passport'	`paa pot`	0.99848
SCV	Deletion	`spaak`	'spark'	`paak`	0.88259
VCS	Deletion	`inc`	'inch'	`in`	0.98150
VSC	Deletion	`tost`	'toast'	`tot`	0.92087

A few comments should be made before moving on to examining the full range of predicted repairs for the current batch of grammars. First, there is a tendency for the grammar to **eliminate** syllables which consist of a single vowel, both in native Cantonese words and phontactically-licit loanwords. This may be due to an unnaturally high weighting of the constraint enforcing syllable bimoracity, BIMORAIC. This may be indicative of having too few examples of this kind of vowel in the training set, and may be expected to not be an issue for grammars trained on data sets of a larger size.

Second, as discussed previously, the model is sometimes unable to select a **single** winning candidate, and instead assigns an identical probability to two competing candidates. This is the case for the loanword [sɐ́p] 'shaft', where the grammar is incapable of predicting which of the final obstruents to preserve.

Third, there is an overall tendency towards segment **deletion** in those loanwords which do not conform to the phonotactics of Cantonese. Furthermore, the segment which is deleted is overwhelmingly the **sibilant**, or the segment which is **furthest from the vowel**. This is contrary to most of the repair strategies used by actual speakers of Cantonese. Empirically, these speakers tend to **retain** all sibilants whenever possible, and they tend to prefer retaining either **both** members of an onset cluster or the **first** member of the cluster. This is an unexpected result, since there should be **no preference** for deletion over vowel epenthesis for these grammars.[5] This is due to the fact that, in the absence of evidence, the constraints governing both deletion and vowel epenthesis might be expected to maintain a **low weight** throughout learning.

It is tempting to speculate that this could be due to the way in which the data in Figure 4-1 are being presented. After all, this figure is limited to only showing the **most probable** candidates for any input, as opposed to the full range of candidates and the probabilities assigned to them. If the probability assigned to all candidates was examined instead, it may be the case that there is more probability assigned to vowel epenthesis candidates than would be expected given the figure. However, the high degree of certainty that the model has in selecting its winning candidates does not make this a viable explanation. As shown in (71), most winning candidates receive 75% or more of the probability for any given input, leaving very little probability to be spread out among all other candidates for that input. This is borne out when the probability assigned to all candidates is examined, as is done in Figure 4-2.

In Figure 4-2, each bar represents the **totality** of the probability assigned to lexical items of that particular category, standardly assumed to be 1.0. Each bar is subdivided into blocks of different sizes, where the size corresponds to the amount of total probability assigned to candidates

[5]Consonant epenthesis candidates are largely ruled out by their violations of highly-weighted *COMPLEX and CODACOND, and are not expected to be considered as viable repairs to loanword inputs.

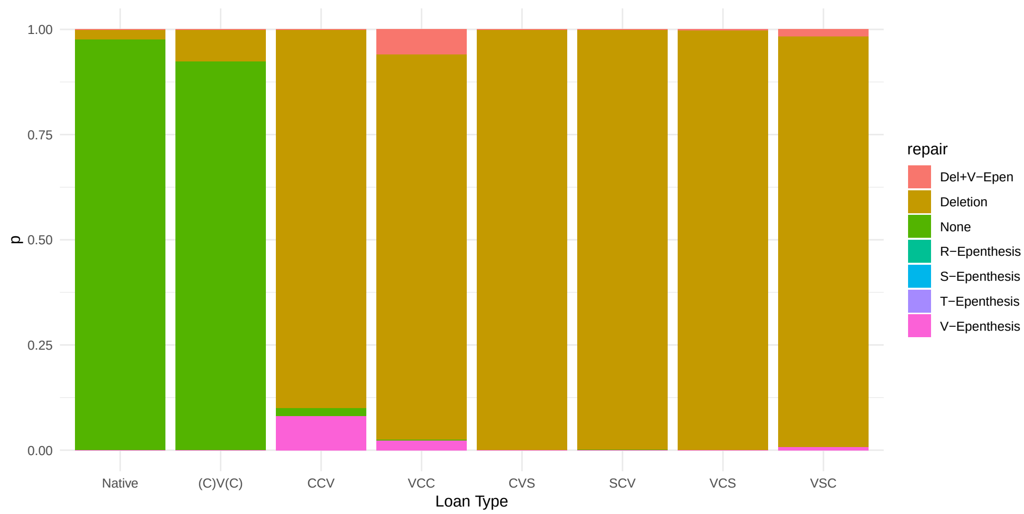


Figure 4-2: Probability of repair by input structure category, representative sample of 1000 Cantonese lexical entries with Markedness $\gg$ Faithfulness bias

of a particular Repair Type. Every candidate present in the test set was included in this figure, regardless of whether that candidate received the majority of the probability for a particular lexical item or not. Probabilities assigned to individual candidates were **normalized** according to how many lexical items were present in their category. This was done to make the graph easier to parse.

While it is the case that there is more overall probability assigned to candidates which show vowel epenthesis – as evidenced by the larger pink and red portions of the graph – the majority of the probability is assigned exclusively to the **deletion** candidates across the board for all phonotactically-illicit loanwords. This is ultimately a consequence of the generation procedure used for the experiments throughout the dissertation, and will be discussed in detail in Chapter 5.

The data presented in Figures 4-1 & 4-2 and example (71) are from a single grammar, trained on only **one** of the five 1000-item training sets used for this experiment. The grammars trained on the remaining four 1000-item training sets were also applied to the test set to predict which of the candidates would be selected as most probable. The results for each grammar were compared by computing the KL-divergence between the predictions made by each possible pair of grammars. The resulting KL-divergence values are presented in (72). The grammar whose predictions are shown in Figures 4-2 & 4-1 corresponds to sample A in the table below, while all others correspond to samples B–E.

(72) **KL-Divergence values for Predictions of 1000-item Models**

Comparison		D_{KL}
Grammar A	vs. Grammar B	0.00407850
	vs. Grammar C	0.00884646
	vs. Grammar D	0.00372557
	vs. Grammar E	0.01696343
Grammar B	vs. Grammar C	0.00521803
	vs. Grammar D	0.00048273
	vs. Grammar E	0.01766905
Grammar C	vs. Grammar D	0.00523841
	vs. Grammar E	0.01211318
Grammar D	vs. Grammar E	0.01529500

While there is not a standard for determining if a particular KL-divergence value should count as "significantly similar" or "significantly different", as there are for p and t values, the fact that all of the values in (72) fall below 0.02 is promising, and indicates that the results of all five models are relatively similar to one another. We can thus say that the results of the grammars trained on sets of 1000 native Cantonese words alone are quite **consistent**, even if the actual predictions made by these grammars are far from accurate.

Before making direct comparisons to the predictions made by the best-case grammar, the results from the grammars trained on the larger five training sets will be examined. It may be the case that the sample size used to train the above grammars was too small, and some crucial – if rare – lexical items may have been left out. If so, then it may be expected that the performance of these grammars should improve.

4.3.3.2 2000-item samples

As was done in the previous section, the **accuracy** for one of the 2000-item grammars is presented in Figure 4-3. As was the case for the 1000-item grammars above, the results ended up being sufficiently similar, so their results will not be discussed directly here.

As was the case for the set of 1000-item grammars, the results for the 2000-item grammars have not improved model performance, as can be shown by examining Figure 4-3. The overwhelming preference of the model is to repair all phonotactially-illicit loanwords by **deletion** of some portion of the problematic structure. A set of representative repairs is presented in (73).

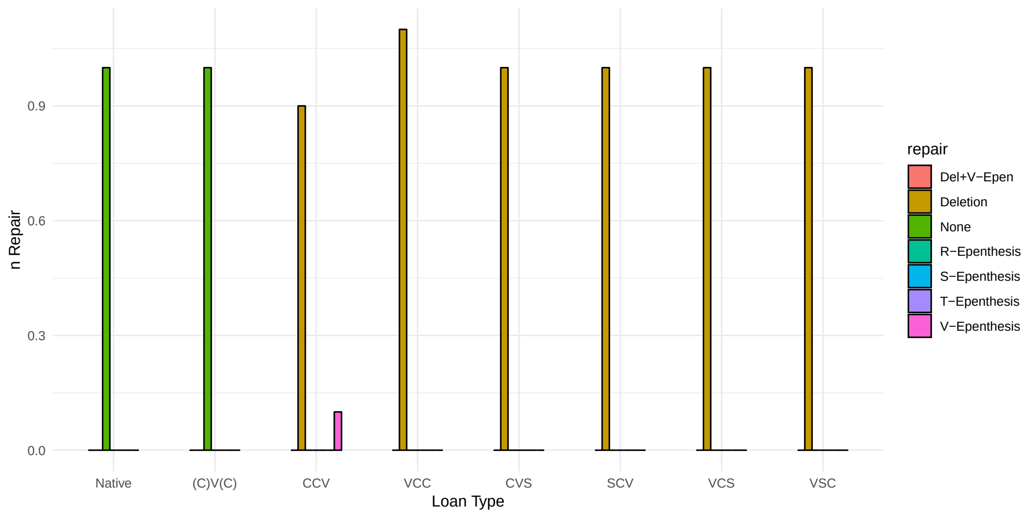


Figure 4-3: Proportion of winning candidates divided by repair and input structure category, representative sample of 2000 Cantonese lexical entries with Markedness ≫ Faithfulness bias

(73) **Sample repairs for a grammar trained on 2000 Cantonese lexical entries with Markedness ≫ Faithfulness bias**

Category	*Repair Type*	*Input*	*Gloss*	*Output*	*p*
Native	None	`o pej`	'to fart'	`o pej`	0.95814
(C)V(C)	None	`e sej`	'essay'	`e sej`	0.95798
CCV	Deletion	`klip`	'clip'	`lip`	0.94679
	V-Epenthesis	`wo ljam`	'volume'	`wo lV jam`	0.62143
VCC	Deletion	`sapt`	'shaft'	`sat`	0.37287
				`sap`	0.37287
CVS	Deletion	`paas pot`	'passport'	`paa pot`	0.99918
SCV	Deletion	`spaak`	'spark'	`paak`	0.91921
VCS	Deletion	`inc`	'inch'	`in`	0.98887
VSC	Deletion	`tost`	'toast'	`tot`	0.96407

The only changes from the results of the 1000-item grammars outlined in section 4.3.3.1 are that words which contain a syllable consisting of a single vowel no longer exhibit deletion of that vowel. Words like the native Cantonese [ó pě̋j] 'to fart' and the English loanword [é sě̋j] 'essay' retain their initial vowel. However, the grammar is still indecisive about which segment to delete if the segments are identical in sonority, as is the case for the loanword [sép̋] 'shaft'. Futhermore, the remaining phonotactically-illicit loanwords still overwhelmingly exhibit **deletion** of a segment, regardless of whether this is the attested repair by native speakers of Cantonese. The deleted segment is still overwhelmingly the **sibilant** or the **furthest** one from the vowel – segments which are often chosen for **retention** by native speakers. These results remain consistent when all candidates in the test set are examined, as shown in Figure 4-4. In fact, this figure shows that the preference for selecting a deletion candidate is **stronger** for the 2000-item grammars than it is for the 1000-item grammars.

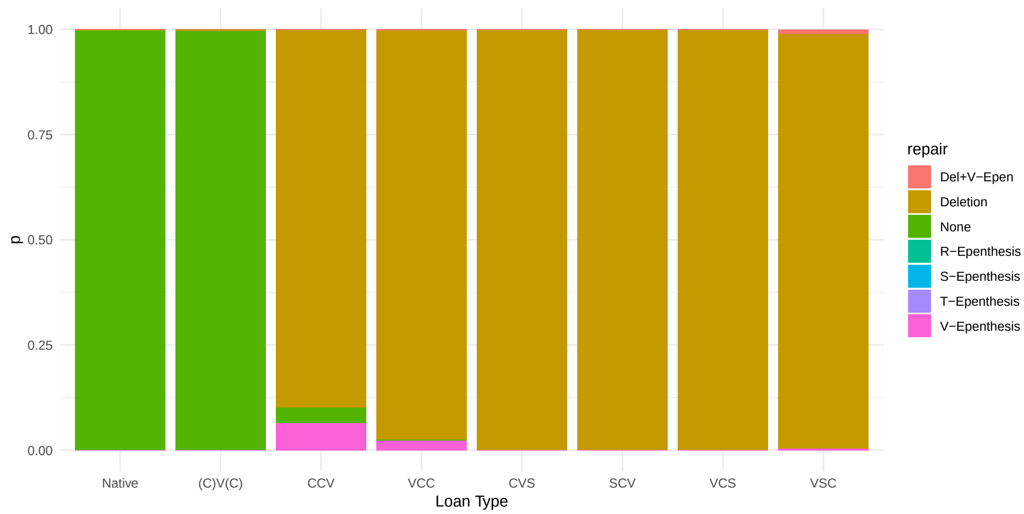


Figure 4-4: Probability of repair by input structure category, representative sample of 2000 Cantonese lexical entries with Markedness $\gg$ Faithfulness bias

As was done for the 1000-item grammars, the predictions made by all five of the 2000-item grammars were compared by calculating their KL-divergence values. The results of each comparison are provided in (74). As was the case for the 1000-item grammars, the 2000-item grammar that was used to generate Figures 4-3 & 4-4 is labeled grammar A, and the remaining 2000-item grammars receive labels B–E.

(74) **KL-Divergence values for predictions of 2000-item models**

Comparison		D_{KL}
Grammar A	vs. Grammar B	0.00307367
	vs. Grammar C	0.09427366
	vs. Grammar D	0.13458792
	vs. Grammar E	0.06835165
Grammar B	vs. Grammar C	0.08770468
	vs. Grammar D	0.13118515
	vs. Grammar E	0.07081258
Grammar C	vs. Grammar D	0.08450342
	vs. Grammar E	0.00948173
Grammar D	vs. Grammar E	0.05039571

While the KL-Divergence values for the models of the 2000-item grammars are still relatively small, they are on average **larger** than the values for the 1000-item grammars. For instance, the smallest KL-Divergence value for the 1000-item grammars is around 0.0004, while the smallest value for the 2000-item grammars is around 0.003, about an order of magnitude more. Similarly, the largest KL-Divergence value for the 1000-item grammars is around 0.017, while for the 2000-item grammars it is around 0.13 – again, about an order of magnitude more. This indicates that the models for the

2000-item grammars are overall **less similar** to one another than those of the 1000-item grammars.

Why this should be so may be due to the way sampling is done for each training set. When the desired number of syllable types are calculated for each size of training set, each value is **rounded down** in order to arrive at a whole number. This may mean that for smaller sample sizes, some rarely-attested syllables may receive a target value of 0, and be excluded from the analysis. This in turn may lead to a more uniform dataset that consists only of the most robustly attested syllable types of Cantonese, to the exclusion of marginal forms.

The fact that syllables which consist of a single vowel are deleted in the 1000-item grammars but retained in the 2000-item grammars bears this hypothesis out. Lexical items with single-vowel syllables must consist of less than 0.1% of the entries in CC-Canto, meaning that they will be excluded when the sample size is 1000 entries or less. When a MaxEnt learner is trained on such a set, it will not receive any evidence that single-vowel syllables are allowed in Cantonese, and will as such predict that these kinds of syllables must be repaired in some way. However, when the sample size is larger than 1000 entries, there will be at least one example of such a syllable that must be included in the training set in order for that set to be considered representative. Since a MaxEnt learner trained on this larger data set will have access to evidence that single-vowel syllables are allowed in Cantonese, it will arrive at a grammar that considers them licit structures. As such, they will not require repairs.

The above results demonstrate that there are two conflicting reasons to select a particular size of training set. One the one hand, a training set of 1000 lexical items is guaranteed to be more **consistent**. On the other hand, a training set of 2000 lexical items is guaranteed to be more **accurate** in its treatment of native Cantonese words. While it may be tempting to prioritize accuracy, this is not always what has been done when constructing MaxEnt learners. For example, Hayes and Wilson (2008) exclude rare but attested onset clusters, such as [sf] in [ˈsfiːɹ], 'sphere' from their English training set. They do this on the grounds that these syllables only occur in a handful of loanwords, and may not reflect the broader phonotactics of English. I have chosen to follow Hayes and Wilson (2008) in this regard, and to use training sets of 1000 items for the following experiments. While they may not be as representative of the native Cantonese vocabulary as desired, the number of errors is relatively small, and should have no bearing on the issue of consonant cluster resolution or sibilant retention. Furthermore, the results obtained from training sets of this size have been shown to be more consistent with one another, so choice of one 1000-item training set over another should not make as large a difference in results.

Regardless of sample size, it is apparent from this experiment that the distribution of segments within the native vocabulary of Cantonese does not automatically give rise to a PMap-compliant ranking of constraints. Rather, this model predicts that **deletion** is the preferred repair in all cases, contrary to what is observed empirically in the loanword corpus. It is thus clear that incorporating some sort of **substantive bias** is required in these situations, in order to arrive at a weighting that is compatible with the observed loanword data. However, in a MaxEnt model, this substantive bias must be encoded in some **quantitative** manner, and at this point, only a qualitative examination of the data has been undertaken. As such, a **best-case** grammar, derived from running the MaxEnt learner on the test set, will be used to provide such a quantitative measure of similarity.

4.3.3.3 Best-Case Grammar

Recall from section 4.3.2 that an eleventh grammar was generated as part of this experiment – the grammar that is trained on the **test set** which consists of both native Cantonese words and English loanwords. This was done not only to provide a **quantitative** measure of how well the other grammars performed, but also to ascertain the **best possible** behaviour of a MaxEnt learner given this test set. After all, MaxEnt learners by definition cannot arrive at categorical predictions. However, both the training and the test data were largely categorical in nature, meaning the best possible grammar will nevertheless be inaccurate with respect to the empirical data.

As was done for the grammars trained exclusively on native Cantonese words, the **accuracy** of the best-case grammar was examined by selecting the candidate assigned the **highest probability** for each input. The proportion of repair types was then calculated for each word category. The results are displayed in Figure 4-5.

As evident from Figure 4-5, the profile of attested repair types is much more in line with empirical patterns of loanword repair. Native Cantonese words and phonotactically-licit loanwords do not exhibit any repairs, indicating that the MaxEnt learner has successfully learned Cantonese syllable restrictions. Loanwords with onset clusters show a split between deletion of one member of a cluster and vowel epenthesis, while coda clusters are repaired exclusively by deleting one member of the cluster. Loanwords with coda sibilants are repaired across the board by vowel epenthesis, as are words with onset SC clusters and coda CS clusters. Crucially, words with coda SC clusters are repaired by deletion and vowel epenthesis. While these results look promising, they do not guarantee that the repair chosen is the same one chosen by native speakers of Cantonese. After all, the test set contains multiple candidates which exhibit deletion, as well as multiple candidates which exhibit vowel epenthesis. As such, a sample of the winning candidates chosen by the best-case grammar is displayed below.

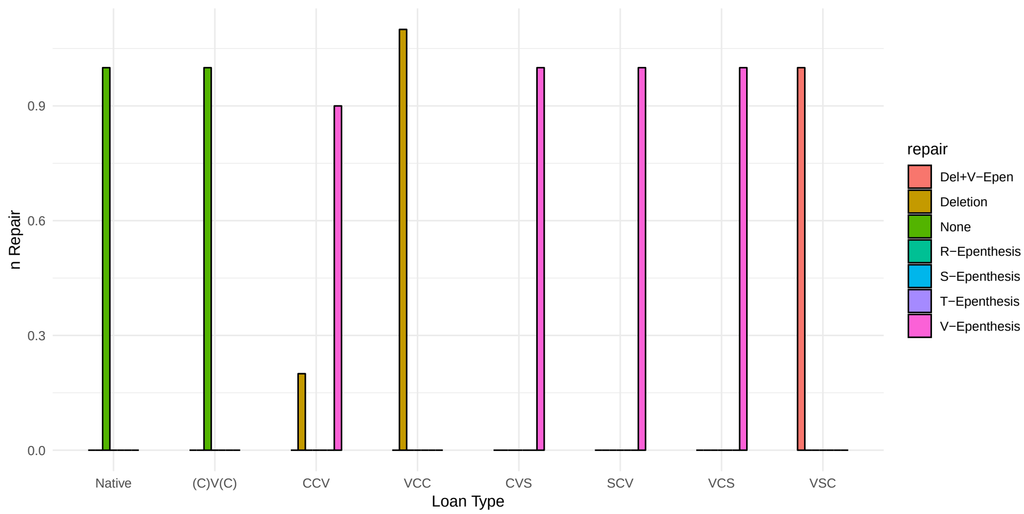


Figure 4-5: Proportion of winning candidates divided by repair and input structure category, test data set with Markedness ≫ Faithfulness bias

(75) **Sample repairs for a grammar trained on the test data set with Markedness ≫ Faithfulness bias**

Category	*Repair Type*	*Input*	*Gloss*	*Output*	*p*
Native	None	`o pej`	'to fart'	`o pej`	0.97426
(C)V(C)	None	`e sej`	'essay'	`e sej`	0.97416
CCV	Deletion	`wo ljam`	'volume'	`wo jam` `wo lam`	0.49263 0.49263
	V-Epenthesis	`klip`	'clip'	`kV lip`	0.65035
VCC	Deletion	`sapt`	'shaft'	`sat` `sap`	0.40399 0.40399
CVS	V-Epenthesis	`paas pot`	'passport'	`paa sV pot`	0.94766
SCV	V-Epenthesis	`spaak`	'spark'	`sV paak`	0.95769
VCS	V-Epenthesis	`inc`	'inch'	`in cV`	0.65098
VSC	Deletion + V-Epenthesis	`tost`	'toast'	`to sV`	0.92236

Overall, by examining the repairs in (75), it seems like the appropriate repair is being selected in most cases, especially in the cases of vowel epenthesis. In all cases, the vowel is inserted at the predicted place – either between members of a consonant cluster, as in *kV lip* 'clip', or after the sibilant, as in *paa sV pot* 'passport', *sV paak* 'spark', and *in cV* 'inch. These correspond as much as possible with the attested English loanwords [kʰì líp] 'clip', [pʰáː sìː pʰɔ̀ːt] 'passport', [sìː pʰáːk] 'spark', and [íːn t͡síː] 'inch', respectively. Crucially, the both the segment deleted and the location of the vowel epenthesis for the VSC words are as expected given the empirical data – the best-case grammar selects forms such as *to sV* for input 'toast', which is analogous to the attested loanword [tʰɔ́ː síː] 'toast'.

However, in the case of other consonant clusters, the best-case grammar has some difficulties. As was the case in the grammars trained only on native Cantonese words above, the model cannot decide which segment to delete if the segments are of identical sonority. This is, as was discussed earlier, likely an artefact of the constraint set used, which does not make distinctions between segments of identical sonority. Figure 4-5 also shows that the grammar appears to prefer **epenthesis** candidates in onset clusters, as opposed to a more even split between epenthesis and deletion candidates. However, a closer examination of the probabilities assigned to these candidates by the grammar makes it apparent that this split **is** occurring, albeit on an input-by-input basis. For instance, the probability assigned to the candidate *kV lip* for input 'clip' is around 0.65, indicating that other candidates may receive a larger chunk of the probability for that particular input. This is borne out by examining how probability is assigned to all candidates, as shown in Figure 4-6. The bar for CCV words is much more closer to evenly split between vowel epenthesis and deletion candidates, indicating that the best-case grammar is approximating the 50-50 split between repairs that can be seen in actual loanwords. The grammar's preference for vowel epenthesis in general will be discussed further in Chapter 5.

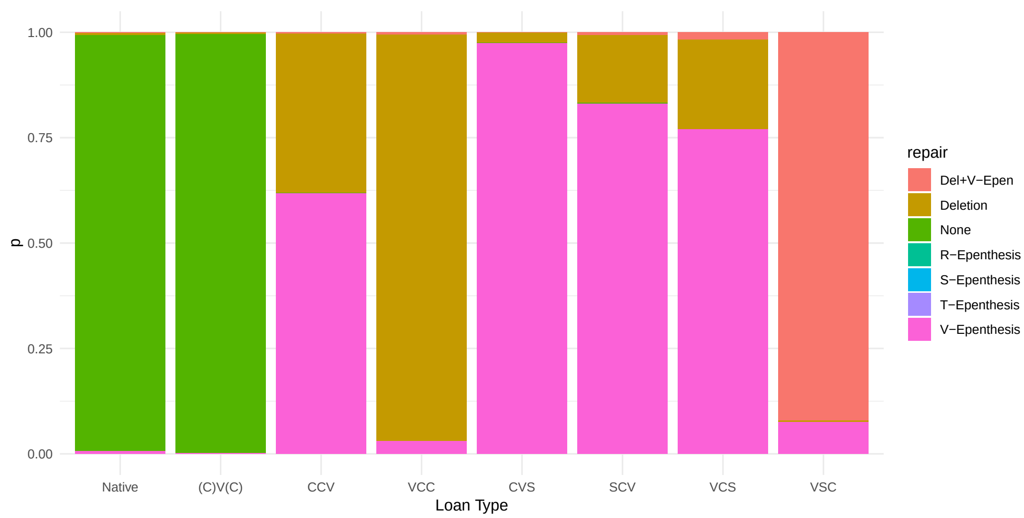


Figure 4-6: Probability of repair by input structure category, test data set with Markedness $\gg$ Faithfulness bias

Now that quantitative predictions have been generated for a grammar which can account for the behaviour of loanwords in Cantonese, it remains to be seen how well the predictions generated by the grammars trained exclusively on Cantonese compare. In each case, the predictions generated by a native Cantonese grammar were compared to the predictions generated by the best-case grammar by calculating the KL-Divergence between them. Since the results of the grammars with identically-sized training sets were extremely similar, only the results from the first grammar from each set are discussed here. The KL-Divergence between the first 1000-item grammar discussed in section 4.3.3.1 and the best-case grammar was $D_{KL} = 1.494433$. Similarly, the KL-Divergence between the first 2000-item grammar discussed in section 4.3.3.2 and the best-case grammar was $D_{KL} = 1.505711$. Both of these KL-Divergence values are quite high, indicating a high degree of dissimilarity between the predictions of the best-case grammar and those of the native Cantonese grammars.

4.3.4 Discussion

The results from running the MaxEnt learner with a simple Markedness $\gg$ Faithfulness bias do not bear out the null hypothesis. That is, the native Cantonese vocabulary does not have **sufficient information** to arrive at a PMap-like asymmetry in the weighting of Faithfulness constraints over the course of learning. Instead, all grammars that were trained exclusively on native Cantonese lexical items predicted that **deletion** should be the preferred repair in **all** phonotactically-illicit loanwords. Furthermore, even if deletion is an attested repair in the loanwords examined, the segment which is deleted by the grammar is **not the same** as the segment deleted by actual speakers of Cantonese. Thus, a substantive bias is deemed **necessary** in order to more accurately predict patterns of loanword adaptation in Cantonese. The first of these substantive bias models will be discussed in the following section.

This experiment has also raised the question of how **large** the training set for the following experiments should be. It was decided that while a training set of 2000 lexical items led to more accurate results with respect to the native Cantonese vocabulary, a training set of 1000 items was preferable, as it provided more consistent results, and the errors produced would likely have no significant effect on the phenomena being examined.

Finally, this experiment introduced the **best-case grammar**, one which was trained on the test set. This was used as a **baseline** by which to compare the grammars derived from native Cantonese words, so that both a qualitative and a quantitative notion of similarity could be reached. This best-case grammar will be explored further in the following experiment, as it contains crucial information regarding the asymmetries a PMap-like bias will require in order to successfully capture the behaviour of Cantonese.

4.4 Experiment 2: Plasticity prior

In the previous section, it was established that a bias consistent with the PMap could not be learned from the native Cantonese data alone, thus disproving the null hypothesis outlined in (50a) at the beginning of this chapter. The purpose of the following two experiments is to test which methods of incorporating a substantive bias will result in accurate predictions of loanword adaptation, also outlined in (50). For the current experiment, the performance of **plasticity prior** will be assessed. Recall from Chapter 3 that the plasticity prior is established by setting the **target weight** of all of the PMap-governed Faithfulness constraints at 0.0 and by **varying** the values of the **constraint plasticities** in a principled way, so that Faithfulness constraints which are predicted to have high weights within the PMap receive a **high plasticity**. This will allow them to more easily arrive at a weight above the target weight of 0.0 than constraints which have a lower plasticity. This will, however, only occur in the **presence of evidence** – if there is a true absence of evidence for how to rank particular Faithfulness constraints, their weight will **remain** at 0.0.

The results from the experiment outlined in 4.3 might, *prima facie*, provide evidence that the distributional facts of Cantonese will never result in a PMap-compliant weighting of constraints. However, recall that in this experiemnt, each constraint is given an **identical plasticity**. Once there is a differential assignment of plasticity values, the amount of data necessary to enforce a difference in weight will also vary across constraints. Drawing on examples (62–64), provided at the outset of section 4.3, say that the learner must arrive at a weighting such that the weight of Max-T/V_R should exceed the weight of Max-R/T_V. This must be the case for a grammar to arrive at [mɐj kow foŋ] for the English input [ˈmʌjkɹəˌfəʊn], 'microphone'. A tableau illustrating this requirement is reproduced in (76).

(76) **Constraints at play when determining which consonant to delete**

/ˈmʌjkɹəˌfəʊn/	*Complex	Max-T/V_R	Max-R/T_V
a. mɐj klow foŋ	*!		
☞ b. mɐj kow foŋ			*
c. mɐj low foŋ		*!	

Also recall from that earlier discussion that, while native Cantonese words will not have onset clusters such as [kl], they **do** have such clusters in the form of coda-onset sequences. A learner who is acquiring Cantonese will thus encounter such forms as [t͡sùː**k l**ɐ̀j] 'customary rule' and [sēj sēj lùː**k** **l**ùːk] 'appropriately or justly'. Earlier, it was shown that a learner attempting to acquire the coda [k] in [t͡sùk lɐ̀j] 'customary rule' could conceivably arrive at a differential weighting of Max-T/V_R and Max-R/T_V, such that Max-T/V_R receives a higher weight than Max-R/T_V. Here, a more nuanced example will be examined, in order to show how a plasticity bias can enforce a differential weighting throughout learning.

As discussed previously, a learner acquiring Cantonese must learn that certain codas are allowed. In other words, this learner must learn that the weight of the common constraint *Coda must be much **lower** than would be assumed under a Markedness ≫ Faithfulness bias. However, this is not the only markedness constraint that the learner will be evaluating throughout learning. As noted by Kenstowicz (2012), there are a number of phonotactic constraints which also drive trends in syllable acceptability in Cantonese. One of the constraints used throughout this dissertation to account for the restrictions on syllable types laid out in Bauer and Benedict (1997) bans sequences of **high vowels** followed by the vowel [u:], written *Tu. A full list of these constraints is provided in Appendix B. However, as belied by the vocabulary item [sēj sēj lùːk lùːk] 'appropriately or justly', this is only a **tendency** in the native Cantonese vocabulary. A learner must therefore also learn that the weight of this *Tu constraint must be reduced throughout learning.

A learner who is attempting to acquire the vocabulary item [sēj sēj lùːk lùːk] 'appropriately or justly' might therefore be tempted to produce errors where a coda [k] is deleted, in order to satisfy an initially highly-weighted *CODA. Likewise, this learner might **also** be tempted to produce errors where an onset [l] is deleted, in order to satisfy *Tu. The constraints which would prevent either of these segments from being deleted would be MAX-T/V_R and MAX-R/T_V, respectively. A learner would then be required to **reduce** the weight of *CODA and **increase** the weight of MAX-T/V_R, and to reduce the weight of *Tu and increase the weight of MAX-R/T_V. This is displayed schematically in the example below.

(77) **Errors produced over the course of learning Cantonese**

/sēj sēj lùːk lùːk/	*COMPLEX	*CODA	*Tu	MAX-T/V_R	MAX-R/T_V
a. ⋯ lùːk lùːk ∼ ⋯ lùː lùːk		↓		↑	
b. ⋯ lùːk lùːk ∼ ⋯ lùːk ùːk			↓		↑

If this learner were equipped with a simple Markedness ≫ Faithfulness bias, they would arrive at a final weighting where the MAX constraints would be roughly **equally weighted**. This is because the plasticity of the constraints under such a bias is **equal**, and the learner is being provided with an **equal** amount of evidence to promote both MAX constraints. However, if this learner were equipped with a PMap-compliant **plasticity** bias, an asymmetry would arise in such a situation.

Say that the learner is equipped with a plasticity bias such that the plasticity of MAX-T/V_R is **double** that of MAX-R/V_R. This means that MAX-T/V_R is able to move **twice as far** away from its target weight as MAX-R/T_V at any point during the learning process. Thus, whenever the learner posits an error where the [k] is deleted, as in (77), the constraint MAX-T/V_R will be promoted twice as much as MAX-R/T_V is when the speaker posits an error where the [l] is deleted. A modified schema is shown in (78), where the differential promotion amount is shown by increasing the number of promotion arrows associated with a particular constraint.

(78) **Errors produced over the course of learning Canontese with a plasticity bias**

/sēj sēj lù:k lù:k/	*Complex	*Coda	*Tu	Max-T/V_R	Max-R/T_V
a. ⋯ lù:k lù:k ∼ ⋯ lù: lù:k		↓		↑↑	
b. ⋯ lù:k lù:k ∼ ⋯ lù:k ù:k			↓		↑

If the learner posits each error the same number of times, it will be the case that Max-T/V_R will have a **higher weight** than Max-R/T_V, since it was promoted twice as much at each step of learning. This is schematized in (79).

(79) **Result of learning from errors like those in (78)**

/sēj sēj lù:k lù:k/	*Complex	Max-T/V_R	Max-R/T_V	*Coda	*Tu
☞ a. sēj sēj lù:k lù:k				****	**
b. sēj sēj lù: lù:k		*!		***	**
c. sēj sēj lù:k ù:k			*!	****	*

This weighting scheme is, of course, consistent with the one required to derive deletion of the [ɹ] from English [ˈmʌjkɹəˌfəʊn] 'microphone' in (76), above.

In this way, although the learner has been provided with an **identical amount** of information from the native vocabulary regarding which Faithfulness constraints must be promoted, an asymmetry still arises due to the incorporation of a plasticity bias. Thus, an asymmetrical weighting scheme can be derived from a symmetrical source of data. This is a much simpler requirement of the training data than the one that would be required for the null hypothesis to be true – all that is necessary is for the learner to produce errors which allow the Faithfulness constraints to be promoted, and the plasticity bias will ensure that they are promoted in a PMap-compliant fashion. The purpose of this experiment is to see whether a learner trained exclusively on native Cantonese data encounters such evidence, and if it is sufficient to arrive at a weighting of constraints that adequately predicts the attested patterns of loanword adaptation.

A secondary purpose of this experiment is to lay out a method by which a **quantitative** PMap for Cantonese may be derived. As noted in Chapter 3, there is not a wealth of experimental evidence about the confusability of segments and their absences that can be used to derive a PMap-compliant plasticity bias. Instead, the weights of the best-case grammar derived in the previous section will be used as a **proxy**. The reasoning behind this choice is laid out in section 4.4.2. Two separate substantive biases will be constructed using the weights of the best-case grammar. One bias – hereafter called the **extreme** bias – will simply use the set of weights from the best-case grammar in order to derive plasticities. This was done in order to test how the MaxEnt learner would behave if it were deriving all of the constraint asymmetries from a substantive bias. The second bias – hereafter called the **pseudo-PMap** bias – will only use those constraint weights which were deemed to be plausibly due to the PMap after comparing the best-case grammar to the predictions made by the cue-based literature in Wright (2004) and Yun (2016). This was done in order to test whether the patterns of loanword adaptation seen in Cantonese were, in fact, due to the PMap in particular, and not some other set of biases.

4.4.1 Data

The training data used for this experiment were **identical** to the training data used for the first five grammars constructed as a part of experiment 1, detailed in section 4.3. Each training set consisted of 1000 lexical items of Cantonese drawn from CC-Canto. This size of training set was chosen since it provided the most internally consistent results.

The test data used for this experiment were also identical to the test data used in experiment 1. This ensured that the results of both experiments could be compared directly, both qualitatively by comparison to the attested loanword data and one another, and quantitatively by calculation of KL-Divergence.

4.4.2 Model parameters

Naturally, as this experiment is meant to test whether a plasticity bias can allow a learner to arrive at a grammar that predicts loanword adaptation patterns, the MaxEnt model used for this experiment made use of a **plasticity bias** as opposed to the more general Markedness $\gg$ Faithfulness bias used in experiment 1. However, as the learners used throughout this thesis are meant to model the PMap knowledge of speakers of **Cantonese**, the confusability data from Miller and Nicely (1955) used by Wilson (2006) and White (2013) to construct their own substantive biases could not be used. Compounding the issue is the **lack** of confusability data for the consonants of Cantonese. Instead, a substantive bias must be constructed through other means.

The best-case grammar, used for the purposes of comparison in experiment 1, gives a reasonable starting point when reasoning about what a substantive bias could look like for Cantonese. After all, if it is assumed that at least some of the asymmetries in the weights of Faithfulness constraints are due to the PMap, then those weights can be used as part of a substantive bias when training new models. Furthermore, as discussed in Chapter 1 and at the outset of the current chapter, the kinds of repairs undertaken by speakers of Cantonese appear to be in line with the PMap. Thus, any weighting asymmetries that account for the behaviour of loanwords in Cantonese can be considered to be candidates for inclusion in the PMap.

The remainder of this section will be dedicated to exploring how the best-case grammar for

Cantonese loanword adaptation can be used to create a substantive bias for use in this experiment. First, the weights of the best-case grammar will be examined in detail, to ascertain whether they are compliant with what we know of the PMap or not. Then, the methods by which the best-case grammar will be converted into a substantive bias are discussed.

4.4.2.1 Examining the Best-Case Grammar

As discussed in Chapter 3, Steriade's (2001) conception of the PMap relies on the notion of availability of **cues to contrast**. For example, given the constraints MAX-T/V_R and MAX-T/V_T, we will expect MAX-T/V_R to be weighted **higher** than MAX-T/V_T, as a following sonorant provides more cues to the presence of an obstruent than does another obstruent. The same methodology will be pursued here, drawing heavily on experimental work summarized by Wright (2004) and expanded upon by Yun (2016).

For reasons of space, only a selection of the relationships between Faithfulness constraints will be examined below, as examining every pairwise relationship between the Faithfulness constraints present in the constraint set would result in 8,128 comparisons. As such, only the constraints that are deemed to be most important for deriving patterns of loanword adaptation are examined – namely, the DEP-V constraints and the MAX constraints. The DEP-C constraints are ignored, as many of them receive very low weights across the board. I attribute their universal low weights to the fact that candidates which exhibit epenthesis of a single consonant will also run afoul of the high weighted *COMPLEX and CODACOND constraints. Since these constraints will already rule out candidates which violate DEP-C constraints, the weight of the DEP-C constraints will not need to be increased much, if at all. In other words, the high weighting of the Markedness constraints **obscures** the weighting of the DEP-C constraints, and a PMap-compliant weighting will not be entertained by the best-case grammar.

4.4.2.1.1 Dep-V Constraints Yun's (2016) work makes it possible to establish parameters on the weighting of DEP-V constraints, such that DEP-V constraints which make reference to environments where there is a **decrease** or **plateau** in overall intensity are weighted higher than those which make reference to environments where there is an **increase** in intensity. For the constraints used in the experiments outlined below, this means that a division can be made as in (80).

(80) **Weighting schema for Dep-V constraints**

$$\left\{\begin{array}{c} \text{Dep-V/\#\_\#} \\ \text{Dep-V/T\_\#} \\ \text{Dep-V/T\_T} \\ \text{Dep-V/R\_\#} \\ \text{Dep-V/R\_T} \\ \text{Dep-V/R\_R} \end{array}\right\} > \left\{\begin{array}{c} \text{Dep-V/\#\_R} \\ \text{Dep-V/T\_R} \end{array}\right\}$$

It should be noted that none of the Dep-V constraints listed in (80) involve environments where a vowel is inserted adjacent to another vowel. This is because these environments were never explored in Yun (2016), and so it is not known if the same phonetic cues apply in these environments. The only remaining Dep-V constraint that is not included in (80) is Dep-V/#_T. Since the obstruent in this constraint could refer either to a sibilant or to a non-sibilant, it could either result in a rise in intensity or a plateau, respectively, meaning it could be weighted in either category. I have as such refrained from placing it into the weighting schema. Finally, it is assumed that all of the Dep-V/R_ constraints involve **unreleased** nasals, as Bauer and Benedict (1997) claim that nasals in Cantonese are pronounced "as in English", and Jun (2004) notes that English nasals are unreleased when located before other consonants. As such, they should lead to either a plateau or a decrease in intensity, and receive a high weight.

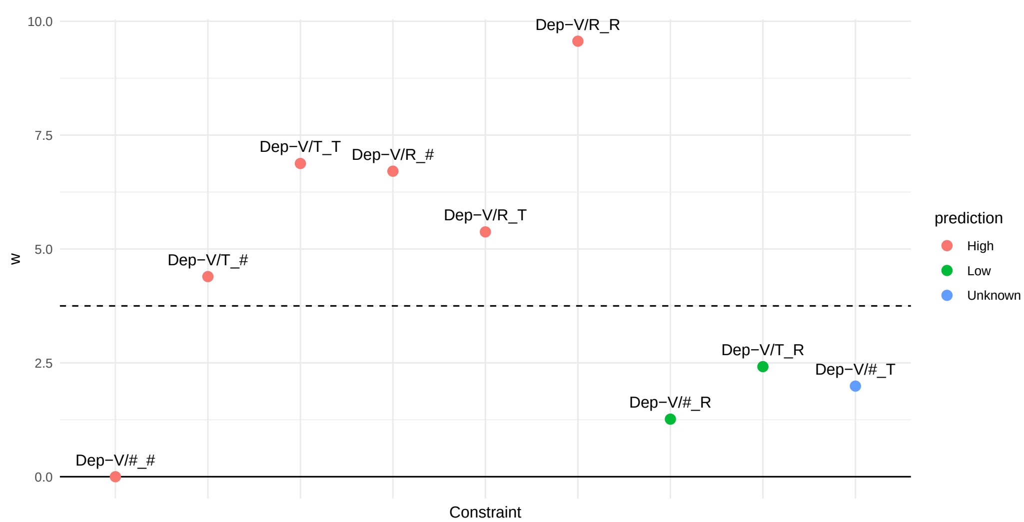


Figure 4-7: Weighting results for Dep-V constraints

The weights learned by the best-case grammar are presented in Figure 4-7. Each constraint is assigned a category based on the predictions outlined in (80), indicated by the colour of the dot associated with each constraint. The height of each dot corresponds to its weight as assigned under the best-case grammar.

By examining Figure 4-7, it appears that the predictions made in (80) are met overall. The weights appear to follow the same basic division, with clusters involving an intensity rise receiving weights near or below 2.0 and clusters involving no intensity rise receiving weights above 3.0. The perceived boundary between the two is indicated by the dotted line in Figure 4-7.

The exception to these predictions occurs with Dep-V/#_#, which receives a low weight in spite of its environment, which should result in an intensity plateau. Its low weight may be attributable to the fact that the candidate generation procedure used for these experiments would never have generated a candidate where this constraint could apply. After all, in order for the candidate generating script to have considered such a candidate, it would have had to consider an input that was **completely devoid** of phonological material. Since the learner would never have encountered a candidate that would violate this constraint, its weight was allowed to stay near the target weight of 0.0.

With regard to the Dep-V constraint which could or could not contain an intensity rise, Dep-V/#_T, it appears to have a relatively low weight. This may indicate that it should be evaluated as if it contains an intensity rise, or it may indicate that there is simply not enough data in the test set to rank it as high as other Dep-V constraints. Further examination of the distribution of environments in the lexical items contained in the test set should be done to determine which explanation is most plausible.

4.4.2.1.2 Max-V Constraints Following discussion contained in Chapter 3, it is assumed that the schema in (80) will also apply to the parallel Max-V constraints. Since both families of constraints involve making comparisons between a particular segment and its absence over a variety of contexts, they should receive similar, if not identical, weights. As such, the predictions for the weights of the Max-V constraints will follow the same weighting schema.

(81) **Weighting schema for Max-V constraints**

$$\left\{\begin{array}{c}\textsc{Max-V}/\#\_\# \\ \textsc{Max-V}/\mathrm{T}\_\# \\ \textsc{Max-V}/\mathrm{T}\_\mathrm{T} \\ \textsc{Max-V}/\mathrm{R}\_\# \\ \textsc{Max-V}/\mathrm{R}\_\mathrm{T} \\ \textsc{Max-V}/\mathrm{R}\_\mathrm{R}\end{array}\right\} > \left\{\begin{array}{c}\textsc{Max-V}/\#\_\mathrm{R} \\ \textsc{Max-V}/\mathrm{T}\_\mathrm{R}\end{array}\right\}$$

As was done for the Dep-V constraints above, the Max-V constraints that reference environments that contain vowels are excluded from the schema. The constraint Max-V/#_T is not included in the schema, although it is included in Figure 4-8, as it is unclear whether it should be classified as having an intensity rise or not.

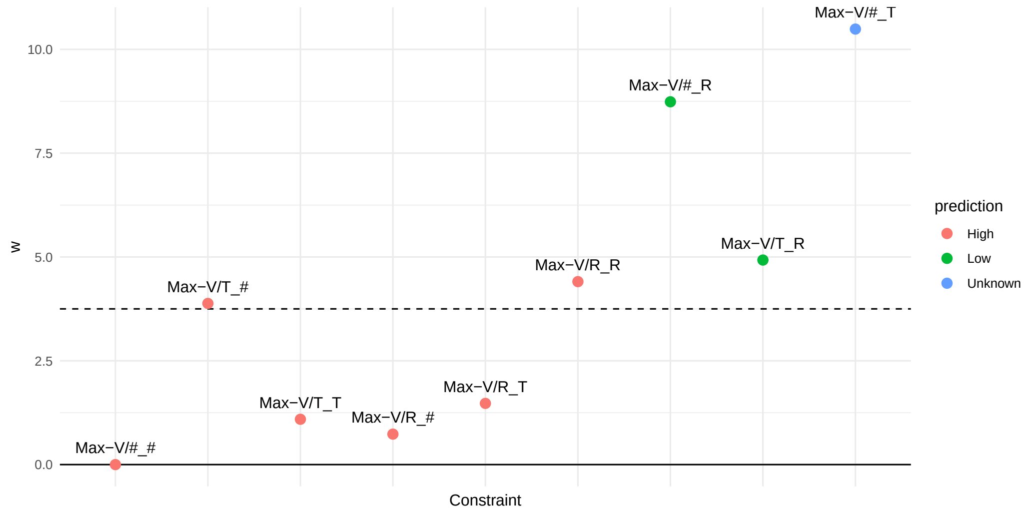


Figure 4-8: Weighting results for Max-V constraints

Contrary to what was seen for the Dep-V constraints, the predictions made in (81) are **not** borne out. In fact, the predictions appear to have been **reversed**, with environments that would result in an intensity rise receiving high weights (above 4.0), and environments that would result in an intensity plateau or fall receiving low weights (around 3.0 or below). This division is approximated by the dotted line in Figure 4-8. It is clear from these results that the phonetic basis for establishing a PMap-compliant weighting for the Max constraints may not be identical to the phonetic basis for establishing a PMap-compliant weighting for the Dep constraints. Additional research should be done into the confusability of segments and their absence in order to determine how these two constraint families could be differentiated by the PMap.

4.4.2.1.3 Max-C Constraints With respect to the Max constraints which target consonants, a comparable study of the discriminability of these segments in various contexts such as Yun's (2016) is not readily available. In lieu of such a study, I will be drawing on a survey of the phonetic cue literature undertaken by Wright (2004) in order to hypothesize what a PMap-compliant weighting of these constraints could be. I will be focusing especially on the section dedicated to the cues to the **manner** of consonants, as these are the main distinctions made by the constraint set used.

The Max-C constraints will be divided into three classes with respect to their **target segment**. This is done on the basis of the strength of the **internal cues** involved, which predominate in Wright's discussion of manner cues. The class of Max-C constraints which is predicted to have the highest average weight, and be less confusable with an absence of that segment in an identical context, is the set of Max-S constraints. These are be followed by the Max-R constraints, which also have strong internal cues to their presence in the form of formant structure (Wright, 2004). The constraints predicted to have the lowest overall weights are the Max-T constraints, which either have very weak cues to their presence throughout, such as the non-sibilant fricatives [f] and [θ]; or have next to no cues independent of context, such as the stops [p, t, k]. These inequalities were discussed in (44) in the previous chapter, and are presented again in (82).

(82) **Weighting schema for Max-C constraint families**

$$\textsc{Max-S} > \textsc{Max-R} > \textsc{Max-T}$$

Within each of these classes of Max-C constraints, additional inequalities in ranking can be inferred based on the **transitional cues** that may be present in each. According to Wright (2004), the transitional cues to manner include the presence of a **release burst** for stops, and the **slope of formant transitions** into and out of the target segment for all manner of consonants. Wright (2004) also observes that the auditory processing system gives attentional preference to cues at the **onset of vowels**, as opposed to cues present at the end of vowels. Thus, it will be assumed that cues at the **right edge** of a segment, when present, will be **better cues** to the presence of a particular consonant than cues at the left edge. Likewise, consonants which have both sets of cues are predicted to be more discernible than those which only have one set of cues. These facts lead to the formation of the general weighting schema for constraint contexts in (83), extrapolated from the discussion of cues to obstruents provided in (45) of Chapter 3.

(83) **Weighting schema for constraint contexts**

$$\left\{\begin{array}{c}\textsc{Max-C/V\_V}\\ \textsc{Max-C/V\_R}\\ \textsc{Max-C/R\_V}\\ \textsc{Max-C/R\_R}\end{array}\right\} > \left\{\begin{array}{c}\textsc{Max-C/\#\_V}\\ \textsc{Max-C/\#\_R}\\ \textsc{Max-C/T\_V}\\ \textsc{Max-C/T\_R}\end{array}\right\} > \left\{\begin{array}{c}\textsc{Max-C/V\_\#}\\ \textsc{Max-C/V\_T}\\ \textsc{Max-C/R\_\#}\\ \textsc{Max-C/R\_T}\end{array}\right\} > \left\{\begin{array}{c}\textsc{Max-C/\#\_\#}\\ \textsc{Max-C/\#\_T}\\ \textsc{Max-C/T\_\#}\\ \textsc{MaxC/T\_T}\end{array}\right\}$$

With this information in mind, each broad category of Max-C constraints (Max-T, Max-R, and Max-S) will be examined in turn. First, the fully contextual constraints will be examined individually to see if the weights assigned to them by the best-case grammar follow the schema outlined in (83). Then the average weight for the entire class will be calculated for the purposes of seeing whether the prediction made in (82) is met.

4.4.2.1.4 Max-T Constraints In order to make the procedure for examining the Max-C constraints clearer, the full schema for the Max-T constraints is provided in (84).

(84) **Weighting schema for Max-T constraints**

$$\left\{\begin{array}{c}\textsc{Max}\text{-T/V\_V}\\ \textsc{Max}\text{-T/V\_R}\\ \textsc{Max}\text{-T/R\_V}\\ \textsc{Max}\text{-T/R\_R}\end{array}\right\} > \left\{\begin{array}{c}\textsc{Max}\text{-T/\#\_V}\\ \textsc{Max}\text{-T/\#\_R}\\ \textsc{Max}\text{-T/T\_V}\\ \textsc{Max}\text{-T/T\_R}\end{array}\right\} > \left\{\begin{array}{c}\textsc{Max}\text{-T/V\_\#}\\ \textsc{Max}\text{-T/V\_T}\\ \textsc{Max}\text{-T/R\_\#}\\ \textsc{Max}\text{-T/R\_T}\end{array}\right\} > \left\{\begin{array}{c}\textsc{Max}\text{-T/\#\_\#}\\ \textsc{Max}\text{-T/\#\_T}\\ \textsc{Max}\text{-T/T\_\#}\\ \textsc{Max}\text{-T/T\_T}\end{array}\right\}$$

This schema predicts that those obstruents that are located between sonorants and vowels should receive the **highest weight**, as they have strong cues to their presence from both the **left transition** out of the preceding vowel or sonorant, and the **right transition** into the following vowel or sonorant. Obstruents which only have strong cues to their presence from the **right transition** are predicted to have the next highest weight. Next highest are those which only have strong cues to their presence from the **left transition**. Finally, those obstruents which do not have **any** strong cues to their presence are predicted to have the lowest weights.

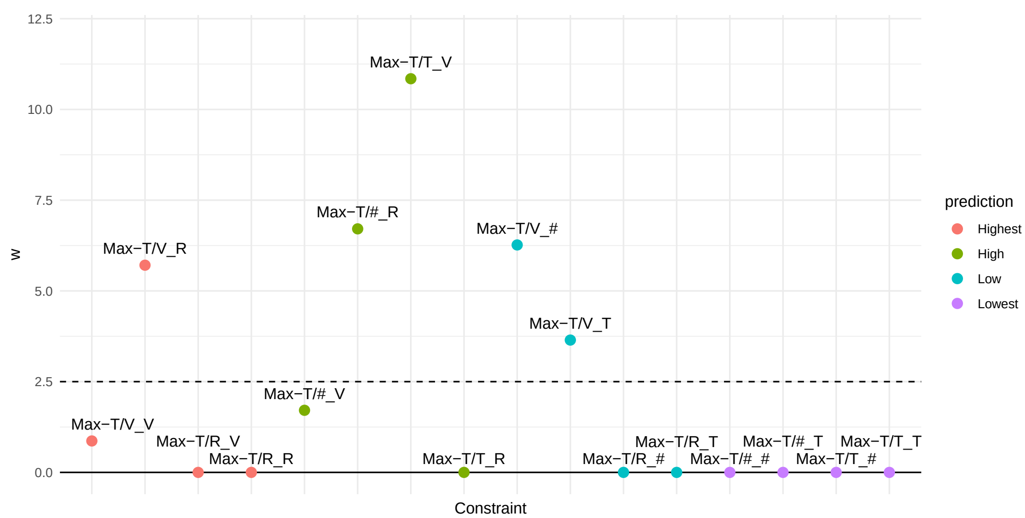


Figure 4-9: Weighting results for Max-T constraints

The actual weights learned by the best-case grammar are presented in Figure 4-9. With regards to the constraints which are predicted to have the lowest weights, predictions seem to be met – that is, constraints which contain environments that are not predicted to host consonantal manner cues very well, such as T_T and T_#, receive very low weights. Similarly, some environments which contain only left-edge transitions, such as R_# and R_T, are also particularly low-weighted. Other left-edge transition-containing constraints, such as Max-T/V_# and Max-T/V_T, receive higher weights than 0.0, but lower weights than some of their right-edge transition-containing counterparts.

However, with regards to the constraints which are predicted to have high weights, results are **mixed**. Constraints which contain environments where consonants are predicted to be best-cued, such as V_V and R_R, overall receive very **low weights**. Constraints which contain environments with only right-edge transitions, such as T_V and #_R, receive the highest weights of all. These results may arise from the **distribution** of stops in these contexts within the test set data – for example, it may be that there are an abundance of forms which contain a stop that is preceded by another stop and followed by a vowel, but relatively few forms where a stop is truly intervocalic. This may very well be true of the native Cantonese vocabulary, since all syllables have a tendency to be **bimoraic** (Kenstowicz, 2012), and may have a greater chance of containing a coda consonant than might be expected based on the data from other languages. Thus, there may not be sufficient evidence from the small set of native Cantonese words present in the test set for these constraints to have been moved sufficiently far from zero.

Overall, the weight of the Max-T constraints μ_T is 2.23484. This value will be compared to the mean weights of the other Max-C constraint classes in the following sections.

4.4.2.1.5 Max-R Constraints When the same environmental criteria are applied to the series of Max-R constraints, predictions are even less well-met, with weights of constraints on all but the end of the scale being **contrary** to what is expected based purely on environmental cues. The weights for each of the Max-R constraints are shown in Figure 4-10.

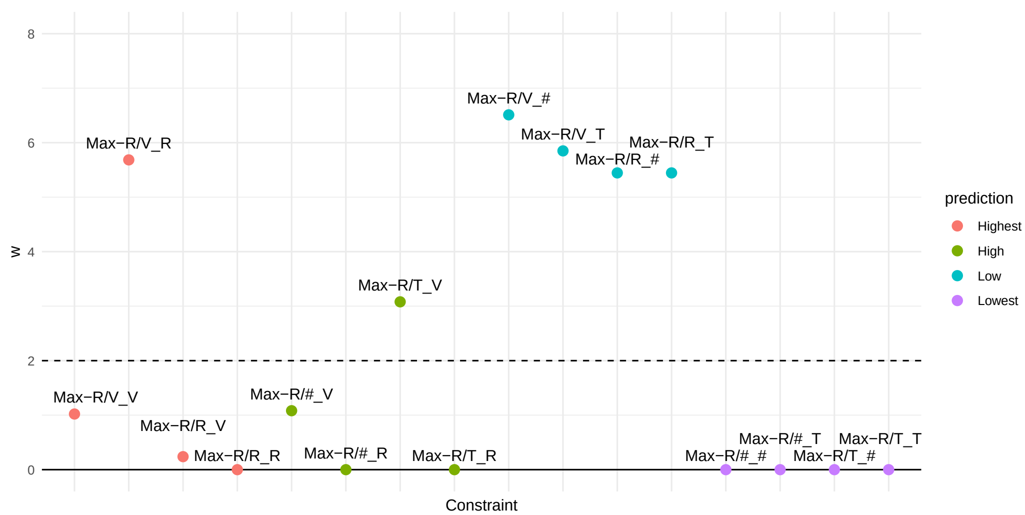


Figure 4-10: Weighting results for Max-R constraints

The worse fit of these weights to the schema in (83) may be due to the fact that the schema is largely designed around the behaviour of **obstruents**, which require strong transitional cues in order to be perceived at all. Sonorants, which have strong internal cues of their own, may not need to rely so strongly on their transitional cues.

It may also be the case that the relative strength of contextual cues is **different** for sonorants, and the contextual schema should be different for this class of segments. For example, take the constraint MAX-R/T_V. According to the schema in (83), this constraint should have a relatively high weight, as it has a source of strong transitional cues at its right edge. However, **intuitions** about the discriminability of a sonorant in this context predict that its weight should be low, as it is often quite difficult to distinguish sonorants from vowels in this context (Wright, 2004).[6] This appears to be more consistent with a model which **integrates** both kinds of cue information in a more sophisticated manner than entertained here.

On average, the weight of the MAX-R constraints μ_R is 2.14743 – slightly **lower** than those of the MAX-T constraints. This is contrary to the prediction made in (82), which predicted that MAX-R constraints should, as a whole, receive higher weights than the MAX-T constraints. Again, this indicates that the naïve approach undertaken here may be too inaccurate to be of substantial use.

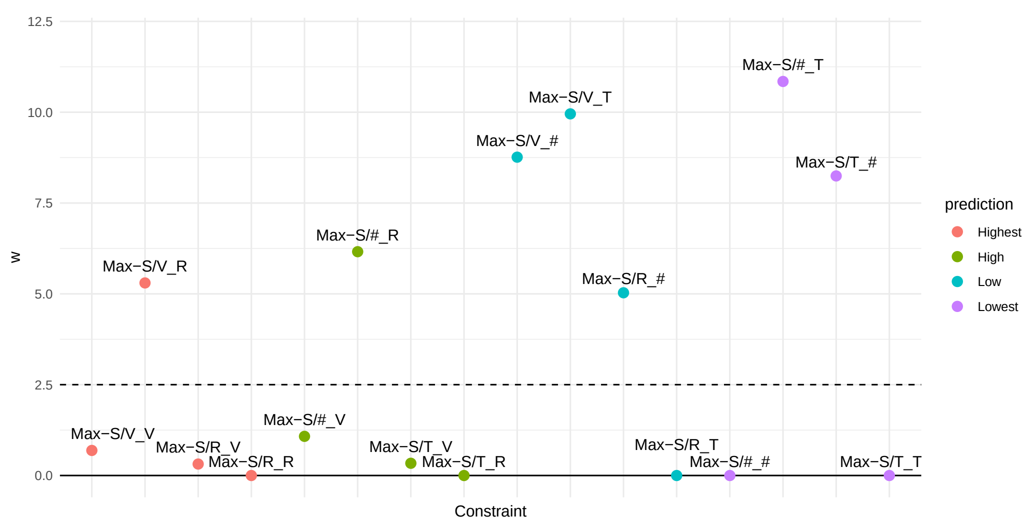


Figure 4-11: Weighting results for MAX-S constraints

4.4.2.1.6 Max-S Constraints The overall fit for the MAX-S constraints within the schema in (83) is the **worst** of the target segments, with environments that provide worse contextual cues being associated with the highest weights, and vice-versa. This is shown in Figure 4-11. However, when it comes to the **overall** weight of the MAX-S constraints, the prediction of the inequality outlined in (82) holds, with the MAX-S constraints receiving an average weight of 3.54523, well in excess of

[6]At least, this is the case for native English speakers. Whether this is true of Cantonese speakers remains to be seen.

the average weights of both other classes of Max-C constraints.

4.4.2.1.7 Discussion of the Max-C constraints With regards to the cues provided by transitions into and out of adjacent segments, it appears that the predictions about the strength of these transitional cues are **flipped** – that is, transitions at the **left edge** of a segment appear to be **more influential** than transitions at the **right edge**, contrary to what is reported in Wright (2004). This could simply be the result of coincidence, or the result of learning directly from the loanword data, which have an abundance of coda segments that must be preserved in the output.

Discrepancies may have also arisen due to the way in which the grammar was **generated**. As discussed in section 4.4.2.1.1, some constraints will have weights at or close to 0 simply because the test set failed to contain candidates that would cause these constraints to be violated. In addition, as discussed with respect to experiment 1 in section 4.3, some constraints may have weights close to 0 because any candidate which might violate them would also run afoul of the more highly-weighted Markedness constraints. Since a learner will always use the highly-weighted Markedness constraints to rule these candidates out, there is no need to adjust the weight of the Faithfulness constraint over the course of learning. It may be that if a different model were trained on the test data set – one without a Markedness $\gg$ Faithfulness bias – some of these constraints might receive higher weights, and be more in line with what is predicted in (82) and (83).

The highly granular and independently-evaluable constraint set may also be to blame for these inconsistent results. For example, it may be that a model which makes use of, say, a single Max-S constraint and a series of contextual Max-T constraints, where "T" includes both sibilant and non-sibilant obstruents, makes better predictions about the relative weightings of the constraints. However, such a model runs afoul of the problem outlined in Chapter 3, where it may not always be transparent how to assign weights to more general constraints with respect to the PMap. This type of solution will be discussed further in Chapter 5.

Finally, it may also be the case that the relative inequalities present in the weights of the best-case grammar are due to **some other bias**, such as a bias towards weighting Max constraints above Dep constraints for loanwords specifically. Regardless of where the bias comes from, speakers will nevertheless have access to it during learning, and this should be maintained in the models to come if at all possible.

4.4.2.2 Constructing a plasticity bias

Now that a little more is understood about how well the weights of the best-case grammar line up with predicted PMap weighting inequalities, it remains to be seen how this information can be used to construct a plasticity bias for the current experiment. While some aspects of the best-case grammar are consistent with the PMap, such as the weighting of Dep-V constraints and the higher overall weight of the Max-S constraints, others are either wholly unexpected or obscured by the manner in which the best-case grammar was generated. Without a better understanding of how confusable native speakers of Cantonese find these particular contrasts, it will be hard to tell exactly how accurate or inaccurate this particular grammar is.

Regardless of the best-case grammar's shortcomings given our current understanding of the PMap, it still has value both as a basis for **a** substantive bias that can more accurately predict patterns of loanword adaptation, and as a source of **plasticities** for constructing a PMap-like bias. After all, the purpose of this experiment is split into two separate questions: 1) whether a learner equipped with a plasticity bias will encounter enough evidence from the native Cantonese vocabulary to arrive at the asymmetrical faithfulness constraint weightings necessary to predict attested patterns of loanword adaptation; and 2) whether such a bias is PMap-compliant. With respect to the first question, **any** substantive bias which can predict the attested loanword patterns will be sufficient. With respect to the second, a more nuanced approach is required.

With these two considerations in mind, I will use **two** separate plasticity biases for this experiment. The first will merely be a **translation** of the best-case grammar from experiment 1 into a plasticity bias, to be outlined below. The second will be an **approximation** of the PMap based on the weighting schema outlined in the previous section.

For the first plasticity bias, or the **extreme** plasticity bias, the weights learned by the best-case grammar from experiment 1 were **converted** into plasticities according to the formula in (85).

(85) **Formula for converting weight to plasticity**

$$\sigma^2 = w^2 + 10^{-8}$$

Recall from Chapter 3 that the variance of a Gaussian distribution, noted as σ^2, is equivalent to the **plasticity** accorded to a particular constraint. The variance is equivalent to the **standard deviation** of the Gaussian distribution, squared. Say now that we are given a constraint – Max-V/#_T – and a weight which that constraint must be able to achieve – 10.49. As per the requirements of a plasticity bias, established by Wilson (2006), any Gaussian prior established over Max-V/#_T must have a mean, or target weight, of 0. The question we must then ask is this: What plasticity should we assign this constraint so that it is not overly penalized if it arrives at a weight of 10.49?

While many plasticities will work, a simple method of ensuring this is possible is to say that the desired value falls within one **standard deviation**, noted as σ, of the mean. In this way, the Gaussian established for the constraint MAX-V/#_T will not greatly penalize any deviation from the mean that is 10.49 or less, but will still penalize higher weights. As the mean is always set at 0 for a plasticity-based prior, and the desired weight lies one standard deviation from the mean, we know that the desired weight of 10.49 must be equal to one standard deviation, or σ. Since the plasticity assigned to the Gaussian bias is defined as σ^2, we know that the plasticity must be equal to the desired weight of a constraint, squared.

If, however, the best-case grammar derived in experiment 1 assigns a particular constraint a weight of 0, something else must be done. If this value were to be used as a plasticity, it would raise a divide-by-zero error during computation. The addition of some value to each trained weight was done to avoid this issue. Its size was kept small to ensure that the resulting plasticities were not too different from the square of the learned weight.

A sample of plasticity values for the DEP-V constraints discussed in the previous section is provided in (86), while the full plasticity bias is provided in Appendix B.

(86) **Sample plasticities for Dep-V constraints, extreme bias**

Constraint	μ	σ^2
DEP-V/#_#	0.0	0.00000001
DEP-V/T_#	0.0	19.931167
DEP-V/T_T	0.0	47.30006
DEP-V/R_#	0.0	44.99862
DEP-V/R_T	0.0	28.89574
DEP-V/R_R	0.0	91.44766
DEP-V/#_R	0.0	1.597567
DEP-V/T_R	0.0	5.840400

The second plasticity bias, the **pseudo-PMap** bias, used the same method of converting weights taken from the best-case grammar of experiment 1, but did not use every weight from the best-case grammar when establishing a Gaussian prior. Rather, a **mix** of the Markedness $\gg$ Faithfulness bias from experiment 1 and a PMap-like bias was undertaken. Constraints which were deemed to be likely candidates for deriving the attested patterns of loanword adaptation in Cantonese (*i.e.*, those discussed in the previous section) were assigned a plasticity derived from the **highest weight in their category**. The highest value was chosen to represent all constraints of a particular category in order to mitigate the concern that constraints with unexpectedly low weights had been overshadowed by the Markedness constraints over the course of learning. However, this value was not altered further, even if it contradicted the schemas outlined in the previous section. This was done in order to reflect the uncertainty about how internal and external segmental cues are integrated in order to arrive at a PMap for segment insertion and deletion. All other constraints were assigned the target weight and plasticity values from experiment 1 – in other words, they followed a general Markedness $\gg$ Faithfulness bias.

A sample of plasticity values for the Max-S constraints discussed in the previous section is provided in (87). In addition, the plasticities and target weights for constraints not considered in the previous section are provided, in order to illustrate that they follow a more general Markedness $\gg$ Faithfulness bias. A full list of plasticities and target weights for this bias is provided in Appendix B.

(87) **Sample plasticities for Max-S constraints, psuedo-PMap bias**

Constraint	μ	σ^2
Max-S/V_V	0.0	28.09776
Max-S/V_R	0.0	28.09776
Max-S/R_V	0.0	28.09776
Max-S/R_R	0.0	28.09776
Max-S/#_V	0.0	37.95856
Max-S/#_R	0.0	37.95856
Max-S/T_V	0.0	37.95856
Max-S/T_R	0.0	37.95856
Max-S/V_#	0.0	99.09944
Max-S/V_T	0.0	99.09944
Max-S/R_#	0.0	99.09944
Max-S/R_T	0.0	99.09944
Max-S/#_#	0.0	117.6548
Max-S/#_T	0.0	117.6548
Max-S/T_#	0.0	117.6548
Max-S/T_T	0.0	117.6548
*Complex	10.0	100
*Coda	10.0	100
Dep-T/V_V	0.0	100
Dep-R/#_T	0.0	100

Aside from the two plasticity priors discussed above, the procedure for constructing MaxEnt grammars for this experiment was **identical** to the procedure for experiment 1. The same 1000-item training sets were used to train two Max-Ent learners, one equipped with the extreme plasticity prior and another with the pseudo-PMap prior. Once the grammars were trained and a set of constraint weights found for each, they were applied to the test set to predict probabilities for each candidate in the test set. The predicted probabilities for an exemplar grammar will be examined in the following section, and compared to the best-case grammar and the exemplar of the 1000-item null hypothesis grammar, both qualitatively and quantitatively via computation of KL-Divergence.

4.4.3 Results

4.4.3.1 Extreme Bias

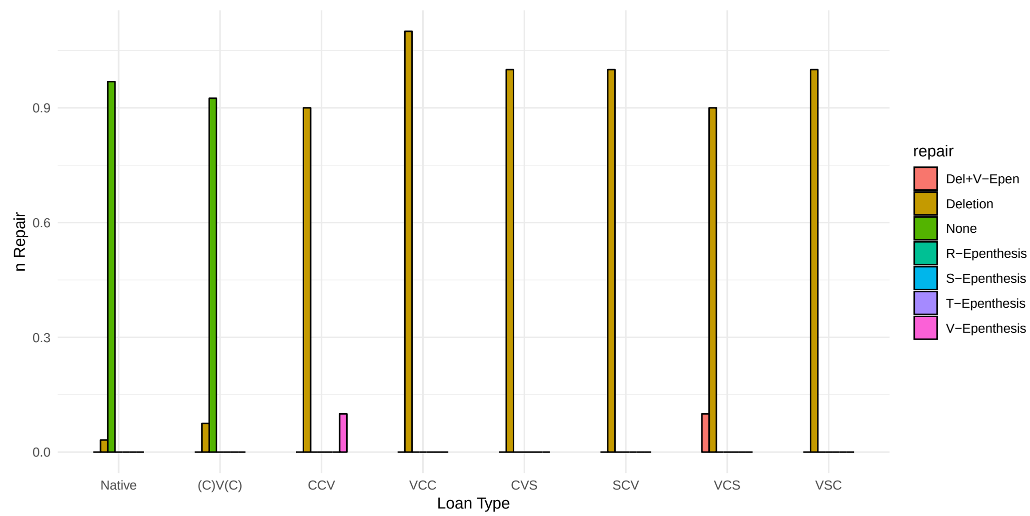


Figure 4-12: Proportion of winning candidates divided by repair and input structure category, representative sample of 1000 Cantonese lexical entries with an extreme plasticity-based substantive bias

As can be seen in Figure 4-12, a MaxEnt learner equipped with a plasticity bias produces results which are **extremely similar** to the results obtained under the null nypothesis, discussed in detail in section 4.3. With respect to the native Cantonese words and the phonotactically-licit loanwords, the candidates which require no changes to the input are overwhelmingly the candidates assigned the highest probability by the model. There are a few exceptions where deletion is favoured in these cases – as is expected when training on a data set that consists of 1000 lexical items, which was shown in section 4.3.3.1 to be lacking lexical items containing single-vowel syllables. With respect to the phonotactically-illicit loanwords, the candidates which receive the highest weights are those which **delete** a segment of the input. A sample of the highest-probability repairs are provided in (88).

(88) **Sample repairs for a grammar trained on a 1000-item data set with an extreme plasticity bias**

Category	*Repair Type*	*Input*	*Gloss*	*Output*	*p*
Native	None	`naap seoj`	'tax'	`naap seoj`	0.99455
	Deletion	`o pej`	'to fart'	`pej`	0.99999
(C)V(C)	None	`kaa low lej`	'calorie'	`kaa low lej`	0.85773
	Deletion	`e sej`	'essay'	`sej`	0.99999
CCV	Deletion	`klip`	'clip'	`lip`	0.54213
	V-Epenthesis	`wo ljam`	'volume'	`wo lV jam`	0.99599
VCC	Deletion	`sapt`	'shaft'	`sat`	0.49851
				`sap`	0.49851
CVS	Deletion	`paas pot`	'passport'	`paa pot`	0.99996
SCV	Deletion	`spaak`	'spark'	`paak`	0.94949
VCS	Deletion	`inc`	'inch'	`in`	0.99854
	Deletion + V-Epenthesis	`o leonc`	'orange'	`leon cV`	0.80359
VSC	Deletion	`tost`	'toast'	`tot`	0.99983

After examining the exact repairs chosen as most likely by the extreme plasticity bias grammar, it can be confirmed that not much has substantially changed from the comparable null hypothesis grammar of experiment 1. Lexical items which contain single-vowel syllables indeed show deletion of this syllable, as expected. If a consonant cluster contains two segments of the same sonority, the grammar still has difficulty in chosing which to select for preservation. Segments which undergo deletion in phonotactically-illicit loanwords are overwhelmingly the **sibilant** or the segment **furthest** from the vowel, contrary to what is attested for native speakers of Cantonese.

One small difference from the null hypothesis grammar is the treatment of the item *o leonc* 'orange'. In the extreme plasticity bias grammar, the initial single-vowel syllable is deleted **and** a vowel is epenthesized after the final sibilant. While an overall preference for deletion would suggest that the preferred candidate would be one where **both** the single-vowel syllable and the offending sibilant were deleted, this is not possible given the current test set. Recall that this test set did **not** consider candidates where two deletions occurred – the script which generated the test set was not set up to generate candidates in this way. As such, a high weight of constraints such as Bimoraic and CodaCond will be sufficient to explain this sequence of repairs – Bimoraic will rule out the single-vowel syllable, and CodaCond will force vowel epenthesis after the sibilant. It may be that the plasticities generated for the extreme bias were **higher** than those used for the null hypothesis experiments, allowing these constraints to receive higher weights over the course of learning. For example, plasticities in excess of 100 are observed in (87), while all constraints had a plasticity of 100 under the Markedness $\gg$ Faithfulness bias. This could explain the discrepancy between the extreme plasticity bias grammar and the null hypothesis grammars.

Another small difference is that words with onset consonant clusters, such as *klip* 'clip' receive a much **lower** probability from the extreme plasticity bias grammar – just over 50%. This could be indicative of one of two things. Either a vowel-epenthesis candidate, such as *kV lip*, is being assigned more probability under this model, or the contrasting deletion candidate *kip* is receiving the probability (or, perhaps, both). If this is the case, it is a sign that the plasticity bias may be having **some** effect in the grammar, albeit slight.

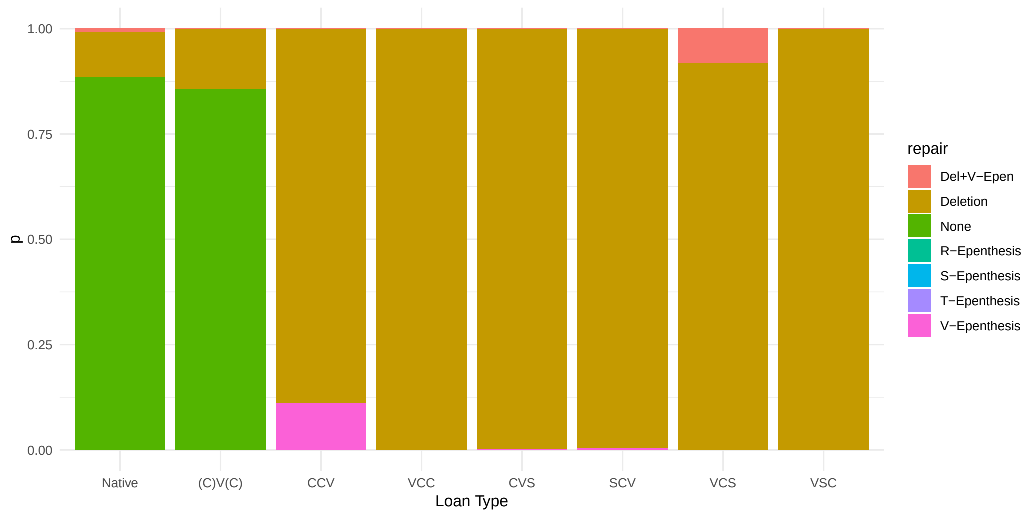


Figure 4-13: Probability of repair by input structure category, representative sample of 1000 Cantonese lexical items with an extreme plasticity-based substantive bias

The results from Figure 4-13 provide some slight support for the hypothesis about how probability is assigned within the onset cluster candidates. There is a small uptick in the total probability assigned to vowel epenthesis candidates, as well as in the total probability assigned to deletion candidates. It may therefore be the case that constraints such as MAX-T/V_R or MAX-T/#_R are able to receive **higher** weights than in the null hypothesis models, and constraints such as DEP-V/T_R are forced to have **lower** weights. However, the overwhelming preference for deletion as a repair across the board is nevertheless maintained.

With respect to KL-Divergence values, the extreme plasticity bias grammar has a KL-Divergence value of $D_{KL} = 2.15023$ when compared with the best-case grammar from experiment 1. This indicates that these two grammars are incredibly **dissimilar**, as is expected. When compared to the null hypothesis grammar discussed in section 4.3.3.1, this grammar has a KL-Divergence value of $D_{KL} = 0.23497$. While this is indeed a value that is quite low, it is not as low as the KL-Divergences between the 1000-item grammars of experiment 1. It can thus be stated that, while introducing an extreme plasticity bias does not lead to attested patterns of loanword adaptation, it **does** make an impact on the grammar learned.

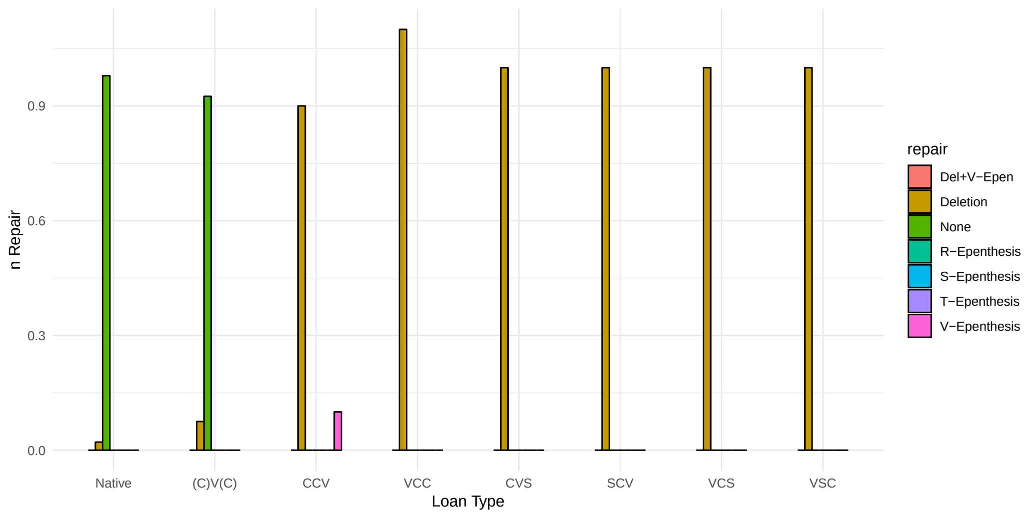


Figure 4-14: Proportion of winning candidates divided by repair and input structure category, representative sample of 1000 Cantonese lexical items with a plasticity-based PMap-style bias

4.4.3.2 Pseudo-PMap Bias

A MaxEnt learner equipped with a pseudo-PMap plasticity bias does not fare any better than a learner equipped with a more extreme plasticity-based substantive bias, as evidenced by the data presented in Figure 4-14. The overwhelming preference is to repair all phonotactically-licit loanwords and other words which contain rare syllablic profiles by **deletion**.

(89) **Sample repairs for a grammar trained on a 1000-item data set with a plasticity-based PMap-style bias**

Category	*Repair Type*	*Input*	*Gloss*	*Output*	*p*
Native	None Deletion	`naap seoj` `o pej`	'tax' 'to fart'	`naap seoj` `pej`	0.99331 0.94207
(C)V(C)	None Deletion	`kaa low lej` `e sej`	'calorie' 'essay'	`kaa low lej` `sej`	0.99652 0.94191
CCV	Deletion V-Epenthesis	`klip` `wo ljam`	'clip' 'volume'	`lip` `wo lV jam`	0.84986 0.99794
VCC	Deletion	`sapt`	'shaft'	`sat` `sap`	0.49522 0.49522
CVS	Deletion	`paas pot`	'passport'	`paa pot`	0.99912
SCV	Deletion	`spaak`	'spark'	`paak`	0.95242
VCS	Deletion	`inc`	'inch'	`in`	0.99228
VSC	Deletion	`tost`	'toast'	`tot`	0.99894

The repairs entertained by the psuedo-PMap plasticity bias grammar are much more consistent with those of the null hypothesis than were the results of the extreme bias grammar. The familiar patterns of deleting single-vowel syllables and of being unable to choose between consonants of identical sonority are still present, as are the patterns of deleting sonorants and peripheral consonants.

These results are also borne out when examining how probability is assigned to candidates in the test set as a whole, displayed in Figure 4-15.

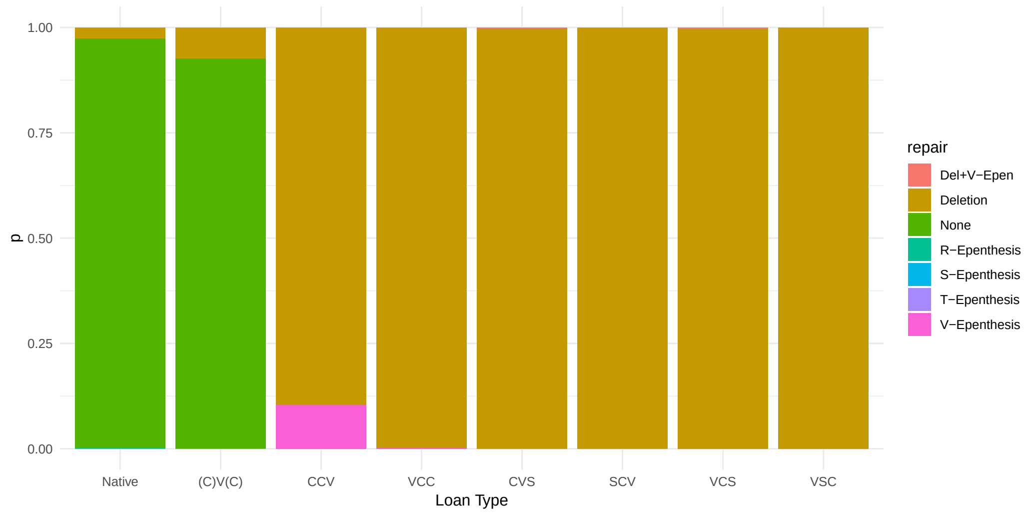


Figure 4-15: Probability of repair by input structure category, training set of 1000 Cantonese lexical items with a plasticity-based PMap-style bias

The KL-Divergence between the predictions made by the pseudo-PMap plasticity bias grammar and the best-case grammar from experiment 1 is $D_{KL} = 1.89153$, indicating that these two distributions are quite dissimilar, although not quite as dissimilar as the results of the extreme plasticity bias grammar. In contrast, the KL-Divergence between these predictions and the comparable null hypothesis grammar is $D_{KL} = 0.08226$, indicating that these two grammars are actually quite similar. Finally, the KL-Divergence between the psuedo-PMap plasticity bias grammar and the extreme plasticity bias grammar is $D_{KL} = 0.43074$, again indicating that these two grammars are quite similar.

In light of how the two plasticity biases were constructed, these results are unsurprising. After all, the extreme bias used entirely different target weights and plasticities from the biases employed by the null hypothesis grammars, while the pseudo-PMap bias was a **mixture** of the two. It would make sense if the results of a grammar trained using a pseudo-PMap bias were more similar to those of a grammar trained using a simple Markedness $\gg$ Faithfulness bias, since the pseudo-PMap bias uses the Markedness $\gg$ Faithfulness bias whenever possible.

Nevertheless, the results of both plasticity bias experiments has shown that there is **no substantial change** in results from those of the null hypothesis. Deletion is still the preferred method for repairing phonotactically-illict inputs, rather than the more sophisticated series of repairs that is attested for speakers of Cantonese.

4.4.4 Discussion

The goal of this experiment was twofold: first, it was meant to test whether encoding a substantive bias of **any kind** in terms of asymmetrical constraint plasticities would allow a MaxEnt learner to arrive at a grammar that would mirror the asymmetries seen in loanword adaptation; and second, it was meant to test whether this substantive bias could – at least in part – be due to the PMap.

It was shown in section 4.4.2.1 that there is as of yet no conclusive answer to the latter question. No perceptual confusability data currently exist for the presence or absence of a segment in Cantonese, so reasoning directly from confusability, as was done for the biases constructed by Wilson (2006) and White (2013), is not possible. Although reasoning about potential inequalities from the presence or absence of acoustic cues to contrast seemed to match up with the relative weighting of the Dep-V constraints and broad categories of Max constraints within the best-case grammar from experiment 1, this was not the case for the remainder of constraints considered. Either the way in which the best-case grammar was learned caused any potential PMap-compliant weights to be **obscured** by the high weights of the Markedness constraints, or the data on which the grammar was trained was not sufficient to alter the weights of the Faithfulness constraints in question. A third possibility is that the way in which the cues to contrast were assumed to interact was incorrect, and that more needs to be understood about the way in which segments of various sonorities behave in a variety of segmental contexts before attempting to reason about the PMap directly from cues.

With respect to the former question, it was also shown that a substantive bias encoded via asymmetrical plasticities is **not sufficient** to derive or even approximate patterns of loanword adaptation in Cantonese. In fact, even a substantive bias that is a **direct translation** of the best-case grammar into a reasonable set of plasticities fares little better than grammars equipped only with a Markedness $\gg$ Faithfulness bias. This indicates that the native Cantonese vocabulary cannot behave as outlined in the introduction to this section, where the evidence from the promotion of Faithfulness constraints over the course of learning interacts with the asymmetrical plasticity bias to arrive at a set of asymmetrical weights. Rather, it is apparent that the native Cantonese vocabulary does not provide adequate evidence to substantially re-weight any of the Faithfulness constraints responsible for the attested patterns of loanword adaptation. While making use of a plasticity bias may be appropriate when modelling the behaviour that learners have when presented with **explicit repair strategies**, as is done in Wilson (2006), it is not appropriate for modelling how the PMap shapes monolingual language learning in the **absence of direct evidence**, as is done here.

4.5 Experiment 3: Weight prior

As opposed to establishing the PMap via a series of Gaussian priors that differ only in the plasticity of each constraint, White (2013) takes a more direct approach and establishes the substantive bias by setting the **plasticity uniformly** but by **varying the target weights** of each constraint that is governed by the PMap. As was done for the previous experiment, this experiment will test the behaviour of this weight-based prior by training a series of MaxEnt learners on native Cantonese words, and then using those grammars to predict patterns of loanword adaptation in Cantonese. As stated in (50c) at the outset of this chapter, it is expected that a learner equipped with a weight-based prior should **always** make accurate predictions about the kinds of repairs which are attested in loanwords, since the target weight for any PMap-governed constraints will be **maintained** in the absence of evidence.

As a quick illustration, recall from the previous experiments that in order to predict deletion of sonorants in English loanwords such as [ˈmʌjkɹəˌfəʊn] 'microphone', there must be a weighting of constraints such that the weight of Max-T/V_R must exceed that of Max-R/T_V. Previous experiments have required that there be lexical items in Cantonese which have a *kl* sequence in order for the learner to adjust the weights of these constraints, either symmetrically in the case of experiment 1, or asymmetrically in the case of experiment 2. However, if the PMap is encoded as an asymmetrical set of target weights, then the weight of Max-T/V_R can be set higher than that of Max-R/T_V even **before** the learner is presented with any evidence. Thus, even if the training data drawn from the native Cantonese vocabulary do not provide any direct evidence that the weight of a particular Faithfulness constraint should be altered, it will remain at a PMap-compliant weighting throughout learning.

4.5.1 Data

Both the training and the test data sets used for this experiment were **identical** to the training and test data sets used in experiments 1 and 2. This was done so that the results of all of the grammars could be compared directly in both a qualitative and a quantitative fashion. As was done for the previous two experiments, discussion is limited to an exemplar grammar.

4.5.2 Model parameters

As was done for experiment 2, above, two separate substantive biases were constructed for the present experiment. The first substantive bias was an **extreme** bias, which simply took the weights learned by the best-case grammar as the target weights to be used throughout learning. This was done in order to confirm whether the learner would maintain target weights set by a weight-based substantive bias throughout learning, regardless of whether this bias was PMap-compliant or not. The plasticity of each constraint was set identically, as was done in White's (2013) model, and was set at the relatively low weight of 1 in order to force weights to stay near their target weight. A sample of target weights for the DEP-V constraints discussed in the previous experiment are provided in (90), and the full weight bias is provided in Appendix B.

(90) **Sample weights for Dep-V constraints, extreme bias**

Constraint	μ	σ^2
DEP-V/#_#	0.000063	1.0
DEP-V/T_#	4.394505	1.0
DEP-V/T_T	6.877504	1.0
DEP-V/R_#	6.708101	1.0
DEP-V/R_T	5.275476	1.0
DEP-V/R_R	9.562827	1.0
DEP-V/#_R	1.263949	1.0
DEP-V/T_R	2.416692	1.0

The second substantive bias was one which aimed to approximate what a true PMap-style bias could look like for Cantonese, in much the same way as was done for experiment 2, above. This bias was constructed to confirm whether the weights of the constraints assumed to be governed by the PMap would, in fact, still predict the correct patterns of loanword adaptation in Cantonese. This bias was constructed such that those constraints discussed in section 4.4.2.1 received target weights which were equivallent to the **highest** weight in their respective category. The Markedness constraints received a target weight of 10, and all other Faithfulness constraints received a target weight of 0. The same plasticity value was assigned to all constraints, so that the only differences between them would be the target weights. As such, the plasticity value was set to 100, as was done for the previous experiments, so that the learner could more quickly arrive at a viable grammar for Cantonese. A sample of target weights for the MAX-S constraints is provided in (91), and the full bias is provided in Appendix B.

(91) **Sample weights for Max-S constraints, pseudo-PMap bias**

Constraint	μ	σ^2
Max-S/V_V	5.30073	100
Max-S/V_R	5.30073	100
Max-S/R_V	5.30073	100
Max-S/R_R	5.30073	100
Max-S/#_V	6.16105	100
Max-S/#_R	6.16105	100
Max-S/T_V	6.16105	100
Max-S/T_R	6.16105	100
Max-S/V_#	9.95487	100
Max-S/V_T	9.95487	100
Max-S/R_#	9.95487	100
Max-S/R_T	9.95487	100
Max-S/#_#	10.84688	100
Max-S/#_T	10.84688	100
Max-S/T_#	10.84688	100
Max-S/T_T	10.84688	100
*Complex	10.0	100
*Coda	10.0	100
Dep-T/V_V	0.0	100
Dep-R/#_T	0.0	100

Aside from these two priors, the procedure for constructing MaxEnt grammars for this experiment was **identical** to that used in previous experiments. Once the learner was trained on the 1000-item training sets and a set of weights was found for each, these grammars were then applied to the test set to predict probabilities for each candidate. The predicted probabilities are examined in the next section, and are compared to the best-case grammar and all other comparable grammars as needed, both qualitatively and quantitatively.

4.5.3 Results

4.5.3.1 Extreme Bias

The **accuracy** of the grammar acquired by a Max-Ent learner equipped with an extreme weight-based bias is, overall, quite high, as shown in Figure 4-16. In terms of the native vocabulary and phonotactically-licit loanwords, all are produced as entirely faithfully, with the candidate that received the highest probability receiving no repairs. With respect to the phonotactically-illicit loanwords, there is a distinct **asymmetrical** pattern evidenced, with different patterns of repair for consonant clusters, sibilants, and most importantly, word-final SC clusters. Furthermore, the candidates assigned the highest probability in these cases mirror the behaviour of the best-case grammar, with **deletion** being preferred for consonant clusters, **vowel epenthesis** being preferred for many of the loanwords with sibilants, and a **combination** of both overwhelmingly used to repair

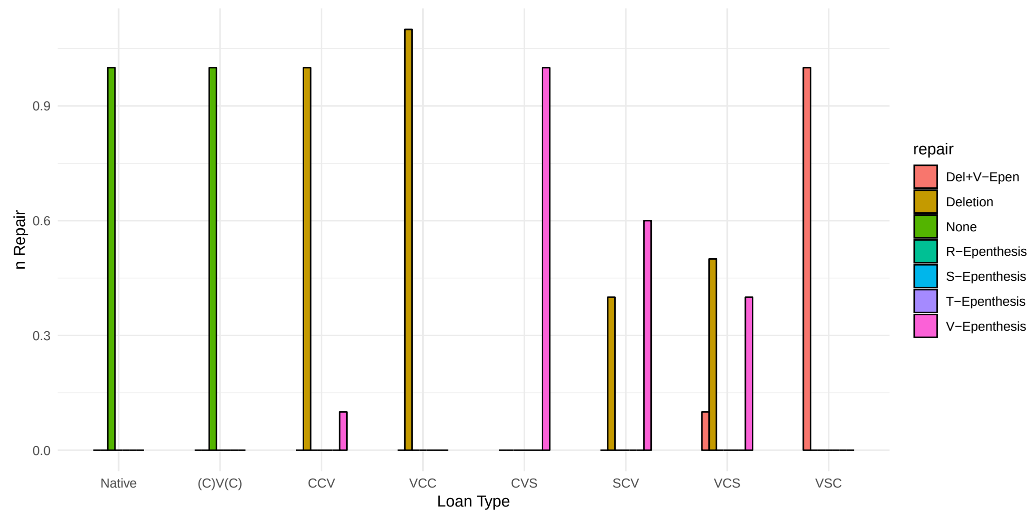


Figure 4-16: Proportion of winning candidates divided by repair and input structure category, representative sample of 1000 Cantonese lexical items with an extreme weight-based substantive bias

words with SC clusters.

Of course, the results displayed above are not a perfect match to the best-case grammar. Deletion is still considered to be a viable option for repairing consonant clusters which contain sibilants, although this is not the case for actual speakers of Cantonese. There is also a preference for onset consonant clusters to undergo deletion, as opposed to showing a 50-50 split between deletion and vowel epenthesis. A sample of repairs selected by the grammar in each category is presented below.

(92) **Sample repairs for a grammar trained on a 1000-item data set with an extreme weight-based bias**

Category	*Repair Type*	*Input*	*Gloss*	*Output*	*p*
Native	None	`o pej`	‘to fart’	`o pej`	0.98494
(C)V(C)	None	`e sej`	‘essay’	`e sej`	0.98291
CCV	Deletion V-Epenthesis	`klip` `plam`	‘clip’ ‘plum’	`kip` `pV lam`	0.61507 0.95119
VCC	Deletion	`sapt`	‘shaft’	`sat` `sap`	0.40223 0.40223
CVS	V-Epenthesis	`paas pot`	‘passport’	`paa sV pot`	0.94102
SCV	Deletion V-Epenthesis	`smak` `spaak`	‘smart’ ‘spark’	`sak` `sV paak`	0.60926 0.95318

VCS	Deletion
	V-Epenthesis
	Deletion + V-Epenthesis
		`o leonc`
		`aats`
		`inc`
			‘orange’
			‘arts’
			‘inch’
				`o leon`
				`aat sV`
				`n cV`
					0.52389
					0.94459
					0.35793

VSC	Deletion + V-Epenthesis	`tost`	‘toast’	`to sV`	0.86189

Interestingly enough, these repairs show that, even though the training set contained no instances of single-vowel syllables, the grammar was nevertheless able to acquire them. This can be attributed to the fact that the weight-based bias was **derived** from the best-case grammar. This grammar, as shown in section 4.3.3.3, was able to correctly predict that all native Cantonese words and phonotactially-licit loanwords should surface without any repairs. If the weights from the best-case grammar were sufficient to achieve this result, then they would be transferred over as target weights in the extreme weight-based bias. Given that there is no evidence in the training set to have re-weighted these constraints, they would then have remained at their target weights. These data alone prove that the grammar will maintain target weights in the absence of evidence, as hypothesized above. However, it is still worthwhile to examine other data, in order to obtain a better understanding of how this process affects the phonotactically-illicit loanwords.

Overall, the predictions for the phonotactically-illicit loanwords match those which are attested by native speakers of Cantonese. Both deletion and vowel epenthesis are attested repairs for onset clusters, while deletion is the only repair predicted for coda clusters. Sibilants in coda are always repaired by vowel epenthesis, and clusters which contain sibilants show an increase in vowel epenthesis repairs as well. Finally, codas which consist of an SC cluster always select a candidate which shows deletion of the final segment, accompanied by epenthesis of a vowel in order to prevent a sibilant from appearing in coda.

The biggest discrepancy between the attested loanword repairs and the predictions of the extreme weight bias grammar is that there is still a slight preference for **deletion**. For example, there is an overwhelming bias towards deletion in onset clusters. However, upon examination of these kinds of repairs, it was shown that the grammar only assigns them $\sim 60\%$ of the overall probability, as opposed to $\sim 90\%$ for other candidates in other categories. It may be that vowel epenthesis candidates receive more of the probability space on an input-by-input basis, and this will become apparent when the overall results are examined.

Also of interest for the loanwords with onset clusters is that while the current grammar prefers deletion, the best-case grammar prefers vowel epenthesis. This is ultimately due to the different **training sets** used when generating these grammars. The method by which the Cantonese training data were generated tends to bias models towards favouring deletion candidates, while the method of generating the test data tends to bias models towards favouring epenthesis candidates. The exact mechanism by which this result is obtained is discussed in Chapter 5, and helps to explain why deletion was favoured more for onset clusters involving sonorants in this experiment as well.

The aggregate results follow the accuracy results in showing the same asymmetries of repair types across various Cantonese words. Candidates which receive no repairs still receive the majority of the probability for native words and phonotactially-licit loanwords. Loanwords with onset clusters

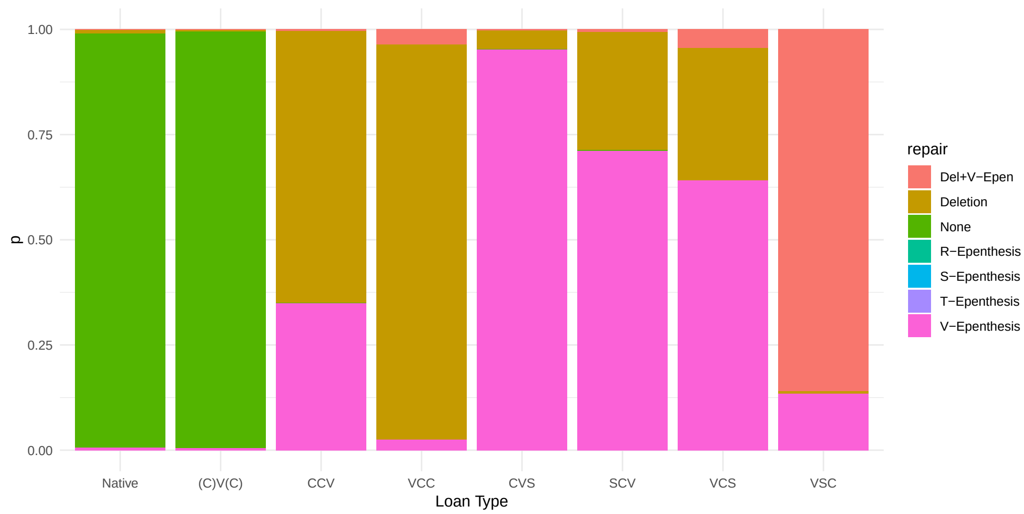


Figure 4-17: Probability of repair by input structure category, representative sample of 1000 Cantonese lexical items with a weight-based substantive bias

.

split the probability between deletion and vowel epenthesis, although deletion is slightly preferred here. Loanwords with coda clusters overwhelmingly prefer deletion only. Words with sibilants prefer vowel-epenthesis, although deletion candidates are more preferred for sibilants in clusters. Finally, a combination of both deletion and vowel-epenthesis is preferred for the coda SC cluster candidates – overall, in line with what is expected given attested patterns of loanword adaptation.

When these results are compared with the results derived from the best-case grammar, they are found to be incredibly similar ($D_{KL} = 0.0399562$). In contrast, when compared with the results found using the comparable null hypothesis grammar, they are found to be relatively different ($D_{KL} = 0.8852479$). Crucially, these results are also found to be extremely different to those of the extreme plasticity bias of the previous section ($D_{KL} = 1.29012$). It can thus be safely assumed that this manner of modelling a substantive bias is **successful** when predicting the behaviour of loanwords in Cantonese.

4.5.3.2 Pseudo-PMap Bias

When a pseudo-PMap bias was used when training a MaxEnt learner, the results obtained were not as accurate as they were for the extreme substantive bias used above, but they still showed evidence of an **asymmetrical** treatment of loanwords. As shown in Figure 4-18, there is a greater tendency for **deletion** to be considered a viable repair for all of the phonotactically-illicit loanwords, although it is not the only legitimate repair entertained by the pseudo-PMap grammar. In fact, vowel

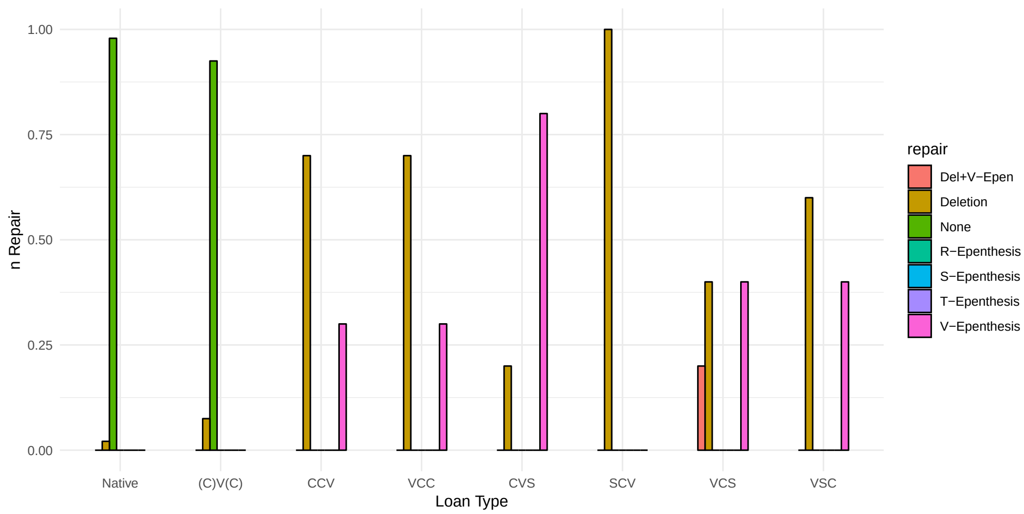


Figure 4-18: Proportion of winning candidates divided by repair and input structure category, representative sample of 1000 Cantonese lexical items with a weight-based PMap-style bias

epenthesis is considered viable for **all types** of phonotactically-illicit loanword with the exception of those which contain onset SC clusters. Combinations of deletion and vowel epenthesis are limited to loanwords with coda CS clusters – contrary to what is attested in actual loanword adaptation, where coda SC clusters are the only kinds of structures repaired in this way. In addition, deletion reappears as a repair strategy to native Cantonese words and phonotactically-licit loanwords, as was seen in the null hypothesis experiments.

However, in contrast to the results from experiment 2, this bias was still capable of treating the different classes of loanwords **asymmetrically**, as is done by native speakers of Cantonese. Furthermore, the vowel epenthesis was entertained as a repair to loanwords which contained sibilants in illicit positions more often than it was entertained as a repair to regular consonant clusters. While segment deletion was preferred overall, it was hypothesized in the previous section to be due to a bias towards deletion that arises through the training set. The fact that vowel epenthesis is chosen as a repair in this asymmetrical way for the current experiment, but that it is never chosen as a repair within experiment 2 indicates that the prior is behaving as expected here, and having some beneficial effect.

(93) **Sample repairs for a grammar trained on a 1000-item data set with a weight-based PMap-style bias**

Category	*Repair Type*	*Input*	*Gloss*	*Output*	*p*
Native	None	`naap seoj`	'tax'	`naap seoj`	0.99861
	Deletion	`o pej`	'to fart'	`pej`	0.53374
(C)V(C)	None	`kaa low lej`	'calorie'	`kaa low lej`	0.99929
	Deletion	`e sej`	'essay'	`sej`	0.53363
CCV	Deletion	`plam`	'plum'	`lam`	0.99823
	V-Epenthesis	`wo ljam`	'volume'	`wo lV jam`	0.79865
VCC	Deletion	`si mant`	'cement'	`si man`	0.99889
	V-Epenthesis	`sapt`	'shaft'	`sap tV`	0.85463
		`singk`	'sink'	`sing kV`	0.93453
CVS	Deletion	`paas pot`	'passport'	`paa pot`	0.94934
	V-Epenthesis	`paas`	'pass'	`paa sV`	0.76081
SCV	Deletion	`spaak`	'spark'	`paak`	0.82832
VCS	Deletion	`aats`	'arts'	`aat`	0.77798
	V-Epenthesis	`zins`	'jeans'	`zin sV`	0.96845
	Deletion + V-Epenthesis	`o leonc`	'orange'	`leon cV`	0.85004
VSC	Deletion	`laast`	'last'	`laat`	0.95150
	V-Epenthesis	`tost`	'toast'	`to sVt`	0.91276

The repairs presented in (93) show that many of the same tendencies present in the null hypothesis grammars have returned. For example, onset consonant clusters often show deletion of the **first** member of the cluster as opposed to the second, as is the case for [ˈpʰlʌm] 'plum'. This is likely due to the fact that not all Faithfulness constraints were assigned weights that could be considered to be compatible with the PMap, and instead defaulted to a Markedness ≫ Faithfulness bias. As use of a Markendess ≫ Faithfulness bias in experiment 1 was shown to lead to these kinds of repairs, it is unsurprising that some of the same repairs are selected here. This could indicate either that the weight of the Faithfulness constraints within the bias was not high enough, and should have been ascribed some value other than 0, or that the weight of the Markedness constraitns was too high. Furthermore, deletion is considered a more viable option for repair overall, much as was the case for the null hypothesis grammars. Discussion of this result is reserved for Chapter 5.

What is interesting about this grammar is that it shows evidence of **more sophisticated** analyses than those entertained by native speakers of Cantonese. For example, vowel epenthesis in onset clusters seems to be restricted to only those loanwords which are **multisyllabic**. This is quite mysterious given the nature of the bias used during training – while Max-T/V_R would have received a weight around 6 by the bias, the weight of Max-T/#_R should have received a much **higher** weight of around 11. Max-R/T_V, in contrast, should have received a moderate weight of around 3. A similar concern arises for the onset SC clusters, where the sibilant is uniformly deleted. The constraint governing its deletion, Max-S/#_T, should also have received a high target weight in the bias. It may be assumed that training on the native vocabulary causing the weights of these constraints to be **adjusted** in some way, although why they are adjusted in this manner is unknown at present.

Also interesting is the behaviour of coda clusters, where vowel epenthesis is considered a viable option for coda clusters of level sonority, such as *sap tV* 'shaft', or of velar consonant sequences, such as *sing kV* 'sink', but no others. This could have arisen through an interaction between the phonotactic constraints and the Faithfulness constraints – after all, there are very few phonotactic constraints which make reference to the velar consonants, while there are a great many which impose restrictions on the distributions of coronals and labials.

With respect to coda sonorants, deletion seems to be the repair preferred for **multisyllabic** loanwords, while vowel epenthesis remains the norm for all others. This indicates that there could be an asymmetry in the weights of the constraints MAX-S/V_# and MAX-S/V_T. In a similar vein, the asymmetry in the behaviour of coda CS clusters also hints at an asymmetry among the MAX-S constraints. Items which contain a sequence of obstruents, such as [ˈɑːts] 'arts', appear to resolve the cluster by deleting the sonorant; while those which contain a nasal followed by an obstruent, such as [ˈd͡ʒiːnz] 'jeans', resolve the cluster via epenthesis. This is likely due to an asymmetry in the weight of constraints MAX-S/T_# and MAX-S/R_#. This is of particular interest because all of the constraints mentioned here would have been assigned **identical** weights by the pseudo-PMap prior.

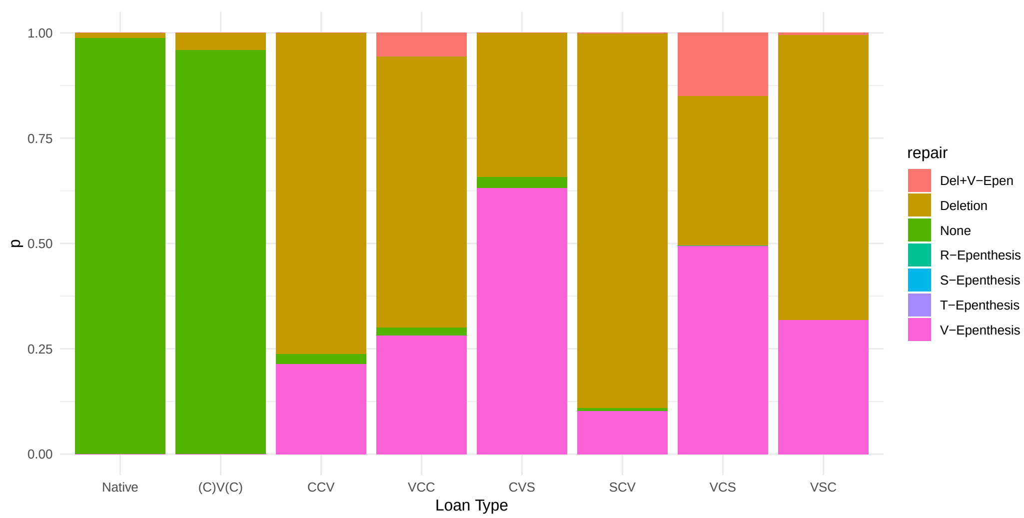


Figure 4-19: Probability of repair by input structure category, representative sample of 1000 Cantonese lexical items with a weight-based PMap-style bias

.

Finally, with respect to loanwords which contain a coda SC cluster, the choice between vowel epenthesis and sonorant deletion appears to be largely based on the nature of the preceding vowel.

Only those words with a preceding [o] show vowel epenthesis, while those with other vowels show deletion of the sonorant. As is the case for other coda clusters, there could be some interaction with the phonotactic constraints which is driving this difference.

Examination of the aggregate results for all candidates in the test set backs up the general trend towards positing asymmetrical but not-always-accurate repairs. An increase in the viability of vowel epenthesis as a loanword repair is seen overall, with the highest proportion of probability assigned to vowel epenthesis candidates occurring for single-sonorant codas, and the next highest occurring for coda clusters also containing sonorants. Deletion is still the preferred repair for other phonotactically-illicit loanwords.

In terms of KL-Divergence, this grammar receives a value of $D_{KL} = 1.02762$ when compared with the results of the best-case grammar from experiment 1. This is quite a high value, indicating that the model is dissimilar to the best-case results. This grammar also receives a value of $D_{KL} = 0.48436$ when compared with the results of the comparable null hypothesis grammar, indicating that these results are moderately similar. Finally, with respect to the pseudo-PMap plasticity bias of the previous section, this grammar receives a value of $D_{KL} = 0.51693$, indicating that this grammar is only slightly more dissimilar to one trained using a comparable plasticity-based bias.

These results indicate that, while the weight-based pseudo-PMap bias has helped the MaxEnt learner to maintain some asymmetries that can begin to account for attested patterns of loanword repair, a more nuanced model must still be constructed. The overall preference for deletion could indicate that some of the Faithfulness constraints assumed not to contribute to the results of the best-case grammar should, in fact, have been hypothesized to contribute. Alternately, it could indicate that the Markedness constraints are interacting with the Faithfulness constraints in unexpected ways, indicating that their weights are too high. In addition, the evidence from the native vocabulary appears to cause the weights of PMap-governed constraints to be **adjusted** in unexpected ways. This, in turn, leads to asymmetries in the weighting of faithfulness constraints which yield unexpected and overly-sophisticated patterns of loanword adaptation. This may be due to a variety of factors, to be discussed in greater detail below.

4.5.4 Discussion

The models generated as part of experiment 3 have confirmed that making use of a **weight-based** substantive bias will allow a MaxEnt learner to successfully arrive at a grammar which can predict attested patterns of loanword adaptation in Cantonese. This is shown especially well by the grammars which used an **extreme** substantive bias, outlined in section 4.5.3.1. The results obtained by the grammars which used a pseudo-PMap model also show an asymmetry in repairs, with vowel epenthesis being preferred for loanwords with sibilants more often than loanwords with consonant clusters. While the results obtained by these grammars were **closer** to what is observed for actual speakers of Cantonese, they also showed evidence that the analysis arrived at by the grammar was **overly complex**, and were influenced to a much greater degree by the influence of the phonotactic constraints.

There are a variety of reasons why this discrepancy among the two substantive biases might have occurred. The discrepancy may be at least partially the fault of the **constraint set used**, which is extremely granular, and allows for a greater degree of distinctions to be made among Faithfulness constraints than is perhaps necessary. However, a grammar that is trained using a more extreme bias does not show such dramatic evidence of these issues, as shown in section 4.5.3.1.

Two alternate reasons why the results from grammars trained on these two biases differed may be due to the way in which the biases were constructed. The extreme bias used **only** the trained weights from the best-case grammar as the target weights, while the pseudo-PMap bias used a **mix** of weights from the best-case grammar and a more general Markedness $\gg$ Faithfulness bias. However, the way in which the Markedness $\gg$ Faithfulness bias was instantiated may have been **inconsistent**. This is because while all Markedness constraints received a target weight of 10, some Faithfulness constraints received even **higher** weights. For example, the target weight of MAX-S/T_T was assigned around 11, meaning that the prior would have broken the general Markedness $\gg$ Faithfulness principle in this case. This may account for why the results of the grammars trained with a pseudo-PMap bias were so sensitive to phonotactic constraints.

The other reason for the discrepancy may lie in the **plasticities** used for each set of experiments. The prior used for the extreme bias experiments used a uniform plasticity of 1, while the prior used for the psuedo-PMap bias used a prior of 100, as was done for the Markedness $\gg$ Faithfulness prior. This may have allowed the constraints in the pseudo-PMap experiments **too much** freedom to deviate from their target weights, resulting in the much larger differences between the weights of constraints that belonged to a single class.

With all caveats about the behaviour of the pseudo-PMap grammars aside, the results from these experiments were nevertheless much more in line with what is expected given attested patterns of loanword adaptation. Indeed, they are substantially closer to what is attested than the results of grammars trained using either a simple Markedness $\gg$ Faithfulness prior, or a PMap-style prior encoded as asymmetrical constraint plasticities. In this respect, Thus, encoding a substantive bias via asymmetrical target weights is the best option for modelling how loanwords are adapted.

4.6 Discussion & Conclusion

The main findings of the experiments run in this Chapter with respect to how substantive biases should be encoded are as follows:

(94) **Main findings of this Chapter with respect to substantive biases**

a. A learner trained **only on native Cantonese words** which has **no substantive bias** only predicts **deletion** within loanwords, contrary to what is observed (section 4.3).

b. A learner trained on native Cantonese words which has a **plasticity bias** also only predicts **deletion**, and fails to capture the observed loanword data (section 4.4).

c. A learner trained on native Cantonese words which has a **weight bias** succeeds in modelling at least some of, if not all of, the attested loanword patterns of Cantonese (section 4.5).

From these results, it can be inferred that modelling a substantive bias such as the PMap should be done by **varying the target weights** of the relevant constraints, rather than varying the plasticities of these constraints. Regardless of how the bias itself is acquired, or whether the bias is entirely the result of the PMap, the information must be **encoded** in this way in order to accurately capture the behaviour of loanword data in the **absence of overt phonological evidence**.

This does not necessarily mean that a substantive bias encoded via plasticity will always fail – as shown in Wilson (2006), this method of encoding a bias works well when modelling a system where **phonological evidence is overt**. It also appears to capture how we believe the PMap is acquired. Since the PMap is not necessarily identical from language to language, it can be inferred that it is **built up over time**. Encoding the PMap by varying plasticities and by assigning all constraints a uniform target weight is one way of enforcing this idea. As will be discussed in the following chapter, presence of **any overt evidence** that a particular constraint needs to be re-weighted can still lead to a PMap-compliant weighting of constraints, indicating that this method of encoding may succeed if a different model of loanword adaptation is used. Nevertheless, it is not the correct encoding for the way in which loanword adaptation has been assumed to proceed throughout this thesis.

With respect to whether the biases used are actually consistent with the PMap, the results are somewhat mixed. A summary of these results is presented in (95).

(95) **Main findings of this Chapter with respect to the PMap in Cantonese**

a. The weights learned by a model trained on the loanword data are in some respects **compatible** with what is known of the PMap in the following ways (section 4.4.2.1):

- The Dep-V constraints pattern as expected from Yun (2016), in that environments which exhibit a rise in intensity lead to lower Dep-V weights, and environments which do not lead to higher weights.
- The overall weight of the Max-S constraints is higher than the weight of other Max-C constraints, as expected from Wright (2004).

b. The weights learned by a model trained on the loanword data are in some respects **contrary** to what is know of the PMap in the following ways (section 4.4.2.1):
   - There appears to be a preference for weighting constraints which make exclusive reference to left-edge cues higher than those which make exclusive reference to right-edge cues, contrary to Wright (2004).
   - Max-R and Max-T constraints receive roughly equivalent overall weights, instead of Max-R constraints being weighted higher.
   - Constraints that make reference to cue-rich environments, such as Max-T/V_V tend to receive unusually low weights.

c. Even under ideal encoding conditions, a true PMap-style bias only **approximates** attested repairs (section 4.5.3.2)

While an attempt at reasoning about what the PMap for Cantonese may look like was undertaken as part of these experiments, it should not be considered a complete model of how the PMap behaves in this language. While some tendencies with respect to the inequality of weight assignment are compatible with what is understood about the PMap at the present time, more minute differences in weighting often **do not emerge** or actively **subvert** what is expected. Ideally, when more is known about the way in which Cantonese speakers compute differences in perceptibility, or when more is known about how various cues are integrated in order to arrive at an approximation of the difference in perceptibility, a better model of the PMap can be constructed and tested against the same set of data. It is expected that regardless of the result of this improved model of the PMap, encoding it through differential assignment of target weights will lead to a better model of the Cantonese loanword data than encoding it through differential assignment of plasticities.

The experiments undertaken as part of this Chapter also make a certain series of **assumptions** about the composition of the native Cantonese grammar. For instance, all experiments described here make use of a **single set** of constraints, which are extremely granular in their definition and can be independently weighted. While this allows the results of these experiments to be more easily understood, they have also been shown to lead to unnecessarily complicated grammars, and may not be the way in which speakers conceive of the PMap to begin with. What might these grammars look like if a **different** constraint set were used? How might the PMap be **encoded** for these grammars? Are there **alternate** ways of encoding the PMap for the constraint set used here that might help solve this issue? I will turn to discussion of these questions and others in the following chapter.

Chapter 5

Discussion

The focus of this thesis thus far has been on the **adequacy** of various models of PMap encoding for MaxEnt-based phonological learners. All methods explored here choose to encode the PMap or other substantive bias as a **separate condition** on the outcome of learning in the form of a **Gaussian distribution**. A separate Gaussian distribution is created for each constraint, characterized by a mean, or **target weight**, and a variance, or **plasticity**. The target weight established by the Gaussian represents the **ideal** weight of the constraint, and the **plasticity** represents how far a constraint weight can differ from this ideal before receiving too high a penalty. The previous chapter explored two extant methods of encoding the PMap: one which assigned asymmetrical **plasticities** to constraints governed by the PMap, *à la* Wilson (2006); and one which assigned asymmetrical **target weights** to the constraints governed by the PMap, *à la* White (2013). It was found that those learners which made use of a weight-based bias fared better when tasked with modelling loanword adaptation in Cantonese than did learners that made use of a plasticity-based bias.

However, as mentioned often in the previous chapter, the simulations run were in may ways **dependent** upon choices made when stetting up an appropriate experimental protocol. These choices, in turn, had significant effects on the results obtained, leading to **mismatches** between attested loanword data from Silverman (1992); Yip (1993, 2006) and Kenstowicz (2012) and model predictions, even in the most successful cases. For example, the results of the best-case grammar from Chapter 4 – which was meant to model the best possible result that could be achieved by the MaxEnt-derived grammars by training a learner on the test data – nevertheless showed discrepancies when compared with attested loanword data. Rather than showing an even split between deletion (*e.g.*, producing [kʰíːp] for English [kʰlɪp] 'clip') and vowel epenthesis (*e.g.*, producing [kʰìː líːn] for English [kʰliːn] 'clean') repairs to onset clusters, the best-case grammar showed a much stronger bias towards **vowel epenthesis**. When this grammar was used to create a substantive bias in experiment 3, the resulting grammar trained on native Cantonese words alone showed a bias in the opposite direction for these same words, towards **deletion**.

The present chapter aims to examine some of these discrepancies in more detail, and to shed light on what aspects of model construction may have led to them. In some cases, these discrepancies are the result of issues that appear to be **inherent** to the experimental procedure adopted, such as the candidate **generation procedures** and the **sensitivity** of the constraints used for evaluating candidates. Other causes of discrepancies or potential model underperformance are more easily remediable and keep the experimental procedure from Chapter 4 largely intact. It was mentioned often in previous chapters that a set of constraints that made use of **overlapping** contexts was also a viable option when constructing a constraint set – one that was, in fact, undertaken by Wilson (2006). This option will be explored further in this chapter by running three new MaxEnt learners equipped with such a constraint set.

Of course, other potential improvements are also possible. The set of Markedness constraints used was, in some ways, quite small, and made use of **broadly-applicable** constraints. The effects that a larger set of more narrowly-defined Markedness constraints could have on a learner acquiring Cantonese will also be discussed here. Each MaxEnt learner was also run only **once** per training set, but this may have been too simplistic a model for how loanword adaptation proceeds for generations of actual speakers. As such, the potential effects that an **iterative** model of learning could have on the simulated grammars and the priors used to create them will also be discussed. Finally, a method by which a **third kind of prior** could be used to better match Steriade's (2001) classical OT-based model of the PMap will be outlined.

5.1 Procedural Anomalies

The more persistent discrepancies between attested Cantonese loanwords and loanword repairs predicted by the MaxEnt grammars in Chapter 4, such as the overall preference for **deletion** and the **uncertainty** about which of two obstruent consonants to delete, can both be attributed to choices made about how to generate candidates. While changes to these generation procedures could be instantiated in future models, doing so would require the current generation script to be substantially re-worked. As such, this section aims to note how each discrepancy may have occurred, and how they may be avoided in the future. It will also discuss some of the other difficulties that arise from the way in which experiments were run, such as the granularity of the constraints used.

5.1.1 Biases in Results

One of the more surprising discrepancies noted in Chapter 4 was that of the influence of the **candidate generation procedure** when training MaxEnt learners. While those grammars that were trained on data sets which had a strict "one change per candidate" generation procedure favoured **deletion** across the board, those which were trained on data sets which allowed a selection of two-change candidates favoured **vowel epenthesis**. For instance, when examining the results of experiment 1, detailed in section 4.3, it was found that a MaxEnt learner equipped with a simple Markedness $\gg$ Faithfulness bias and trained exclusively on native Cantonese lexical items assigned the majority of the available probability space to **deletion candidates** in the test set, regardless of whether these candidates were expected to win given the attested patterns of loanword repair. In contrast, the best-case grammar – which was trained on the test set – was more heavily biased towards **vowel epenthesis** candidates in the test set.

These results were **unexpected**, especially given what was predicted to happen if the learner truly did not alter the weight of the relevant Faithfulness constraints at all. It was predicted that the grammar would be largely **agnostic** about which repair to prefer, and should choose repairs **randomly**. With respect to a preference for deletion or vowel epenthesis, then, it would be expected that **both** should be selected close to 50% of the time for **every** phonotactically-illicit loanword.

By examining the candidate generation procedures in detail, this discrepancy can be explained, although finding an adequate solution to the issue may prove somewhat difficult. Recall that the training sets which consisted only of native Cantonese lexical items were generated by a procedure which only considered candidates which exhibited **one change** from the fully-faithful candidate. The generation procedure is repeated in full in (96).

(96) **Candidates Generated for Native Cantonese Words**

- The **fully faithful** candidate, with any potential tonal information removed.
- All **single deletion** candidates, excluding tones.
- All **single epenthesis** candidates, where the epenthetic segment could be:
  - A non-sibilant obstruent, T
  - A sibilant obstruent, S
  - A sonorant, R
  - A vowel, V

This procedure meant that, for a lexical item such as *ze1 jyun3* 'to sigh or whine', the following candidates referencing either vowel **epenthesis** or vowel **deletion** would be generated.

(97) **Entries for *ze1 jyun3* 'to sigh or whine' which make reference to vowels**

input	output	p	Max-V/T_T	Max-V/T_R	Dep-V/#_T	Dep-V/V_R	···
ze1 jyun3	z jyun	0	1	0	0	0	···
ze1 jyun3	ze jn	0	0	1	0	0	···
ze1 jyun3	V ze jyun	0	0	0	1	0	···
ze1 jyun3	zV e jyun	0	0	0	0	0	···
ze1 jyun3	ze V jyun	0	0	0	0	1	···
ze1 jyun3	ze jV yun	0	0	0	0	0	···
ze1 jyun3	ze jyu Vn	0	0	0	0	1	···
ze1 jyun3	ze jyu nV	0	0	0	0	0	···

As should be evident from the table in (97), the overall **number** of candidates that contain an epenthetic vowel is much **larger** than those that show evidence of a deleted vowel. This means that, in a broad sense, the Dep-V constraints will have more opportunities to have their weights altered than the Max-V constraints. For the null hypothesis grammars, this means that while all constraints will ideally have a weight of 0, the Dep-V constraints will be more likely to have **higher** weights than the Max-V constraints, simply because there are more vowel epenthesis candidates to rule out. This remains true when other kinds of segments are examined, such as Dep-R and Max-R, or Dep-T and Max-T. Those candidates which contain segment deletions will always be **fewer** in number than those which contain epenthetic segments.

A quick calculation of the number of candidates able to be generated for any given input will reveal why this is guaranteed to be the case. Say that an input consists of n non-tonal segments. Any given single-change deletion candidate will be identical to the input, with the exception that the i^{th} segment is not present in the output. We can therefore say that the number of single-change deletion candidates produced by the procedure in (96) is also equivalent to n – one for each segment that could be deleted. These n deletion candidates will likewise be **divided up** among the various types of Max constraints, depending on how many of each kind of segment appears in the fully-faithful candidate.

However, when the epenthesis candidates are considered, the formula changes. To make reasoning through the formula simpler, we will first consider only the vowel epenthesis candidates. Any given single-change vowel epenthesis candidate will be identical to the input, with the exception that **before** the i^{th} segment, there will be an epenthetic vowel. So, if the input is *ze1 jyun3*, and we are considering the third epenthesis candidate in our list of epenthesis candidates, it would be *ze V jyun*, since the epenthetic vowel is inserted before the third input segment. As above, this will give us n vowel epenthesis candidates. This will, however, not cover *ze jyu nV*, where the **final** segment is an epenthetic vowel. We must therefore add one to the count, so that there are a total of $n + 1$ vowel epenthesis candidates.

Vowels are not the only segment considered for epenthesis, however. For every vowel epenthesis candidate generated, there will also be a sonorant epenthesis candidate, an obstruent epenthesis candidate, and a sibilant epenthesis candidate. There will therefore be, in total, $4(n+1)$ epenthesis candidates generated for every input that passes through the procedure in (96). In other words, there are overall roughly **four times** as many epenthesis candidates as deletion candidates. The total number of candidates generated is given by the formula in (98).

(98) **Proportions of candidates generated for training sets**

$$N_{\text{cand}} = \underset{\text{fully faithful}}{1} + \underset{\text{deletion}}{n} + \underset{\text{epenthesis}}{4(n+1)}$$

According to this formula, there will always be $3n+4$ more epenthesis candidates than deletion candidates for an input of length n. However, the exact imbalance between these two types of candidates will be determined by the segments present in the input, and it will not be the case that there are always roughly four times as many vowel epenthesis candidates as vowel deletion candidates for every input. However, even when examining the **lower limit** of this difference for each type of segment considered, there will still be more epenthesis candidates than deletion candidates.

Say, hypothetically, that there exists an input that consists only of **vowels**. Thus, every deletion candidate for this input will be assessed by the MAX-V constraints, and they will receive n chances to have one of their weights altered in the final grammar. However, when considering vowel epenthesis candidates, the generation procedure will always construct $(n+1)$ new candidates, meaning the set of DEP-V constraints will receive $(n+1)$ chances to have one of their weights altered. This is schematized in (99).

(99) **Minimum difference between number of Max- and Dep-assessed candidates**

Input	*Deletion*	*V-Epenthesis*
$v_1v_2\cdots v_n$	$v_2\cdots vn$	$Vv_1v_2\cdots v_n$
	$v_1\cdots v_n$	$v_1Vv_2\cdots v_n$
	$\cdots$	$\cdots$
	$v_1v_2\cdots v_{n-1}$	$v_1v_2\cdots Vv_n$
		$v_1v_2\cdots v_nV$
Total	n	$n+1$

Thus, it is guaranteed that there will always be at least one more chance to alter the weights of the DEP constraints than the MAX constraints for any given segment type. This, in turn, leads to the conclusion that the DEP constraints will have, overall, a **higher** weight than the MAX constraints, leading to a preference for deletion.

While this is an outcome that arises naturally out of the candidate generation procedure combined with the MaxEnt algorithm, it is a somewhat **unnatural** idea when it comes to human learning of a first language. After all, it would be odd if a language cared about the exact number of potential epentheses and deletions, as it is odd for languages to care about exact numbers at all. In fact, most modern grammatical frameworks attempt to write out any notion of counting altogether (Prince and Smolensky, 1993, *a.o.*). A potential solution to this issue would be to **limit** the number of candidates considered in some principled way, such as is done for error-driven learning algorithms (Tesar and Smolensky, 2000; Boersma and Levelt, 2000; Hayes, 2004; Prince and Tesar, 2004; Magri, 2012, *a.o.*). This would also require a **stochastic** or **gradual** version of the MaxEnt algorithm to be instantiated, rather than the **batch** learners used here, an undertaking which will require a not insignificant amount of additional effort.

Now that it is understood how the majority of the grammars in Chapter 4 have acquired a bias towards deletion candidates, it remains to be seen how the best-case grammar has arrived at a bias towards vowel epenthesis. Unfortunately, a simple difference between the number of candidates that contribute towards the re-weighting Max versus those that contribute towards the re-weighting of Dep constraints will not be possible in this case, as there is still predicted to be an asymmetry between them in favour of the Dep constraints. The reasoning behind this claim is as follows:

The best-case grammar was derived by using the **test set** as the training set. However, the generation procedure for the test set differed in one crucial aspect from the other training sets in that it included all candidates which contained both **one deletion** and **one vowel epenthesis**. This generation procedure is repeated in full in (100).

(100) **Candidates Generated for English Loanwords**

- The **fully faithful** candidate, with any potential tonal information removed.
- All **single deletion** candidates, excluding tones.
- All **single epenthesis** candidates, where the epenthetic segment could be:
  - A non-sibilant obstruent, T
  - A sibilant obstruent, S
  - A sonorant, R
  - A vowel, V
- All combinations of **one deletion** and **one vowel epenthesis**.

This procedure would mean that input items like *soeng5* 'superior' would end up with the following list of vowel epenthesis and deletion candidates:

(101) **Entries for *soeng5* 'superior' which make reference to vowels**

input	output	p	Max-V/T_R	Dep-V/#_T	Dep-V/T_V	Dep-V/V_R	Dep-V/R_#	···
soeng5	sng	0	1	0	0	0	0	···
soeng5	V soeng	0	0	1	0	0	0	···
soeng5	sV oeng	0	0	0	1	0	0	···
soeng5	soe Vng	0	0	0	0	1	0	···
soeng5	soe ngV	0	0	0	0	0	1	···
soeng5	V oeng	0	0	0	0	0	0	···
soeng5	oe Vng	0	0	0	0	1	0	···
soeng5	oe ngV	0	0	0	0	0	1	···
soeng5	V sng	0	1	1	0	0	0	···
soeng5	sVng	0	1	0	0	0	0	···
soeng5	s ngV	0	1	0	0	0	1	···
soeng5	V soe	0	0	1	0	0	0	···
soeng5	sV oe	0	0	0	1	0	0	···
soeng5	soe V	0	0	0	0	0	0	···

As should be evident from the list above, calculating exactly how many candidates show epenthesis or deletion of a vowel is more complicated for this generation procedure. The above example will be run through in detail, in order to understand how the procedure is functioning on a small scale, moving through each section in the table in turn. Then, the more general calculation of the difference between numbers of epenthesis and deletion candidates will be undertaken.

Of course, as is the case for the generation procedure used for the training sets, there will always be as many single-change deletion candidates as there are segments in the word. When limiting our attention to vowels in the lexical item *soeng5* 'superior', there is only **one** – *oe* – so there will only be one single-change candidate which will contribute any change in weight to at least one of the MAX-V constraints. Similarly, there will always be $n + 1$ single-change epenthesis candidates that involve vowels. Since the lexical item *soeng5* 'superior' consists of three segments – *s*, *oe*, and *ng* – there will be **four** single-change vowel epenthesis candidates, and four candidates which can contribute to any change in weight in the DEP-V constraints. Thus far, the number of candidates which could influence the weight of the MAX-V constraints is **one**, and the number of candidates which could influence the weight of the DEP-V constraints is **four**.

We must now turn our attention to the dual-change candidates, which will consist of a single segment deletion plus a single vowel epenthesis. There are three potential single deletions, as discussed above, which will lead to three possible two-segment candidates – *oeng*, *sng*, and *soe*. The two of these which contain vowels – *oeng* and *soe* – will contribute nothing towards the weight of the MAX-V constraints, as the input vowel is still present. Likewise, any candidates which insert a vowel into these strings will not contribute anything to the weight of the MAX-V constraints. They will, however, **exclusively** contribute to the weight of the DEP-V constraints – once per epenthesis. This will occur **three** times each for *oeng* and *soe*, as each consists of two segments and will therefore have three possible epenthesis sites. At this point, there is still only **one** candidate which could contribute to the weight of the MAX-V constraints, and **ten** which could contribute to the weight of the DEP-V constraints.

However, we have still not considered the dual-change candidates where the vowel **is** deleted, those which take as a base *sng*. As above, there will be exactly **three** epenthesis candidates which can be generated using this base, meaning there will be three additional chances for the weights of the DEP-V constraints to be altered. However, there will **also** be three candidates which will contribute to the weight of the MAX-V constraints, as all candidates generated using this base will have the original vowel deleted. Thus, the total number of candidates which could contribute to the weight of the DEP-V constraints is **thirteen**, while the number of candidates which could contribute to the weight of the MAX-V constraints is **four**.

In a more general sense, this means that, in addition to the numbers obtained for the single-change generation procedure, there will always be n^2 additional dual-change candidates created through the procedure outlined in (100). This is derived by taking the total number of deletion candidates – previously shown to be n – and **multiplying** it by the number of vowel epenthesis candidates that can be generated from each deletion candidate. As each deletion candidate is of length $(n-1)$, and the epenthesis procedure generates $(x+1)$ candidates, substituting $(n-1)$ for x gives $(n-1)+1=n$ vowel-epenthesis candidates per deletion candidate. As there are n deletion candidates, this gives n^2 new candidates generated under this procedure.

Of these n^2 new candidates, **all** of them will contribute to the weights of the DEP-V constraints **and** all of them will contribute to the weights of the MAX constraints as a whole, although only a selection of them will contribute directly to the weights of the MAX-V constraints. However, this still means that there will always be **more** candidates which will contribute to the weight of the DEP constraints, as the number of single-change epenthesis candidates still contributes to the overall total. These proportions are shown in (102), while the number of candidates which contribute to each set of Faithfulness constraints is given in (103).

(102) **Proportions of candidates generated for the test set**

$$N_{\text{cand}} = \underbrace{1}_{\text{fully faithful}} + \underbrace{n}_{\text{deletion}} + \underbrace{4(n+1)}_{\text{epenthesis}} + \underbrace{n^2}_{\text{deletion \textbf{and} epenthesis}}$$

(103) **Number of candidates per input which contribute to Faithfulness constraint weightings**

Fully faithful	Deletion	Epenthesis
1	n^2+n	n^2+4n+4

Subtracting the number of deletion-containing candidates from the epenthesis-containing candidates in (103) gives $3n + 4$ additional epenthesis candidates generated by the procedure used to create the test set. This is, in fact, **identical** to the difference in each kind of candidate generated by the single-change procedure. As such, it might have been expected for a learner trained on either set to prefer **higher weights** for the Dep constraints, and to keep the weights of the Max constraints relatively low, as there are always more chances to re-weight the Dep constraints. Such a difference in weighting will cause the resulting grammar to be biased toward **deletion**. Why, then, is the grammar trained on the test set not also a deletion-preferring grammar?

The answer lies in **which candidates** are deemed the winners in each set. After all, although the grammar has more chances to re-weight the Dep constraints, there is no guarantee that these constraints will always see an **increase** in their weight. Some candidates which are predicted to win, such as *pV lam* for input *plam* 'plum', would cause the grammar to more readily promote the Max constraints at the **expense** of the Dep constraints, as this winning candidate tells the learner that a vowel must be epenthesized in favour of retaining all input segments. Of the winning candidates in the test set, 22 preferred deletion, 38 preferred epenthesis, and the remaining 130 preferred neither. This accounts for the bias toward epenthesis as a repair, as the majority of the informative winners contained an instance of vowel epenthesis.

It is of note that the test set also contained 130 lexical items which preferred **no repair** at all. These items can be presumed to act similarly to the items present in the training sets, which also required no repairs, and would most likely contribute a bias towards deletion. The fact that a learner trained on the test set – with only 38 of the total 190 candidates favouring vowel epenthesis – was nevertheless able to acquire a bias towards vowel epenthesis speaks to the strength of the influence of overt evidence on the learner.

In sum, while it was shown that the **generation procedure** was likely the cause of the bias towards deletion in the **training sets**, this could not be used to explain the bias towards vowel epenthesis when a learner was trained on the test set. Instead, this was shown to be due to the difference in the overall **number of winners** that preferred deletion versus those which preferred epenthesis. While both of these kinds of candidates only contributed to a fraction of the test data, they nevertheless showed a strong influence on the resulting grammar. This indicates that the MaxEnt learner used for the experiments above was quite sensitive to the overt evidence presented to it throughout the learning process.

5.1.2 Identical Probabilities

Another persistent discrepancy between attested loanwords and the grammars generated through the experiments in Chapter 4 was that the grammars often could not discriminate between segments of **identical sonority**. The most persistent issue in this respect occurred with the loanword [ˈʃæft] ‘shaft’, where the MaxEnt grammars almost universally could not decide whether to produce output *sat* or the attested *sap*.

This was ultimately an error in the way the generation scripts applied the Max constraints to this item. Somehow, the generation script assigned both candidates **identical** violation profiles, meaning that no grammar could have differentiated them in any case. It is, at present, unknown whether this was an issue exclusive to the lexical entries included in the test data, or if it occurred elsewhere in the training data.

5.1.3 Granularity of Constraints

Finally, it must be noted that the constraint set used in the experiments above is, in some ways, too granular; and in others, not granular enough. As discussed at the end of Chapter 4, the extreme granularity of constraints may be to blame for the inability of the MaxEnt learner to adequately mirror patterns of loanword adaptation in the absence of an extremely detailed prior. This issue will be explored in section 5.2, below. However, as will be discussed here, the constraint set used is in some ways not granular **enough**, or makes divisions among constraints in the wrong places.

This issue was noted in section 4.3 when discussing how native speakers of Cantonese may be able to reason about contextual constraints that on the surface appear to be unique to English loanwords, such as Max-T/V_R. Presumably, no monosyllabic word of Cantonese would contain the VTR sequence necessary for this constraint to be evaluated, as it would require there to be a consonant cluster in coda position. In Cantonese, coda clusters are banned by highly-weighted *Complex. However, if the context were allowed to span multiple syllables, then many lexical items of Cantonese would meet the evaluation criteria of this constraint, as one syllable could host the obstruent as a coda and another could host the sonorant as an onset. If a learner were to produce some sort of deletion error when attempting to pronounce these kinds of lexical items, they would have direct evidence about how its weight should be adjusted.

There are a few issues with this line of reasoning. The first – as noted in Chapter 4 – is that the **articulations** of sonorants may not be identical across languages, even if they occur in the same contexts. For example, stops in the environment V_R in Cantonese are always **unreleased**, as they must exist as codas in order for this sequence to remain unrepaired. However, in English, depending on the syllable structure of the word or phrase being produced, stops in this environment may be unreleased or released (if in coda), or even **aspirated** (if in onset). Cantonese speakers will no doubt be able to hear this difference in articulation and use it to reason about where syllable/word boundaries should occur within a lexical item, loan or not.

This issue could, perhaps be amended if the constraint set were **expanded** to include constraints that are sensitive to the **featural identity** of the segments being deleted. This could be done either by including the set of IDENT constraints into the grammar, or by including a series of feature- or cue-specific MAX and DEP constraints (Pater, 1996b; Flemming, 2008, *a.o.*). While this would greatly increase the total number of constraints that a MaxEnt learner would be required to reason about and learn from, it would most likely improve model performance and be more in line with how we believe speakers make use of acoustic information. One reason this solution is not pursued here is because it would require much more time for the learner to arrive at a solution in each case.

The solution proposed above also has the advantage that making direct reference to features and/or cues would allow the training **data** to be vastly simplified. Say, for instance, that the learner, rather than attempting to reason about a set of MAX-T constraints, is instead provided with a series of constraints such as MAX[+s.g.], MAX[−son], MAX[coronal], or their cue-based equivalents, such as MAX[release burst], MAX[voicing], MAX[F_1 transition], *etc.* This learner would be able to make direct reference to the way the obstruent is articulated, rather than using the obstruent's position as a proxy. If, say, MAX[+s.g.] had received a high weight after learning from the native Cantonese vocabulary (a likely scenario, as aspiration is used to differentiate between types of onset obstruents), then a learner could use this high weight to predict that aspirated obstruents in onset clusters should be preserved. Crucially, the learner would not need to know the exact segmental context in order to arrive at this conclusion, and could presumably arrive at it by learning from **monosyllables alone**.

While this particular solution is attractive, it would require a much more detailed corpus of Cantonese over which to evaluate the proposed constraints. Even if these constraints were limited to MAX[feature] and DEP[feature] constraints, the corpus used for this study would not be able to accommodate them, as it was limited to Jyutping transcriptions. In the future, if a full featural matrix could be provided for each syllable and additional acoustic or articulatory information about Cantonese could be incorporated, a more adequate experiment could be run.

A second issue with the idea of using clusters resulting from syllable contact to "bootstrap" a learner into reasoning about syllable clusters is that, often, the set of admissible syllable onsets or codas is **not identical** to the set of admissible consonant clusters, especially if those consonant clusters are split across multiple words or syllables. In fact, the set of admissible clusters which occur through **syllable contact** is often much larger than the set of admissible onsets or codas. As such, any learner which is attempting to reason about admissible onset or coda sequences from sequences which occur as a result of contact may end up with a grammar that allows **too many** combinations of consonants.

This latter issue would require a much better understanding of syllable contact in Cantonese than is presently available. As Cantonese does not allow complex onsets or codas, the set of admissible onsets and codas is already extremely limited. It is unknown how a monolingual native speaker might identify or treat a complex onset upon first listen, especially if their entire experience with consonant clusters has been through syllable contact. This is unlikely to be solved by experimenting with different constraint sets, and requires a much better understanding of how Cantonese speakers perceive and judge consonant clusters.

5.1.4 Interim Conclusion

In sum, while many choices made while training the MaxEnt learner on Cantonese were either 1) requirements for the mechanization of the learning process, 2) errors, or 3) simplifications, it is hoped that future research will allow substantial improvements to be made to the experimental process. These improvements could be implemented by developing a different candidate generation procedure, or by making reference to a different and more nuanced set of Faithfulness constraints.

The following section, rather than attempting to introduce nuance into the set of Faithfulness constraints by increasing their number, attempts to introduce nuance by taking advantage of the **ganging-up effects** inherent to the MaxEnt framework, as discussed in Chapter 2. In other words, rather than making use of a set of purely independent Faithfulness constraints, constraints were allowed to **overlap**. As this led to a **reduction** in the number of Faithfulness constraints needed in order to characterize the data, and as it was a solution already considered in Wilson (2006), it was undertaken as part of this thesis.

5.2 Experiment 4: Alternate Faithfulness Constraints

All of the experiments undertaken in Chapter 4 made use of a set of Faithfulness constraints which could be **independently** weighted. This required each constraint to make reference to a **unique context**, so that no input sequence could be evaluated by multiple Faithfulness constraints. As such, each constraint was required to have both a **preceding** and a **following** context.

The advantages of using this constraint set were twofold. First, use of this constraint set eliminated **ganging-up** effects, which could have complicated analysis of the results. Second, use of this constraint set allowed for a more **transparent** understanding of how the grammar was treating segments in a variety of contexts. As the weight of each constraint could only be altered in highly specific environments, it meant that the underlying distribution of segments and contexts in Cantonese could be inferred directly from results. Likewise, as constraints were written so as to make reference to environments which were hypothesized to contribute most to segment perceptibility, it was easier to reason about how these highly specific environments could contribute to a particular segment's place within the PMap.

However, as discussed along with the results of experiment 3, presented in section 4.5 of Chapter 4, use of this constraint set was hypothesized to be one reason why the learner was able to acquire an overly **complex** pattern of loanword adaptation. The MaxEnt learner equipped with a weight-based PMap-style bias ended up making **more** divisions among loanwords than is attested. For instance, it posited a difference in the behaviour of **monosyllabic** versus **multisyllabic** loanwords, where SC clusters were repaired by vowel epenthesis in monosyllabic words but by sibilant deletion in multisyllabic words. This was assumed to be due to **asymmetric weights** of the constraints MAX-S/#_T and MAX-S/V_T, where the former constraint had a weight higher than DEP-V/S_T and the latter a lower weight. If a different constraint set were used – one which collapsed across contexts so that only the relevant portion of the context was present – then both kinds of words would contribute to the weight of this single constraint, and would receive the same analysis in the final grammar.

As mentioned above and in Chapter 4, collapsing across contexts in this way will lead to a grammar which **allows** ganging-up effects. This will lead to a grammar which is a bit more difficult to reason about, as the same candidate can force the weights of multiple Faithfulness constraints to be altered. However, as assumed by Wilson (2006), this could be an advantage for the model, provided the constraints overlap in a way that would make sense for the PMap. For example, as discussed in Steriade (2001) and Wright (2004), it is known that obstruents are best cued and most perceptible when located **intervocalically**, but can receive adequate cues from a single **following** vowel, and some cues from a single **preceding** vowel. They receive the fewest cues when not flanked by any vowels or sonorants.

(104) **Scale of cue robustness of obstruents by context** (Steriade, 2001)

V_V > _V > V_ > T_T

Of course, the intervocalic context can be considered to be a **combination** of the pre-vocalic and post-vocalic contexts, and both of these are known to contribute to a consonant's perceptibility in some way. If the constraint set were limited to MAX-T/_V and MAX-T/V_, any obstruent located between two vowels would be evaluated by **both**. Any obstruent deleted in this context would likewise **violate** both constraints, and would be penalized in proportion to each constraint weight. However, obstruents which occurred in a different context, such as simply pre-vocalic, would only be evaluated (and potentially penalized) by **one** of these constraints. These constraints should also receive **asymmetric** weights, as obstruents in prevocalic position are better-cued than in postvocalic position (Wright, 2004). Assuming this asymmetry, these two constraints would effectively re-create the full scale in (104), simply through constraint overlap.

(105) **Overlapping constraints leads to cue robustness scale for obstruents**

Contexts	Intervocalic V_V		Prevocalic _V		Postvocalic V_		Neither T_T
Constraints	MAX-T/_V & MAX-T/V_		MAX-T/_V		MAX-T/V_		—
Penalty	$w_{\text{prevoc}} + w_{\text{postvoc}}$	$>$	w_{prevoc}	$>$	w_{postvoc}	$>$	0

This solution does have one additional potential disadvantage when compared to the constraint set used in Chapter 4. While constraints which contain overlapping contexts can gang up to make sensible predictions about the behaviour of segments which occur in both, this overlap is assumed to be **linear**. That is, the overlap is assumed to be adequately modelled by simple **addition** of penalties incurred by each constraint. This is an assumption that is inherent to the MaxEnt model, and any model made in a framework which contains constraint overlap will behave in this way. It is unknown at the present time whether this matches how human speakers conceive of contextual overlap. After all, it may be the case that deleting an obstruent between two vowels incurs a penalty that is **in excess** of the sum of the penalties incurred by deleting an obstruent in other vocalic contexts. Likewise, it could also be the case that deleting an obstruent between two vowels incurs a penalty that is somehow **less than** the sum of the penalties incurred by deleting an obstruent in other vocalic contexts. Use of the larger, non-overlapping constraint set avoided this issue, as gang up effects were not possible with that constraint set.

The following experiment aims to test whether use of an overlapping constraint set will lead to more **uniform** (and potentially more **accurate**) results when modelling loanword adaptation in Cantonese. In order to make these experiments as comparable to those undertaken in Chapter 4, the same sets of training and test data were used when constructing these models, and the same general methodology was used. The main differences were in the set of Faithfulness **constraints** used, and the kinds of priors that could be generated using these constraints. First, the set of constraints will be discussed, followed by the procedures used for three separate experiments. The first two made use of a **weight-based** prior, derived from the best-case grammar for the current experiment and the best-case grammar from experiment 1, respectively. The third and final experiment made use of a **plasticity-based** prior. The data, model parameters, and results of each experiment will be presented, and a general discussion of the results will conclude.

5.2.1 Constraints

There are a few crucial differences between the constraint set used in experiments 1–3 and the experiment undertaken here. The most notable is the fact that constraints were written so that contexts were allowed to **overlap**. As opposed to fully specifying a context, as done in Chapter 4, contexts were limited to occurring either at the **left-edge** or **right-edge** of a target segment. The list of target segments is provided in (106), and remains largely unchanged from the target segments considered in Chapter 4. The list of contexts is provided in (107).

(106) **Classes of Target Segments for Overlapping Faithfulness Constraints**

a. **Obstruent** consonants, represented by T
b. **Sibilant** obstruents, represented by S
c. **Sonorant** consonants, represented by R
d. **Vowels**, represented by V

(107) **Classes of Contexts for Overlapping Faithfulness Constraints**

a. Context **segments**, consisting of:
   i. **Obstruent** consonants, represented by T
   ii. **Sonorant** consonants, represented by R
   iii. **Vowels**, represented by V
b. Context **positions**, consisting of:
   i. **Left-edge** segments, represented by X_
   ii. **Right-edge** segments, represented by _X

As contexts were defined with respect to the preceding or following segment, it was deemed unnecessary to include word edges as a separate contextual category. After all, a word-initial obstruent which is followed by a vowel can be assessed by the constraint MAX-T/_V, as can an obstruent which is preceded by another obstruent in coda position and followed by a vowel. The only difference between these two scenarios is that the second obstruent would **also** be able to be assessed by the constraint MAX-T/T_, while the first could not. In this way, these two cases can still be differentiated by the grammar without requiring reference to a word edge.

One other crucial difference between the way constraints were defined in Chapter 4 and the way they are defined here is in how **sibilant** obstruents are treated. In the present grammar, there are only the extremely general Max-S and Dep-S constraints, without their contextual variants. The contexts are provided through the Max-T and Dep-T constraint families, on the assumption that **all** obstruents utilize the same transitional cues when located next to vowels, sonorants, and obstruents. The general sibilant constraints are included in order to model the idea that the **internal cues** to sibilants are significantly stronger than other obstruents, as noted in Wright (2004).

Constraints were constructed in the following way: The context **segments** listed in (107a) were combined with the **positions** listed in (107b) to yield $3 \times 2 = 6$ different possible contexts. Each context was then applied to each **broad** segmental category listed in (106) – with the exception of sibilants – to give $6 \times 3 = 18$ Faithfulness constraints. One additional constraint was added to cover the set of sibilant obstruents as a whole, to give $18 + 1 = 19$ total target-context combinations. Finally, each combination was then used to construct one Max and one Dep constraint, for a total of $19 \times 2 = 38$ Faithfulness constraints. A sample set of constraints is provided in (108).

(108) **Schema for Overlapping Faithfulness Constraints**

$$\left\{\begin{array}{c}\textsc{Dep}\\ \textsc{Max}\end{array}\right\} - \left\{\begin{array}{c}\mathrm{T}\\ \mathrm{R}\\ \mathrm{V}\end{array}\right\} / \left\{\begin{array}{c}\mathrm{T}\\ \mathrm{R}\\ \mathrm{V}\end{array}\right\}\_ \;\cup\; \left\{\begin{array}{c}\textsc{Dep}\\ \textsc{Max}\end{array}\right\} - \left\{\begin{array}{c}\mathrm{T}\\ \mathrm{R}\\ \mathrm{V}\end{array}\right\} /\_ \left\{\begin{array}{c}\mathrm{T}\\ \mathrm{R}\\ \mathrm{V}\end{array}\right\} \;\cup\; \left\{\begin{array}{c}\textsc{Dep}\\ \textsc{Max}\end{array}\right\} - \mathrm{S}$$

which generates constraints such as:

- Dep-V/_T
- Dep-S
- Max-R/_T
- Max-T/R_
- *etc.*

The remainder of the constraints considered as part of this experiment were the same Markedness constraints used in previous experiments. A full list of constraints is included in Appendix B.

5.2.2 Data

As was the case for the experiments discussed in Chapter 4, the data used were split into a set used for **training** the MaxEnt models and a set used for **testing** how well the resulting grammars matched attested patterns of loanword adaptation. The training sets were **identical** to those used in the first half of experiment 1, discussed in section 4.3.3.1. Each consisted of a representative sample of 1000 lexical items from the native vocabulary of Cantonese, as represented by the content of the open-source CC-Canto dictionary (Pleco Software, 2016). The test set was also identical to that used in experiments 1–3. It consisted of a set of 190 items, 95 of which were drawn from the CC-Canto dictionary, and 95 of which were loanwords of various types drawn from examples provided in Silverman (1992); Yip (1993, 2006) and Kenstowicz (2012). Exact details of which kinds of words were included in the test set are provided in section 4.3.1.2 of Chapter 4.

The test data sets were passed through the same tableau generation procedures discussed in section 4.2.2 of Chapter 4, with the exception that the set of constraints used for evaluation was the one outlined in the previous section. All of the training sets were passed through the generation procedure which produced only single-change candidates, and the test set was passed through the generation procedure which generated additional deletion + vowel epenthesis candidates, in order to give the models a chance to predict the correct repairs for words ending in SC clusters. The resulting files were then read into the `solve.R` MaxEnt learner in order to run the appropriate models.

5.2.3 Model parameters

As the constraint set was changed from the one used in Chapter 4, none of the priors used in that chapter could be used for the present experiment. As such, **three** new priors were constructed – two of which were weight-based priors, as was done in White (2013); and one of which was a plasticity-based prior, as was done in Wilson (2006). The main focus was on the weight-based priors, as these were shown in the previous chapter to be able to predict **asymmetries** in preferred repair strategies among loanwords. Plasticity-based priors were shown to predict **uniform** repair strategies across the board, and so were not the focus here. However, one such prior was constructed in order to test whether the behaviour of a grammar equipped with this prior would show **improved** performance on a more compact set of constraints. After all, the learner will be able to adjust the weights of more constraints for each candidate in the training set, and the constraint set is reduced enough that these adjustments will not be spread out over a wide variety of nearly-identical contextual Faithfulness constraints. The learner may, then, be given more opportunities to learn a more PMap-compliant grammar, and the relative weakness of the plasticity-based prior may be at least partially due to the constraint set used.

The two weight-based priors were largely identical, except that the target weights of the Faithfulness constraints came from two different sources. The first source was the weights learned by a best-case grammar trained using the current constraint set. This was done in order to provide a parallel to the extreme prior tested in experiment 3, outlined in section 4.5.3.1 of Chapter 4. The results of this best-case grammar are presented in more detail in section 5.2.4, below. The second source was the weights learned by the best-case grammar from experiment 1 in Chapter 4. Since it is already known that the weights learned by this grammar can account for attested patterns of loanword adaptation, it was deemed a good starting point for a grammar with a condensed version of the constraint set used previously. This was meant to stand as a parallel to the pseudo-PMap bias tested in experiment 3, outlined in section 4.5.3.2.

The grammar trained using the first weight-based prior will be discussed in section 5.2.5. This prior is the one that is based directly on the weights obtained from the best-case grammar outlined in section 5.2.4. As such, it was constructed by using the weights from the best-case grammar as the **target weights** of the Faithfulness constraints in the prior. The target weights of the Markedness constraints were set at 10, in order to establish a Markedness ≫ Faithfulness bias. The plasticities of all constraints were set at 100, as was done for all priors which had a Markedness ≫ Faithfulness bias present. A sample of target weights set by this prior is provided below.

(109) **Sample target weights for overlapping Dep-V constraints, extreme bias**

Constraint	μ	σ^2
Dep-V/T_	2.51559	100
Dep-V/R_	8.30560	100
Dep-V/V_	1.84606	100
Dep-V/_T	0.90250	100
Dep-V/_R	1.67878	100
Dep-V/_V	1.45191	100

With respect to the weight-based prior based on the best-case grammar from Chapter 4, this was derived by **averaging** across all constraints which matched the context of one of the constraints in the present set. So, for example, the current constraint Max-V/_T would correspond to the constraints Max-V/#_T, Max-V/T_T, Max-V/R_T, and Max-V/V_T from experiments 1–3. Likewise, the current constraint Dep-S would correspond to the **entire family** of Dep-S constraints used in those experiments. Once each correspondence group was found, the weights of each constraint present in the group were averaged to find the target weight of the condensed constraint. As above, the target weights of the Markedness constraints were set at 10, and the plasticity of every constraint in the set was set at 100. A sample of target weights for the second weight-based prior is provided in (110).

(110) **Sample target weights for overlapping Dep-V constraints, averaged from previous experiments**

Constraint	μ	σ^2
Dep-V/T_	4.22284	100
Dep-V/R_	5.81124	100
Dep-V/V_	1.27883	100
Dep-V/_T	4.01173	100
Dep-V/_R	3.86859	100
Dep-V/_V	1.39974	100

This prior was meant to mirror the pseudo-PMap bias used in experiment 3, in order to better parallel the structure of that experiment here. However, rather than using the **highest** weight in each category, weights were instead averaged. This was done in order to **mitigate** any ganging-up effects that would occur with the present constraint set. After all, the pseudo-PMap bias used in experiment 3 already occasionally interfered with the Markedness constraints in unexpected ways, which was attributed to a high weight of key Faithfulness constraints. If constraints were allowed to gang up, this effect would be much more pronounced, and might cause the grammar to fail to repair phonotactically-illicit loanwords altogether.

An alternate solution to this issue lies in assigning each constraint in the condensed set the **lowest** corresponding constraint weight from the previous best-case grammar. This option was pursued by Wilson (2006) when constructing a bias for a set of overlapping PMap-governed constraints.[1] However, as many of the weights of the Faithfulness constraints were deemed to be unnaturally low through being **obscured** by highly-weighted Markedness constraints during the learning process, this was considered to be an inadequate solution for the present experiment. As such, the averaged weight was instead chosen, in order to give the highly-weighted independent constraints an opportunity to contribute to the weight of the condensed constraint without contributing so much as to interfere with or undo the effects of the Markedness constraints.

The above prior was used to create the plasticity-based bias used in the third experiment discussed here. As was done in experiment 2 from Chapter 4, the averaged target weights were **converted** into plasticities by squaring them and adding some small value to the result to prevent divide-by-zero errors in the `solve.R` script. A list of sample plasticities is provided in (111).

(111) **Sample plasticities for overlapping Dep-V constraints, averaged from previous experiments**

Constraint	μ	σ^2
Dep-V/T_	0.0	17.83236
Dep-V/R_	0.0	33.77053
Dep-V/V_	0.0	1.63540
Dep-V/_T	0.0	16.09399
Dep-V/_R	0.0	14.96596
Dep-V/_V	0.0	1.959283

Aside from the priors, the methodology used to construct each of the MaxEnt models for this experiment was identical to previous experiments. The `solve.R` MaxEnt learner objective function was paired with one of the Gaussian priors listed above, and then **trained** on each of the 1000-item sets of native Cantonese training data. Once training was complete, there existed a MaxEnt grammar for each combination of prior and training data set. A best-case model was also run, which used the test set as the training set in order to arrive at the set of weights which would best model the training data given the constraint set. Each grammar was then **tested** by applying it to the set

[1] As discussed in Chapter 3, this is not exactly the case. As Wilson (2006) made use of low-weighted Markedness constraints to model the PMap, he was instead forced to select the highest value in these cases. Translating his analysis to be applicable to Faithfulness constraints forces this selection preference to be reversed.

of loanwords and native Cantonese items and obtaining probabilities for each candidate therein.

These probabilities were then analyzed according to two rough measures: whether the grammar was **accurate** in assigning the most probability to the attested repair for each item in the test set; and how the grammar assigned probability among **all attested** repair types. The second measure was included in order to help assess how each grammar was treating loanwords which have multiple attested repairs. Finally, a quantitative measure of the **KL-Divergence** between the grammar's predictions and those of the best-case grammar was included, in order to assess the similarity between grammars. For reasons of space and clarity, only the results from the first 1000-word training set are presented here.

The remainder of this section will present and briefly discuss the results from the best-case grammar, followed by presentation and discussion of the MaxEnt grammar trained using the extreme weight bias. Then the results of the MaxEnt grammar trained using the averaged target weight bias will be examined, followed by a brief discussion of the grammar trained using the averaged plasticity bias. An overall discussion of the merits and downsides to using an overlapping constraint set will round out the section, before other potential alterations to the experimental protocol are discussed.

5.2.4 Best-Case Grammar

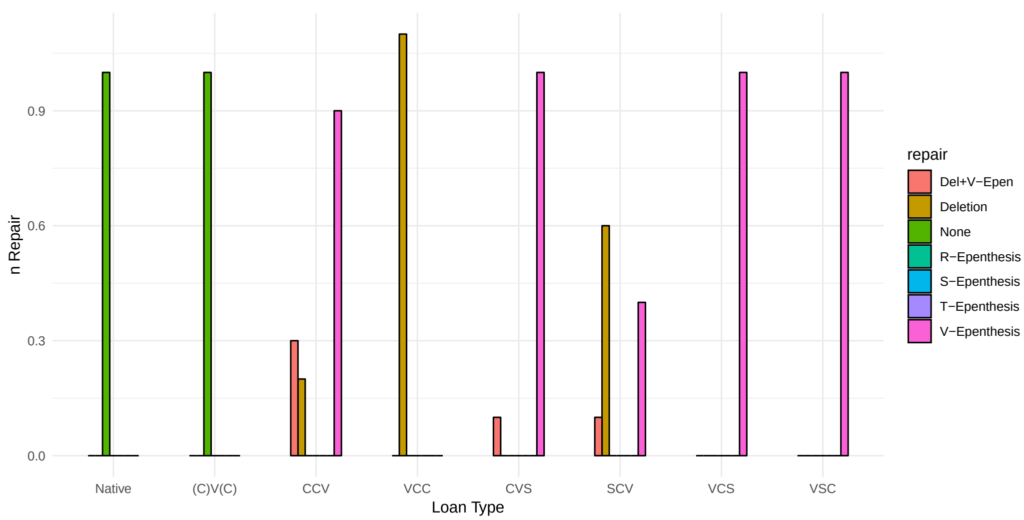


Figure 5-1: Proportion of winning candidates divided by repair and input structure category, trained on the test set with Markedness $\gg$ Faithfulness bias and overlapping constraints

The kinds of repairs associated with the candidates assigned the highest probability by the best-case grammar are shown in Figure 5-1. As with previous figures, each bar corresponds to one combination of loanword category and repair type, and is **proportional** to the number of items present in each loanword category. On occasion, the grammar will assign multiple candidates an **identical** probability, and so the proportion of "winning candidates" may be in excess of 1.0 for some categories.

Figure 5-1 shows that the best-case grammar using this more condensed constraint set is, overall, quite good at mirroring the attested patterns of loanword adaptation in Cantonese, However, it is still more **inaccurate** than the best-case grammar using the expanded constraint set of experiments 1–3. While the grammar is able to correctly learn the phonotactics and syllable structure of Cantonese – as evidenced by the lack of repairs present for all native lexical items and phonotactically-licit loanwords – it is not able to adequately mirror the behaviour of loanword adaptation for native speakers of Cantonese. Some aspects of the attested pattern are captured, such as the preference for deletion in coda consonant clusters, and the general preference for vowel epenthesis in loanwords which contain sibilants outside of onset position. However, the best-case grammar has failed to capture the behaviour of onset clusters, in that it prefers vowel epenthesis for simple onset clusters, but deletion for clusters which contain a sibilant. This is the reverse of what is attested for native speakers of Cantonese. Likewise, the best-case grammar has failed to predict that loanwords ending in SC clusters should show a **combination** of epenthesis and deletion – instead, vowel epenthesis is predicted to be the preferred repair here.

This grammar is also more **inconsistent** when compared with the best case-grammar that was trained on datasets which made use of the more expanded constraint set. Most loanword categories had at least one item about which the grammar was **indecisive**, as opposed to the persistent coda consonant issue mentioned in section 5.1.2 above. This is shown by the proportions in excess of 1 in Figure 5-1, and by the examples presented in the table below.

(112) **Sample repairs for the best-case grammar with a Markedness ≫ Faithfulness bias and overlapping constraint set**:

Category	*Repair Type*	*Input*	*Gloss*	*Output*	*p*
Native	None	`o pej`	'to fart'	`o pej`	0.97659
(C)V(C)	None	`e sej`	'essay'	`e sej`	0.97659
CCV	Deletion	`wo ljam`	'volume'	`wo jam` `wo lam`	0.47618 0.47618
	V-Epenthesis Deletion + V-Epenthesis	`so flej`	'soufflé'	`so fV lej` `so fV le`	0.38134 0.38134
VCC	Deletion	`sapt`	'shaft'	`sat` `sap`	0.34317 0.34317
CVS	V-Epenthesis Deletion + V-Epenthesis	`paas pot`	'passport'	`paa sV pot` `paa sV po`	0.48093 0.48093
SCV	Deletion	`spaak`	'spark'	`saak`	0.54180
	V-Epenthesis Deletion + V-Epenthesis	`swej`	'sway'	`sV wej` `sV we`	0.38129 0.38129
VCS	V-Epenthesis	`inc`	'inch'	`in cV`	0.74933
VSC	V-Epenthesis	`tost`	'toast'	`to sVt`	0.81472

As can be seen by examining the highest-probability candidates in detail, many of the odd instances where a combination of deletion and vowel epenthesis is attested outside of the VSC loanword category are those where a vowel is epenthesized at the correct position, such as between consonants in onset clusters such as *so fV lej* 'soufflé', but where the final **coda** consonant is deleted. This seems to be due to the **identity** of the preceding vowel, which is always a mid vowel which can be considered to be **bimoraic**. This indicates that the influence of the Markedness constraints is interfering in some ways with the expected constraint weightings. The set of Markedness constraints used throughout these experiments will be discussed further in section 5.3.

Aside from the unexpected preference for deletion in the words with onset SC clusters, this preference is fortunately accompanied by a preference for retaining the **sibilant**. Previous experiments have shown that grammars which prefer deletion in these contexts prefer instead to retain the segment closest to the vowel. This does indicate that treating the Max-S constraint as an additional condition to be imposed on top of the contextual Max-T constraints is having a beneficial effect in at least this case. The fact that the probability assigned to this form is close to 50% could indicate that a vowel epenthesis candidate is close behind, further bolstering this idea.

The aggregate results presented in Figure 5-2 seem to bear this out, with roughly half of the attested repairs for sibilant-containing onset clusters involving deletion and the other half involving vowel epenthesis. Similarly, the repairs to onset clusters of other types are split roughly half-half as well, although the preference is for simple vowel epenthesis in these cases. There is an overwhelming preference for deletion as a repair for coda clusters, while this preference switches to epenthesis for codas which contain a sibilant. Unfortunately, while all other categories fall in line with attested

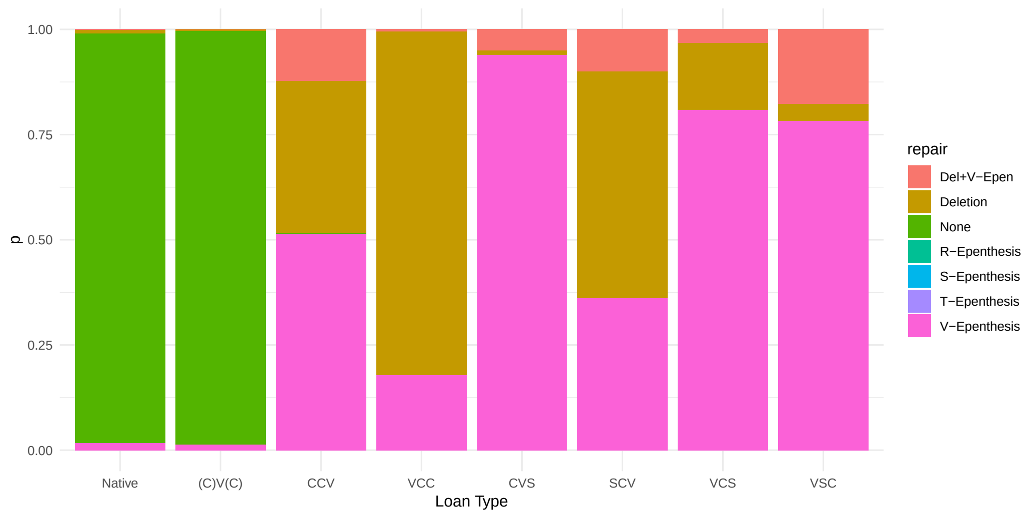


Figure 5-2: Probability of repair by input structure category, trained on the test set with Markedness $\gg$ Faithfulness bias and overlapping constraints

repairs, coda SC clusters still overwhelmingly prefer vowel epenthesis as a repair, instead of the predicted combination of epenthesis and deletion.

Overall, while there are nevertheless still discrepancies between attested repairs and the results predicted by the best-case grammar, the grammar nevertheless gets the broad strokes of the attested loanword repair pattern correct. The lone exception is the set of coda SC clusters, which fared better under an analysis which eliminated gang-up effects for these segments. Why this should be so is discussed in detail in section 5.2.8.

5.2.5 Extreme Weight-Based Prior

Figure 5-3 shows the results of training a MaxEnt model on the native Cantonese vocabulary **alone**. Instead of a simple Markedness $\gg$ Faithfulness bias, as was used fo train the best-case model above, the resulting weights from this grammar were used as the target weights of the Faithfulness constraints. A learner which encoded the PMap as a series of asymmetrical target weights was chosen as the first test of a MaxEnt learner using the more compact constraint set as it was shown in Chapter 4 that this was the method that yielded the best results when it came to predicting the behaviour of loanwords.

While in the previous chapter, incorporating the best-case weights of the Faithfulness constraints as target weights in the model bias parameter led to a fairly good match between model results and the best-case results, this is not the case here. The model is quite **uncertain** about what kinds of repairs to prefer overall, with most preferred repairs among the highest probability repairs being

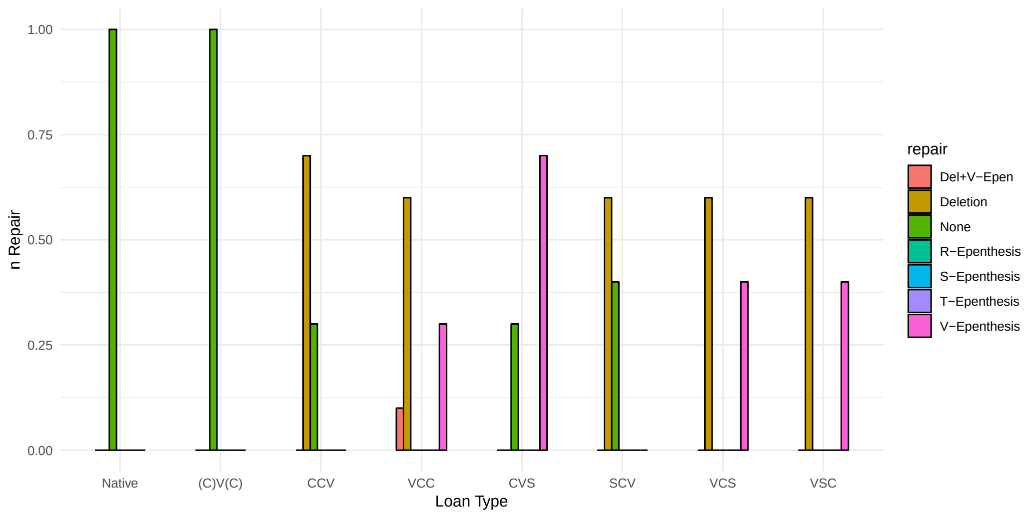


Figure 5-3: Proportion of winning candidates divided by repair and input structure category, representative sample of 1000 Cantonese lexical items with an extreme weight-based bias and overlapping constraints

chosen at most 75% of the time. More troubling is the fact that certain loanword structures are not repaired **at all**, indicating that some Faithfulness constraints may have achieved weights which were too high, that the gang-up effects of Faithfulness constraints were too strong, or that the crucial Markedness constraints may have been demoted too low. Deletion also appears to be the preferred repair for most kinds of phonotactically-illicit structures, with the exception of coda sibilants, which are repaired most often by vowel epenthesis.

(113) **Sample repairs for a grammar trained on a 1000-item data set with an extreme weight-based bias and overlapping constraint set**

Category	*Repair Type*	*Input*	*Gloss*	*Output*	*p*
Native	None	`o pej`	'to fart'	`o pej`	0.99668
(C)V(C)	None	`e sej`	'essay'	`e sej`	0.99668
CCV	None	`wo ljam`	'volume'	`wo ljam`	0.38059
	Deletion	`plam`	'plum'	`lam`	0.87360
VCC	Deletion	`si mant`	'cement'	`si man`	0.99908
	V-Epenthesis	`singk`	'sink'	`sing kV`	0.65788
	Deletion + V-Epenthesis	`pamp`	'pump'	`am pV`	0.85712
CVS	None	`paas pot`	'passport'	`paas pot`	0.94065
	V-Epenthesis	`paas`	'pass'	`paa sV`	0.52865
SCV	None	`swej`	'sway'	`swej`	0.48028
	Deletion	`spaak`	'spark'	`saak`	0.69206
VCS	Deletion	`inc`	'inch'	`in`	0.95759
	V-Epenthesis	`aats`	'arts'	`aat sV`	0.99086
VSC	Deletion	`jist`	'yeast'	`jit`	0.96588
	V-Epenthesis	`tost`	'toast'	`to sVt`	0.43806

Examining the exact repairs chosen does not shed much light on the results in general, although a few additional notes can be made. First, many of the items which show **no repairs** despite containing phonotactically-illicit sequences of segments are **multisyllabic**. Second, there appears to be a return to the repairs predicted by the null hypothesis from Chapter 4, where segments which are deleted are largely those which native speakers choose to retain. Namely, this grammar predicts that the first member of an onset cluster is deleted, and that sibilant consonants are deleted, when possible.

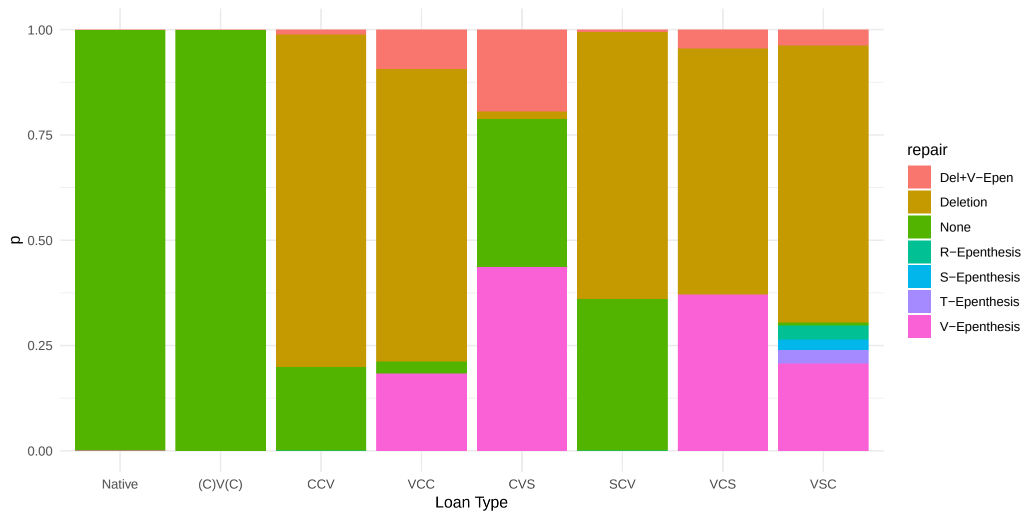


Figure 5-4: Probability of repair by input structure category, representative sample of 1000 lexical items with an extreme weight-based bias and overlapping constraints

Figure 5-4 shows much the same pattern, with the majority of the probability space assigned to deletion candidates in all phonotactically-illict loanwords except those which contain coda sibilants. These kinds of words slightly prefer vowel epenthesis as a repair. There is an overall bias towards assigning more of the probability space to vowel epenthesis in loanwords with coda clusters of all types, while there is a preference for retaining both members of a cluster in onset position.

With respect to KL-Divergence values, the difference between this grammar and the best-case grammar discussed in the previous section is $D_{KL} = 0.88683$. This indicates that the two grammars are moderately dissimilar, although not quite as dissimilar as grammars examined in Chapter 4.

The exact reasons for these discrepancies will be discussed further in section 5.2.8, but it is apparent from these data that the bias may have allowed the weights of the Faithfulness constraints to remain too high, and failed to adjust the Markedness constraints. If, for instance, the combined weight of Max-T/_V and Max-S were to **exceed** the weight of CodaCond, then this would predict the failure to repair input structures such as *paas1 pot2* 'passport'. Although Cantonese has native words which could influence the weighting of Max-T/_V and Max-S, it is unlikely that the weights of either constraint would decrease over the course of learning, as these segments are always **preserved** in native Cantonese lexical items.

In the following section, the results of a model trained using a different weight-based prior are examined, in order to see if altering the way in which the prior is derived has any significant influence on results.

5.2.6 Averaged Weight-Based Prior

As opposed to simply taking the weights of the current best-case grammar and using them as the target weights for the White (2013)-style prior, the grammar explored in this section used a prior which obtained the target weights from a different source. As discussed in section 5.2.4, the current best-case grammar is not quite as accurate or as consistent as the best-case grammar from experiments 1–3, although differences are relatively minor. The previous simulation, outlined in section 5.2.5, has shown that using the best-case grammar trained using the condensed constraint set is likewise less accurate and consistent when predicting how loanwords will be adapted into Cantonese. As such, this simulation aims to utilize the weights from the previous best-case grammar in an attempt to improve results for grammars which utilize an overlapping constraint set.

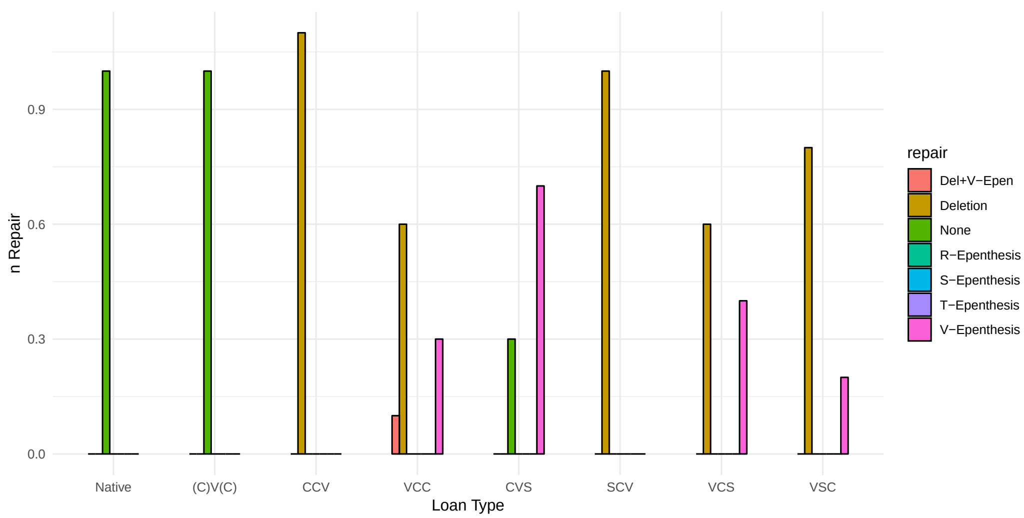


Figure 5-5: Proportion of winning candidates divided by repair and input structure category, representative sample of 1000 Cantonese lexical items with an averaged weight-based bias and overlapping constraints

Figure 5-5 shows the relative proportions of the repair types for candidates assigned the highest probability by a MaxEnt grammar equipped with this averaged weight-based prior. The most notable improvement in results lies in the fact that phonotactically-illicit loanwords are rarely left as-is. However, this grammar still shows notable discrepancies with attested patterns of loanword adaptation. Onset clusters of all types are exclusively repaired by deletion of one member of the cluster, limiting vowel epenthesis repairs to coda clusters in general. Still, coda clusters are also overwhelmingly repaired by deletion, when possible. Finally, coda SC clusters are repaired either by deletion or by vowel epenthesis, but not by both.

(114) **Sample repairs for a grammar trained on a 1000-item data set with an averaged weight-based bias and overlapping constraint set**

Category	*Repair Type*	*Input*	*Gloss*	*Output*	*p*
Native	None	`o pej`	'to fart'	`o pej`	0.96162
(C)V(C)	None	`e sej`	'essay'	`e sej`	0.96162
CCV	Deletion	`wo ljam`	'volume'	`wo jam` `wo lam`	0.36129 0.36129
VCC	Deletion	`si mant`	'cement'	`si man`	0.99932
	V-Epenthesis	`sapt`	'shaft'	`sap tV`	0.49910
	Deletion + V-Epenthesis	`pamp`	'pump'	`am pV`	0.74789
CVS	None	`paas pot`	'passport'	`paas pot`	0.94516
	V-Epenthesis	`paas`	'pass'	`paa sV`	0.50449
SCV	Deletion	`spaak`	'spark'	`saak`	0.65142
VCS	Deletion	`inc`	'inch'	`in`	0.94459
	V-Epenthesis	`aats`	'arts'	`aat sV`	0.98923
VSC	Deletion	`jist`	'yeast'	`jit`	0.97151
	V-Epenthesis	`toast`	'toast'	`to sVt`	0.25136

Examining the highest probability candidates in more detail shows that many familiar errors are still present in this model. For instance, when a segment is deleted, it is most often the leftmost segment of an onset cluster or the coda sibilant. Segments of identical sonority are incapable of being distinguished when it comes to deletion, as shown by the two candidates selected for input *wo1 ljam2* 'volume'. What is somewhat promising about the above results is that sibilants are more often selected for preservation, even in unattested deletion candidates such as *saak* for *spaak1* 'spark'. This indicates that the gang-up effects introduced by the MAX-T and MAX-S constraints are having some sort of beneficial effect.

The aggregate results presented in Figure 5-6 are likewise quite similar to the results of the previous experiment, with the exception that a greater portion of the probability assigned to each category is dedicated to deletion candidates in most cases.

With respect to KL-Divergence, this grammar has a divergence value of $D_{KL} = 1.157753$ when compared the the best-case grammar discussed in section 5.2.4. This indicates that this grammar is more dissimilar to the best-case grammar than the extreme weight-based grammar. When compared to the previous weight-based grammar, the divergence value is $D_{KL} = 0.06677$, indicating that these two grammars are actually quite **similar**.

This is, in some respects, surprising. In previous experiments, changing the prior had a much

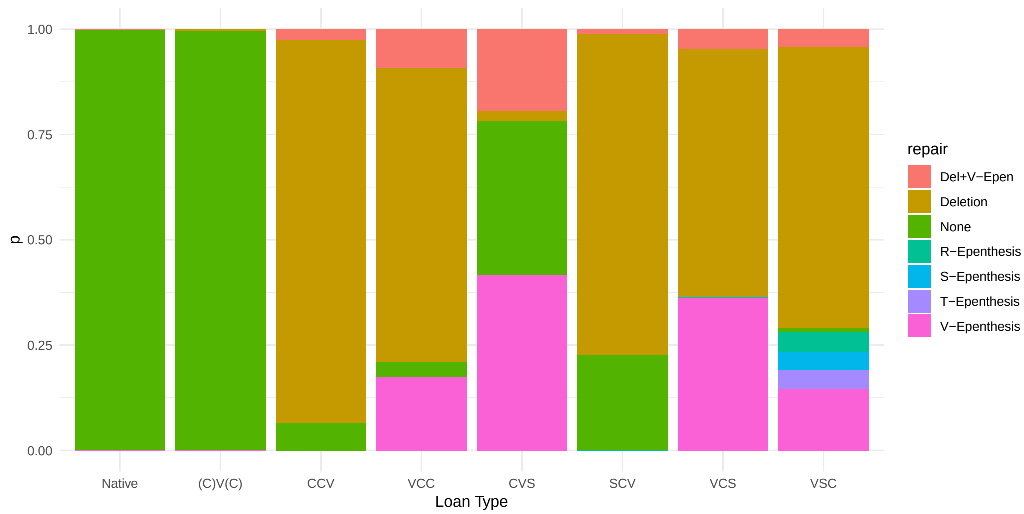


Figure 5-6: Probability of repair by input structure category, representative sample of 1000 lexical items with an averaged weight-based bias and overlapping constraints

larger effect on the resulting grammar than the current experiments. This may be due to the fact that there are substantially **fewer** constraints to consider in this grammar, which also means substantially fewer weight differences that can be maintained. Constraint interactions are likewise confined to linear combination of constraint weights, and so there is no opportunity for the grammar to treat, say, intervocalic contexts as any more or less privileged than their pre- or post-vocalic counterparts. Further discussion of why this particular constraint set had such a greater influence on results is undertaken in section 5.2.8.

5.2.7 Plasticity-Based Prior

While the results of the plasticity priors explored in experiment 2 did not predict any sort of asymmetry in loanword repairs and were deemed to be inadequate for the purposes of modelling how learners generalize in the absence of direct phonological evidence, a plasticity prior was nevertheless included in the present experiment. This was done in order to ascertain whether the constraint set used had an undue influence on this result, or whether it was indeed entirely due to the lack of overt phonological evidence in the native Cantonese lexical items. As such, a version of the averaged weight-based prior was used to construct a plasticity-based variant. This manner of constructing the prior was chosen in an attempt to avoid a situation where the Faithfulness constraints received weights high enough to interfere with the behaviour of the Markedness constraints.

Unfortunately, the question of whether the behaviour of a plasticity-based prior is improved with a more condensed constraint set cannot be definitively answered at present, as the model trained on the plasticity-based prior did not converge within 1,000 iterations. Any results presented below should thus be considered to be **preliminary**, and may not be identical to what the model would find if it were given enough time to converge.

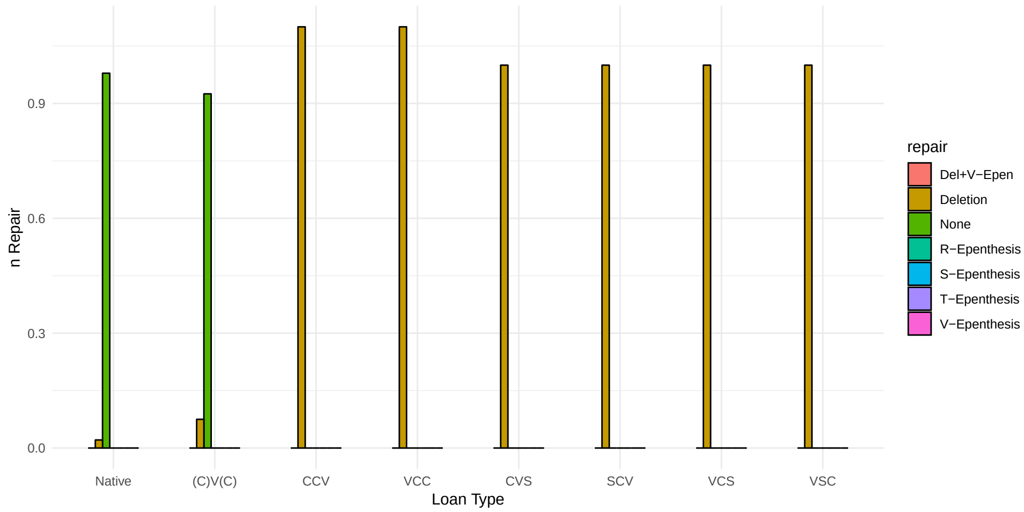


Figure 5-7: Proportion of winning candidates divided by repair and input structure category, representative sample of 1000 Cantonese lexical items with an averaged plasticity-based bias and overlapping constraints

As shown in Figure 5-7, the interim plasticity grammar found using the condensed set of constraints prefers deletion for **all** phonnotactially-illicit loanwords **exclusively**. Furthermore, it also prefers deletion as a repair to some phonotacically-licit loanwords and native Cantonese words.

(115) **Sample repairs for a grammar trained on a 1000-item data set with an averaged plasticity-based bias and overlapping constraint set**

Category	*Repair Type*	*Input*	*Gloss*	*Output*	*p*
Native	None	`naap seoj`	'tax'	`naap seoj`	0.98941
	Deletion	`o pej`	'to fart'	`pej`	0.83278
(C)V(C)	None	`kaa low lej`	'calore'	`kaa low lej`	0.99797
	Deletion	`e sej`	'essay'	`sej`	0.83278
CCV	Deletion	`wo ljam`	'volume'	`wo jam`	0.49999
				`wo lam`	0.49999
VCC	Deletion	`sapt`	'shaft'	`sat`	0.49985
				`sap`	0.49985
CVS	Deletion	`paas pot`	'passport'	`paa pot`	0.99999
SCV	Deletion	`spaak`	'spark'	`saak`	0.75579
VCS	Deletion	`inc`	'inch'	`in`	0.97226
VSC	Deletion	`tost`	'toast'	`tot`	0.98348

As expected from previous results, most attested deletions present in this model are those which are **contrary** to what might be expected from the perspective of native speaker adaptation and the PMap. Segments which are deleted are those located **furthest** from the vowel or **sibilants**.

The one exception is the words which contain SC onset clusters, where the sibilant is the retained, indicating that the gang-up effect of the MAX-T and MAX-S constraints is having some effect.

As in previous models that used a plasticity bias, single-vowel syllables are **deleted** under this model as well, indicating that the lack of evidence present in the training set is preventing any of the relevant Faithfulness constraints from outweighing Markedness constraints such as BIMORAIC. Likewise, the model is incapable of deciding which segment should be retained if it belongs to a cluster of level sonority.

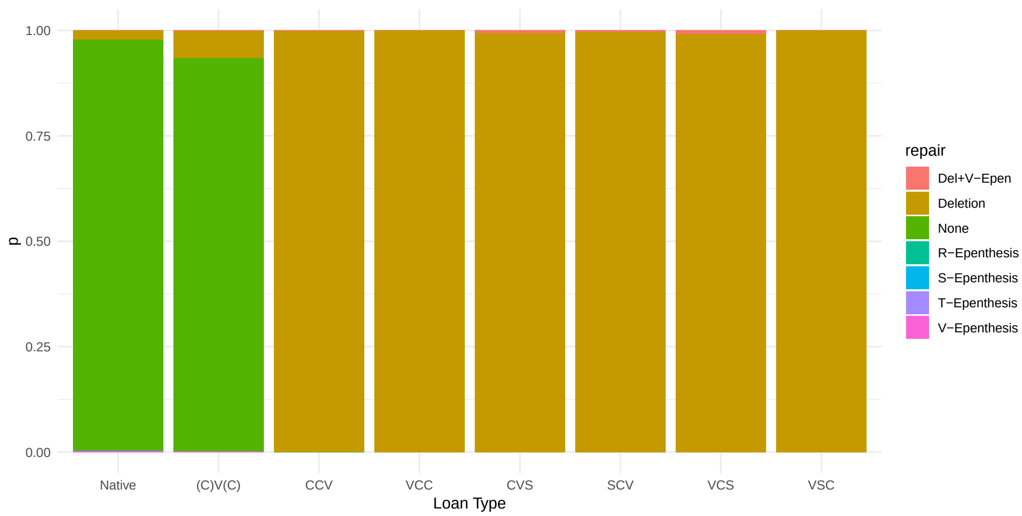


Figure 5-8: Proportion of winning candidates divided by repair and input structure category, representative sample of 1000 Cantonese lexical items with an averaged plasticity-based bias and overlapping constraints

With respect to KL-Divergence, this interim grammar receives a value of $D_{KL} = 3.20052$ when compared with the best-case grammar from section 5.2.4. This indicates that these grammars are incredibly **dissimilar** – the most dissimilar of all comparisons made thus far. With respect to the most analogous weight-based grammar generated in this experiment, discussed in section 5.2.6, this plasticity-based grammar receives a value of $D_{KL} = 1.48523$. This indicates that this grammar – while still dissimilar to the current grammar – is not quite as dissimilar as the best-case grammar. This is likely due to the fact that the same training set was used for the grammars which made use of priors, while the best-case grammar was trained on the test set.

Overall, this simulation has shown that the exact constraint set used has **little effect** on the grammar of a MaxEnt learner equipped with a plasticity-based substantive bias. Rather, the failure of this kind of learner is due to the **lack of overt evidence** in the training set about how to re-weight the Faithfulness constraints necessary to account for the attested patterns of loanword adaptation.

5.2.8 Discussion

The three simulations discussed above were meant to test whether changing the constraint set to allow for overlapping Faithfulness constraints led to an improvement in predicting attested patterns of loanword adaptation in Cantonese. The results obtained indicated that no substantial improvement was obtained – if anything, the results of the best-case grammar from section 5.2.4 indicated that the current constraint set is in some ways **inadequate**. The broad patterns of loanword adaptation were capable of being obtained, in that there was an asymmetry in how onset and coda clusters were repaired, and that clusters involving sibilants showed an increased preference for vowel epenthesis. However, the most interesting of the attested repairs, involving simultaneous epenthesis and deletion, was **absent** from this grammar. This seems to indicate that a grammar that uses this constraint set is unable to acquire the correct entire set of attested loanword repairs for Cantonese.

This may ultimately be due to the error in the test set generation procedure whereby clusters of equal sonority are assigned identical constraint violations. ST clusters are, under the current constraint set, considered to be of the **same** sonority class instead of two separate sonority classes. As such, the grammar is unwilling to delete **either**, and opts to epenthesize a vowel in order to keep both. If this error were repaired, then it would likely be the case that this constraint set could still capture the behaviour of loanwords.

Another issue raised by the current experiment is that it is more difficult to maintain a Markedness $\gg$ Faithfulness bias using the set of overlapping constraints. This was shown to be the case for the grammars which were trained using weight-based biases, and especially so for the extreme bias discussed in section 5.2.5. This prior showed that the gang-up effects of constraints such as MAX-S and MAX-T/_V were sufficient to **exceed** the weights of the Markedness constraints which drive loanword adaptation, CODACOND and *COMPLEX. This indicated that the target weight of the Markedness constraints in the prior was too low.

One potential solution to this issue would be to set the Markedness constraints at some **fixed distance** above the highest Faithfulness constraint, rather than at some fixed point. Another more sophisticated solution would be to maintain **as wide a distance as possible** between both sets of constraints. This is, in fact, a solution that is available within the `solve.R` script for this particular bias, and will be discussed in further detail in section 5.5, below.

Finally, this set of experiments showed that using a more condensed set of overlapping constraints yielded more **consistent** results, even when different priors were used. This was not the case for the experiments discussed in Chapter 4, which showed quite a bit of variation depending on which prior was used, especially for the weight-based priors of experiment 3. This is due to the substantial decrease in the number of constraints to be considered, as they will be applicable to a much greater number of candidates and receive many more opportunities for their weights to be altered. This, in turn, means that the data used to train the model have a greater effect on the resulting grammar, and reliance on the prior is minimized. This does not seem to be an effect that is easily mitigated, or even necessarily detrimental to the model. Rather, it is a consequence of the trade-off between the amount of data used to train the model and the parameters that the model can utilize to account for the patterns in the data.

In sum, examining the effects of a model that makes use of overlapping Faithfulness constraints does not seem to have a substantial effect on the results obtained in Chapter 4. The best method of encoding a substantive bias into a MaxEnt model is by encoding the bias as a series of asymmetrical target weights, rather than as plasticities. No substantial improvement to the accuracy of results is made by condensing the constraint set in this way, and a choice of constraint sets and priors may be up to the preference of the researcher conducting the simulations, provided they are careful about how violations are assessed.

In the following sections, additional potential improvements to the experimental procedure used in Chapter 4 will be discussed, although similar tests of the repairs will not be undertaken. This is largely due to the amount of preparation that would be necessary in order to implement each improvement, although they were still deemed to be possible under the current procedure. The first – and potentially simplest – improvement will be in making use of a different set of Markedness constraints, to be discussed below.

5.3 Alternate Markedness Constraints

While the PMap as conceived of in Steriade (2001) is framed as an issue of determining the relative rankings of contextual Faithfulness constraints, these are not the only kinds of constraints which can be sensitive to segment contexts. Contextual Markedness constraints have also been proposed to account for contextually-conditioned phonological behaviour, and in many cases a phonological phenomenon can be analyzed using either kind of constraint. For example, the phenomenon of partial cluster acquisition in children learning English as a first language can be analyzed either as being due to a highly-ranked MAX-C/$\acute{\sigma}$ constraint, preventing consonants from being deleted in a stressed syllable (Gnanadesikan, 2004; Jesney and Tessier, 2008; Jesney, 2011; Tessier, 2009); or as being due to a highly-ranked *Complex/$\breve{\sigma}$ (Olson, 2016), preventing clusters from surfacing in an unstressed syllable.

The constraint set used in the previous chapter exclusively made use of Markedness constraints which were **as broad as possible**. *Complex banned both complex onsets and codas, and CodaCond enforced all of the restrictions on what is admissible as a coda of Cantonese. Thus, the number of constraints which caused English loanwords to be repaired in Cantonese was only **two**.

This may be too simplistic of a tactic to take in these cases, and splitting these constraints up into a larger number of independently weighted constraints may be what is required in order to more adequately model how native speakers of Cantonese conceive of their language. *Complex may be split into *Complex-Onset and *Complex-Coda, and CodaCond may be split into the set of constraints in (116).

(116) **Constraints which can enforce the conditions on coda consonants in Cantonese**
*Liquid-Coda, *Voiced-Obst-Coda, *Labiovelar-Obst-Coda, *Fric-Coda ≫
*Vclss-Obst-Coda, * Nas-Coda, *Glide-Coda

These constraints must be ranked or weighted according to the schema outlined above in order for the coda condition of Cantonese to be met.

There is, however, a potential issue that is raised by introducing this broader set of contextual Markedness constraints into the grammar. As noted earlier, contextual Markedness constraints can often do the **same work** as contextual Faithfulness constraints. In fact, for phonological frameworks which allow gang up effects, such as Harmonic Grammar and MaxEnt, only **one** set of contextual constraints is required in order to adequately capture contextual effects in any language (Jesney, 2011). And, as the MaxEnt learners utilized here are largely all equipped with a Markedness ≫ Faithfulness bias, this means that this set of contextual Markedness constraints may be able to **mimic** the behaviour of the PMap and further obscure proposed contextual Faithfulness weight asymmetries.

Take, for example, the proposed division of *Complex into *Complex-Onset and *Complex-Coda. As noted in Chapter 1, native Cantonese speakers show an **asymmetry** between how onset and coda consonant clusters are repaired. Speakers show a 50–50 split between whether they choose to repair an onset consonant cluster via deletion of the second member of the cluster, or via vowel epenthesis between both members of the cluster. However, speakers universally prefer deletion of the second member of the cluster when one occurs in coda position. Previous grammars have assumed that this is a result of the contextual Faithfulness constraints, but this could also be accommodated by a differential weighting of *Complex-Onset and *Complex-Coda.

Splitting up the Markedness constraints in this way also raises the question of whether the PMap can or should be modelled as done in Wilson (2006), where it is the contextual **Markedness** constraints which are governed by the PMap. After all, if either kind of contextual constraint is adequate for modelling contextual behaviour in a MaxEnt-based grammar, then utilizing contextual Markedness constraints in this way should be possible. It may, then, be worthwhile to run an experiment where this is implemented, and to see whether modelling the PMap as an asymmetrical treatment of Markedness constraints leads to an improvement in how loanword adaptation is modelled. As Wilson (2006) also makes use of such a constraint set when implementing a plasticity-based substantive bias, it may also be of interest to see whether such a system also allows for improved performance of a grammar that incorporates this kind of bias.

A second issue raised by the set of Markedness constraints used above is whether all of the included Markedness constraints are always **necessary** in order to model the loanword behaviour examined here. For example, some phonotactic constraints, such as those banning a syllable containing both a labial onset and a labial coda, were shown to influence loanword adaptation in a way that was contrary to how native speakers chose to adapt the word. For example, the loanword [ˈpʰʌmp] 'pump' was adapted as *am pV* in the experiments discussed in section 5.2. It is likely that the model chose this repair over the attested *pam* due to the high weighting of the Markedness constraint banning a syllable that begins with [p] and ends in [m].

The phonotactic constraints were included in the model in an attempt to **prevent** the model from accounting for distributional asymmetries by altering the weights of the Faithfulness constraints. However, this may instead be **obscuring** asymmetries in the weighting of Faithfulness constraints which should be ascribed to the PMap. In the future, additional experiments which eliminate these phonotactic constraints from the grammar should be undertaken, in order to see whether this is true.

5.4 Iterative Models

A key assumption made about the experiments conducted throughout this thesis was that loanword adaptation is a simple process by which **monolingual** native speakers of one language use the phonological system they have acquired over their lifetime to **adapt** novel loanwords from a second language. It is assumed that these speakers have **next to no knowledge** of this second language, but that they are nevertheless able to hear the language fairly accurately. It is also assumed that this knowledge is entirely **auditory**. This adaptation process is assumed to be **uniform** across speakers, and it is assumed that it occurs within **one generation** after contact between the two groups of speakers is initiated.

Many of these assumptions appear to be quite naïve when considered against proposals of how loanword adaptation proceeds among communities of speakers. Rather than being introduced by speakers of a second language, loanwords have been hypothesized to be introduced by **bilingual** individuals within the community, who have a solid knowledge of the phonological systems of both languages (Paradis and LaCharité, 1997). These speakers may not always agree on a single repair, and introduce **idiosyncratic** repairs that must then be **standardized** by the broader linguistic community. They may also encounter loanwords via the **written word**, in addition to speech (Smith, 2006; Dohlus, 2010). Loanwords do not necessarily have to enter the language at the **same time**, and younger generations of speakers may adapt new loanwords in **novel ways**. Younger generations may also have grown up speaking a language which already **contains** loanwords, and may have an idea about what repairs are preferred before encountering new loanwords.

Many of the assumptions made above were done for the sake of constructing the simplest models possible in order to assess how various biases perform in the absence of overt phonological evidence. A more accurate model of loanword adaptation will take these theoretical considerations into account. However, there is one potential improvement listed here which may be easier to implement under the current system: constructing a model for **multi-generational** loanword adaptation.

As discussed in Chapter 4, it was assumed that the models constructed thus far were models of **monolingual** speakers of Cantonese. It was also acknowledged that this is **not** an accurate model when modern day speakers of the language are considered, as many of them are bilingual or trilingual in English and/or Mandarin. Rather, the model should be considered to be an approximation of how loanword adaptation may have worked in the past, for speakers who were encountering English on a day-to-day basis for the first time. Since then, many loanwords from English would have entered the language, and later generations would have encountered these loanwords. Later generations may also have learned the language that the loanwords came from, and been able to standardize any potentially inconsistent loanword repairs.

These later generations could, under the current experimental protocol, be modelled by constructing an **iterative** series of MaxEnt learners. The first generation would be modelled as done in the experiments discussed previously, where the learner is not provided any knowledge of the loanwords, and is asked to **predict** loanword repairs. The second generation would be modelled in the same way, with the same substantive bias, but would **include** the loanword repairs predicted by the first generation. These learners would then be tested on the same set of loanwords to see if their predictions were **closer** to the predictions attested in the literature.

Constructing an iterative model like this for a grammar that makes use of a weight-based substantive bias is fairly straightforward, and is predicted to look much the **same** across generations, with potential improvements after the loanwords are included directly in the training sets. Recall that the target weights introduced by the bias term in the first-generation grammar will remain unchanged if there is no overt evidence about how to alter their weights from the training data. This will keep a large portion of the target weights intact, especially if the more expanded, independent set of Faithfulness constraints is used. As such, loanwords will be much closer to their attested forms, and will be fed to the second-generation grammar. This learner will then have **direct evidence** about how to select the weights of a **larger number** of constraints, and this evidence is not likely to contradict the substantive bias, as it was the substantive bias which helped predict the loanword repairs in the first place.

While the effects of this multi-generational model are not predicted to affect the results obtained through a series of learners equipped with a weight-based substantive bias, they **could** affect the results obtained through a series of learners which make use of a **plasticity-based** bias. Say, for instance, that the results of the first-generation MaxEnt learner trained using a plasticity-based substantive bias were in line with what was initially predicted for these kinds of learners. That is, the learner was truly **agnostic** about which kind of repair to prefer, and **alternated** evenly between deletion and vowel epenthesis for all phonotactically-illicit loanwords. While this particular MaxEnt learner would not have acquired the attested loanword repair pattern, a second-generation learner would at least be provided with the information that loanwords had entered the language, and that there were multiple potential repairs to each. This second-generation learner – equipped with the same plasticity bias – would then receive **overt evidence** about loanword repairs, and the constraints responsible for each would be able to be have their weights altered. Since the plasticity bias dictates **how far** each constraint is able to move from its target weight, and since this bias is by definition **asymmetrical**, the grammar acquired by the second-generation learner will have a chance to move closer to the one needed to predict loanword repairs.

In this way, constructing an iterative, multi-generational model of learners could allow MaxEnt grammars equipped with a plasticity-based prior to be able to adequately model loanword adaptation. This result is obtained by allowing the second-generation learners to obtain **direct evidence** about loanword adaptation from the first-generation learners. It is also predicated on the idea that there are **multiple** repairs predicted by the first-generation grammar. This, unfortunately, was not the case for the experiments undertaken in this thesis – when equipped with plasticity-based priors, these grammars overwhelmingly preferred **deletion**. If a second-generation learner were presented with an outcome from experiment 2 in section 4.4, it would learn that deletion was the preferred repair, and perpetuate the weighting necessary to maintain this result. This preference for deletion was shown in section 5.1.1 to be due to the number of deletion versus epenthesis candidates present in the training set. If some way of circumventing this issue is later found, it may be worthwhile to test this iterative approach.

5.5 Alternate Encoding

One final potential improvement to the experimental procedures used throughout this thesis lies in improving the model of the **PMap encoding** itself. Both the plasticity-based prior used by Wilson (2006) and the weight-based prior used by White (2013) make use of an encoding which establishes a **static asymmetry** between the constraints which are governed by the PMap. Each constraint is assigned a **single value** that governs how that constraint will behave throughout the learning process. For the plasticity-based prior used by Wilson (2006), it is how far that single constraint can move from its target weight without incurring a strong penalty; for the weight-based prior used by White (2013), it is what target weight that single constraint should ultimately have. However, the PMap, as originally conceived of in Steriade (2001), is inherently **comparative**, while the encodings used here are only capable of imposing restrictions on one constraint at a time.

As a brief illustration, let us consider the constraints MAX-T/#_V and MAX-T/V_#. If a PMap-like bias were to be established for these constraints using a plasticity-based prior, each would receive its own Gaussian distribution, where the plasticity of MAX-T/#_V would receive some value x, and the plasticity of MAX-T/V_# would receive some value y. In order for the prior as a whole to be PMap-compliant, x would have to be **greater than** y, if it is assumed that the strength of cues discussed in Wright (2004) contributes to this asymmetry. However, over the course of learning, each constraint will have its weight altered **independently**, and the ending weight of MAX-T/V_# is not guaranteed to be lower than that of MAX-T/#_V. The prior only establishes that this **tends** to happen across grammars. The table below gives a small illustration of this process.

(117) **Subverting a plasticity prior**

		Plasticity prior		*Learning*	
	w_{init}	μ	σ^2		w_{final}
MAX-T/#_V	1.0	0.0	$x = 10$	$\cdots$	2.0
MAX-T/V_#	1.0	0.0	$y = 5$	$\cdots$	5.0

Likewise, if a PMap-like bias were to be established using a weight-based prior, each would receive its own Gaussian distribution with its own target weight. MAX-T/#_V would receive some target weight m, and MAX-T/V_# would receive some target weight n. In order for the prior as a whole to be PMap-compliant, m would be greater than n. But again, over the course of learning, the MaxEnt learner will adjust the weights of each constraint independently, and the final weights of each are not guaranteed to be the equal to or even **parallel to** the target weights. Again, the prior only establishes what the **tendency** is across grammars.

(118) **Subverting a weight prior**

	w_{init}	*Plasticity prior* μ	σ^2	*Learning*	w_{final}
MAX-T/#_V	1.0	$m = 5.0$	100.0	$\cdots$	4.0
MAX-T/V_#	1.0	$n = 2.0$	100.0	$\cdots$	4.5

The issue illustrated above is not necessarily that the PMap is capable of being **subverted** in the presence of evidence, but that the priors established over individual constraints make it easier to do so. Of course, depending on how **strict** one may believe the PMap to be, this may be a desirable outcome. For example, under White's (2013) analysis of saltation phenomena, the PMap **must** be capable of being subverted in this manner. If it were otherwise, saltation would not be possible in any phonological system. The desirability of this result is discussed in more detail in Chapter 3.

However, the original conception of the PMap is not couched in terms of the properties of **individual constraints**. It does not make claims about any one constraint in isolation, along the lines of "Constraint A can be x distance from its target weight" (for the plasticity-based prior) or "The weight of constraint B is y" (for the weight-based prior). Rather, it is always stated in terms of constraint **comparison**: "constraint A has a higher weight (or ranking) than constraint B". Both of the priors used thus far fail to directly model this state of affairs. Instead, it is up to the researcher to set a prior which is PMap-compliant – the grammar has no **independent** way of maintaining this asymmetry. How then, if we still desire that the PMap be capable of being subverted, can we construct a prior which captures the comparative nature of the bias?

One solution used in the `solve.R` script of Pater and Staubs (2013) is to establish a **difference bias** over the constraints which are required to be separated by the grammar. This is implemented in order to establish a Markedness $\gg$ Faithfulness bias in `solve.R`, but could be extended to apply to PMap-governed constraints in order to maintain the comparative nature of the PMap. In order to establish how such a bias works, a brief review will be made of how Pater and Staubs (2013) implement the Markedness $\gg$ Faithfulness bias. A sketch of how this might be extended to the set of PMap-governed constraints will then be provided.

The difference bias used by Pater and Staubs (2013) to maintain separation between the weights of the Markedness and the Faithfulness constraints is given in (119). In the equation below, the term F refers to the **boolean function** which determines whether constraint C_i belongs to the class of Faithfulness constraints or not. It is 1 if the constraint belongs to this class, and 0 otherwise. Similarly, the term M does the same for the set of Markedness constraints. The term α is a parameter which can be set by the researcher, and determines how **strong** the bias should be when incorporated into the overall objective function.

(119) **Difference bias**

$$\alpha \sum_{i=1}^{n} (F(C_i)w_i - M(C_i)w_i)$$

Say that this bias is applied to a set of only two constraints: constraint 1 is the Faithfulness constraint MAX-C, and constraint 2 is the Markedness constraint *COMPLEX. Thus, $F(C_1)$ should be equal to 1 and $M(F_1)$ should be equal to 0, since MAX-C is a Faithfulness constraint. Likewise, $F(C_2)$ should be 0 and $M(C_2)$ should be 1, since *COMPLEX is a Markedness constraint. When calculating what MAX-C contributes to the difference bias as a whole, it will contribute $F(C_1)w_1 - M(C_1)w_1 = 1 \cdot w_1 - 0 \cdot w_1 = w_1$. That is, it contributes its current weight directly into the bias. However, the constraint *COMPLEX will contribute $F(C_2)w_2 - M(C_2)w_2 = 0 \cdot w_2 - 1 \cdot w_2 = -w_2$. That is, it contributes the **negative** of its weight into the bias. These two are then **summed** to determine how far apart the two constraints are at the current point in the grammar. If the weight of MAX-C is quite high and the weight of *COMPLEX low, it will yield a **positive** bias term; if the reverse is true, it will yield a **negative** one.

This appears at first glance to be counterintuitive – after all, the goal is to establish the weight of the Markedness constraints as being **higher** than the Faithfulness constraints. However, the effect of the bias in (119) appears to be the reverse, with positive values assigned when the Faithfulness constraints as a whole exceed the weights of the Markedness constraints as a whole. This is ultimately due to the way in which the algorithm is structured. Recall from Chapter 2 that the entire MaxEnt algorithm is often **negated**, meaning that the goal is to arrive at the **smallest possible value** rather than the highest possible one. As such, the Markedness $\gg$ Faithfulness bias will also need to be reversed in order to be compatible with the larger algorithm. The fact that the bias in (119) assigns **negative** values when the weight of the Markedness constraints exceeds that of the Faithfulness constraints means that it will **drive down** the value of the algorithm when this condition is met, helping the algorithm achieve the smallest possible value. It is, therefore, compatible with the algorithm used, and has the desired effect.

The advantage of using such a bias lies in the fact that it makes a **direct comparison** between the Markedness and Faithfulness constraints at each point in the learning process, and will at each point aim to **increase** the distance between the two sets of constraints while maintaining a higher weight for the Markedness constraints. An additional advantage to using this bias is that it does not attempt to maintain a difference between **every** Markedness and Faithfulness constraint, but between the sets of constraints as a **whole**. Thus, while there may occasionally be Faithfulness constraints which exceed the weight of some Markedness constraints, the overall bias is maintained as much as possible.

While it will not be possible to establish a **single bias** to implement the PMap in this manner, the same concept could be used to implement **portions** of the PMap. For example, say that the following schema from section 4.4.2.1.1 in Chapter 4 is to be used to implement at least a portion of the PMap:

(120) **Weighting schema for Dep-V constraints**

$$\left\{\begin{array}{c} \textsc{Dep}\text{-V}/\#\_\# \\ \textsc{Dep}\text{-V}/\text{T}\_\# \\ \textsc{Dep}\text{-V}/\text{T}\_\text{T} \\ \textsc{Dep}\text{-V}/\text{R}\_\# \\ \textsc{Dep}\text{-V}/\text{R}\_\text{T} \\ \textsc{Dep}\text{-V}/\text{R}\_\text{R} \end{array}\right\} > \left\{\begin{array}{c} \textsc{Dep}\text{-V}/\#\_\text{R} \\ \textsc{Dep}\text{-V}/\text{T}\_\text{R} \end{array}\right\}$$

Recall that this schema follows research by Yun (2016), which showed that constraints which reference environments where there is a decrease or plateau in overall intensity should be weighted or ranked **higher** than those which show an increase in intensity. As such, the Dep-V constraints can be separated into **two groups**, one of which must maintain a higher overall weight than the other.

This weight difference could be enforced not by a difference in the target weights or plasticities assigned to each group, but instead by a **difference bias**. Establishing this difference bias would require formulation of two **functions** which would evaluate every constraint in the grammar to determine which class of constraints they fell into. One function, here termed IF for "intensity fall", would return 1 if the constraint belonged to the category of Dep-V constraints which referenced an intensity fall or plateau (the leftmost set of constraints in (120)), and 0 otherwise. The other, termed IR for "intensity rise", would return 1 if the constraint belonged to the category of Dep-V constraints which referenced an intensity rise (the rightmost set of constraints in (120)).

Once these two functions are defined, they can then be used to create a **difference bias** that would establish a persistent difference in their cumulative weights throughout the MaxEnt learning process. This bias would follow the equation in (121), defined below.

(121) **Difference bias for Dep-V constraints**

$$\alpha \sum_{i=1}^{n} (IR(C_i)w_i - IF(C_i)w_i)$$

As was done for the Markedness $\gg$ Faithfulness bias above, the constraints which are predicted to have the higher weights are **subtracted** from the constraints which are predicted to have the lower weights. This will ensure that the bias is compatible with the MaxEnt learner as a whole.

One issue raised by this method of constructing difference biases is **how small** each constraint category is allowed to be. If a category can be as small as a **single constraint**, then any differences between this constraint and any others with which it has an asymmetrical PMap-style relationship with will be fairly **rigid**. Since any potential subversions to the PMap will be mitigated by the constraints in the same category which adhere to the PMap, any subversions introduced by a single-constraint category will be **unmitigated**, and be penalized much more heavily within the grammar. For grammars which require that the PMap be subverted, encoding it in this way may be detrimental. However, for grammars which would require the PMap to be more rigid, this would be an advantage of this kind of encoding.

The difficulty in implementing the PMap in this way is that it would require a **separate** difference bias to be constructed for **each** inequality present in the PMap, along with a classification function for each set of constraints to be used to establish the difference bias. For the present experiment, this would require a much better understanding of the perceptual differences between segment presence and absence across a variety of contexts in Cantonese, a requirement which is currently lacking, as shown in section 4.4.2.1. While a proxy PMap could conceivably be implemented by theorizing about what cues may or may not be present in these contexts, as was done for experiments 2 and 3 in Chapter 4, this would require additional code to be written for the `solve.R` script, which was not feasible at the time of writing.

Incorporating these biases into a model which attempts to account for patterns of loanword adaptation is predicted to show an improvement with respect to the accuracy of results. With respect to the Markedness $\gg$ Faithfulness bias, incorporating a difference bias for this particular asymmetry will ensure that the weights of the Markedness constraints will, where possible, always be weighted higher than the Faithfulness constraints, no matter the **exact value** of the weights involved. With respect to the PMap, this will not only ensure that the PMap will be directly encoded as a difference in **weights**, as is necessary for a MaxEnt grammar, but that any single **change** in weights will attempt to maintain the distance between these two classes of constraints, regardless of the exact value of the weights at that step. In this way, the PMap can be maintained more readily **throughout learning**, even in the absence of direct phonological evidence. It is thus expected that the patterns of loanword adaption seen in Cantonese should be able to be re-created within the MaxEnt model.

5.6 Conclusion

Throughout this chapter, a variety of improvements and alternate solutions to issues raised over the course of experiments 1–3 in Chapter 4 were discussed. First, the most persistent **discrepancies** between the predictions of the MaxEnt grammars and the attested loanword repairs were examined, and the causes of each were outlined. Some of the discrepancies were due to choices made in order to run simulations. For example, the overall preference for the MaxEnt grammars to prefer **deletion** candidates when trained only on the native Cantonese vocabulary was shown to be due to the lack of deletion repairs in comparison to epenthesis repairs, as provided by the generation script. As the learners had more access to information about how to re-weight the Dep constraints, and as the native Cantonese vocabulary does not repair native lexical items via epenthesis, the Dep constraints were allowed to achieve much higher weights than the Max constraints in these cases. This bias towards deletion was found to be easily overcome when the learner was presented with **direct** evidence of a repair, as was shown when the best-case grammar was examined in more detail. As this grammar had access to the test set, which contained direct evidence about which Dep and Max constraints were required to account for deletion and epenthesis phenomena, and as there were more epenthesis-preferring winners than deletion-preferring winners, this grammar ended up with a bias towards vowel epenthesis. Other discrepancies were due to coding errors in the generation script, and the extent of these coding bugs is unknown at the present time.

The exact nature of the **constraint set** was also examined in detail in this chapter, with three alternate constraint sets discussed. Two of these were not explored in further detail, as they would have substantially altered the set of constraints used in the experiments of Chapter 4, and would have taken time to properly include in the tableau generation script. For example, a set of Faithfulness constraints which made more direct reference to segment features or cues would have required a completely different set of Faithfulness constraints to be defined, and would have greatly expanded the constraint set. Similarly, expanding the set of Markedness constraints was also deemed to be too involved to pursue at the present time. This solution also raised the issue of whether expanding the set of Markedness constraints would compound the issue of PMap **obstruction** discussed in

Chapter 4.

However, a new experiment was run using a set of **overlapping** Faithfulness constraints, and the results of this experiment were presented here. This set of overlapping constraints was **not** found to improve performance of the MaxEnt learners with respect to loanword adaptation, and instead led to less consistent and less accurate results, even for the learners which made use of a weight-based prior. This was deemed to be most likely due to discrepancies already discussed in this chapter, such as small bugs in the generation script, or inclusion of an inconsistent Markedness $\gg$ Faithfulness bias. It was also shown that results from these experiments were more **internally consistent**, irrespective of the bias used. This was found to be due to the condensed constraint set, which reduced the model's reliance on the bias and increased its reliance on the training data. The training data had many more opportunities to influence the weight of any individual constraint, and so the resulting model was more sensitive to the training data.

Finally, improvements to model parameters which would lead to a better match to phonological **theory** were proposed. Some improvements proposed would better model how loanword adaptation proceeds across **multiple generations** of speakers. It was shown that such a model may aid the performance of MaxEnt learners equipped with a plasticity bias, as later generations of speakers would have access to **overt** evidence about loanword adaptation over which a plasticity bias could have effect. Other improvements proposed would better model the original conception of the **PMap** from Steriade (2001). Rather than implementing the PMap as a series of asymmetrical conditions on individual constraints, it was proposed that the PMap be encoded as a series of **difference biases**. This would enable the PMap to be encoded as a series of **weight differences**, and would more directly model the comparative nature of the PMap as implemented in classical OT. It is hoped that, in the future, many of these suggestions and potential improvements may be undertaken, and that they will lead to improved MaxEnt models of loanword adaptation.

Chapter 6

Conclusion

This dissertation has primarily been concerned with testing how computational models of the PMap fare when tasked with accounting for patterns of loanword adaptation. This was carried out in Chapter 4 through a series of experiments which used the native phonology and adaptation processes of Cantonese as a case study. In section 4.3, it was found that, as predicted from the literature on loanword adaptation within Cantonese (Silverman, 1992; Yip, 1993, 2006; Kenstowicz, 2012), the PMap or some other substantive bias is necessary to account for patterns of loanword adaptation within Cantonese. Section 4.4 was concerned with testing the computational model outlined in Wilson (2006), which incorporated the PMap as a series of Gaussian distributions on constraint weights with asymmetrical plasticities. As predicted, these models were not capable of replicating the patterns of loanword adaptation, although they did not entirely behave as expected. Rather, they predicted that deletion should be the preferred repair across the board, a result which was found to be due to the way in which the computational simulation was run. Section 4.5 tested the computational model outlined in White (2013), which incorporated the PMap as a series of Gaussian distributions with asymmetrical means, or target weights. This method of encoding the PMap was found to fare better than that of Wilson (2006), in that it predicted asymmetrical repair strategies for the loanwords of Cantonese even when the bias only approximated the knowledge that a speaker of Cantonese would need to have to account for loanword adaptation in their language. These results were shown in section 5.2 of Chapter 5 to be robust across different constraint sets.

Over the course of performing and reviewing the experiments listed above, it was shown that the study of loanword adaptation proved to be useful when exploring how computational models of grammar will behave. Most specifically, it provides an abundance of real-world data about how speakers treat novel phonological forms in the absence of direct evidence, and it also provides additional insights into the full grammar of these speakers.

The experimental protocol constructed for simulating loanword adaptation in the experiments above has also made more explicit what kinds of assumptions there are about how loanword adaptation proceeds within speakers and across communities. While the experiments pursued here made the (perhaps naïve) assumption that this is a process which occurs uniformly for each speaker of Cantonese within a single generation, this was shown in Chapter 5 to be only one potential model of loanword adaptation. Other models, which take into account additional biases and sources of loanwords, will necessarily require a different methodology when being implemented computationally. The way in which the simulation is set up will have consequences for the efficacy of various kinds of PMap-style biases, and may prove to be more useful for models which make use of a plasticity-based bias.

In addition to the effects of the choices made about the exact method for simulating loanword adaptation within a MaxEnt framework, this thesis has also shown that much more research must be undertaken to flesh out the behaviour of the PMap in languages other than English. While the cue-based literature from Wright (2004) is quite comprehensive, the predictions it makes were not found to be easily translatable into an adequate PMap-style bias for Cantonese, as discussed in section 4.4.2.1. Rather, the weights found by the MaxEnt learner for the best-case grammar were, in many ways, inconsistent with the majority of these predictions, even when the obscuring effects of the Markedness constraints were taken into account. They were much more in line with those explored in Yun (2016), but this covered only a small portion of the whole grammar, and could not aid in explaining why the weights of Max-C constraints varied in the ways that they did.

While these factors have been proven to be complications for the present study, they have nevertheless led back to or opened up new veins of inquiry to be pursued in future research. For instance, while the biases constructed throughout this thesis could only be safely deemed to be approximations of or proxies for the PMap, some of the biases found here could prove to be in line with data from perceptual confusability data, and lead to a better understanding of how various segmental cues are perceived cross-linguistically.

In a similar vein, the fact that biases which incorporated information which was not explicitly due to the PMap fared better than those which only incorporated asymmetrical weights for those constraints which were governed by the PMap raises the question of what other kinds of biases might exist throughout language learning. One such bias – the Markedness $\gg$ Faithfulness bias – has been a mainstay of the phonological learning literature (Gnanadesikan, 2004; Hayes, 2004; Prince and Tesar, 2004). The study of loanword data has also led to the proposal of a set of biases unique to loanword adaptation, such as the Preservation Principle of Paradis and LaCharité (1997), or the bias towards MAX constraints when the orthography of the loanword is taken into account (Smith, 2006; Dohlus, 2010). How might these biases be incorporated into a computational model? Should they be included in the same bias as the PMap, or should different biases be constructed for each?

Another further quesition raised here and not touched on in this thesis is exactly **how** the learner acquires these biases. As discussed in Steriade (2001), the PMap is not considered to be **strictly universal** – speakers of different languages may perceive the same phonetic contrast differently, depending on the grammar they employ. This, in turn, implies that these speakers must acquire certain PMap rankings throughout their lifetime – in other words, that they are learned. Every model explored here has simply treated the PMap as a separate, **static** condition on the MaxEnt learner. Future studies should, if possible, account for how this knowledge can be built up over time, rather than simply focusing on how best to encode it.

Finally, while this thesis has explored the effects that a particular **kind** of extra-grammatical bias can have on the results of modelling language learning through MaxEnt, it has not explored the full range of possibilities in this regard. As discussed in Chapter 5, other biases have been used in conjunction with a MaxEnt learner, such as the L1 bias or the simpler difference bias (Pater and Staubs, 2013). Future studies could examine whether use of these different bias terms has a beneficial effect on modelling loanword adaptation, or whether the approach pursued thus far is adequate. I, for one, look forward to seeing what kinds of results can be obtained through these varied new approaches.

Appendix A

List of English Loanwords in Cantonese

Included here is a list of the loanwords used to construct the **test set** of data used throughout Chapters 4 and 5. The set of loanwords is taken from four main sources: Silverman (1992); Yip (1993, 2006) and Kenstowicz (2012). Many of the words included in these sources come from the phonological grammars of Cantonese compiled by Hashimoto (1972) and Bauer and Benedict (1997), but are supplemented with elicitations performed by the authors. As such, each loanword included here will list the source or sources in which it is cited. S92 refers to Silverman (1992), Y93 to Yip (1993), Y06 to Yip (2006), and K12 to Kenstowicz (2012), respectively.

Each loanword will also be accompanied by an IPA transcription of its pronunciation in RP, as given by the Oxford English Dictionary (Various, 2020). This is done in order to make predictions about how words with vocalized rhotics will behave in Cantonese more transparent. Silverman (1992) and Yip (1993) make the claim that coda rhotics show a unique pattern of deletion within Cantonese loanwords, since these segments are often the ones selected for deletion, despite their being closest to the vowel. However, this pattern assumes that Cantonese speakers treat the vocalized rhotics as **consonants**, which is not guaranteed. A simpler analysis of the loanword adaptation pattern can be established if it is instead assumed that Cantonese speakers treat these as **vowels**, thus obviating the need for an alternate analysis of these segments. Of course, such an analysis also assumes that the variety of English being adapted is RP, as opposed to some other dialect of English, which may not be entirely accurate. However, without additional information about how language contact may have occurred within Cantonese, more accurate transcriptions cannot be provided.

There are a set of loanwords which have been claimed to follow a particular **truncation template** (Silverman, 1992; Yip, 1993; Kenstowicz, 2012). To the best of my knowledge, these words appear to be earlier borrowings (Lo, p.c.), and deal largely with vocabulary that may be used in a university or school setting. As these words appear to make up less of the data than previously supposed, and as the truncation procedure may be exclusive to a particular sociolinguistic domain, these truncated loans were not considered when selecting which data points to include in the test set. They may nevertheless be of interest to future researchers, and so I have included them here.

This appendix will be divided into sections based on the type of structure present in the loanword, as has been done during the discussion of experimental results in Chapters 4 and 5. Each section lists the loanwords alphabetically according to their gloss, grouped by whether they show evidence of deletion or epenthesis, respectively. Loanwords which show truncation patterns are displayed in grey. Any words which have variation in their pronunciation display both, separated by a comma. The page numbers for sources are also given, except if they come from the appendix of Kenstowicz (2012).

A.1 Phonotactically-licit loanwords

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
accoun[tancy]	əˈk^{h}aʊn[tənsi]	aː k^{h}aːŋ	K12 (?)
auntie	ˈɑːnt^{h}i	aːn t^{h}iː	K12
argue	ˈɑːgjuː	aː kiw	K12
baby	ˈbeɪbi	piː piː	K12
baccarat	ˈbakəˌɹɑː	paːk kaː lɔːk	K12
ballet	ˈbaleɪ, ˈbali	paː lej	K12
bar	ˈbɑː	páː	Y93 p.267; K12
barbeque	ˈbɑːbɪˌk^{h}juː	paː piː k^{h}iːw	K12
BBC	ˌbiːˌbiːˈsiː	piː piː siː	K12
bearing	ˈbɛːɹɪŋ	pɛ liŋ, pɛː leŋ	S92 p.297; K12
beer	ˈbiə	pɛː	K12
berry	ˈbɛɹi	pɛː lej	K12
bio[logy]	ˌbʌɪˈɒ[ləd͡ʒi]	pāj ɔ́ː	S92 p.308; K12
body	ˈbɒdi	pɔ́ː tí́ː	S92 p.303; Y93 p.288; K12
boot	ˈbuːt	puːt	K12
bowling	ˈbəʊlɪŋ	pow leŋ	K12
bowtie	ˌbɛʊˈt^{h}ʌɪ	pow taːj	K12
boxing	ˈbɒksɪŋ	pɔ́k sí́ŋ, pɔːk seŋ	Y93 p.266; K12
boy	ˈbɔɪ	pɔːj	K12
boycott	ˈbɔɪˌk^{h}ɒt	puːj k^{h}ɔːt	K12
buffet	ˌbʊˈfeɪ	pōw féj, pow fɛj	S92 p.304; K12
bumper	ˈbʌmpə	pɐ́m pá́ː	S92 p.298; Y93 p.266; K12
bun	ˈbʌn	pɐn	Y06, p.953; K12
bye-bye	bʌjˈbʌj	paːj paːj	K12
cake	ˈk^{h}eɪk	k^{h}ík, k^{h}ek	Y06 pp.968, 969; K12
calorie	ˈk^{h}ałəɹi	k^{h}aː low lej	K12
cancer	ˈk^{h}ansə	k^{h}ɛ́ːn sá́ː	Y93 p.266; Y06 pp.961–2; K12
cap	ˈk^{h}ap	k^{h}iːp	K12

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
captain	ˈkʰaptɪn	kʰɛ́ːp tʰø̋n	Y93 p.266; K12
(car-)coat	ˈkʰəʊt	kʰʊ́k	Y06 p.969
card	ˈkʰɑːd	kʰáːt, kʰɐt	S92 pp.302, 325; Y93 p.267; Y06 p.969; K12
cartoon	ˌkʰɑːˈtʰuːn	kʰaː toŋ	K12
cassette	kʰʌˈsɛt	kʰaː sek	K12
CD	ˌsiːˈdiː	siː tiː	K12
cello	ˈt͡ʃɛləʊ	t͡sʰɛ́ː lő̋w	S92 p.303; K12
cha-cha	ˈt͡ʃɑːt͡ʃɑː	t͡saː t͡saː	K12
chalk	ˈt͡ʃɔːk	t͡sʰɔ́ːk	Y06 p.969; K12
cheap	ˈt͡ʃiːp	t͡sʰíːp	Y06 p.969; K12
check	ˈt͡ʃɛk	t͡sʰɛːk	K12
chem[istry]	ˈkʰɛm[ɪstɹi]	kʰɛːm	S92 p.314; K12
cherry	ˈt͡ʃɛɹi	t͡sʰɛ́ lē̄j	S92 p.313
chocolate	ˈt͡ʃɒkələt	t͡syː kuː lik, t͡syː kuː lek	S92 p.298; K12
chowder	ˈt͡ʃaʊdə	t͡sɐw taː	K12
CID	ˌsiːˌʌɪˈdiː	siː aːj tiː	K12
cider	ˈsʌɪdə	saːj taː	K12
cigar	sɪˈgɑː	sȳːt káː	S92 pp.302–3; K12
cocoa	ˈkʰəʊkʰəʊ	kʰok kʰuː	K12
coffee	ˈkʰɒfi	kʰaː fɛː	K12
cognac	ˈkʰɒnjak	kʰɔːn jɐp	K12
coin	ˈkʰɔɪn	kʰɔːn	K12
cola	ˈkʰəʊlə	kʰɔː laː	K12
cologne	kʰəˈləʊn	kʰuː loŋ	K12
commission	kʰəˈmɪʃən	kʰɔ̄m míː sø̋n, kʰɐ̄m míː sø̋n	S92 p.308; Y93 pp.274, 288; Y06 p.962; K12
com[parative] lit[erature]	kʰəm[ˈparətɪv] ˈlɪt[əɹət͡ʃə]	kʰɐm liːt	K12
compo[sition]	ˌkʰɑmpə[zɪʃən]	kʰɐ́m pʰő̋w	S92 pp.310, 314; K12
computer	kʰəmˈpʰjuːtə	kʰɐm pʰiːw tʰaː	K12
condenser	kʰənˈdɛnsə	kʰɔ̄ːn tɛ́ːn sa̋ː	S92 p.308; K12
cone	ˈkʰəʊn	kʰʊŋ, kʰoŋ	Y93 p.288; K12
cookie	ˈkʰʊki	kʰok kʰiː	K12
copy	ˈkʰɒpi	kʰɐ́p pʰı̋ː	S92 p.323; Y93 p.274; K12
corner	ˈkʰɔːnə	kʰɔːn naː, kʰɔː naː	Y06 p.968; K12
court	ˈkʰɔːt	kʰɔːt	K12
cousin	ˈkʰʌzən	kʰaː sɐn	K12
coxswain	ˈkʰɒksən	kʰɔːk sɐn	K12
curry	ˈkʰʌɹi	kʰː lɛj	K12
cushion	ˈkʰʊʃən	kʰuː søn, kʰuː san	Y06 p.962; K12
cut	ˈkʰʌt	kʰɐ́t	S92 p.298; Y93 p.267; K12
cutlet	ˈkʰʌtlɪt	kʰɐt liːt	K12
cutter	ˈkʰʌtə	kʰɐt tʰaː	K12
daddy	ˈdadi	tɛ́ː tı̋ː	Y93 p.288; Y06 p.961; K12
darling	ˈdɑːlɪŋ	taː leŋ	K12
DDT	ˌdiːˌdiːˈtʰiː	tíː tíː tʰíː	Y93 p.267; K12
dictaphone	ˈdiktəˌfəʊn	tík tā fúŋ	S92 p.306
dinner	ˈdɪnə	tiːn naː	K12
DJ	ˈdiːd͡ʒeɪ	tiː t͡sej	K12

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
dockyard	ˈdɒkˌjaːd	tɔ́ːk jáː	S92 p.304; Y93 p.267; K12
doctor	ˈdɒktə	tɔːk taː	K12
dollar	ˈdɒlə	taː laː	K12
donut	ˈdəʊnʌt	toŋ lɐt	K12
dynamo	ˈdʌɪnəməʊ	taː laːm mow	K12
econ[omics]	ˌiːkəˈn[ɒmiks]	jīː kʰɔ̄ːn, īː kʰɔ̄ːn	S92 p.308; K12
elec[trical]	ɪˈlɛk[tɹɪkəɫ]	jiː lɛːk	K12
encore	ˈɒːŋkɔː	ɛːn kʰɔː	K12
engine	ˈɛnd͡ʒɪn	ɛ́n t͡sɪ́n, ɛː t͡siːn	Y93 p.266; K12
Eng[lish] Lit[erature]	ˈɪŋ[glɪʃ] ˈlɪt[əɹət͡ʃə]	iŋ lit, ɛːŋ liːt	S92 p.316; K12
essay	ˈɛseɪ	ɛː sej	S92 pp.298, 324; K12
fare	ˈfɛː	fej	K12
fashion	ˈfaʃən	fáː sɛ̄́n	S92 p.303; Y06 p.962; K12
Ferrari	fəˈɹɑːɹi	faːt laj lej	K12
fibre	ˈfʌɪbə	faːi paː	K12
fit	fɪt	fit	S92 p.292
foreman	ˈfɔːmən	fɔ́ː mɛ̄́n	Y93 p.274; Y06 p.962; K12
fun	ˈfʌn	fɐn	K12
function	ˈfʌŋʃən	fɐŋ søn	K12
fussy	ˈfʌsi	fɐt siː	K12
gallon	ˈgaɫən	kaː løn	Y06 p.962; K12
game	ˈgeɪm	kɛ́ːm	S92 pp.292, 297, 306; Y93 pp.266–8; K12
garberdine	ˌgabəˈdiːn, ˈgaəˌdiːn	kaː paː tiːn	K12
gay	ˈgeɪ	kej	K12
geo[graphy]	d͡ʒɪˈɒ[gɹəfi]	t͡sɔ́k kā́	S92 p.310
gin	ˈd͡ʒɪn	t͡sín	S92 pp.302, 306
goodbye	ˌgʊdˈbʌɪ	kuːt paːj	K12
guard	ˈgɑːd	kɐt	K12
guitar	gɪˈtʰɑː	kīt tʰáː	S92 pp.302, 323; Y93 p.270; K12
Gurkha	ˈgəːə, ˈgʊəkə	kœː kaː	K12
hello	hə'ləʊ	haː low	K12
hi-fi	ˈhʌɪfʌɪ	hɐj fɐj	K12
high	ˈhʌɪ	haːj	K12
horn	ˈhɔːn	hɔːn	K12
hum	ˈhʌm	hɐm	K12
import	ˈɪmˌpɔːt	ím pʰṓt	S92 p.305
insu[rance]	ɪnˈʃʊə[ɹəns]	īn sɔ́	S92 pp.308, 314
insure	ɪnˈʃɔː, ɪnˈʃəʊ	jiːn sɔː	K12
IQ	ˌʌɪˈkʰjuːw	aːj ksuperhiːw	K12
Jack	ˈd͡ʒak	t͡sɪk, t͡sek	Y06 pp.954, 961; K12
jacket	ˈd͡ʒakɪt	t͡sɛːk kʰɛːt	K12
jam	ˈd͡ʒam	t͡sɛ́ːm	S92 p.306; Y06 p.961; K12
jeep	ˈd͡ʒiːp	t͡siːp	K12
jelly	ˈd͡ʒɛli	t͡sɛː lej	K12
jersey	ˈd͡ʒəːzi	t͡sɛː siː	Y06 p.962; K12
jockey	ˈd͡ʒɒki	t͡sɔːk kʰiː	K12
jumbo	ˈd͡ʒʌmbəʊ	t͡sɐn pow	K12
kid	ˈkʰɪd	kʰiːt	K12

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
king	ˈkʰɪŋ	kʰeŋ	K12
KMB	ˌkʰejˌɛmˈbiː	kʰej ɛːm piː	K12
Kodak	ˈkʰəʊˌdak	kʰɔː taːt	K12
lacquer	ˈlakə	lɪk kʰaː, lek kʰaː	Y06 p.961; K12
laser	ˈleɪzə	løy seː	K12
lemon	ˈlɛmən	liːŋ muːŋ	K12
letter	ˈlɛtə	lɛ́t tʰa̋ː	S92 p.323; Y93 p.274; K12
linen	ˈlɪnɪn	liːn jɐn	K12
lit[erature]	ˈlɪt[əɹət͡ʃə]	liːt	K12
lorry	ˈlɒɹi	lɔː lej, lɔː liː	S92 p.297; K12
lotion	ˈləʊʃən	lɔː sɐn	K12
madam	ˈmadəm	mɛː tuːm	K12
major	ˈmeɪd͡ʒə	mɛ́ː t͡sa̋ː	S92 p.323; Y93 p.274; Y06 p.962; K12
margarine	ˌmɑːd͡ʒɛˈɹiːn	maː t͡siː liːn	K12
margin	ˈmɑːd͡ʒɪn	maː t͡siːn	K12
mark	ˈmɑːk	maːk (v), mɐk (n)	S92 pp.298, 325; K12
market	ˈmɑːkɪt	máː kʰɛ́ːt	S92 p.323; Y93 p.274; K12
market[ing]	ˈmɑːkɪt[ɪŋ]	má kʰɛt	S92 pp.310, 314
maxi	ˈmaksi	mɐt sɐt	Y06 p.961
mechan[ical]	mɪˈkʰan[ɪkəɫ]	mɛːk kʰɛːn	K12
(milk-)shake	ˈʃeɪk	sík, sek	Y06 p.969; K12
mommy	ˈmɒmi	maː miː	K12
motor	ˈməʊtə	mɔ́ː tə̋	S92 pp.298, 304, 323; Y93 p.274; K12
movie	ˈmuːvi	muː fiː	K12
mummy	ˈmʌmi	mɐ́ mı̋	Y93 p.288
NG	ˌɛnˈd͡ʒiː	ɛːn t͡siː	K12
number	ˈnʌmbə	nɐm paː, lam paː	K12 (?)
number one	ˈnʌmbə ˈwʌn	nɐ́m pà wɐ́n, lɐ́m pà wɐ́n	Y93 p.267; K12
nylon	ˈnʌɪˌlɒn	naː loŋ	K12
OK	əʊˈkʰeɪ	ow kʰej	K12
omelette	ˈɒmlɪt	ɐm liːt	K12
oral	ˈɔːɹəɫ	ɔ́ː lőw	Y93 p.273; K12
order	ˈɔːdə	ɔː taː	K12
orlon	ˈɔːlɒn	ɔː løn	K12
OT	ˌəʊˈtʰiː	ow tʰiː	K12
over	ˈəʊvə	ɔː faː	K12
pair	ˈpʰɛː	pʰɛː	S92 p.298; K12
pan	ˈpʰan	pʰaːŋ	Y06 p.961; K12
pancake	ˈpʰanˌkʰeɪk	pʰaːn kʰek	K12
paper	ˈpʰeɪpə	pʰej pa	K12
park	ˈpʰɑːk	pʰaːk	K12
partner	ˈpʰɑːtnə	pʰɐ́t la, pʰaːt naː	Y06 p.969; K12
party	ˈpʰaːti	pʰaːt tʰiː	K12
pea	ˈpʰiː	pʰiː	K12
polit[ical]	pʰəˈlɪt[ɪkəɫ]	pow liːt	S92 p.314; K12
penny	ˈpʰɛni	pʰɛːn niː	K12
philo[sophy]	fɪˈlɒ[səfi]	fiː lɔː	K12
photon	ˈfəʊˌtɒn	fów tʰa̋n	S92 p.305

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
phy[sics]	ˈfɪ[zɪks]	fiː	K12
phy[sics]-chem[istry]-bi[ology]	ˈfɪ[ziks]-kʰɛm[ɪstɹi]-ˌbʌɪ[ˈɒləd͡ʒi]	fíː kʰɛm pāj	S92 p.310
pie	ˈpʰʌɪ	pʰáj	S92 p.298; Y93 pp.266–7; Y06 p.953; K12
pin	ˈpʰɪn	pʰiːn	K12
poker	ˈpʰəʊkə	pʰɔk ka	K12
port	ˈpʰɔːt	pʰuːt	K12
porter	ˈpʰɔːtə	pʰɔ tʰa	Y93 p.274
powder	ˈpʰʊdə	pʰaːw tʰa	K12
psycho[logy]	ˌsʌɪˈkʰɒ[ləd͡ʒi]	saj kʰɔː	S92 p.314; K12
PVC	ˌpʰiːviːˈsiː	pʰiː wiː siː	K12
quali[fication]	ˌkʷʰɒlɪ[fɪˈkʰeɪʃən]	kʰwɔː liː	K12
quart	ˈkʰwɔːt	kʷʰɐt	Y06 p.957; K12
queen	ˈkʰwiːn	kʷʰiːn	Y06 p.957
quinella	kʰwɪˈnɛɫə	kʷʰiːn nɛː laː	K12
ream	ˈɹiːm	liːm	K12
resi[dent]	ˈɹɛzɪ[dənt]	lɛ si	S92 p.314
rifle	ˈɹʌɪfəɫ	láj fő̋w	Y93 p.273
roller	ˈɹəʊlə	low laː	K12
rum	ˈɹʌm	lɐ̋m	Y93 p.267; Y06 p.953; K12
rumba	ˈɹʌmbə	lœːn paː	K12
rupee	ɹuːˈpʰiː, ɹʊˈpʰiː	low pʰej	K12
salad	ˈsaɫəd	sáː lø̋t	S92 p.297; Y93 p.267; Y06 p.961; K12
salmon	ˈsamən	saː muːŋ	K12
saloon	səˈluːn	saː loŋ	K12
sardine	sɑːˈdiːn	saː tiːn	K12
satay	ˈsateɪ	saː tɛː	K12
sauna	ˈsɔːnə	sɔːŋ naː	K12
saxophone	ˈsaksəˌfəʊn	síːk sìː fuːŋ, sek siː foŋ	S92 p.306; K12
set	ˈsɛt	søt	K12
sexy	ˈsɛksi	sɛ́k sí	Y93 p.266; Y06 p.969
sharp	ˈʃɑːp	saːp	S92 p.298; K12
shirt	ˈʃəːt	sœːt	K12
shoot	ˈʃuːt	søt	K12
show	ˈʃəʊ	sow	S92 p.298; K12
shutter	ˈʃʌtə	sɐ́t tʰáː	S92 p.323; Y93 p. 274; K12
sir	ˈsəː	sœː, aː sœː	Y06 p.962; K12
sirloin	ˈsəːlɔɪn	sɐj laːŋ	K12
soci[ology]	ˌsəʊʃɪ[ˈɒləd͡ʒi]	sōw sīː	S92 p.308; K12
socket	ˈsɒkɪt	sɔː kʰiːt	K12
soda	ˈsəʊdə	sɔ́ː táː	S92 pp.298, 304, 323; Y93 p.274; K12
sofa	ˈsəʊfə	sɔː faː	K12
soli[citor]	səˈlɪ[sɪtə]	sow liːt	K12
sorry	ˈsɒɹi	sɔː liː	K12
sugar	ˈʃʊgə	suː kaː	K12
sundae	ˈsʌndeɪ, ˈsʌndi	sɐn taj	K12
T-shirt	ˈtʰiːʃəːt	tʰiː søt	K12

Gloss	English Loan (RP)	Cantonese Adaptation	Sources
tart	ˈtʰɑːt	tʰaːt	K12
taxi	ˈtʰaksi	tɪk siː, tek siː	Y06 p.961; K12
TB	ˌtʰiːˈbiː	tʰiː piː	K12
telephone	ˈtʰɛlɪˌfəʊn	tʰík lōt fúŋ	S92 p.306
tenderloin	ˈtʰɛndəˌlɔɪn	tʰiːn taː løn	K12
thank you	ˈθaŋkˌjuː	teŋ kiːw	K12
thin	ˈθɪn	fin	Y06 p.952
this	ˈðɪs	tis	Y06 p.952
tick	ˈtʰɪk	tʰek	K12
tie	ˈtʰʌɪ	tʰáːj	S92 p.298; Y93 p.267; K12
ton	ˈtʰʌn	tʰøn	K12
tonic	ˈtʰɒnɪk	tʰɔːn nek	K12
tuto[rial]	ˌtʰjuːˈtʰɔː[ɹɪəɫ]	tʰiw tʰɔ	S92 p.316; K12
tutor	ˈtʰjuːtə	tʰíːw tʰa̋ː	Y93 pp.274, 288; Y06 p.957; K12
TV	ˌtʰiːˈviː	tʰiː wiː	K12
twee	ˈtʰwiː	t͡sʰyːj	Y06 p.957
twin	ˈtʰwɪn	tʰyːn	Y06 p.957
u[niversity] li[brary]	ˌjuː[nɪˈvəːsɪti] ˈlʌɪ[bɹəɹi]	ju laj	S92 p. 316
un[derstand]	ˌʌn[dəˈstand]	ɐn	K12
van	ˈvan	wɛːn	K12
vanilla	vəˈnɪlə	wēn lɛ́ la̋, wɐn niː laː	S92 p.323; Y93 p.274; K12
vaseline	ˈvasəˌliːn	faː siː leŋ	K12
very good	ˌvɛɹiˈɡʊd	weː liː kuːt	K12
vitamin	ˈvɪtəmɪn, ˈvʌɪtəmɪn	wɐː taː miŋ	K12
volley	ˈvɒli	wɔː lej	K12
wafer	ˈweɪfə	wɐj faː	K12
waiter	ˈweɪtə	wɐ́j tʰaː	Y06 p.968; K12
Walkman	ˈwɔːkmən	wɔːk mɐn	K12
watt	ˈwɒt	wɔːk	K12
win	ˈwɪn	win	S92 p.292
winner	ˈwɪnə	wɐn naː	K12
wire	ˈwʌɪə	wɐj jaː	K12

Some words which do not run afoul of the highly weighted Markedness constraints *Complex and CodaCond nevertheless do show evidence of being repaired, although there does not appear to be any particular pattern to the repairs undertaken. These words are included here for the same of completeness.

A.1.1 Deletion repairs

Gloss	English Loan (RP)	Cantonese Adaptation	Sources
ATM	ˌeɪtʰiːˈɛm	ej tʰiː ɛː	K12
carat	ˈkʰaɹət	kʰaː	Y06 p.961; K12
dozen	ˈdʌzən	taː	K12
gin sling	ˈd͡ʒɪnˌslɪŋ	t͡siː siː leŋ	K12
loud	ˈlaʊd	laːw	Y06 p.954

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
MC	ˌɛmˈsiː	mə	K12
metre	ˈmiːtə	mɐj	K12
mic	ˈmʌɪk	maːj	Y93 p.288; Y06 p.954; K12
notebook	ˈnəʊtˌbʊk	lɔː bok	K12
pizza	ˈpʰiːtsə, ˈpʰɪtsə	pʰej saː	K12
quarter	ˈkʰwɔːtə	kʷʰɐt	K12
tire	tʰʌɪə	tʰaːj	K12 (?)

A.1.2 Epenthesis repairs

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
cover	ˈkʰʌvə	kʰɐ́p fa̋ː	Y93 p.267; K12
ether	ˈiːθə	jí ta̋	Y93 pp.267, 274
hormone	ˈhɔːməʊn	hoː jiː muŋ	K12
MTR	ˌɛːmˌtʰiːˈaː	ɛːm tʰiː aː low	K12
shilling	ˈʃɪlɪŋ	siːn leŋ	K12

A.2 Loanwords with onset clusters

A.2.1 Deletion repairs

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
blender	ˈblɛndə	pɛ́n ta̋	S92 p.318
broker	ˈbɹəʊkə	pʊ́k kʰa̋ː	S92 pp.290, 318; Y93 p.267; K12
class	ˈkʰlɑːs	kʰaː siː	K12
click	ˈkʰlɪk	kʰek	K12
clip	ˈkʰɪp	kʰiːp	K12
cream	ˈkʰɹiːm	kʰiːm	Y06 p.953
creep	ˈkʰɹiːp	kʰiːp	Y06 p.953
floorshow	ˈflɔːˌʃəʊ	fɔ́ː sów	S92 pp.299, 304; Y93 p.267; K12
freezer	ˈfɹiːzə	fíː sa̋ː	S92 pp.290, 318; Y93 p.267; K12
friend	ˈfɹɛnd	fɛːn	S92 p.324; Y06 p.953; K12
gross	ˈgɾo ˈgɹəʊs	lɔ lɔː	S92 p.301 K12
high-class	ˌhʌɪˈkʰlɑːs	hɐ́j kʰáː sı̋ː, háːj kʰáː sı̋ː	S92 p.320; K12
Jaguar	ˈd͡ʒagjʊə	t͡sɛː kaː	K12
microphone	ˈmʌjkɹəˌfəʊn	mɐj kow foŋ	K12
place	ˈpʰleɪs	pʰej siː	S92 p.318; Y93 p.267; K12
price	ˈpʰɹʌɪs	pʰáj sı̋ː	S92 p.318
printer	ˈpʰɹɪntə	pʰɛ́n tʰa̋	S92 pp.290, 318–9
professor	pʰɹəˈfɛsə	pōw fá sa̋, pʰow fɛː saː	S92 pp.308, 318; K12
prom	ˈpʰɹɒm	pɒŋ	Y06 p.953
proton	ˈpʰɹəʊˌtɒn	pów tʰa̋n	S92 pp.305, 318

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
spleen	ˈspliːn	siː pin	Y06 p.953
spring	ˈspɹɪŋ	siː pɛŋ	Y06 p.953
strawberry	ˈstɹɔːbɛɹi	sìː tʰáw pɛ́ː lɛ̋j	Y93 pp.267, 270; Y06 p.953; K12
twee	ˈtʰwiː	tʰiː	Y06 p.957
twin	ˈtʰwɪn	t͡sʰiːn	Y06 p.957
volume	ˈvaɒljuːm	wɔ́ː lɛ̋m	S92 p.299; Y93 p.267; K12

A.2.2 Epenthesis repairs

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
blonde	ˈblɒnd	pí lán	S92 p.317
blouse	ˈblaʊz	pow lɐw siː	K12
brandy	ˈbɹandi	pɐ̀t láːn tɛ̋j	S92 p.318; K12
break	ˈbɹeɪk	pìk lík, pek lek	S92 pp.290, 303, 317; Y93 p.270; K12
clean	ˈkʰliːn	kʰì lín	S92 p.324
clutch	ˈkʰlʌt͡ʃ	kìk lík t͡sīː, kek lek t͡siː	S92 p.318; Y93 pp.270, 285; K12
cracker	ˈkʰɹakə	haːk lek kaː	K12
cream	ˈkʰɹiːm	kʰèj líːm	S92 pp.390, 303, 317; Y93 p.270; Y06 p.953; K12
Dacron	ˈdakɹɒn	tek kʰɔːk lœːŋ	K12
deuce, duce	ˈdjuːs	tiw siː	K12
flange	ˈfland͡ʒ	fɐt laːn	K12 (?)
flannel	ˈflanəɫ	faːt laːn	K12 (?)
flea	ˈfliː	fù lí	S92 p.324
fluke	ˈfluːk	fù lúk, fuː lok	S92 pp.303–4, 317, 324; Y93 p.270; K12
frank	ˈfɹaŋk	faːt lɔːŋ	K12
grand	ɡɹãn	kak lan	Y93 p.270
percent	pʰəˈsɛnt	pʰœː sɛːn, pʰaː siːn	Y06 p.962; K12
place	ˈpʰlʌɪs	pʰej lej siː	Y93 p.285; K12
pleat	ˈpʰliːt	pʰí lít, pʰe lit	S92 p.324; Y06 p.953
plum	ˈpʰlʌm	pōw lɐ́m	Y93 p.270; Y06 p.953; K12
print	ˈpʰɹɪnt	pʰì lín	S92 pp.290, 317, 319
soufflé	ˈsuːˌfleɪ	sɔː fuː lej	K12
spring	ˈspɹɪŋ	si pit liŋ	S92 p.318; Y93 p.285
straight	ˈstɹeɪt	siː tek lek	K12
twee	ˈtʰwiː	tʰiː wiː	Y06 p.957
twin	ˈtʰwɪn	tʰiː wɪn	Y06 p.957

A.3 Loanwords with coda clusters

A.3.1 Deletion repairs

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
alongside	əˌlɑŋˈsʌɪd	aː lɔːŋ sɐj	K12
assignment	əˈsʌɪnmənt	āː sáj mɐ̋n	S92 p.308; Y06 p.962; K12
band	ˈband	pɛ́ːn	S92 pp.292, 324; Y93 p.267; Y06 p.954; K12
blonde	ˈblɒnd	pí lán	S92 p.317
cement	sɪˈmənt	si mɐn	Y93 p.267
frank	ˈfɹaŋk	faːt lɔːŋ	K12
friend	ˈfɹɛnd	fɛːn	S92 p.324; Y06 p.953; K12
function	ˈfʌŋkʃən	fɐŋ søn	K12
husband	ˈhʌzbənd	haː siː pen	K12
jump ball	ˈd͡ʒʌmpˌbɔːɫ	t͡sɐm pɔː	K12
length	ˈlɛŋkθ	lɛn	S92 p.324
lift	ˈlɪft	liːp	S92 pp.299, 324; Y93 p.282; Y06 p.953; K12
mild	ˈmʌɪɫd	mɐj	K12
mink	ˈmɪŋk	miːŋ	K12
mould	ˈməʊɫd	mow	K12
pound	ˈpʰaʊnd	pʰɔːŋ	K12
pump	ˈpʰʌmp	pʰám	S92 p.292; Y93 p.267; K12
round	ˈɹaʊnd	laːn	K12
sandwich	ˈsandwɪd͡ʒ, ˈsandwɪt͡ʃ	saːm mɐn t͡siː	K12
sergeant	ˈsɑːd͡ʒənt	saː t͡sin	K12
shaft	ˈʃɑːft	sɐ́p, sʌ́p	S92 pp.298–9, 324; Y06 pp.954, 969; K12
sideboard	ˈsʌɪdˌbɔːd	sáj púːt	S92 pp.297, 304; Y93 p.288; K12
sink	ˈsɪŋk	síŋ, seŋ	S92 p.306; Y93 p.267; Y06 p.954; K12
slide	ˈslʌɪd	siː laːj	K12
stamp	ˈstamp	sìː táːm	S92 p.301; Y93 pp. 267, 270; K12
volt	ˈvəʊɫt, ˈvɒɫt	fɔːt	K12
warrant	ˈwɒɹənt	wɔː løn, wɔː leŋ	S92 p.297; K12
wide angle	ˈwʌɪdˌaŋgəɫ	wáːj ɛ́ːŋ kőw	S92 p.320; Y93 p.273; K12
zinc	ˈzɪŋk	sɪŋ	Y06 p.952

A.3.2 Epenthesis repairs

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
film	ˈfɪɫm	féj lɐ̋m, fi lɐm	S92 pp.299, 304, 326; Y93 p.270; K12
golf	ˈɡɒɫf	koː jiː fuː	K12
kiln	ˈkʰɪɫn	kʰí lőn	S92 p.326
valve	ˈvaɫv	wáː lőw	S92 pp.299–301; Y93 pp.267, 270; K12

A.3.3 Deletion and epenthesis repairs

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
soft	ˈsɒft	sɔ́ː fűː	S92 p.325; Y93 pp.270, 282; Y06 p.954; K12

A.4 Loanwords with illicit singleton codas

A.4.1 Sibilant coda deletion repairs

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
gross	ˈɡɹəʊs	lɔː	K12

A.4.2 Sibilant coda epenthesis repairs

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
ace	ˈeɪs	ɛ́j sīː	Y93 p.270; K12
blouse	ˈblaʊz	pow lɐw siː	K12
boss	ˈbɒs	pó sí, pɔː siː	Y93 p.270; K12
bus	ˈbʌs	páː síː	S92 pp.300–1, 304; Y06 p.954; K12
case	ˈkʰeɪs	kʰej siː	K12
cash	ˈkʰaʃ	kʰɛː syː	Y06 p.961; K12
cashmere	ˈkʰaʃˌmɪə	kʰɛ́ː sìː mɛ́ː	Y93 p.270; Y06 p.961; K12
cheese	ˈt͡ʃiːz	t͡siː siː	S92 p.298; K12
cherries	ˈt͡ʃɛɹiz	t͡sʰɛː lɛj t͡siː	K12
class	ˈkʰlɑːs	kʰaː siː	K12
clutch	ˈkʰlʌt͡ʃ	kìk lík t͡sīː, kek lek t͡siː	S92 p.318; Y93 pp.270, 285; K12
deuce, duce	ˈdjuːs	tiw siː	K12
disco	ˈdɪskəʊ	tík sìːków, tek siː kow	S92 p.306; Y93 p.270; K12
face	ˈfeɪs	fej siː	K12
fuse	ˈfjuːz	fiːw siː	Y93 p.288; Y06 p.957; K12
gas	ˈɡas	kɛː siː	K12
gin fizz	ˈd͡ʒɪnˌfɪz	t͡sinː fej siː	K12

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
high-class	ˌhʌɪˈkʰlɑːs	hɐ́j kʰáː síː, háːj kʰáː síː	S92 p.320; K12
husband	ˈhʌzbənd	haː siː pen	K12
king-size	ˈkʰɪŋˌsʌɪz	kʰeŋ saːj siː	K12
lace	ˈleɪs	léj síː	S92 p.304; K12
LC	ˌɛɫˈsiː	ɛː low siː	K12
Miss	ˈmɪs	miː siː	K12
Morris	ˈmɒɹɪs	mɔː lej siː	K12
office	ˈɒfɪs	ɔː fej siː	K12
pass	ˈpʰɑːs	pʰáː síː	S92 p.306; K12
passport	ˈpʰɑːsˌpʰɔːt	pʰáː sìː pʰɔ́ːt	S92 pp.306, 320; Y93 p.270; K12
place	ˈpʰlʌɪs	pʰej lej siː	Y93 p.285; K12
		pʰej siː	S92 p.318; Y93 p.267; K12
pose	ˈpʰəʊz	pʰów síː	Y93 p.267; K12
price	ˈpʰɹʌɪs	pʰáj síː	S92 p.318
raze	ˈɹeɪz	lej siː	Y06 p.952
Royce	ˈɹɔɪs	lɔːj siː	K12
sandwich	ˈsandwɪd͡ʒ, ˈsandwɪt͡ʃ	saːm mɐn t͡siː	K12
size	ˈsaɪz	sáːj síː	S92 p.298; Y93 p.267; Y06 p.954; K12
switch	ˈswɪt͡ʃ	sìː wíː t͡síː	Y93 p.270; Y06 p.957; K12
tennis	ˈtʰɛnɪs	tʰɛːŋ nej siː, tʰɛːn niː siː	K12
whiskey	ˈwɪski	wɐj siː kej	K12

A.4.3 Unrepaired singleton codas

While sibilant codas are the most noticeable of the codas which are banned within the phonology of Cantonese, there are additional codas which are excluded from coda position by CodaCond. The following words belong to this class. Most are repaired, as predicted, although words which end in [ɫ] show variation between deleting the [ɫ] or epenthesizing a vowel after it.

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
poster	ˈpʰəʊstə	pʰɔː sta	K12
zeal	ˈziːɫ	siːl	Y06 p.952

A.4.4 Deletion repairs

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
apple	ˈapəɫ	ɛː poː	K12
ball	ˈbɔːɫ	pɔː	S92 p.297; K12
ball-bearing	ˈbɔːɫˌbɛːɹɪŋ	pɔː pɛː leŋ	K12
call	ˈkʰɔːɫ	kʰɔː	K12
callback	ˈkʰɔːɫˌbak	kʰɔː pɛːk	K12
carnival	ˈkʰɑːnɪvəɫ	kʰaː niːn waː	K12

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
civil	ˈsɪvəɫ	siː fow	K12
cocktail	ˈkʰɒkˌtʰeɪɫ	kʰɔːk tɛː	K12
flannel	ˈflanəɫ	faːt laːn	K12 (?)
foul	ˈfaʊɫ	fɐw	K12
hustle	ˈhʌsəɫ	há sów	Y93 p.273
jump ball	ˈd͡ʒʌmpˌbɔːɫ	t͡sɐm pɔː	K12
label	ˈleɪbəɫ	lej pow	K12
Lysol	ˈlʌɪsɒɫ	laːj sow	K12
mile	ˈmʌɪɫ	mɐj	K12
nickel	ˈnɪkəɫ	nik kow	K12
Ovaltine	ˈəʊvəɫˌtʰiːn	ow waː tʰiːn	K12
ping-pong ball	ˈpʰɪŋˌpʰɐŋˌbɔːɫ	pʰeŋ pʰɐm pɔː	K12
potential	pʰəʊˈtʰɛnʃəɫ	pʰow tʰɛːn sow	K12
rouble	ˈɹuːbəɫ	low pow	K12
sample	ˈsɑːmpəɫ	saːm pʰow	K12
social	ˈsəʊʃəɫ	sow sow	K12
table	ˈtʰeɪbəɫ	tʰej pow	K12
TOEFL	ˈtʰowfəɫ	tʰow fow	K12
uncle	ˈʌŋkəɫ	ɐŋ kʰow	K12
wide angle	ˈwʌɪdˌaŋgəɫ	wáːj ɛ́ːŋ kőw	S92 p.320; Y93 p.273; K12

A.4.5 Epenthesis repairs

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
coil	ˈkʰɔɪɫ	kʰɔːj low	S92 p.301; K12
fail	ˈfeɪɫ	fej low	Y93 p.288; K12
file	ˈfʌɪɫ	fáːj lőw	S92 pp.301, 309; Y93 p.270; Y06 p.954; K12
Hillman	ˈhɪɫmən	hej low mɐn	K12
LC	ˌɛɫˈsiː	ɛː low siː	K12
off	ˈɒf	ɔː fuː	K12
style	ˈstʌɪɫ	siː taːj low	K12
wife	ˈwʌɪf	wɐj fuː	K12

A.5 Loanwords with onset sibilant-consonant clusters

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
gin sling	ˈd͡ʒɪnˌslɪŋ	t͡siː siː leŋ	K12
hysteria	hɪˈstɪəɹɪə	kiːt siː taj ej	K12
score	ˈskɔː	siː kɔː	K12
slick	ˈslɪk	siː lek	K12
slide	ˈslʌɪd	siː laːj	K12
smart	ˈsmɑːt	sìː mék, siː maːt	Y93 p.270; K12
snooker	ˈsnuːkə	siː lok kaː	K12
spanner	ˈspanə	sìː pʰáː lắː	Y93 p.270; K12
spare	ˈspɛː	sìː péː	Y93 pp.267, 270; K12
spark	ˈspɑːk	siː pʰaːk	K12

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
spleen	ˈspliːn	siː pin	Y06 p.953
sports-shirt	ˈspɔːtsˌʃəːt	siː puːt søt	K12
spring	ˈspɹɪŋ	si pit liŋ	S92 p.318; Y93 p.285
		siː pɛŋ	Y06 p.953
stamp	ˈstamp	sìː táːm	S92 p.301; Y93 pp. 267, 270; K12
start	ˈstɑːt	siː tʰaːt	K12
stati[stics]	stəˈtʰɪ[stɪks]	siː tɛːt	K12
steam	ˈstiːm	siː tiːm	K12
stick	ˈstɪk	sìː tík, siː tek	S92 pp.297, 303–4; Y93 p.270; K12
store	ˈstɔː	síː tɔ̋ː	S92 p.301; Y93 p.270; K12
straight	ˈstɹeɪt	siː tek lek	K12
strawberry	ˈstɹɔːbɛɹi	sìː tʰáw pɛ́ː lɛ̋j	Y93 pp.267, 270; Y06 p.953; K12
style	ˈstʌɪɫ	siː taːj low	K12
sway	ˈsweɪ	siː wej	Y06 p.957
switch	ˈswɪt͡ʃ	sìː wíː t͡sı̋ː	Y93 p.270; Y06 p.957; K12
swoon	ˈswuːn	siː wuːn	Y06 p.957

A.6 Loanwords with coda consonant-sibilant clusters

A.6.1 Deletion repairs

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
flange	ˈfland͡ʒ	fɐt laːn	K12 (?)
jeans	ˈd͡ʒiːnz	t͡siːn	K12
license	ˈlʌɪsəns	laːj sɐn	Y93 p.288; K12
sports-shirt	ˈspɔːtsˌʃəːt	siː puːt søt	K12

A.6.2 Epenthesis repairs

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
arts	ˈɑːts	aː t͡siː	K12
Benz	ˈbɛnz	pɛːn siː	K12
billiards	ˈbɪɫjədz	piː lɛːt	K12
chance	ˈt͡ʃɑːns	t͡sʰaːn siː	K12
fans	ˈfanz	feːn siː, faːn syː	K12 (?)
guts	ˈgʌts	kɐt siː	K12
inch	ˈɪnt͡ʃ	íːn t͡sı̋ː	Y93 p.270; K12
maths	ˈmaθs	mɛːt siː	K12
notes	ˈnəʊts	nók sı̋ː, nɔ́ːk sı̋ː	Y06 p.969; K12
orange	ˈɒɹɪnd͡ʒ	ɔː løn t͡siː	K12
ounce	ˈaʊns	ɔːn siː	K12
pence	ˈpʰɛns	pʰiːn siː	Y06 p.954; K12

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
postcard	ˈp^{h}əʊstˌk^{h}ɑːd	p^{h}ów sì k^{h}ák, p^{h}ow siː k^{h}aːt	S92 p.306; K12
sex	ˈsɛks	sɛːk siː	K12
tips	ˈt^{h}ɪps	t^{h}íːp síː	S92 pp.301, 319; Y93 p.270; K12
Volks[wagon]	ˈvəʊks[ˌwɑːgən]	fok siː	K12
X	ˈɛks	ék sìː	S92 p.313

A.6.3 Deletion and epenthesis repairs

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
Beatles	ˈbiːtəɫz	p^{h}ej t^{h}aw siː	K12
cents	ˈsənts	siːn siː	K12
Rolls	ˈɹəʊɫz	low siː	K12

A.7 Loanwords with coda sibilant-consonant clusters

Gloss	*English Loan (RP)*	*Cantonese Adaptation*	*Sources*
cast	ˈk^{h}ɑːst	k^{h}aː siː	Y93 pp.267, 270; Y06 p.954; K12
forecast	ˈfɔːˌk^{h}ɑːst	fɔ́ː k^{h}áː síː	S92 p.320; K12
last	ˈlɑːst	laː siː	K12
mask	ˈmɑːsk	maː siː	K12
minced	ˈmɪnst	miːn t͡siː	K12
post	ˈp^{h}əʊst	p^{h}ów síː	S92 p.306; Y93 p.270; K12
Sunkist	ˈsʌnˌk^{h}ɪst	sɐn k^{h}ej siː	K12
toast	ˈt^{h}əʊst	t^{h}ɔː siː	K12
waste	ˈweɪst	wɐj siː	S92 pp.301, 303; K12
yeast	ˈjiːst	jiː siː	K12

Appendix B

List of Constraints and Model Parameters

The tables below contain a list of the constraints used throughout this dissertation, divided according to whether they are Markedness or Faithfulness constraints, and according to whether they make use of unique segmental trigrams or overlapping contexts. A general definition format for each kind of constraint is provided at the outset of each section. The table will contain the model parameters used for each bias as well.

B.1 Markedness constraints

The Markedness constraints used here can be roughly divided into **prosodic** constraints and **phonotactic** constraints. Some prosodic constraints are familiar from classical OT (Prince and Smolensky, 1993), while others are adapted for use with a MaxEnt model. The two most unique are CodaCond (Silverman, 1992) and Bimoraic (Kenstowicz, 2012). Each of these constraints is defined in the table below.

Constraint	*Description*
Onset	Every syllable has an onset.
*Coda	No syllable has a coda.
*Complex	No syllable onset or coda contains more than one segment.
CodaCond	No coda may contain segments outside the set { p, t, k, m, n, ŋ, j, w }.
Peak(Nas)	Every syllable peak is of sonorancy Nasal or higher.
Peak(Liq)	Every syllable peak is of sonorancy Liquid or higher.
Peak(Vowel)	Every syllable peak is a vowel.
Bimoraic	Every syllable is bimoraic.

The phonotactic constraints *OP, *oT, *IK, and *TOT are drawn from Kenstowicz (2012), and the others are inferred from the typology of syllables present in Bauer and Benedict (1997). Each of these constraints takes the form of a **ban** on certain sequences of segments. They are defined in the table below.

Constraint	*Description*
*OP	[+round] vowels are not followed by [Labial] consonants.
*oT	[o] is not followed by a consonant that is [+anterior].
*IK	[+high] vowels are not followed by [Velar] consonants.
*#V	Vowels { i, u, ø, œ, y } are not allowed syllable-initially.
*POE	A [Labial] consonant is not followed by a [−back, +round] vowel.
*NOE	A [+nasal] consonant is not followed by a [−back, +round] vowel.
*Tu	A [Coronal] consonant is not followed by [u].
*KWI	A labialized velar consonant is not followed by a [+high] vowel.
*NGI	A [ŋ] is not followed by a [+high] vowel.
*NGE	A [ŋ] is not followed by a [−back, −round] vowel.
*PVP	No syllable has an onset and a coda that are both [Labial].
*TOT	No syllable has a vowel that is [+back, +round] and an onset and coda that are both [Coronal].

The constraint *ET from Kenstowicz (2012) was not included here, due to the fact that the syllabary contained within Bauer and Benedict (1997) noted that this constraint would rule out a large number of syllables which are observed only in loanwords. In order to prevent the model from enforcing this constraint, it was left out of the current analysis.

The biases assigned to the Markedness constraints are provided in the table below. Unless otherwise noted, the **default** Markedness ≫ Faithfulness bias is assumed to be the one used in a given experiment.

	M≫F (default)		Plasticity (extreme)		Weight (extreme)	
Constraint	μ	σ^2	μ	σ^2	μ	σ^2
Onset	10.0	100.0	0.0	44.20053113	6.648348	1.0
*Coda	10.0	100.0	0.0	$1.00E-08$	0.0	1.0
*Complex	10.0	100.0	0.0	153.995864	12.409507	1.0
CodaCond	10.0	100.0	0.0	142.601309	11.94157901	1.0
Peak(Nas)	10.0	100.0	0.0	99.9781012	9.99890501	1.0
Peak(Liq)	10.0	100.0	0.0	0.969809375	0.98478901	1.0
Peak(Vowel)	10.0	100.0	0.0	4.484551641	2.11767601	1.0
Bimoraic	10.0	100.0	0.0	16.47465297	4.05889801	1.0
*OP	10.0	100.0	0.0	99.98872032	9.99943601	1.0
*oT	10.0	100.0	0.0	3.161308892	1.77800701	1.0
*IK	10.0	100.0	0.0	$1.00E-08$	0.0	1.0
*#V	10.0	100.0	0.0	96.97987388	9.84783601	1.0
*POE	10.0	100.0	0.0	99.98872032	9.99943601	1.0
*NOE	10.0	100.0	0.0	99.98872032	9.99943601	1.0
*Tu	10.0	100.0	0.0	31.68781383	5.62919301	1.0
*KWI	10.0	100.0	0.0	99.98872032	9.99943601	1.0
*NGI	10.0	100.0	0.0	99.98872032	9.99943601	1.0
*NGE	10.0	100.0	0.0	99.98872032	9.99943601	1.0
*PVP	10.0	100.0	0.0	38.8869851	6.23594301	1.0
*TOT	10.0	100.0	0.0	4.195094565	2.04819301	1.0

B.2 Faithfulness constraints (unique trigrams)

The Faithfulness constraints used in experiments 1–3 all made use of contextual constraints which were capable of being weighted **independently** of one another. As discussed in Chapter 4, this was achieved by creating a unique segmental **trigram** for each constraint. Constraints were defined according to the following constraint templates.

(122) **Template for Dep constraints (unique trigrams)**

Dep-X/Y_Z: Segments of type X which are preceded by Y and followed by Z in the output correspond to segments of type X in the input.

(123) **Template for Max constraints (unique trigrams)**

Max-X/Y_Z: Segments of type X which are preceded by Y and followed by Z in the input correspond to segments of type X in the output.

The full list of constraints is provided in the tables below, along with the biases used in experiments 1–3. The plasticity-based biases are presented first, and then the weight-based biases are presented. The default Markedness $\gg$ Faithfulness bias is provided in both, for comparison.

B.2.1 Plasticity biases

	M≫F (default)		Plasticity (extreme)		Plasticity (psuedo)	
Constraint	μ	σ^2	μ	σ^2	μ	σ^2
Dep-T/#_#	0.0	100.0	0.0	$1.00E-08$	0.0	100.0
Dep-S/#_#	0.0	100.0	0.0	$3.969E-09$	0.0	100.0
Dep-R/#_#	0.0	100.0	0.0	$3.969E-09$	0.0	100.0
Dep-V/#_#	0.0	100.0	0.0	$3.969E-09$	0.0	91.44766024
Dep-T/#_T	0.0	100.0	0.0	$3.6E-09$	0.0	100.0
Dep-S/#_T	0.0	100.0	0.0	$3.6E-09$	0.0	100.0
Dep-R/#_T	0.0	100.0	0.0	$3.6E-09$	0.0	100.0
Dep-V/#_T	0.0	100.0	0.0	3.958734978	0.0	3.958734988
Dep-T/#_R	0.0	100.0	0.0	$3.844E-09$	0.0	100.0
Dep-S/#_R	0.0	100.0	0.0	$3.844E-09$	0.0	100.0
Dep-R/#_R	0.0	100.0	0.0	$3.844E-09$	0.0	100.0
Dep-V/#_R	0.0	100.0	0.0	1.597567075	0.0	5.840400233
Dep-T/#_V	0.0	100.0	0.0	$7.48023E-06$	0.0	100.0
Dep-S/#_V	0.0	100.0	0.0	$7.48023E-06$	0.0	100.0
Dep-R/#_V	0.0	100.0	0.0	$7.48023E-06$	0.0	100.0
Dep-V/#_V	0.0	100.0	0.0	0.630956972	0.0	100.0
Dep-T/T_#	0.0	100.0	0.0	$4.096E-09$	0.0	100.0
Dep-S/T_#	0.0	100.0	0.0	$4.096E-09$	0.0	100.0
Dep-R/T_#	0.0	100.0	0.0	$4.096E-09$	0.0	100.0
Dep-V/T_#	0.0	100.0	0.0	19.3116742	0.0	91.44766024
Dep-T/T_T	0.0	100.0	0.0	$6.084E-09$	0.0	100.0
Dep-S/T_T	0.0	100.0	0.0	$6.084E-09$	0.0	100.0
Dep-R/T_T	0.0	100.0	0.0	$6.084E-09$	0.0	100.0
Dep-V/T_T	0.0	100.0	0.0	47.30006127	0.0	91.44766024

Constraint	M≫F (default) μ	σ^2	Plasticity (extreme) μ	σ^2	Plasticity (psuedo) μ	σ^2
DEP-T/T_R	0.0	100.0	0.0	$3.844E-09$	0.0	100.0
DEP-S/T_R	0.0	100.0	0.0	$3.844E-09$	0.0	100.0
DEP-R/T_R	0.0	100.0	0.0	$3.844E-09$	0.0	100.0
DEP-V/T_R	0.0	100.0	0.0	5.840400223	0.0	100.0
DEP-T/T_V	0.0	100.0	0.0	$4.624E-09$	0.0	100.0
DEP-S/T_V	0.0	100.0	0.0	$4.624E-09$	0.0	100.0
DEP-R/T_V	0.0	100.0	0.0	$4.624E-09$	0.0	100.0
DEP-V/T_V	0.0	100.0	0.0	10.25697983	0.0	5.840400233
DEP-T/R_#	0.0	100.0	0.0	$3.721E-09$	0.0	100.0
DEP-S/R_#	0.0	100.0	0.0	$3.721E-09$	0.0	100.0
DEP-R/R_#	0.0	100.0	0.0	$3.721E-09$	0.0	100.0
DEP-V/R_#	0.0	100.0	0.0	44.99861903	0.0	91.44766024
DEP-T/R_T	0.0	100.0	0.0	$3.969E-09$	0.0	100.0
DEP-S/R_T	0.0	100.0	0.0	$3.969E-09$	0.0	100.0
DEP-R/R_T	0.0	100.0	0.0	$3.969E-09$	0.0	100.0
DEP-V/R_T	0.0	100.0	0.0	28.89574223	0.0	91.44766024
DEP-T/R_R	0.0	100.0	0.0	0.001004256	0.0	100.0
DEP-S/R_R	0.0	100.0	0.0	0.001004256	0.0	100.0
DEP-R/R_R	0.0	100.0	0.0	0.001004256	0.0	100.0
DEP-V/R_R	0.0	100.0	0.0	91.44766023	0.0	91.44766024
DEP-T/R_V	0.0	100.0	0.0	$5.329E-09$	0.0	100.0
DEP-S/R_V	0.0	100.0	0.0	$5.329E-09$	0.0	100.0
DEP-R/R_V	0.0	100.0	0.0	$5.329E-09$	0.0	100.0
DEP-V/R_V	0.0	100.0	0.0	2.555406862	0.0	100.0
DEP-T/V_#	0.0	100.0	0.0	$3.844E-09$	0.0	100.0
DEP-S/V_#	0.0	100.0	0.0	$3.844E-09$	0.0	100.0
DEP-R/V_#	0.0	100.0	0.0	$3.844E-09$	0.0	100.0
DEP-V/V_#	0.0	100.0	0.0	1.159317345	0.0	100.0
DEP-T/V_T	0.0	100.0	0.0	$4.489E-09$	0.0	100.0
DEP-S/V_T	0.0	100.0	0.0	$4.489E-09$	0.0	100.0
DEP-R/V_T	0.0	100.0	0.0	$4.489E-09$	0.0	100.0
DEP-V/V_T	0.0	100.0	0.0	3.255458796	0.0	100.0
DEP-T/V_R	0.0	100.0	0.0	$3.6E-09$	0.0	100.0
DEP-S/V_R	0.0	100.0	0.0	$3.6E-09$	0.0	100.0
DEP-R/V_R	0.0	100.0	0.0	$3.6E-09$	0.0	100.0
DEP-V/V_R	0.0	100.0	0.0	4.976803266	0.0	100.0
DEP-T/V_V	0.0	100.0	0.0	$3.969E-09$	0.0	100.0
DEP-S/V_V	0.0	100.0	0.0	$3.969E-09$	0.0	100.0
DEP-R/V_V	0.0	100.0	0.0	$3.969E-09$	0.0	100.0
DEP-V/V_V	0.0	100.0	0.0	$1.17718E-05$	0.0	100.0
MAX-T/#_#	0.0	100.0	0.0	$3.969E-09$	0.0	$1.40E-08$
MAX-S/#_#	0.0	100.0	0.0	$3.969E-09$	0.0	117.6548274
MAX-R/#_#	0.0	100.0	0.0	$3.969E-09$	0.0	$1.40E-08$
MAX-V/#_#	0.0	100.0	0.0	$3.969E-09$	0.0	19.41979811
MAX-T/#_T	0.0	100.0	0.0	$3.969E-09$	0.0	$1.40E-08$
MAX-S/#_T	0.0	100.0	0.0	117.6548274	0.0	117.6548274
MAX-R/#_T	0.0	100.0	0.0	$3.969E-09$	0.0	$1.40E-08$
MAX-V/#_T	0.0	100.0	0.0	110.0213027	0.0	110.0213027

	M≫F (default)		Plasticity (extreme)		Plasticity (psuedo)	
Constraint	μ	σ^2	μ	σ^2	μ	σ^2
Max-T/#_R	0.0	100.0	0.0	45.02473074	0.0	117.6542634
Max-S/#_R	0.0	100.0	0.0	37.95856175	0.0	37.95856176
Max-R/#_R	0.0	100.0	0.0	$3.969E-09$	0.0	9.484786159
Max-V/#_R	0.0	100.0	0.0	76.35002262	0.0	76.35002263
Max-T/#_V	0.0	100.0	0.0	2.928013789	0.0	117.6542634
Max-S/#_V	0.0	100.0	0.0	1.162881857	0.0	37.95856176
Max-R/#_V	0.0	100.0	0.0	1.170325858	0.0	9.484786159
Max-V/#_V	0.0	100.0	0.0	1.186927449	0.0	100.0
Max-T/T_#	0.0	100.0	0.0	$1.00E-08$	0.0	$1.40E-08$
Max-S/T_#	0.0	100.0	0.0	67.99598826	0.0	117.6548274
Max-R/T_#	0.0	100.0	0.0	$3.969E-09$	0.0	$1.40E-08$
Max-V/T_#	0.0	100.0	0.0	15.06550661	0.0	19.41979811
Max-T/T_T	0.0	100.0	0.0	$3.969E-09$	0.0	$1.40E-08$
Max-S/T_T	0.0	100.0	0.0	$3.969E-09$	0.0	117.6548274
Max-R/T_T	0.0	100.0	0.0	$3.969E-09$	0.0	$1.40E-08$
Max-V/T_T	0.0	100.0	0.0	1.192092749	0.0	19.41979811
Max-T/T_R	0.0	100.0	0.0	$3.969E-09$	0.0	117.6542634
Max-S/T_R	0.0	100.0	0.0	$3.969E-09$	0.0	37.95856176
Max-R/T_R	0.0	100.0	0.0	$3.969E-09$	0.0	9.484786159
Max-V/T_R	0.0	100.0	0.0	24.27118064	0.0	76.35002263
Max-T/T_V	0.0	100.0	0.0	117.6542634	0.0	117.6542634
Max-S/T_V	0.0	100.0	0.0	0.11356563	0.0	37.95856176
Max-R/T_V	0.0	100.0	0.0	9.484786149	0.0	9.484786159
Max-V/T_V	0.0	100.0	0.0	0.312718061	0.0	100.0
Max-T/R_#	0.0	100.0	0.0	$1.00E-08$	0.0	39.27208037
Max-S/R_#	0.0	100.0	0.0	25.3011515	0.0	99.09943673
Max-R/R_#	0.0	100.0	0.0	29.65539799	0.0	42.39687143
Max-V/R_#	0.0	100.0	0.0	0.538381725	0.0	19.41979811
Max-T/R_T	0.0	100.0	0.0	$1.00E-08$	0.0	39.27208037
Max-S/R_T	0.0	100.0	0.0	$3.969E-09$	0.0	99.09943673
Max-R/R_T	0.0	100.0	0.0	29.65539799	0.0	42.39687143
Max-V/R_T	0.0	100.0	0.0	2.172699584	0.0	19.41979811
Max-T/R_R	0.0	100.0	0.0	$3.66025E-07$	0.0	32.59125378
Max-S/R_R	0.0	100.0	0.0	$3.969E-09$	0.0	28.09775975
Max-R/R_R	0.0	100.0	0.0	$3.969E-09$	0.0	32.31280128
Max-V/R_R	0.0	100.0	0.0	19.4197981	0.0	19.41979811
Max-T/R_V	0.0	100.0	0.0	$4.356E-09$	0.0	32.59125378
Max-S/R_V	0.0	100.0	0.0	0.099821243	0.0	28.09775975
Max-R/R_V	0.0	100.0	0.0	0.057034992	0.0	32.31280128
Max-V/R_V	0.0	100.0	0.0	0.036222083	0.0	100.0
Max-T/V_#	0.0	100.0	0.0	39.27208036	0.0	39.27208037
Max-S/V_#	0.0	100.0	0.0	76.78086298	0.0	99.09943673
Max-R/V_#	0.0	100.0	0.0	42.39687142	0.0	42.39687143
Max-V/V_#	0.0	100.0	0.0	$3.9854E-05$	0.0	100.0
Max-T/V_T	0.0	100.0	0.0	13.29697684	0.0	39.27208037
Max-S/V_T	0.0	100.0	0.0	99.09943672	0.0	99.09943673
Max-R/V_T	0.0	100.0	0.0	34.21894329	0.0	42.39687143
Max-V/V_T	0.0	100.0	0.0	0.251645699	0.0	100.0

Constraint	M≫F (default) μ	σ^2	Plasticity (extreme) μ	σ^2	Plasticity (psuedo) μ	σ^2
MAX-T/V_R	0.0	100.0	0.0	32.59125377	0.0	32.59125378
MAX-S/V_R	0.0	100.0	0.0	28.09775974	0.0	28.09775975
MAX-R/V_R	0.0	100.0	0.0	32.31280127	0.0	32.31280128
MAX-V/V_R	0.0	100.0	0.0	1.771662157	0.0	100.0
MAX-T/V_V	0.0	100.0	0.0	0.750732137	0.0	32.59125378
MAX-S/V_V	0.0	100.0	0.0	0.476262854	0.0	28.09775975
MAX-R/V_V	0.0	100.0	0.0	1.042959733	0.0	32.31280128
MAX-V/V_V	0.0	100.0	0.0	$3.969E-09$	0.0	100.0

B.2.2 Weight biases

Constraint	M≫F (default) μ	σ^2	Weight (extreme) μ	σ^2	Weight (psuedo) μ	σ^2
DEP-T/#_#	0.0	100.0	0.0	1.0	0.0	100.0
DEP-S/#_#	0.0	100.0	$6.30E-05$	1.0	0.0	100.0
DEP-R/#_#	0.0	100.0	$6.30E-05$	1.0	0.0	100.0
DEP-V/#_#	0.0	100.0	$6.30E-05$	1.0	9.562827	100.0
DEP-T/#_T	0.0	100.0	$6.00E-05$	1.0	0.0	100.0
DEP-S/#_T	0.0	100.0	$6.00E-05$	1.0	0.0	100.0
DEP-R/#_T	0.0	100.0	$6.00E-05$	1.0	0.0	100.0
DEP-V/#_T	0.0	100.0	1.989657	1.0	1.989657	100.0
DEP-T/#_R	0.0	100.0	$6.20E-05$	1.0	0.0	100.0
DEP-S/#_R	0.0	100.0	$6.20E-05$	1.0	0.0	100.0
DEP-R/#_R	0.0	100.0	$6.20E-05$	1.0	0.0	100.0
DEP-V/#_R	0.0	100.0	1.263949	1.0	2.416692	100.0
DEP-T/#_V	0.0	100.0	0.002735	1.0	0.0	100.0
DEP-S/#_V	0.0	100.0	0.002735	1.0	0.0	100.0
DEP-R/#_V	0.0	100.0	0.002735	1.0	0.0	100.0
DEP-V/#_V	0.0	100.0	0.794328	1.0	0.0	100.0
DEP-T/T_#	0.0	100.0	$6.40E-05$	1.0	0.0	100.0
DEP-S/T_#	0.0	100.0	$6.40E-05$	1.0	0.0	100.0
DEP-R/T_#	0.0	100.0	$6.40E-05$	1.0	0.0	100.0
DEP-V/T_#	0.0	100.0	4.394505	1.0	9.562827	100.0
DEP-T/T_T	0.0	100.0	$7.80E-05$	1.0	0.0	100.0
DEP-S/T_T	0.0	100.0	$7.80E-05$	1.0	0.0	100.0
DEP-R/T_T	0.0	100.0	$7.80E-05$	1.0	0.0	100.0
DEP-V/T_T	0.0	100.0	6.877504	1.0	9.562827	100.0
DEP-T/T_R	0.0	100.0	$6.20E-05$	1.0	0.0	100.0
DEP-S/T_R	0.0	100.0	$6.20E-05$	1.0	0.0	100.0
DEP-R/T_R	0.0	100.0	$6.20E-05$	1.0	0.0	100.0
DEP-V/T_R	0.0	100.0	2.416692	1.0	2.416692	100.0
DEP-T/T_V	0.0	100.0	$6.80E-05$	1.0	0.0	100.0
DEP-S/T_V	0.0	100.0	$6.80E-05$	1.0	0.0	100.0
DEP-R/T_V	0.0	100.0	$6.80E-05$	1.0	0.0	100.0
DEP-V/T_V	0.0	100.0	3.202652	1.0	0.0	100.0

	M≫F (default)		Weight (extreme)		Weight (psuedo)	
Constraint	μ	σ^2	μ	σ^2	μ	σ^2
Dep-T/R_#	0.0	100.0	$6.10E-05$	1.0	0.0	100.0
Dep-S/R_#	0.0	100.0	$6.10E-05$	1.0	0.0	100.0
Dep-R/R_#	0.0	100.0	$6.10E-05$	1.0	0.0	100.0
Dep-V/R_#	0.0	100.0	6.708101	1.0	9.562827	100.0
Dep-T/R_T	0.0	100.0	$6.30E-05$	1.0	0.0	100.0
Dep-S/R_T	0.0	100.0	$6.30E-05$	1.0	0.0	100.0
Dep-R/R_T	0.0	100.0	$6.30E-05$	1.0	0.0	100.0
Dep-V/R_T	0.0	100.0	5.375476	1.0	9.562827	100.0
Dep-T/R_R	0.0	100.0	0.03169	1.0	0.0	100.0
Dep-S/R_R	0.0	100.0	0.03169	1.0	0.0	100.0
Dep-R/R_R	0.0	100.0	0.03169	1.0	0.0	100.0
Dep-V/R_R	0.0	100.0	9.562827	1.0	9.562827	100.0
Dep-T/R_V	0.0	100.0	$7.30E-05$	1.0	0.0	100.0
Dep-S/R_V	0.0	100.0	$7.30E-05$	1.0	0.0	100.0
Dep-R/R_V	0.0	100.0	$7.30E-05$	1.0	0.0	100.0
Dep-V/R_V	0.0	100.0	1.598564	1.0	0.0	100.0
Dep-T/V_#	0.0	100.0	$6.20E-05$	1.0	0.0	100.0
Dep-S/V_#	0.0	100.0	$6.20E-05$	1.0	0.0	100.0
Dep-R/V_#	0.0	100.0	$6.20E-05$	1.0	0.0	100.0
Dep-V/V_#	0.0	100.0	1.076716	1.0	0.0	100.0
Dep-T/V_T	0.0	100.0	$6.70E-05$	1.0	0.0	100.0
Dep-S/V_T	0.0	100.0	$6.70E-05$	1.0	0.0	100.0
Dep-R/V_T	0.0	100.0	$6.70E-05$	1.0	0.0	100.0
Dep-V/V_T	0.0	100.0	1.804289	1.0	0.0	100.0
Dep-T/V_R	0.0	100.0	$6.00E-05$	1.0	0.0	100.0
Dep-S/V_R	0.0	100.0	$6.00E-05$	1.0	0.0	100.0
Dep-R/V_R	0.0	100.0	$6.00E-05$	1.0	0.0	100.0
Dep-V/V_R	0.0	100.0	2.230875	1.0	0.0	100.0
Dep-T/V_V	0.0	100.0	$6.30E-05$	1.0	0.0	100.0
Dep-S/V_V	0.0	100.0	$6.30E-05$	1.0	0.0	100.0
Dep-R/V_V	0.0	100.0	$6.30E-05$	1.0	0.0	100.0
Dep-V/V_V	0.0	100.0	0.003431	1.0	0.0	100.0
Max-T/#_#	0.0	100.0	$6.30E-05$	1.0	$6.30E-05$	100.0
Max-S/#_#	0.0	100.0	$6.30E-05$	1.0	10.846881	100.0
Max-R/#_#	0.0	100.0	$6.30E-05$	1.0	$6.30E-05$	100.0
Max-V/#_#	0.0	100.0	$6.30E-05$	1.0	4.40679	100.0
Max-T/#_T	0.0	100.0	$6.30E-05$	1.0	$6.30E-05$	100.0
Max-S/#_T	0.0	100.0	10.846881	1.0	10.846881	100.0
Max-R/#_T	0.0	100.0	$6.30E-05$	1.0	$6.30E-05$	100.0
Max-V/#_T	0.0	100.0	10.489104	1.0	10.489104	100.0
Max-T/#_R	0.0	100.0	6.710047	1.0	10.846855	100.0
Max-S/#_R	0.0	100.0	6.161052	1.0	6.161052	100.0
Max-R/#_R	0.0	100.0	$6.30E-05$	1.0	3.079738	100.0
Max-V/#_R	0.0	100.0	8.73785	1.0	8.73785	100.0
Max-T/#_V	0.0	100.0	1.711144	1.0	10.846855	100.0
Max-S/#_V	0.0	100.0	1.07837	1.0	6.161052	100.0
Max-R/#_V	0.0	100.0	1.081816	1.0	3.079738	100.0
Max-V/#_V	0.0	100.0	1.089462	1.0	0.0	100.0

	M≫F (default)		Weight (extreme)		Weight (psuedo)	
Constraint	μ	σ^2	μ	σ^2	μ	σ^2
Max-T/T_#	0.0	100.0	0.0	1.0	$6.30E-05$	100.0
Max-S/T_#	0.0	100.0	8.245968	1.0	10.846881	100.0
Max-R/T_#	0.0	100.0	$6.30E-05$	1.0	$6.30E-05$	100.0
Max-V/T_#	0.0	100.0	3.881431	1.0	4.40679	100.0
Max-T/T_T	0.0	100.0	$6.30E-05$	1.0	$6.30E-05$	100.0
Max-S/T_T	0.0	100.0	$6.30E-05$	1.0	10.846881	100.0
Max-R/T_T	0.0	100.0	$6.30E-05$	1.0	$6.30E-05$	100.0
Max-V/T_T	0.0	100.0	1.09183	1.0	4.40679	100.0
Max-T/T_R	0.0	100.0	$6.30E-05$	1.0	10.846855	100.0
Max-S/T_R	0.0	100.0	$6.30E-05$	1.0	6.161052	100.0
Max-R/T_R	0.0	100.0	$6.30E-05$	1.0	3.079738	100.0
Max-V/T_R	0.0	100.0	4.926579	1.0	8.73785	100.0
Max-T/T_V	0.0	100.0	10.846855	1.0	10.846855	100.0
Max-S/T_V	0.0	100.0	0.336995	1.0	6.161052	100.0
Max-R/T_V	0.0	100.0	3.079738	1.0	3.079738	100.0
Max-V/T_V	0.0	100.0	0.559212	1.0	0.0	100.0
Max-T/R_#	0.0	100.0	0.0	1.0	6.266744	100.0
Max-S/R_#	0.0	100.0	5.030025	1.0	9.95487	100.0
Max-R/R_#	0.0	100.0	5.445677	1.0	6.511288	100.0
Max-V/R_#	0.0	100.0	0.733745	1.0	4.40679	100.0
Max-T/R_T	0.0	100.0	0.0	1.0	6.266744	100.0
Max-S/R_T	0.0	100.0	$6.30E-05$	1.0	9.95487	100.0
Max-R/R_T	0.0	100.0	5.445677	1.0	6.511288	100.0
Max-V/R_T	0.0	100.0	1.474008	1.0	4.40679	100.0
Max-T/R_R	0.0	100.0	0.000605	1.0	5.708875	100.0
Max-S/R_R	0.0	100.0	$6.30E-05$	1.0	5.300732	100.0
Max-R/R_R	0.0	100.0	$6.30E-05$	1.0	5.684435	100.0
Max-V/R_R	0.0	100.0	4.40679	1.0	4.40679	100.0
Max-T/R_V	0.0	100.0	$6.60E-05$	1.0	5.708875	100.0
Max-S/R_V	0.0	100.0	0.315945	1.0	5.300732	100.0
Max-R/R_V	0.0	100.0	0.23882	1.0	5.684435	100.0
Max-V/R_V	0.0	100.0	0.190321	1.0	0.0	100.0
Max-T/V_#	0.0	100.0	6.266744	1.0	6.266744	100.0
Max-S/V_#	0.0	100.0	8.762469	1.0	9.95487	100.0
Max-R/V_#	0.0	100.0	6.511288	1.0	6.511288	100.0
Max-V/V_#	0.0	100.0	0.006313	1.0	0.0	100.0
Max-T/V_T	0.0	100.0	3.646502	1.0	6.266744	100.0
Max-S/V_T	0.0	100.0	9.95487	1.0	9.95487	100.0
Max-R/V_T	0.0	100.0	5.849696	1.0	6.511288	100.0
Max-V/V_T	0.0	100.0	0.501643	1.0	0.0	100.0
Max-T/V_R	0.0	100.0	5.708875	1.0	5.708875	100.0
Max-S/V_R	0.0	100.0	5.300732	1.0	5.300732	100.0
Max-R/V_R	0.0	100.0	5.684435	1.0	5.684435	100.0
Max-V/V_R	0.0	100.0	1.331038	1.0	0.0	100.0
Max-T/V_V	0.0	100.0	0.866448	1.0	5.708875	100.0
Max-S/V_V	0.0	100.0	0.690118	1.0	5.300732	100.0
Max-R/V_V	0.0	100.0	1.021254	1.0	5.684435	100.0
Max-V/V_V	0.0	100.0	$6.30E-05$	1.0	0.0	100.0

B.3 Faithfulness constraints (overlapping)

The Faithfulness constraints used in experiment 4 made use of contextual constraints whose contexts **overlapped**, leading to ganging-up effects in the MaxEnt grammar. Constraints were defined according to the following constraint templates.

(124) **Templates for Dep constraints**

a. Dep-X/Y_: Segments of type X which are preceded by Y in the output correspond to segments of type X in the input.

b. Dep-X/_Y: Segments of type X which are followed by Y in the output correspond to segments of type X in the input.

(125) **Templates for Max constraints**

a. Max-X/Y_: Segments of type X which are preceded by Y in the input correspond to segments of type X in the output.

b. Max-X/_Y: Segments of type X which are followed by Y in the input correspond to segments of type X in the output.

The full list of constraints and biases used in experiment 4 are provided below.

B.3.1 Weight biases

	M≫F (default)		Weight (extreme)		Weight (psuedo)	
Constraint	μ	σ^2	μ	σ^2	μ	σ^2
Dep-S	0.0	100.0	0.097893408	100.0	0.002208188	100.0
Dep-T/T_	0.0	100.0	0.00116172	100.0	$6.80E-05$	100.0
Dep-T/R_	0.0	100.0	0.198971592	100.0	0.00797175	100.0
Dep-T/V_	0.0	100.0	0.000863538	100.0	$6.30E-05$	100.0
Dep-T/_T	0.0	100.0	0.000726904	100.0	$6.70E-05$	100.0
Dep-T/_R	0.0	100.0	0.185139453	100.0	0.0079685	100.0
Dep-T/_V	0.0	100.0	0.02082007	100.0	0.00073475	100.0
Dep-R/T_	0.0	100.0	0.000992599	100.0	$6.80E-05$	100.0
Dep-R/R_	0.0	100.0	0.12004136	100.0	0.00797175	100.0
Dep-R/V_	0.0	100.0	0.000762772	100.0	$6.30E-05$	100.0
Dep-R/_T	0.0	100.0	0.000662542	100.0	$6.70E-05$	100.0
Dep-R/_R	0.0	100.0	0.112153161	100.0	0.0079685	100.0
Dep-R/_V	0.0	100.0	0.011850524	100.0	0.00073475	100.0
Dep-V/T_	0.0	100.0	2.515595834	100.0	4.22283825	100.0
Dep-V/R_	0.0	100.0	8.305602493	100.0	5.811242	100.0
Dep-V/V_	0.0	100.0	1.846057199	100.0	1.27882775	100.0
Dep-V/_T	0.0	100.0	0.90250205	100.0	4.0117315	100.0
Dep-V/_R	0.0	100.0	1.678779361	100.0	3.86858575	100.0
Dep-V/_V	0.0	100.0	1.451907182	100.0	1.39974375	100.0
Max-S	0.0	100.0	3.621023114	100.0	3.54523375	100.0
Max-T/T_	0.0	100.0	1.209034177	100.0	2.71174525	100.0
Max-T/R_	0.0	100.0	0.0	100.0	0.00016775	100.0
Max-T/V_	0.0	100.0	6.084566003	100.0	4.12214225	100.0

Constraint	M≫F (default) μ	σ^2	Weight (extreme) μ	σ^2	Weight (psuedo) μ	σ^2
MAX-T/_T	0.0	100.0	1.209034177	100.0	0.911657	100.0
MAX-T/_R	0.0	100.0	8.201963722	100.0	3.1048975	100.0
MAX-T/_V	0.0	100.0	0.0	100.0	3.35612825	100.0
MAX-R/T_	0.0	100.0	2.323446965	100.0	0.76998175	100.0
MAX-R/R_	0.0	100.0	3.739465076	100.0	2.78255925	100.0
MAX-R/V_	0.0	100.0	6.439702961	100.0	4.76666825	100.0
MAX-R/_T	0.0	100.0	8.148275122	100.0	2.82387475	100.0
MAX-R/_R	0.0	100.0	3.739465076	100.0	1.421156	100.0
MAX-R/_V	0.0	100.0	2.349584968	100.0	1.355407	100.0
MAX-V/T_	0.0	100.0	2.891437681	100.0	2.614763	100.0
MAX-V/R_	0.0	100.0	2.11125331	100.0	1.701216	100.0
MAX-V/V_	0.0	100.0	1.098501093	100.0	0.45976425	100.0
MAX-V/_T	0.0	100.0	10.63804322	100.0	3.38914625	100.0
MAX-V/_R	0.0	100.0	9.125958178	100.0	4.85056425	100.0
MAX-V/_V	0.0	100.0	1.098100974	100.0	0.4597645	100.0

B.3.2 Plasticity bias

Constraint	M≫F (default) μ	σ^2	Plasticity (pseudo) μ	σ^2
DEP-S	0.0	100.0	0.0	$4.89E-06$
DEP-T/T_	0.0	100.0	0.0	$1.46E-08$
DEP-T/R_	0.0	100.0	0.0	$6.36E-05$
DEP-T/V_	0.0	100.0	0.0	$1.40E-08$
DEP-T/_T	0.0	100.0	0.0	$1.45E-08$
DEP-T/_R	0.0	100.0	0.0	$6.35E-05$
DEP-T/_V	0.0	100.0	0.0	$5.50E-07$
DEP-R/T_	0.0	100.0	0.0	$1.46E-08$
DEP-R/R_	0.0	100.0	0.0	$6.36E-05$
DEP-R/V_	0.0	100.0	0.0	$1.40E-08$
DEP-R/_T	0.0	100.0	0.0	$1.45E-08$
DEP-R/_R	0.0	100.0	0.0	$6.35E-05$
DEP-R/_V	0.0	100.0	0.0	$5.50E-07$
DEP-V/T_	0.0	100.0	0.0	17.8323629
DEP-V/R_	0.0	100.0	0.0	33.77053359
DEP-V/V_	0.0	100.0	0.0	1.635400424
DEP-V/_T	0.0	100.0	0.0	16.09398964
DEP-V/_R	0.0	100.0	0.0	14.96595572
DEP-V/_V	0.0	100.0	0.0	1.959282576
MAX-S	0.0	100.0	0.0	12.56868235
MAX-T/T_	0.0	100.0	0.0	7.353562311
MAX-T/R_	0.0	100.0	0.0	$3.81E-08$
MAX-T/V_	0.0	100.0	0.0	16.99205674
MAX-T/_T	0.0	100.0	0.0	0.831118496
MAX-T/_R	0.0	100.0	0.0	9.640388496
MAX-T/_V	0.0	100.0	0.0	11.26359684

	M≫F (default)		Plasticity (pseudo)	
Constraint	μ	σ^2	μ	σ^2
Max-R/T_	0.0	100.0	0.0	0.592871905
Max-R/R_	0.0	100.0	0.0	7.74263599
Max-R/V_	0.0	100.0	0.0	22.72112622
Max-R/_T	0.0	100.0	0.0	7.974268614
Max-R/_R	0.0	100.0	0.0	2.019684386
Max-R/_V	0.0	100.0	0.0	1.837128146
Max-V/T_	0.0	100.0	0.0	6.836985556
Max-V/R_	0.0	100.0	0.0	2.894135889
Max-V/V_	0.0	100.0	0.0	0.211383176
Max-V/_T	0.0	100.0	0.0	11.48631231
Max-V/_R	0.0	100.0	0.0	23.52797355
Max-V/_V	0.0	100.0	0.0	0.211383405

Bibliography

Adler, Allison N. 2006. Faithfulness and perception in loanword adaptation: A case study from hawaiian. *Lingua* 116:1024–1045.

Bauer, Robert S., and Paul K. Benedict. 1997. *Modern cantonese phonology*. Berlin: Walter de Gruyter.

Boersma, Paul, and Silke Hamann. 2009. Loanword adaptation as first-language phonological perception. In *Loan phonology*, ed. A. Calabrese and L. Wetzels, 11–58. Amsterdam: Benjamins.

Boersma, Paul, and Clara Levelt. 2000. Gradual constraint-ranking learning algorithm predicts acquisiton order. In *Proceedings of 30th Child Language Research Forum*, ed. EV Clark, 299–337. Stanford, CA: CSLI Publications.

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu. 1995. A limited memory algorithm for bound constrained optimization. *SIAM Journal of Scientific Computing* 16:1190–1208.

Delgutte, B., and N. Y. S. Kiang. 1984a. Speech coding in the auditory nerve: Iii. voiceless fricative consonants. *Journal of the Acoustical Society of America* 75:887–896.

Delgutte, B., and N. Y. S. Kiang. 1984b. Speech coding in the auditory nerve: Iv. sounds with consonant-like dynamic characteristics. *Journal of the Acoustical Society of America* 75:897–907.

Della Pietra, Stephen, Vincent J. Della Pietra, and John D. Lafferty. 1997. Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19:380 – 393.

Do, Young Ah. 2013. Biased learning of phonological alternations. Doctoral Dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Dohlus, Katrin. 2010. The role of phonology and phonetics in loanword adaptation: German and french rounded vowels in japanese. Doctoral Dissertation, Frankfurt.

Eguchi, Shinto, Osamu Komori, and Atsumi Ohara. 2014. Duality of maximum entropy and minimum divergence. *Entropy* 16:3552–3572.

Flemming, Edward. 1995. Perceptual features in phonology. Doctoral Dissertation, University of California Los Angeles, Los Angeles, CA.

Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18:7–44.

Flemming, Edward. 2008. Asymmetries between assimilation and epenthesis. Paper presented at the 82nd Annual Meeting of the LSA.

Gnanadesikan, Amalia. 2004. Markedness and faithfulness constraints in child phonology. In *Constraints in phonological acquisition*, ed. René Kager, Joe Pater, and Wim Zonneveld, chapter 3, 73–108. Cambridge: Cambridge University Press.

Goldwater, Sharon, and Mark Johnson. 2003. Learning ot constraint rankings using a maximum entropy model. In *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, ed. Jennifer Spenader, Anders Eriksson, and Östen Dahl, 111–120. Stockholm, Sweden: Stockholm University, Department of Linguistics.

Guion, S. G. 1998. The role of perception in the sound change of velar palatalization. *Phonetica* 55:18–52.

Hashimoto, Oi-kan Yue. 1972. *Phonology of cantonese*. Cambridge: Cambridge University Press.

Hayes, Bruce. 2004. Phonological acquisiton in optimality theory: the early stages. In *Constraints in phonological acquisition*, ed. René Kager, Joe Pater, and Wim Zonneveld, chapter 5, 158–203. Cambridge: Cambridge University Press.

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonetic learning. *Linguistic Inquiry* 39:379–440.

I., Miller. M., and M. B. Sachs. 2983. Representation of stop consonants in the discharge patterns of auditory nerve fibres. *Journal of the Acoustical Society of America* 74:502–517.

Jäger, Gerhard, and Anette Rosenbach. 2006. The winner takes it all — almost: Cumulativity in grammatical variation. *Linguistics* 44:937–971.

Jesney, Karen. 2011. Cumulative constraint interaction in phonological acquisition and typology. Doctoral Dissertation, University of Massachusetts Amerhst, Amherst, MA.

Jesney, Karen. 2016. Positional constraints in optimality theory and harmonic grammar. In *Harmonic grammar and harmonic serialism*, ed. John J. McCarthy and Joe Pater. Equinox Publishing.

Jesney, Karen, and Anne-Michelle Tessier. 2008. Gradual learning and faithfulness: Consequences of ranked vs. weighted constraints. In *Proceedings of NELS 38*, ed. M. Abdurrahman, A. Schardl, and Martin Walkow.

Jun, Jongho. 2004. Place assimilation. In *Phonetically based phonology*, ed. Bruce Hayes and Donca Steriade, chapter 3, 58–86. Cambridge: Cambridge University Press.

Kenstowicz, Michael J. 2005. The phonetics and phonology of korean loanword adaptation. Ms, MIT. To appear in *Proceedings of First European Conference on Korean LInguistics*, ed. S.–J. Rhee.

Kenstowicz, Michael J. 2007. Salience and similarity in loanword adaptation: a case study from fijian. *Language Sciences* 29:316–340.

Kenstowicz, Michael J. 2012. Cantonese loanwords: Conflicting faithfulness in vc rime constraints. *Catalan Journal of Linguistics* 11:65–96.

Kullback, Solomon, and Richard A. Leibler. 1951. On information and sufficiency. *Annals of Mathematical Statistics* 22:79–86.

Legendre, Géraldine, Yoshiro Miyata, and Paul Smolensky. 1990. Harmonic grammar. a formal multi-level connectionist theory of linguistic well-formedness: An application. In *Proceedings of the 12^{th} Annual Conference of the Cognitive Science Society*, 884–891. Hillsdale, NJ: Lawrence Erlbaum.

Magri, Giorgio. 2012. Convergence of error-driven ranking algorithms. *Phonology* 29:213–269.

McCarthy, John J. 2010. An introduction to harmonic serialism. In *Selected works of john j. mccarthy*. Amherst, Massachusetts: bepress.

McPherson, Laura. 2016. Cumulativity and ganging in the tonology of awa suffixes. *Language: Phonological Analysis* .

Merrill, John. 2015. Nasalization as a repair for voiced obstruent codas in noon. *LSA Annual Meeting Extended Abstracts* 6:1–5.

Miller, G. A., and P. E. Nicely. 1955. An analysis of perceptual confusions among some english consonants. *Journal of the Acoustical Society of America* 27:338–352.

Nosofsky, R. M. 1986. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115:39–57.

O'Hara, Charlie. 2016. Harmony in harmonic grammar by reevaluating faithfulness. In *Proceedings of NELS 46*, ed. Christopher Hammerly and Brandon Prickett. CreateSpace Independent Publishing Platform.

Olson, Erin. 2016. Intermediate markedness. Ms. Massachusetts Institute of Technology.

Pakendorf, B., and I. Novgorodov. 2009. *Sakha*. Leipzig: Max Plank Institude for Evolutionary Anthropology.

Paradis, Carole, and LaCharité. 1997. Preservation and minimality in loanword adaptation. *Journal of Linguistics* 33:379–430.

Pater, Joe. 1996a. Austronesian nasal substitution and other nC̥ effects. In *The prosody-morphology interface*, ed. René Kager, Harry van der Hulst, and Wim Zonneveld, chapter 8, 310–343. Cambridge: Cambridge University Press.

Pater, Joe. 1996b. Austronesian nasal substitution and other nC̥ effects. In *The prosody-morphology interface*, ed. René Kager, Harry van der Hulst, and Wim Zonneveld, chapter 8, 310–343. Cambridge: Cambridge University Press.

Pater, Joe. 2009. Weighted constraints in generative linguistics. *Congitive Science* 33:999–1035.

Pater, Joe, and Robert Staubs. 2013. Modeling learning trajectories with batch gradient descent. Paper presented at NECPhon, MIT, October 2013.

Peperkamp, Sharon, Inga Vendelin, and Kimihiro Nakamura. 2008. On the perceptual origin of loanword adaptations: experimental evidence from japanese. *Phonology* 25:129–164.

Pleco Software. 2016. Cc-canto. Http://cantonese.org/.

Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt, and Michael Becker. 2010. Harmonic grammar with linear programming: from linear systems to linguistic typology. *Phonology* 27:77–117.

Prince, Alan S., and Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. Technical Report 2, Rutgers University Center for Cognitive Science.

Prince, Alan S., and Bruce Tesar. 2004. Learning phonotactic distributions. In *Constraints in phonological acquisition*, ed. René Kager, Joe Pater, and Wim Zonneveld, chapter 8, 245–291. Cambridge: Cambridge University Press.

R Core Team. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Shinohara, Shigeko. 2006. Perceptual effects in final cluster reduction patterns. *Lingua* 116:1046–1078.

Silverman, Daniel. 1992. Multiple scansions in loanword phonology: evidence from cantonese. *Phonology* 9:289–328.

Sinex, D. G. 1995. Auditory nerve fibre representation of cues to voicing in syllable-final stop contexts. *Journal of the Acoustical Society of America* 90:2441–2449.

Sinex, D. G., and C. D. Geisler. 1983. Responses of auditory-nerve fibres to consonant-vowel syllables. *Journal of the Acoustical Society of America* 73:602–615.

Smith, Jennifer L. 2006. Loan phonology is not all perception: evidence from japanese loan doublets. In *Japanese/korean linguistics*, ed. T. J. Vance, volume 14, 63–74. Stanford, CA: CSLI Publications.

Smolensky, Paul. 1996. On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27:720–731.

Smolensky, Paul, and Géraldine Legendre. 2006. *The harmonic mind: From neural computation to optimality-theoretic grammar*. Cambridge, MA: MIT Press.

Steriade, Donca. 2001. The phonology of perceptibility effects: the p-map and its consequences for constraint organization. Ms., UCLA.

Tesar, Bruce, and Paul Smolensky. 2000. *Learnability in optimality theory*. Cambridge, MA: MIT Press.

Tessier, Anne-Michelle. 2009. Frequency of violation and constraint-based phonological learning. *Lingua* 119:6–38.

Uffmann, Christian. 2015. Loanword adaptation. In *The oxford handbook of historical phonology*, ed. Patrick Honeybone and Joseph Salmons, 644–666. Oxford, UK: Oxford University Press.

Various. 2020. *Oxford english dictionary*. Oxford, UK, 3 edition.

Vasilyeva, L. 2010. Vowel harmony for russian loanwords into yakut. Master's thesis, California State University, Fresno.

Wang, M. D., and R. C. Bilger. 1973. Consonant confusions in noise. *Journal of the Acoustical Society of America* 54:1248–1266.

White, James C. 2013. Bias in phonological learning: Evidence from saltation. Doctoral Dissertation, University of California Los Angeles, Los Angeles, CA.

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30:945–982.

Wright, Richard. 2004. A review of perceptual cues and cue robustness. In *Phonetically based phonology*, ed. Bruce Hayes and Donca Steriade, chapter 2, 34–57. Cambridge: Cambridge University Press.

Yip, Moira. 1993. Cantonese loanword phonology and optimality theory. *Journal of East Asian Linguistics* 2:261–291.

Yip, Moira. 2006. The symbiosis between perception and grammar in loanword phonology. *Lingua* 116:950–975.

Yun, Suyeon. 2016. A theory of consonant cluster perception and vowel epenthesis. Doctoral Dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Zuraw, Kie. 2007. The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from tagalog. *Language* 83:277–316.

Zwicky, Arnold. 1976. Well, this rock and roll has got to stop. junior's head is hard as a rock. In *Papers from the 12th Meeting of the Chicago Linguistic Society*. University of Chicato, Chicago, IL: Chicago Linguistic Society.