ALTERNATION-SENSITIVE PHONEME LEARNING: IMPLICATIONS FOR CHILDREN'S DEVELOPMENT AND LANGUAGE CHANGE

Caitlin Richter

A DISSERTATION

in

Linguistics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2021

Supervisor of Dissertation

Charles Yang, Professor of Linguistics & Computer Science

Supervisor of Dissertation

Don Ringe, Professor of Linguistics

Graduate Group Chairperson

Eugene Buckley, Associate Professor of Linguistics

Dissertation Committee:

Daniel Swingley, Professor & Chair of Psychology

ABSTRACT

ALTERNATION-SENSITIVE PHONEME LEARNING: IMPLICATIONS FOR CHILDREN'S DEVELOPMENT AND LANGUAGE CHANGE

Caitlin Richter

Charles Yang, Don Ringe

This dissertation develops a cognitive model describing when children learn to group distinct sound segments (allophones) into abstract equivalence classes (phonemes). The allophones an individual acquires are arbitrary and determined by their particular input, yet intricately involved in language cognition once learned. The acquisition model characterises the role of surface segment alternations in children's input by using the Tolerance Principle (Yang 2016) to evaluate the cognitive cost of possible phoneme inventory structures iteratively as a child's vocabulary grows. The model traces the emergence of abstract representations from concrete speech stimuli, starting from a default representation where underlying contrasts simply mirror surface-segment contrasts (Invariant Transparency Hypothesis, Ringe & Eska 2013).

A longitudinal corpus study of four children's alveolar stop and flap productions establishes that English medial flap allophony follows a U-shaped acquisition course, which is characteristic of learning linguistic rules or generalisations. The cognitive model is validated by accurately predicting the timing of changes in each child's productions, which signal allophone acquisition. A second case study models the historical process of secondary split in Menominee mid and high back vowels. The acquisition model serves as an independently motivated quantitative test for the occurrence of phonemic split, providing an alternative to traditional reliance on linguists' case-specific subjective judgements about when it might occur. The third case study examines the phonemic status of the velar nasal in German, showing how the acquisition model can discriminate between tolerable grammars and the subset of tolerable grammars that are learnable, with implications for the relationship between formal language description and psychological representation. This dissertation's approach synthesises insights from computational modelling, naturalistic corpus data, historical linguistics, and experimental research on child language acquisition.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The basic units of phonetic and phonological representation are integral to linguistic theory, and identifying their correlates in psychological reality is essential to understanding them. Children's discovery of phonemes and allophonic relations, as surface alternations accumulate in their growing vocabulary, is a vital element of their language acquisition. Allophonic processes in a grammar (1.1, 1.2, 1.3) represent learned systematic generalisations about distributional patterns in the surface forms of lexical items. These productive generalisations are the alternative to a lexicon filled with memorised sets of surface-based representations, where predictable regularities in relationships between some set members (1.1b, 1.2b) would appear incidental.

(1.1) a. Medial alveolar flapping (English):

[ɾ] when intervocalic preceding unstressed vowel.

[t]/[d] otherwise

b.

light	lighting	nesting	study	boulder	attempt
[lʌɪt]	[lʌɪɾɪŋ]	[nɛstɪŋ]	[stʌɾi]	[boʊldəɹ]	[ətəmpt]

(1.2) a. tt ∼ t consonant gradation (Finnish):

[tː] as onset of open syllable (strong grade)

[t] otherwise (weak grade)

b. juttu jutut jutun juttujen 'story'
soitto soitot soiton soittojen 'musical sound'

tyttö	tytöt	tytön	tyttöjen	'girl'
NOM.SG	NOM.PL	GEN.SG	GEN.PL	

(1.3) a. Final obstruent devoicing in German:

[t]/[k]/[s]/etc. only, when in a morpheme-final syllable coda.

Both voiced and unvoiced variants, elsewhere.

b.

Tag	sag-t	Kind	Haus	gab
[taːk]	[zaːkt]	[kɪnt]	[haus]	[gaːp]
day	*says*	*child*	*house*	*gave*
Tag-e	sag-en	kind-isch	Häus-er	geb-en
[taːgə]	[zaːgn̩]	[kɪndɪʃ]	[hɔyzɐ]	[geːbn̩]
days	*say*	*childish*	*houses*	*give*
Decke	trink-en	natür-lich	Füß-e	Treppe
[dɛkə]	[trɪŋkn̩]	[natyːɐ̯liç]	[fyːsə]	[trɛpə]
cover	*drink*	*of-course*	*feet*	*staircase*

Phonetic transcriptions from Duden Ausspracheworterbuch (Mangold 1990)

The allophones an individual acquires are arbitrary and determined by their particular input, yet intricately involved in language cognition once learned (Whalen et al. 1997). Adults can have difficulty consciously discriminating between perceptually related allophones that infants in the same language environment easily distinguish (Pegg and Werker 1997). An allophonic grammar also enables productive generalisation from one form of a lexical item to unheard forms, and seemingly allows language users to more efficiently process allomorphs (Bekemeier 2016; Reifegerste and Clahsen 2018).

1.1 Alternation-sensitive phoneme learning (APL)

This dissertation proposes a cognitively motivated Alternation-sensitive Phoneme Learning (APL) model describing the pivotal first step of allophone acquisition, when input causes a child to treat surface-distinct segments as members of a single linguistic class (phoneme). In this model, allophone acquisition is driven by the development of a lexicon that includes alternating forms, and representations emerge from the distributions of surface sounds in a child's input considered in the context of ongoing lexical, semantic, and morphological development (Stoel-Gammon 2011). Children learn allophonic relationships when alternations as in (1.1b, 1.2b) are sufficiently prevalent, but conversely, without direct alternations children will not unite multiple allophones into one abstract phoneme, even if that would also be compatible with the distributional facts of their lexicon.

APL models initial allophone acquisition by (i) proposing an early stage in which surface distinctions specify underlying contrasts (Peperkamp et al. 2006; Ringe and Eska 2013), and (ii) applying the Tolerance Principle (Yang 2016), a quantitative grammar evaluation model motivated by processing efficiency, to assess this proposal given the child's accumulating input. If it is not tolerable, a correspondence between some surface-distinct segments must be learned instead; evaluation iterates with increasing child vocabularies. Grounding the acquisition of phonology in surface distinctions, with a defined criterion for departure from the surface-based grammar, addresses perennial questions concerning how abstract an abstract phonology might be (Kiparsky 1982). APL sheds light on the critical developmental process of resolving surface percepts from acoustic stimuli into a system of abstract equivalence relations, and describes how gradually accumulating input can bring about more sudden transitions in language development (Jarosz 2019; Vihman 2014).

The predictive ability of APL is illustrated with a case study of learning to group North American English alveolar flaps [ɾ] together with segments like [t,d,ʔ]. The flap, as in *water* and *pudding*, can alternate with another realisation: *eat*∼*eating*, *dad*∼*daddy*. Children develop articulatory control of these distinctions several years before fully mastering their phonological conditioning (Eddington 2007; Klein and Altman 2002; Song et al. 2015). This development is observed in a longitudinal corpus study of an understudied age range (Chapter 4), finding typical U-shaped learning, 'getting better phonologically by getting worse phonetically' (Macken 1995: §2.5). Children's first words often show U-shaped development of word and syllable structures as they acquire and mis-apply productive generalisations (Vihman 2014, 2019), and overgeneralisation of regular affixes is also common in morphological acquisition (Cazden 1968); evidently phoneme/allophone segments are yet another linguistic level with nonmonotonic development. APL accounts for this by predicting an initial period when memorising surface contrasts is tolerable, until a critical point when alternations become prevalent enough to require learning a systematic relationship. Input vocabularies can be estimated with corpora of child-directed speech, and the model's predictions for alveolar flap allophone learning accord with independent developmental data as well as our corpus analysis of the flap acquisition course.

Now that APL has been validated as a synchronic acquisition model, it can be used to research how historical language change is shaped through acquisition (Kiparsky 1982). For instance, secondary split is a common and not fully explained way that languages gain new phonemes from former allophones (Hoenigswald 1960; Kiparsky 2015; Ringe and Eska 2013). APL can account for this as a regular outcome of child language acquisition, if independent changes affect the appearance of conditioned allophony in the vocabulary: when input loses sufficient cognitively accessible evidence of alternations, a default maximally-contrastive grammar remains tolerable throughout acquisition. This process is possible to model as in Menominee high back vowels, where collapse of the short vowel system reduced the prevalence of obvious [o] ∼ [u] surface alternations; these were once predictably conditioned allophones but are now two phonemes (Bloomfield 1962; Richter 2020). While many historical instances of secondary split are qualitatively explained by identifying the independent language changes thought to have critically destroyed the allophones' conditioning environment, APL precisely quantifies the relationship between the presence of alternations and the learnability of phoneme systems, and can measure the development of this relationship in historical corpora.

APL also offers a learnability-based perspective on extensionally equivalent competing analyses of language. For example, the framework proposes that some dialect variation in German devoicing and velar nasal assimilation (Vennemann 1970) is most parsimoniously understood as change in underlying phonological inventory structure (gaining a velar nasal phoneme), rather than change solely in the ordering of productive rule application, because the 're-ordered' system has a striking lack of observable alternations to help children identify the necessary velar nasal assimilation process.

These applications show that APL may model learning in 'perhaps the most interesting and challenging' acquisition scenarios, involving variable, inconsistent, or uninformative input (Jarosz 2019: 3). These properties are extremely common and may produce either language change or stability across generations. Considering the general relationship between language variation and change, the model's freedom from artificially homogeneous input assumptions is a significant benefit. Just as predictions about flap allophone development are verified by comparison to children's behaviour, predicted outcomes from historical corpora may be verified by comparison with actual language change. APL can additionally illuminate the relationship between traditional descriptive grammars, which may encode a deep explanation of a language's history, and the cognitive representations that directly guide speakers' behaviour.

1.2 Structure of the dissertation

Chapter 2 first reviews empirical evidence on adults' representation and processing of surface-segment alternations, to motivate APL as a model of learning generalisations about these alternations. The model is then placed in the context of proposals for related stages of language acquisition, particularly the development of an initial system of surface-segment contrasts which may precede the stage that APL models. Chapter 3 develops the formal model itself, integrating both psycholinguistics and historical linguistics to inform how particular evidence from input bears on evaluation of a learner's early phoneme representations, as well as discussing possible alternative assumptions about acquisition and why they are not supported.

Following the introductory chapters, three case studies are presented in Chapters 4-7. These apply APL with different type of linguistic input to yield different acquisition outcomes, and show what the model has to contribute to research in different synchronic and historical scenarios. Throughout these chapters, notation is intended only to provide clear illustration, not to imply theoretical commitments or requirements of the model other than those discussed in Chapters 2-3.

First, Chapters 4-5 validate the learning model by showing how it accurately traces individual children's learning trajectories, accounting for changes in their productions as a function of vocabulary growth. Chapter 4 introduces North American English flapping and its acquisition course. A longitudinal study of four children (Providence corpus, Demuth et al. 2006) provides a more detailed view of flap allophone development, including U-shaped learning in word-medial alveolar flaps. Then, in Chapter 5, APL is tested on alveolar flap acquisition using multiple methods of estimating children's vocabularies, with generally convergent results. Model predictions are interpreted with reference to the known course of flap acquisition as reviewed in Chapter 4, including individual implementations for the four children with production data. There is a reliable relationship between the composition of children's vocabularies and their alveolar flap productions, validating the model; furthermore, at the point when APL predicts flap allophone acquisition, independent measures of overall language development are also consistent across children.

Chapter 6 models development of the phoneme /u/ in the hitory of Menominee. Phonemic /u/ originated from secondary split of a single historical phoneme /o/ whose allophones [o] ∼ [u] were conditioned by surrounding vowels (Bloomfield 1962). Secondary split is modelled with APL as learners' failure to reject their initial surface-contrast-based grammar by the end of acquisition. Gradual collapse of the Menominee short vowel system reduced the prevalence of obvious [o] ∼ [u] surface alternations until, in the present-day language, APL predicts that it is not necessary for learners to acquire this allophonic relationship. However, when modelling an approximated earlier pre-split stage of the language in which both short vowels and long vowels were informative as to vowel quality, the non-allophonic grammar is correctly rejected for productive allophone learning.

Chapter 7 applies APL to patterns in present-day and historical German that result from interactions between devoicing and velar nasal assimilation processes. Several speech varieties are traditionally accounted for with no underlying velar nasal (instead, an underlying two-segment sequence like /ng/) along with a series of ordered rules whose re-ordering is said to explain historical/dialect differences (Vennemann 1970). The alternative analysis, observing a lack of alternations to learn the two-segment representation from in supposedly 'reordered' input, proposes that there has been a change in the underlying phonological inventory such that surface velar nasal segments are not obligatorily represented as an underlying two-segment sequence including a velar stop. APL supports this second analysis of modern Standard German, as well as a major modern dialect variant, based on what is learnable from the present-day input. In addition, historical corpus data is used to model a stage of the language that does uncontroversially have a productive association of surface velar nasals and adjacent underlying velar stops. This case study shows how APL is useful to distinguish between tolerable grammars, which are not incompatible with the language's surface forms, and the subset of tolerable grammars that are learnable. It also shows how APL provides an unbiased method to investigate whether a language has undergone restructuring or rule reordering.

Finally, Chapter 8 concludes by returning to the scope of the model and its place in the broader context of research on how children acquire abstract representations without direct access to them in the input. This includes a brief summary of the current results before discussing their implications all together, as well as considering current limitations of the proposed model. One important consideration is an inherent asymmetry in how informative the predictions of APL are. A specific acquisitional outcome (independent phonemic status of segments) is predicted when the early surface-based phonology is tolerable. However, there is not such a specific prediction when this early grammar is rejected. In that case, APL leaves it underdetermined whether the size of the language's overall phonological inventory reduces, or if the new grammar resembles English n and m in that [m] is a conditioned realisation of /n/ in some cases, but is also a realisation of an underlying phonemic /m/ != /n/). However, regardless of this limitation, APL makes concrete and reliable predictions within the scope of the model and is capable of providing an informative account of allophone learning in a variety of different language environments.

Chapter 2

Background

This chapter sets out some necessary background on phonemes, allophones, and native language acquisition, before proceeding to the learning model proposed in Chapter 3.

First, §2.1 reviews psychological evidence for the influence of allophonic relationships on adult language processing. This establishes the need to learn such relationships involving percepts that are distinct on some phonetic level while undifferentiated in another level of processing. Discussion of discrete segment learning in §2.2 then clarifies the acquisition task APL must account for. Purely for convenience, phonemes and allophones are usually indicated with simple symbolic notation ([ɾ], /d/, etc.). This dissertation adopts an operational definition for ‘allophones’: they are whatever mental representations have behavioural effects in which segments appear simultaneously contrashtive and non-contrastive, and APL is hypothesised to model acquisition of knowledge that produces such behaviours. This operational definition can be empirically refined with observable data from experiments and modelling, without assuming the nature of phonemes or surface segments (as e.g. exemplar-based categories, feature bundles, phonetic/motor implementations), and acknowledging that multiple types of ‘allophones’ or process for learning them might exist (Liberman 1984; Seidl and Cristia 2012).

All of the topics mentioned in this chapter have been the subject of extensive discussion over many years, and involve issues where reasonable consideration of the currently-available evidence still gives rise to disagreement on some key points. It seems neither feasible nor necessary for this dissertation to try to cover all this ground; instead, this chapter focuses on only what is needed to support the proposal of APL, specifying what it is a model of, and establishing that what it models is part of native language acquisition.

2.1 Allophones in language processing

While questions of phonological representation including phonemes and allophones pertain especially to the structural description of language, the importance of these acquired representations is also strongly motivated by observing their influence on adults' language processing. Allophonic status can make perceptually related segments notably difficult to consciously discriminate for adults and older infants, as compared to younger infants up to 8 months who easily discriminate between the same sounds (Pegg and Werker 1997). Adults also have a lower ability to intentionally discriminate between contrasts that are allophonic in their language, compared to adult speakers of a different language where the same sound contrast marks a phonemic distinction (Whalen et al. 1997). However, the acoustic difference of allophonic segments is still perceived and used, merely without apparent intent or control by the perceiver. For instance, allophones whose presence is conditioned by a phonetic property of a following segment are taken as predictive of that property (Lahiri and Marslen-Wilson 1991). Listeners' ability to consciously discriminate sounds is also mediated by linguistic context: for example, Peperkamp et al. (2003) found that the French segments [χ] and [ʁ] are easily distinguished when presented in artificial isolation, but with each embedded in its naturally occurring conditioning context they become difficult to discriminate. This context-sensitivity is present already by 14 months for French infants (Fort et al. 2017). Therefore, there is strong evidence that people learn representations for conditioned alternation of variants that are both distinct and united at different levels.

In addition to biasing acoustic perception, systematic alternations also interact with the lexicon (Clahsen et al. 2001b). Based on evidence including EEG and lexical decision studies, a general distinction has emerged between three cases: (2.1a) fully suppletive items, for which each distinct surface form is memorised as a whole phonological form; (2.1b) default or regular items with either no stem alternation or predictable systematic alternations, which have a decomposed stem+affix representation with a single stem per lexical item;

and (2.1c) sub-regular paradigms where regular inflectional morphology is affixed to unpredictably alternating stems, which also have a decomposed stem+affix representation but with multiple distinct stems memorised to account for the distinct surface stem forms. The description of these three types, and in particular a decomposed stem+affix representation of both sub-regular multiple stems (2.1c) and regular single-stem types (2.1b), is supported by neuroimaging, speeded production, and lexical decision experiments in children and adults (e.g. Bekemeier 2016; Clahsen 1999; Clahsen et al. 2010; Fleischhauer 2013; Leminen et al. 2016; Münte et al. 1999). Notably, processing lexical items with multiple surface realisations that are predictably conditioned (type 2.1b) appears to involve only one memorised stem form just like non-alternating stems, but when multiple stem realisations take unpredictable forms (2.1c) the lexicon stores multiple forms of the stem. This indicates a way that specifically systematic alternations shape language cognition (Clahsen et al. 2001a; Fleischhauer 2013).

(2.1) a. German: bist ∼ sind 'you are' ∼ 'we/they are'
English: is ∼ are; good ∼ better

b. Spanish: habl-ar habl-é habl-a habl-aban
'to speak' INF 1SG.PRET 3SG.PRS 3PL.IPFV

c. Finnish: haaksi-∅ haahde-t haahte-na haaksi-na
'shipwreck' NOM.SG NOM.PL ESS.SG ESS.PL

In light of this evidence, it is plausible that acquiring allophony involving some surface segments [X] and [Y] could affect the number of forms stored for items in which [X] and [Y] alternate with predictable conditioning: a learner who has not acquired [X]∼[Y] allophony memorises both an [X]-stem and a [Y]-stem, while productive allophony lets a single memorised stem account for both conditioned forms. This relationship between learned allophones and the contents of the lexicon, along with the consequences that the number of memorised stem forms has on lexical access, motivate the proposal expanded on in §3.3.2 that allophone acquisition is driven by surface alternations within paradigms.

2.2 Acquiring representations of segment alternations

Accounts of early phonological development seek to describe acquisition of a segmental inventory from the beginning of speech perception through the induction of abstract knowledge about linguistic representations, in particular equivalence classes (multi-allophone phonemes) defining which percepts that are considered distinct in some mental representation are not treated as distinct in another more abstract representation. A widely recognised framework for models of phoneme acquisition identifies qualitatively and/or temporally distinct processes for (i) acquisition of language-specific discrete surface segments (out of a continuous acoustic signal), and (ii) the way that these segments coalesce into necessarily abstract complex phonemes that are not directly based in the acoustic signal; several approaches supporting this two-stage phoneme acquisition process include Boersma et al. 2003; Martin et al. 2013; Peperkamp et al. 2003, 2006; Pierrehumbert 2003; Samuels et al. 2020 and Ringe and Eska 2013.[1] The initial stage of extracting a finite set of surface contrasts from continuous acoustic input has received substantial attention, but there is less experimental and modelling work concerning the second stage of forming higher-level structures involving these contrasting segments (Pierrehumbert 2003).

Developing the surface-segment inventory takes at least 6 months (Kuhl et al. 1992) though likely around 10-12 months (Werker and Tees 1984). Most models of allophone learning in native language acquisition learn segment categories from the distribution of acoustic values in the input (e.g. Guenther and Gjaja 1996; Maye et al. 2002). Already by two months, infants can at least perceive the existence of differences from acoustic cues that will indicate relatively subtle allophonic distinctions rather than highly salient phonemic distinctions in the language they are learning (Hohne and Jusczyk 1994). However, given the massive acoustic overlap of tokens from different categories in learners' actual input, top-down information like early lexical or semantic knowledge may augment a purely acoustic picture of the input environment (Cui 2020; Feldman et al. 2011; Martin et al. 2013; Swingley 2009; Swingley

[1] However, Dillon et al. (2013) argue against the possibility of learning phonetic segments before abstract relations between them, based on a clustering algorithm's failure to sort out surface-segment categories in Inuktitut vowel acoustic data without representing underlying structure.

and Alarcon 2018; Yeung and Nazzi 2013). These views, embedding 'stage 1' segment acquisition within general linguistic and cognitive development, coordinate with the APL 'stage 2' approach in which children access a variety of emerging linguistic knowledge while they learn equivalence classes over a language's surface-derived segments. Pierrehumbert (2003) provides additional comprehensive review of relevant experimental evidence and proposed models for contrastive segment learning, covering a detailed breakdown of the child's task in preparation for phoneme/allophone learning as well as the context, information sources, and mutual feedback loops potentially involved in this complex process.

APL models acquisition during the period after children have already made appreciable progress towards resolving some discrete surface-segment system for their language (6-12 months as above). Distributional or statistical learning is a popular proposal for this second stage of phoneme learning, relying on complementary distribution to discover segment pairs that appear in non-overlapping environments and identify these as allophones. The classical distributional/statistical approach requires that complementary distributions consistently exist in the input, and that the child's learning process has the ability to identify them and exploit them for phonological learning. For example, Peperkamp and Dupoux (2002) propose a model for allophone learning without semantic comprehension, by assuming that the child has previously acquired a surface segment inventory (stage 1) and word segmentation, and then organises a phonological system from the statistical distribution of the surface segments. Similar models, when intended to be implemented on naturalistic input data, tend to supplement a pure complementary-distribution strategy with additions like (sub)lexical contrast information or phonetic naturalness constraints on possible allophone clusters (Martin et al. 2013; Peperkamp et al. 2006). Phonological learning in laboratory settings has sometimes appeared sensitive to purely distributional information (e.g. Chambers et al. 2003; White et al. 2008), but results are inconsistent across replications, and meta-analysis of infants' distributional learning of conditioned phonotactic generalisations finds no effect (in contrast to positive meta-analytic findings of effects for other relevant phenomena like discrimination of native vs. non-native sound contrasts and mutual exclusivity word-learning bias; Bergmann and Cristia 2016; Bergmann et al. 2018; Cristia 2018; Lewis et al. 2020; Liu et al. 2020; Tsuji and Cristia 2014; Tsuji et al. 2020). Following the practise of Pierrehumbert (2003: 117) among others, APL will 'presuppose some of the consensus features' of theories of language processing, in particular assuming that children can acquire some representation of their language's surface sounds, have pairs in their vocabulary like *Kind* [kɪnt] ∼ *kindisch* [kɪndɪʃ] ('child' ∼ 'childish') that are pronounced with different sounds, and have semantic awareness indicating how the different alternating segments correspond with the same lexical item. This model particularly addresses the role of cognitive cost in shaping the emerging phonology of naturalistic learners as language users, and the potentially non-linear acquisition course emerging in response to continuously accumulating input.

2.3 Scope and context: APL within the complete acquisition process

APL quantitatively predicts whether surface-distinct segments are allophonic or simply phonemically contrastive, which is evidently a critical factor in the development of a learner's phoneme system. §2.1 established that allophony, as conventionally understood, exists in language as something that has to be learned and then observably influences language processing and behaviour. The same section also established that, in isolation, alternations between forms of a lexical item do not necessarily imply allophony, since there are also cases where instead multiple phonological forms are memorised for a particular item with no wider systematicity. Therefore, a key part of understanding phonological learning is to determine what kinds of alternation patterns in a language do and do not lead people to acquire representations with a productive relationship involving the alternants. As shown in the discussion of distributional 'stage 2' learning theories in §2.2, this question is also especially pertinent when some segments are in complementary distribution in a language but rarely or never actually alternate in different forms of the same lexical item; §3.2 will discuss this case in depth.

Application of APL throughout development identifies the moment when abstract representations, grouping distinct surface segments into a higher-level equivalence class, emerge from the learner's experience with concrete stimuli. The model can therefore be tested by observing behavioural changes related to this acquired knowledge. This proposal is deliberately agnostic about what sort of mental representation is learned when allophones are acquired: it is hypothesised to be whatever would result in the kind of behaviour that indicates allophony (e.g. §2.1), and therefore, to evaluate the modelling framework it is just an empirical matter whether such behaviour is observed when predicted. As an acquisition model, APL fundamentally requires that children can represent the form of words they know, are sensitive to lexical items whose surface realisations require representing more than one form, and can tell what the (segmental, featural, articulatory, etc.) difference between the forms is; this is compatible with many theories of language representation and acquisition, and the illustrations presented in Chapters 5-7 are framed with certain formal commitments only for the sake of a concrete and readily understandable implementation. It may also be possible that there is more than one kind of of allophone-like representation behind the various behaviours that indicate, in some way, treating sound segments as simultaneously distinct and equivalent on different levels of language processing. If this is the case, it would be an open question which of these representations APL is relevant to, but if different kinds of 'allophones' can be behaviourally differentiated then it should be possible to investigate in the same way which of them are or are not acquired in a way consistent with predictions of this model.

The present version of APL is concerned only with the allophone-learning event, and does not model other parts of the acquisition process. Therefore, organising continuous acoustic speech input into a fairly well developed language-specific system of discrete sound segments, however this may be achieved as discussed in §2.2, is a necessary prerequisite to APL. In addition, after the point of predicting allophone acquisition, the model establishes only that an initial non-allophonic grammar is rejected, without specifying what particular generalisation involving allophones replaces it or whether the lexicon is exhaustively reanalysed with the generalisation (Dressler 1979; Menn 1971). Longitudinal data on English alveolar flap acquisition, presented in Chapters 4-5, justifies this limitation of the model: APL accounts for one identifiable and predictable significant developmental event, which is surrounded both before and after by a large amount of unpredictable individual variation. A child's first allophonic structure may not be precisely adult-like; APL makes predictions about when children acquire increasingly abstract grammars, regardless of their relationship to adult grammars. However, §5.5.1 illustrates one way that a grammar induction model can be integrated with APL to provide a more detailed hypothesis space at the time when children are predicted to learn some allophonic grammar.

Like many accounts for specific well-defined elements of language acquisition, APL aims to complement other ways of understanding what goes on before, after, and during the period this model describes (Albright and Hayes 2003; Martin et al. 2013; Peperkamp and Dupoux 2002; Stoel-Gammon 2011). APL therefore pinpoints a crucial qualitative change in representations, and then leaves space for any of the 'highly individual and notoriously variable' (Vihman 2014: 269) paths that children can follow beyond this point on their way to adult competence (Ferguson and Farwell 1975; Macken 1995; McAllister Byun and Tessier 2016).

Chapter 3

Model proposal: APL

3.1 Phoneme inventory acquisition model

The Alternation-sensitive Phoneme Learning (APL) model proposes firstly that learners pass through an early developmental stage in which their phonological grammar is roughly 'what you hear is what you get' in terms of the language's surface distinctions; and secondly, that this early grammar is evaluated and rejected if the child's linguistic environment contains too much conflicting input in the form of surface alternations. Initial allophone acquisition arises in this model because, if a grammar in which all relevant surface distinctions map to underlying contrasts is rejected, then by logical necessity any alternative acceptable grammar will be one in which this is not the case. When the initial grammar remains tolerable throughout acquisition, allophonic relationships are not acquired. Evaluation of grammars given available evidence is modelled quantitatively with the Tolerance Principle (TP, Yang 2016), a criterion for productivity of linguistic generalisations in the face of some amount of conflicting data. It defines for the general case a tipping point at which any grammar is cognitively unsustainable for a child to hypothesise, given the linguistic items that the child knows a grammatical generalisation should apply to, and how many of those items evidently violate it. Figure 3.1 gives a schematic view of the basic APL model.

This modelling approach builds on the implications of patterns in language change, that underlying contrasts reflect surface distinctions if possible (even if not necessary), in coordination with a broad synthesis of child acquisition research as observed by Stoel-Gammon (2011: 15,17): 'Postulate III: lexical development and phonological development tend to

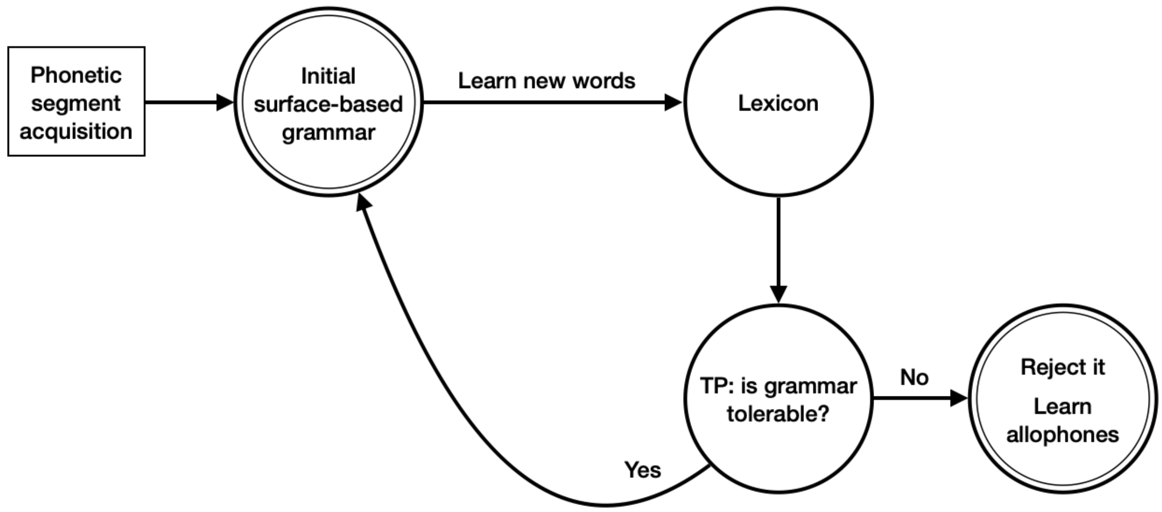


Figure 3.1: Schematic diagram of Alternation-sensitive Phoneme Learning model.

be commensurate' and 'Postulate IV: underlying representations change as the vocabulary increases'. Integrating the TP grammar evaluation mechanism with a principled specification for the phonological grammar that learners evaluate at a certain early acquisition stage (Ringe and Eska 2013), it becomes possible to trace a portion of the developmental course of the phonological inventory.

§3.2 describes and motivates the initial grammar, which is derived from the Invariant Transparency Hypothesis (ITH; Ringe and Eska 2013). §3.2.3 clarifies how this grammar affects adults' phoneme systems, as Invariant Transparency and consequently APL predict that in some circumstances learners fail to acquire descriptively adequate generalisations. §3.2.4 continues by considering how either ITH or substantially different alternatives might explain the process of secondary phonemic split, finding the alternatives implausible given facts of language variation and change. Finally, §3.2.5 presents additional evidence of learners likely acquiring allophonic outputs as distinct phonemes when not distributionally required. Application of the Tolerance Principal to evaluate phonological inventory representations is explored in §3.3.1, while §3.3.2 presents psycholinguistic arguments for this method; in this case, the learner's potential linguistic evidence against their initial maximally-contrastive phonology comes from direct surface alternations like German *Tag* [taːk] ∼*Tage* [taːgə]. Finally, §3.4 recaps how the model as a whole relates to mental and observable events in acquisition.

3.2 Initial grammar: Invariant Transparency

The APL model initialises with a one-to-one mapping between discrete surface segments and underlying segments, which represents a grammar where the listener assumes basic identity between any surface segment they perceive and a phonemic unit. This default system is suggested by Invariant Transparency, a proposal observing that learners posit complex abstract representations (allophones) only when motivated to account for observable surface alternations, and otherwise maintain a more concrete representation:

(3.1) *Native learners in the developmental window for NLA do not posit abstract forms if there is no alternation to account for; instead they project surface segments, defined by the language's system of distinctive features, into underlying forms.* Ringe and Eska 2013: 92

3.2.1 An early surface-based grammar

The contrastive grammar proposed as the initial state of APL is not a child's earliest state of (proto)-phonological knowledge, as it requires already having organised language-specific knowledge from continuous acoustic input. However, the idea of a relatively early stage of surface-based segment representations, such as identity mapping in OT (e.g. Tesar 2008) or Invariant Transparency as implemented in APL (Ringe and Eska 2013), complements a substantial body of experimental work addressing this prerequisite stage, as discussed in §2.2. APL enters the picture at a point in development where a surface segment inventory is reasonably established (probably no sooner than 10 months; Pegg and Werker 1997). It specifies a particular phonological commitment about the grammar of these segments as they become clustered into abstract phonemes, yielding empirical predictions about the relationship between input and acquired phonology.

Learners' tendency towards limited abstraction, and the role of alternations in inducing whatever abstraction there is, has been discussed from a variety of theoretical perspectives, motivated by the relevance of this learnability question to historical sound change as well as understanding the synchronic acquisition process. For example, Kiparsky (1982: 17) observes that in non-alternating lexemes children 'have no reason to set up underlying forms' that differ from the sole surface realisation. Beyond being consistent with a variety of experimental work on acquisition (Peperkamp et al. 2006; Werker and Tees 1984; §2.2), limited abstraction is a common element in Optimality Theory accounts for acquisition, where constraint-ranking grammars are learnable from the interaction between surface alternations within paradigms and a preference to generate a paradigm's surface forms from the same underlying stem (Tesar and Smolensky 2000; Gress-Wright 2010). Prince and Tesar (2004) identify an Identity Map property where 'well-formed outputs should map to themselves' (Tesar 2008: 1), such that infants initially 'adopt underlying forms that are identical to observed surface forms' (Tesar 2008: 54), which could lead them to 'mistake [allophones] for outright phonemic contrast' when the alternations are conditioned by morphology they have not yet acquired (Hayes 2004: 188). Concerning cases when 'a learner must choose among candidate underlying forms which are equivalent in that they all produce the same phonetic output', Prince and Smolensky (2008: 207) propose a principle of Lexicon Optimisation that implies a learning bias to generally minimise disparities between each item's underlying and surface form, but may be overridden by pervasive global morphophonemic alternation patterns (with the details of this 'global lexicon optimization' remaining an open question); this is considered to reflect a 'principle for inducing underlying forms in general' (e.g., in syntax; Prince and Smolensky 2008: 225). Tesar and Smolensky (1996) propose that output-output faithfulness constraints, i.e. similarity between surface forms of different members within a paradigm, are a means to extend Lexicon Optimisation to paradigms rather than single forms, and Benua (1997: 82) describes this type of Lexicon Optimisation as 'a system by which speakers fix rich noncontrastive properties in the underlying forms of morphemes', selecting between underlying forms when there is more than one that could map to the same surface output.

3.2.2 Persistence of default grammar

Regardless of theoretical orientation, the relationship between surface-based default grammars and observable alternations becomes especially important when surface segments occur in distinct environments, yet do not participate in alternations. Traditional descriptive grammars would represent these as conditioned allophones, minimising the grammar's phoneme inventory while increasing explanatory power (Janda 1999; Kiparsky 1982). Children might do likewise if they are also motivated to reduce the size of the phoneme inventory in its own right and can do so from comprehensively analysing the distributions of segments' environments. However, if acquisition tends towards a more direct relationship between surface and phonological forms, non-alternating segments may fail to be learned as allophones despite contextually predictable distributions.

Repeated and thoroughly attested patterns of historical change, in particular secondary split of allophones into distinct phonemes, attest that learners do not always acquire descriptively possible allophones (Gress-Wright 2010; Hayes 2004; Kiparsky 1982; Ringe and Eska 2013; Vennemann 1970). Secondary split occurs when the original conditioning environment of some surface-contrastive allophones becomes lost or obscured, yet the surface-contrasting outputs of the allophonic process persist in place (Hoenigswald 1960). For example, front rounded vowels in Old High German had originated as allophones of back vowels subject to umlaut from following /i/. By Middle High German, this inherited /i/ was no longer distinct from some other vowels that did not condition umlaut, but the contrastive outputs of the old umlaut process were still present. Since these outputs (front vs. back rounded vowels) were now found in the same environments that could not condition a contrast, they had to be distinct phonemes (Kiparsky 1982).

Persistence of a newly-unconditioned contrast like this is taken to indicate that some time earlier there has been a covert phonemic split where 'at least some native learners learn non-alternating outputs of phonological rules as underlying segments' rather than allophones of the same segment (Ringe and Eska 2013: 98); this innovated representation subsequently becomes visible secondary to the segments' environment degrading too much to condition the persisting surface alternation (Janda 1999). The first stage of a secondary split, the covert phonemic split, is a process of restructuring that can change the representation of individual items in the lexicon as well as the language's phoneme system, but does not immediately change the observable language that is produced or accepted (Kiparsky 1982). This may come about when input is compatible with more than one phonemic representation, like when some segments exist in complementary distribution but rarely alternate. It has been difficult to know precisely when restructuring does and does not occur, but further developments in a language can eventually show that it must have happened at some point. It is conceivable that the disappearance of a trigger for a certain allophone might lead to absence of the segment conditioned by that trigger, rather than secondary split preserving it unconditioned; the only way for conditioned 'output' surface segments to remain present is if they were actually no longer dependent on being generated by an allophonic process, but rather were already contrastive underlyingly as well as in surface pronunciation. Since it is quite normal for surface-distinct segments to remain 'frozen' when their conditioning environment degrades, evidently the segments must often be learned as underlyingly contrastive some time before the loss of visible conditioning environments (Janda 1999; Ringe and Eska 2013). APL as developed in this dissertation describes a specific process of acquisition whose outcome matches this expectation, by adopting the ITH proposal of a default grammar with underlying contrast of surface-distinct segments, where allophonic relationships are acquired only when this default grammar fails to account for input.

The following sections take a closer look at the proposal that adults' phonological inventories as predicted by APL may be larger than descriptively necessary, and investigate why and how strongly the ITH is motivated by observed secondary split.

3.2.3 Note on phoneme inventory size

It is first necessary to clarify what 'larger than necessary' or redundant inventories might look like, and what alternative grammars they stand in contrast to. Ideally, it would be possible to specify criteria that can define any proposed inventory as being redundant or not with respect to certain input data.

One possibility is that any inventory with more underlying segment types than a reasonable descriptive phonologist would normally use in an account of the language is redundant. Formal descriptions of a language's phonological system tend to encode a considerable amount of the language's history, which may or may not be reflected in synchronic cognitive representations of the language; a redundant inventory in this sense would flatten some of the explanatory power of a phonologist's description (Kiparsky 1982). This sense of redundancy will be intuitive and decidable for many simple cases, but it is not well defined, and is particularly unclear in edge cases where phonological analyses disagree.

Other possibilities may refer to well-defined mathematical measures of how concise a certain description of a language is, or how much computation is required to parse the language's forms with a given grammar. These tend to be very general measures that can apply to many levels of linguistic description, and represent various ways to formalise the basic intuitions that [lingustically significant] 'real generalizations shorten the grammar' (Chomsky 1965: 42) and 'the grammar should be selected in such a way as to minimize the amount of "computation" that is necessary' (Chomsky and Halle 1968: 63). If such a measurement of economy is adopted, a minimal phoneme inventory might reasonably be understood as the phoneme structure of the most compact grammar and/or the grammar with the shortest derivations (Prince and Smolensky 2008).

One example would be to use minimum description length (MDL) as a reference point (e.g. Rissanen and Ristad 1994), with any phoneme inventory larger than the MDL grammar's inventory considered to be larger than necessary. MDL for phonology is intended to incorporate balanced pressures towards both smaller lexicons and simpler grammars (Rasin and Katzir 2016), with grammar complexity represented by the length of stating the grammar itself rather than e.g. the number of steps in derivations. MDL is fundamentally measured as the number of bits needed to describe certain data with a particular encoding scheme (Rissanen 1978). When the data is a corpus of some language, the original corpus is one possible description of itself, but if the language contains some predictable patterns then a shorter description may be found by combining some general instructions for these patterns (a grammar) with reduced representations of only the remaining non-predictable material. MDL accommodates both rule-based and constraint-based phonology, as well as other levels of language (Rasin and Katzir 2016; Rasin et al. 2018).

Although MDL is a well defined standard measure in one sense, the choice of which structures are made available for representation 'has immediate and perhaps quite drastic empirical consequences' on the results (Chomsky 1965: 45); it is conceivable that either of two different basic descriptions of a language could be identified as the MDL grammar under different encoding schemes, making the conclusions drawn from MDL potentially highly variable.

Besides length-based measures like MDL, another possibility is to define the minimal possible phonemic inventory according to some model of efficient parsing of input (e.g. Berwick and Weinberg 1982), considering an inventory to be larger than necessary if it is larger than that of the smallest possible reasonably-efficient grammar for the input. This definition might likewise vary considerably depending on the model of processing or even peripheral conventions of the grammar chosen for evaluation.

This dissertation will appeal to the Tolerance Principal to characterise phoneme inventories; a 'redundant' inventory is one that has more distinct underlying phonemes than the smallest inventory for a tolerable grammar of the full/adult language. This specification is useful to have available for discussion, but it is not a proposal about some real or meaningful category of 'redundant grammars' in the world, for several reasons. This measure again depends on the formalism used for the representation of phonemes, although its outcome may still be more reliable than measuring inventory redundancy with a grammar-internal measure like MDL, because TP evaluation is grounded in a count of surface differences with less opportunity for arbitrary encoding decisions. Furthermore, this measure (like all others discussed previously in this section) acts as an 'evaluation procedure', not a 'discovery procedure', in the sense of Chomsky (1957); it cannot positively identify a minimal inventory, only make relative comparisons between whatever grammars it is applied to. For present purposes it is most important to be able to designate some cases when there are redundant inventories, to point out when (early) evidence of secondary split creates a problem for theories that do not anticipate such inventories. For many input/grammar pairings, it is very straightforward to use the TP criterion to determine that a particular grammar definitely does not have a minimal inventory, because it is easy to devise some other grammar with a smaller inventory that is also tolerable. However, the choice of this criterion is not critical; for all of the cases presented in this dissertation, it is clear from any perspective that the so-called 'non-minimal' inventories truly are not the smallest possible descriptively adequate systems. Finally, the positive definition of a non-redundant or minimal grammar is probably best left to frameworks that explicitly propose such grammars; this dissertation will generally instead refer to 'non-compact' grammars for the representations expected to result from an acquisition process where underlying contrasts by default derive from surface distinctions, with these underlying contrasts only being lost (in favour of one phoneme with conditioned realisations) when required by obvious alternations in the input.

3.2.4 Alternatives to non-compact grammars

This section evaluates alternatives to the existence of non-compact inventories, or roughly speaking the possibility that first language learners could be more similar to 'historically-minded generative phonologists' (Janda 1999: 330; Kiparsky 1982), in light of the observation of secondary split as a mechanism of language change.

Instead of a non-compact inventory representing a covert phonemic split long before it becomes visible in surface forms or necessary to account for them, perhaps a phonemic split could occur in exactly the same generation as when the changing input has made that split a necessary rather than non-compact description of the language. This view is reflected for example in Twaddell's (1938) characterisation of a German umlaut split, in which $-i$ and $-a$ reduced to [ə] where $-i$ (but not $-a$) had conditioned fronting of [u] to [y]: 'As soon as the difference between -i and -a has ceased to be significant, then (and not until then) does the difference between [y] and [u] become a significant difference'. Keeping in mind the possible ways to characterise a non-redundant grammar in §3.2.3, this 'crucial moment of sound change' might be expected 'when the two new sounds cease to be in complementary distribution' (Fertig 1996: 169), in the generation in which the conditioning environment becomes opaque, or the generation in which the input data is such that a descriptive phonologist would represent both of the relevant segments as distinct underlying phonemes. To use the TP criterion, if learners never acquire non-compact inventories then the phonemic split would occur at the first generation in which the smallest possible inventory for a tolerable adult grammar requires the underlying split, or in other words, the first generation in which no fully allophonic grammar is tolerable. As mentioned above, it is in principle difficult to detect whether the inventory has been restructured much before this at a point when the restructuring would result in a non-compact inventory, because the introduction of an 'unnecessary' phoneme is a silent reanalysis that 'involves no actual change in the language, but merely the substitution [...] of a simpler grammar which generates the same language' (Kiparsky 1982: 53). The remainder of §3.2.4-3 will discuss the evidence from which the ITH, and the acquisition target of APL, can be inferred even without directly observing the first learners to acquire a new phoneme.

Overall, accounts of secondary split without non-compact phonologies seem very brittle, generalising poorly to common scenarios on the community/individual level. An implausibly precise conspiracy of mechanisms would be required to bring about 'two absolutely simultaneous but independent changes: phonemic split, and loss of one or more conditioning factors' (Janda 1999: 330) which together need more explanatory support than the hypothetical non-compact inventory they manage to avoid proposing. For example, if learners generate surface forms from a compact inventory by encoding as much of the language's history of transformations as they can possibly recover from input, then many observed phonemic splits could not have happened because the input from a productively allophonic grammar would never render the allophonic process all that undiscoverable to the very next generation. Even hypothesising somewhat less powerful learners, §3.2.5 shows that evidence of phonemic split can already be found long before the time when an allophonic process is opaque or an allophonic grammar is not tolerable, so input with these characteristics cannot possibly be the cause of learners first reanalysing allophonic outputs as phonemic contrasts. One historical pathway towards secondary split involves a series of different, potentially independent, changes that have the cumulative effect of destroying a conditioning environment. This can eventually result in secondary split without any one of the changes simply neutralising or deleting all instances of the original allophonic trigger; various changes might each affect some different but partially-coinciding morphological sub-class or restricted phonetic environment. For example, if a purely phonological alternation happens to often occur stem-finally conditioned by something in an inflectional suffix, the language may have intermediate stages of morphophonemic conditioning following events of lenition, vowel quality reduction, sound mergers introducing sporadic or pervasive syncretism, apocope, etc. until distinct conditioning environments cannot be recovered for all surface realisations of the inherited allophones. Changes can also directly affect (subsets of) the outputs of alternation instead of their conditioning environments, either by further sound change, or by reshaping the language's inventory of underlying forms (as in paradigm levelling or losing an inflectional category). Since this kind of change influences the prevalence of observable surface alternations, it is relevant to the learnability of allophones with APL even if it does not result in any segments surfacing outside their expected conditioning environment.

Some attested splits, like the change of English fricative voicing from allophonic to phonemic, are known to have occurred with dialect contact and mixed grammars in the population, where clearly there was not a homogeneous language community proceeding together through gradual change until reaching a generation that required a phonemic split (Nielsen 1994). Furthermore, straightforward merger of conditioning contrasts can occur when gradient sound change makes one allophone's environment become indistinguishable from a different allophone's environment, but even classic gradual sound changes always have speakers in the population at different places along the course of the change (Labov 1994). In cases like this, variation in the speech community precludes the possibility that all of the conditioning tokens are just barely distinguishable for one generation (last allophone learners), and all at once become indistinguishable with a small phonetic incrementation for the next generation (first non-redundant split learners). Instead, learners for many generations hear a mixture of input at various degrees of merger, and the mixture eventually contains little enough distinct input that evidently the allophonic relationship is not learned even though it is present in some subset of speakers and can still be identified by linguists.

Besides the ITH and the alternative minimal-inventory hypotheses discussed in this section, there do not seem to be any other possibilities to account for observed language change. The only potential proposal would be that secondary split does not actually happen, purported observations of it are in some way an illusion driven by something else, and therefore no mechanism for secondary split should exist in the first place. However, this logically possible option is not at all supported (or even considered as a worthwhile position) in other research.

Overall, secondary split without non-compact inventories is made implausible because language change never occurs outside the context of language variation. Across several mechanisms of language change, it is not possible to realistically picture the whole relevant environment making the same critical shift at the same time in approximately every instance in a learner's input. A clean transition from an allophonic grammar to a split grammar exactly at the point when the allophonic grammar fails to account for the language (when the split grammar does have the smallest possible phonemic inventory) will in practise have no opportunity to occur (Foulkes and Vihman 2013).

Janda (1999: 330) reports that theoretical objections to the idea of non-compact inventories stem from the lack of a concrete proposal for 'where and how can one draw the line' between actual allophones and redundant separate phonemes, if not the traditional proposal of requiring the smallest descriptively adequate inventory. APL in §3.3 offers an objectively measurable answer to this, which receives support from a large body of acquisitional and psycholinguistic research. First, the following section presents some evidence of early (covert) split grammars, supporting the existence of non-compact inventories by showing that outputs of a historical allophonic process can behave as distinct underlying phonemes even while the allophonic nature of their relationship is still plainly evident.

3.2.5 Synchronic evidence for non-compact inventories

This section reinforces the proposal that non-compact phoneme inventories are a normal outcome of acquisition. These inventories can be detected because 'covert' splits, that precede and enable the loss of an allophonic process in secondary split, are not always so covert that synchronic evidence of them is completely absent. In particular, languages can innovate or borrow words that require the availability of an underlying phonemic contrast, with loan words often becoming the first direct evidence of an underlying split. For example, Ringe and Eska (2013: 131) question why speakers of Menominee, where [u] only occurred as an allophone of [o] before high (semi)vowels in inherited vocabulary, could borrow words with [u] outside of that conditioning environment; they conclude that the borrowing was possible because distinct underlying /u/ was already part of the language in speakers' grammars. When languages borrow loanwords using their native set of surface segments in the wrong distribution like this, the allophone that is borrowed outside of its conditioning environment hints that the language can have that segment without generating it through an allophonic transformation; that is, the segment seems to be available as a distinct underlying form as well as a distinct surface form.

Signs of these 'new' phonemes can be seen with allophones whose alternation is still very much alive, showing that the underlying contrast has apparently been acquired long before it would seem necessary to account for the input. Distribution-violating borrowings appear in populations before the conditioning environment is opaque, when a fully allophonic grammar is otherwise still descriptively accurate. However, if these borrowed forms are held as evidence for underlying splits at the time of borrowing, it is worth questioning why borrowings into the 'wrong' environment are not considered possible when segments have purely allophonic status, and why such loanwords therefore establish phonemic status. As an example appealing to the intuitions of English speakers, today North American English does not seem able to borrow a word with a palatalised (fronted) velar stop in a phonotactically illegal position as a minimal pair of an existing English word with the non-palatalised stop legally occurring in that same position. Given the existing words *keep* [cip] and *cup* [kʌp], demonstrating palatalisation before front vowels, it seems doubtful that a hypothetical loanword [cʌp] could be borrowed as [cʌp]; surely it would end up borrowed as [kʌp], and likewise hypothetical [kip] without fronting would be borrowed as [cip]. English has no phonemic contrast between [k] and [c] for loanwords to make use of.

More persuasively, the inception of a covert underlying split may be tracked by observing changes over time in how loanwords from the same source are borrowed into an evolving language: more recent loans can take advantage of an underlying split to violate an allophonic generalisation, appearing in environments where earlier borrowings from the same language could not (because presumably the underlying contrast to enable such borrowings was not yet available, like English [c]∼[k]). For example, Northern Chiapas Zoque has a phonological voicing alternation of stops and affricates, which are voiceless except when directly preceded by nasals. Words borrowed from Spanish long ago are changed as needed to fit this generalisation (Zoque *mandeka*>Spanish *manteca* 'fat, lard'), but more recent loans (e.g. Zoque & Spanish *golondrina* 'swallow (bird)', Zoque & Spanish *manta* 'blanket') preserve Spanish surface forms in violation of the native Zoque distribution (Ringe and Eska 2013).

It could be argued that co-phonologies are critically relevant, as they are known to often be associated with borrowed vocabulary anyway (Itô and Mester 1999), and then perhaps each co-phonology could have a minimal inventory for the subset of vocabulary that it applies to. However, this merely recasts the question that was previously formulated as defining when a non-compact inventory could be learned: if some speakers have at least one 'foreign' system with an underlying split alongside at least one 'native' system without, presumably at some earlier stage of the language no adult speakers had a split system, and the question now is to determine when a co-phonology with underlying split can take hold. Maintaining the idea that this acquisition will be driven by (in)compatibility of input with grammars, it seems possible to coordinate with the APL model already established. Either way, changes in the model's predictions should correspond to observable events such as the transition between Zoque borrowings like *mandeka* from *manteca*, vs. *manta* from *manta*.

Overall, positing that all inherited allophonic rule outputs are synchronically allophonic rule outputs (right up until a phonologist or effective distributional learner could not generate them by rule) seems to push the occurrence of underlying split far too late – especially if one or two exceptions are not considered fatal for the allophonic generalisation – compared to when languages show evidence of already having an expanded phonemic inventory.

3.3 Evaluation: the cognitive cost of a phonological grammar

3.3.1 Modelling evaluation: Tolerance Principle

Valid linguistic rules or generalisations are rarely absolute, but beyond some amount of exceptions a supposed descriptive 'generalisation' cannot really be considered operational (Chomsky and Halle 1968; Sapir 1921; Yang 2016). The Tolerance Principle (TP, Yang 2016) defines a critical threshold θ_N at which a productive grammar, generalisation, or rule R is possible, given input where the generalisation is applicable to N candidates and violated by e exceptions among those candidates:

(3.2) $$e < \theta_N \text{ where } \theta_N := \frac{N}{\ln(N)}$$

This threshold is a result of processing efficiency for the language user, derived from assumptions of frequency-correlated lexical access and an Elsewhere Condition (exceptional forms processed first before broader generalisations are otherwise applied; Yang 2016, 2017). The Tolerance Principle's predictions for productivity of generalisations, emphasising concrete links between evidence in the learner's input and the abstract grammar acquired from this, have been widely applied and rigorously validated in empirical case studies spanning children's acquisition of phonology, morphology, and syntax in relation to their input environment (Yang 2016). Cognitive validity of the Tolerance Principle as not only a broad population model but also a precise model for individual learners is empirically established by Schuler et al. (2016); by tracking individuals' acquisition in an artificial vocabulary, it is possible for a 'personal Tolerance Principle' to predict (not/)learning a generalisation based on each child's own particular vocabulary composition, with a close match between these predictions and individuals' (generally categorical) non/learning behaviour (Schuler 2017).

In the APL model, learners initially evaluate a concrete mapping between surface segments and underlying forms (§3.2). Exceptions are evident violations of this mapping, such as alternations *leipänä* ∼ *leivästä* (Finnish 'bread', essive/elative) where it will be clear to a child who knows both words and the regular case morphology -nä, -stä relating them that the two different pronunciations [p] and [v] are in some way the same thing, corresponding like [v] ∼ [v] in *päivänä* ∼ *päivästä* 'day'. Labelling something in input as an 'exception' to a certain formal grammar does not necessarily illuminate just how the identified exceptions would impede language processing. APL views the relevant consideration as the cognitive cost that surface alternations specifically incur for a the language user, discussed in the following section. A learner's transition between different grammars is prompted by finding a current hypothesis too costly and trying a better-fitting alternative. Psycholinguistic experimental evidence (§2.1, §3.3.2) validates applying the TP for this model, because processing surface-alternating items either with or without productive allophones is behaviourally distinguishable, with seemingly higher cognitive cost associated to non-allophonic alternations (e.g. Münte et al. 1999; Reifegerste and Clahsen 2018). Therefore, accumulated surface alternations can make an early phonological hypothesis no longer tolerable.

3.3.2 Cost of processing alternations

A learner could in theory acquire their input language perfectly by memorising every form and making no generalisations; however, as this is clearly not what children do, acquisition modelling may offer concrete accounts of how learners identify certain patterns from input (Jarosz 2019). For example, in principle a learner without English intervocalic [t] ∼ [ɾ] allophony might learn *eat* and *eating* as /it/ and /iɾɪŋ/ analogous to *go* ∼ *went* suppletion, or learn /it/ plus morphologically decomposed /iɾ/ alongside -/ɪŋ/, with duplicate labels /iɾ/ and /it/ violating mutual exclusivity. The decomposed analysis in both adult and child lexicons is supported by evidence in Spanish, English, and German (§2.1; Bermúdez-Otero 2016; Linares et al. 2006; Smolka et al. 2007, 2015). A third possibility might derive *eat* ∼ *eating* from the same stem (perhaps /it/), but still hypothesise that surface forms straightforwardly match underlying forms; then, a child would expect pronunciations like [itɪŋ] that are violated by their input [iɾɪŋ]. For at least the specific case of English alveolar flap alternations, data in Chapters 4 - 5 does not suggest that children initially have targets like [itɪŋ], so the two-stem representation is more plausible. Either way, alternations will make exceptions to the initial concrete grammar.

If a child memorises the surface [t] and [ɾ] alternation by storing duplicate forms /it/ and /iɾ/ in their lexicon, this representation violates mutual exclusivity, a learning bias against multiple labels for the same referent (Markman and Wachtel 1988; Markman et al. 2003)(Markman & Wachtel 1988, Markman et al. 2003). Mutual exclusivity is a strong tendency but not absolute, and the mechanism(s) that give rise to it are not entirely known (Liittschwager and Markman 1994; Savage and Au 1996; meta-analysis by Lewis et al. 2020). It may particularly be violable with similar-sounding labels (e.g. [fɛʃ] similar to $fish$; Creel 2012). Children who are given a novel label with minimal phonological difference from a familiar label tend to apply the new label to the already-labelled familiar referent, rather than to a novel unlabelled referent as would be predicted by strict adherence to mutual exclusivity, while still showing phonological awareness of the segmental difference in labels (Swingley 2016). Still, the surface segments belonging to a phoneme can be arbitrarily dissimilar, and surface similarity is also not an absolute requirement for violating mutual exclusivity or learning duplicate labels (Markman et al. 2003; Peperkamp et al. 2006).

Storing multiple-stem lexical representations may have processing disadvantages relative to single-stem representations. For example, Reifegerste and Clahsen (2018) find that older speakers of German are slower to access or use the morphosyntactic features expressed by 'lexically-conditioned suppletive stems' (unpredictably mutating verb stems that take regular affixes) as compared to regular stems, including regular stems that participate in systematic allophonic alternations and therefore have multiple (but predictable) surface forms. Fleischhauer (2013) finds similar results for 7-9 year olds' lexical retrieval. The constraint grammar concept of output-output correspondence at the phoneme level also captures this observation, by restricting surface alternations across phonemic contrasts while permitting sub-phonemic variation that is not represented by multiple alternants in the lexicon (Benua 1997; Raffelsiefen 2016). The cognitive cost implied by these findings, if its accumulation exceeds the TP threshold, pushes learners away from non-allophonic surface-based representations if alternating input would otherwise require memorising an inflated number of stem forms.

3.4 Phonological development model summary

APL posits that learners have an initial default grammar where underlying contrasts are isomorphic to surface segment contrasts, and learning a productive allophonic alternation results from rejecting this grammar. Although a child may not immediately move on to their language's adult allophonic system, when the one-to-one concrete mapping is rejected, any new hypothesis must involve a distinction on one level mapping to a lack of distinction on another level. This property of increased abstraction in phonological representation, identified with allophone creation, is a logical necessity for the special case of the learner's departure from the Invariant Transparency early grammar.

Non-creation of allophones is also modelled by APL; the initial distinction of true phonemes, like those projected from English [h] and [ŋ], persists because the input provides no evidence (alternations) to reject these contrasts. If outputs of historical allophonic processes eventually do not participate in many direct surface alternations, then a covert underlying split may occur, where learners fail to acquire (still descriptively accurate) allophones and instead allow their early surface-contrast mapping to survive into adult grammar.

APL does not specify a learner's exact new allophonic grammar when one is learned, but this model is compatible with a range of grammar induction models, like some using maximum entropy or minimum description length (Berger et al. 1996; Hayes and Wilson 2008; Solomonoff 1964), with the potential to join into a larger cohesive picture of phonological development. Inducing grammars from representations like phoneme strings, features, etc., to potentially bring some particular allophonic grammar into the learner's hypothesis space, is a well-explored problem elsewhere (Johnson 1984; Touretzky et al. 1990; Yip and Sussman

1997; Calamaro and Jarosz 2015; Jarosz 2019). As one example, Chapter 5 will implement the inductive learning model of Albright and Hayes (2003) to generate candidate grammars about English alveolar flapping, based on the same input that APL predicts children will learn this from.

Furthermore, while this dissertation uses a variety of modelling assumptions due to both empirically-motivated hypotheses about language and practical necessity for and concrete description and implementation, many of these elements could be modified while using APL as a more general framework for exploring how children learn about sound alternations. For example, surface alternations are characterised throughout as processes or transformations applied to underlying forms, suggesting that children may acquire such processes when learning about allophones, but the model could accommodate instead a view 'in which processes are not primitives of the theory, but are at best descriptive commentaries on the theory' (Tesar 2008: 2). Since the model already uses pairs of surface forms to evaluate grammars, it could be used with grammars that account for alternations not by referring to underlying forms but by defining output-to-output constraints on the 'mutual resemblance of the surface realisations of a morpheme' (Hayes 2004: 186; Benua 1997). In addition, some other theorised model for evaluation could be tried in place of the TP evaluation.

Changes to a learner's underlying representations are, as postulated by Stoel-Gammon (2011), a function of expanding vocabulary and the cognitive load of maintaining an increasingly inefficient initial phonology (Yang 2016). Vocabulary development is a highly active process for several years, so learners' exposure to the alternations that can inform them about allophone distributions occurs together with acquisition of complementary components of linguistic knowledge, such as systems of morphology (Brown 1973; Hart and Risley 2003; Segbers and Schroeder 2017). The Tolerance Principle alone could model whether a phonological system is possible for adult speakers; however, identifying a specific early phonological hypothesis for learners increases the resolution at which the TP can be applied, by modelling iteratively from this initial state as the child's exposure to language increases. Given infinitely many possible abstract grammars compatible with a learner's input, APL specifies which of these are cognitively plausible, as only some tolerable grammars are reachable with the learning procedure.

Chapter 4

Documenting U-shaped allophone learning

This chapter presents an examination of alveolar flap-context productions for four North American English-learning children in the Providence corpus, age 11-48 months at various stages of articulatory, phonological, and lexical development (Demuth et al. 2006). This is the first longitudinal study of naturalistic English flap productions at this scale, tracking changes during an understudied developmental period that is not covered by existing studies of only younger and older groups (e.g. Song et al. 2015; Klein and Altman 2002). The results demonstrate that allophone acquisition can show a pattern of U-shaped or non-linear learning, similar to what is typical for the acquisition of other types of linguistic rules or generalisations.

§4.1 first presents background on English alveolar flapping, including its nature as a phonological process and what is already known about its acquisition course. Following this, §4.2 describes the present corpus study in which canonical medial flap-context productions are observed for four children, along with two children's word final flap-permitting contexts, based on perceptual coding. In this study's analyses, Mean Length of Utterance (MLU) and Type-Token Ratio (TTR) are used instead of age to match children for comparison across individuals, since these measures provide independent indications of general language development. TTR is a reliable measure of overall lexical diversity in language samples, and can describe individual differences in young children's vocabulary development (Fergadiotis et al. 2015; Wu et al. 2019). MLU describes grammatical rather than lexical aspects of language use, and is also suitable as a 'useful summary of the result of the development process' in longitudinal studies (Cazden 1968: 435; Wieczorek 2010).

In §4.3, results show clear non-monotonic or U-shaped development of word-medial flaps for three children, which manifests with both a decrease in production of adult targets and an increase in innovative errors. Statistical analysis fitting broken line models to each child's rates of target production confirms this non-monotonic acquisition course. In general, children's frequency of adultlike targets progressively increases until MLU ≈ 3.5, after which it decreases while the rate of non-target productions increases. One child's MLU does not yet reach 3.5 in the Providence corpus sample, and within this available data, the child's medial flap productions monotonically approach an adult target distribution. Furthermore, word-boundary flaps are shown to develop distinctly from those in medial contexts.

Finally, §4.4 discusses these corpus results in the context of previous research on both other U-shaped phonological acquisition and English flap acquisition. The four children's flap acquisition trajectories presented in §4.3 will also provide data to validate individualised learning models for each of these children, developed in Chapter 5. In short, the APL results in §5.4 converge around predicting allophone acquisition when MLU is around the same point of 3.5 for each of the three children whose productions in this chapter show U-shaped development, while the child whose productions show no 'regression' and whose MLU does not exceed 3.5 in this data is also not predicted to know flap allophony.

4.1 Background

Alveolar flapping is a phonological phenomenon realising /t/ and /d/ as a flap or tap [ɾ] in certain environments. In North American English, a word-medial allophonic process (4.1) causes intervocalic alveolar stops to surface as flaps before unstressed syllables (De Jong 2011), while intervocalic flapping at word boundaries (4.2) is optional and not governed by a stress condition (Eddington and Channer 2010; Yun et al. 2020). This phenomenon has been an attractive test case for a range of studies due to its prevalence across dialects, high frequency, and measurable change along many dimensions (phonetic, social, etc.). However, it is characterised by variation at all levels and proves to be a complex object of study, so much of the research to date aims to clarify the nature of alveolar flapping itself.

(4.1)	water	putting, pudding	alligator
	[wɔɾɚ]	[pʊɾɪŋ]	[ælɪgeɪɾɚ]

(4.2)	a.	not only	eat enough
		[nɑɾ ˈonli]	[iɾ ɪˈnʌf]
	b.	me to	you don't
		me [ɾ]*o*	*you* [ɾ]*on't*

TIMIT, Yun et al. 2020 (full transcriptions not available for 4.2b

The following sections briefly discuss the relevant previous findings necessary to support accurate application of APL as a model of flap acquisition. §4.1.1 justifies treating [ɾ] as a categorically distinct sound segment from [t] (or [d], [ʔ], etc.) - either as allophones or separate phonemes, according to developmental progress, but not just varied acoustic realisations of a single segment. §4.1.2 reviews child acquisition of English flaps, both providing context to apply APL in this situation (Chapter 5) and motivating the need for the present longitudinal study.

The experimental findings summarised in §4.1.1–4.1.2 often do not distinguish between a range of possible phonological formalisms, such as the flap process being represented as either a generative transformation rule or a system of ranked constraints. Likewise, this literature includes a range of theories about the basic nature of phonetic and/or phonological segments, and empirical studies have not always found it necessary to specify any particularly concrete position. Reflecting this, APL is broadly compatible with different representations, provided that they are extensionally equivalent with respect to alveolar flaps and have some internal structure aligning with allophones.

4.1.1 Phonological status of flapping

Three important questions to consider before modelling flap acquisition as an instance of allophone acquisition are:

(i) Is this really an allophonic process, or is it a type of non-phonological variation?

(ii) If flaps are truly allophones of /t/ and /d/, how are they distinguished from the other allophones?

(iii) Does the flap process neutralise /t/ and /d/, or is their underlying contrast recoverable?

A variety of methodologies and theoretical interests converge on the assessment that English flaps are allophones of alveolar stops, that is, they are not a non-phonologised epiphenomenon like one extreme on a continuum of articulatory reduction. De Jong concludes that despite recent efforts (including his own; 1998), the articulatory, acoustic, and perceptual data on American English flapped /t/ cannot be fully explained by within-category gradient articulatory/motor 'processes that would yield the appearance of flapping as a byproduct' (De Jong 2011: 5)(de Jong 2011: 5). Multiple articulatory processes influence various acoustic outcomes, but there are no identifiable mechanical processes that systematically correspond with flap/stop occurrence (De Jong 2011; Fukaya and Byrd 2005). Other intervocalic stops are subject to gradient lenition in English, to the greatest degree in casual speech styles; however, flap-context /t,d/ articulation is frequently reduced even in careful speech when other consonants are minimally reduced, implying that flapping is a distinct phonological process (Warner 2005). Priming effects across flap/stop alternants also show a different profile than priming among acoustic variants within a single phonetic class, in repetition and lexical decision tasks (Luce et al. 2003).

In its typical intervocalic post-stress context, North American word-medial flapping may apply very consistently. Zue and Laferriere (1979) and Byrd (1994) find flapping rates of 99% here. Patterson and Connine (2001) find a rate of 95% for high-frequency words, in a slightly broader context; however, only 76% of similar lower-frequency words in their sample contained medial flaps. Word-final /t/ and /d/ are more variable, with several potential realisations (Eddington and Channer 2010). For example, in the TIMIT corpus, utterance-final 'that' is produced with 23% released stop, 67% unreleased stop, and 9% glottal stop (Byrd 1994), while overall TIMIT /d/ and /t/ are each realised as flaps in 74% of possible intervocalic instances (combining both word-edge and word-medial flap contexts; Yun et al. 2020). The alternation-sensitive learning model will easily incorporate these additional variants, because any obvious alternation of a flap with something that is not a flap contributes to learning allophonic status of the flap (see Chapter 3).

Although flapping is evidently a categorical alternation, there are no straightforward acoustic measures of whether a /t/ or /d/ token is a flapped realisation. Some of the difficulty may derive from methodological limitations of study scenarios, where acoustic and articulatory data of isolated segments is viewed without the contextual information (speaker, voice, room/channel, etc.) normally available to listeners. Fukaya and Byrd (2005), echoed by Son (2008), refer to complex articulatory and acoustic data to suggest that a constellation of individually gradient articulatory phenomena gives rise to perceptually quantised acoustic signals, which listeners in naturalistic environments interpret categorically as stop or flap allophones.

In addition, while /t/ and /d/ can both be realised as [ɾ], this does not necessarily neutralise the voicing distinction, as the contrast of these underlying consonants is still expressed in the length of the sonorants that precede the flap (Fox and Terbeek 1977; Patterson and Connine 2001; Treiman et al. 1994). Voicing distinctions among medial alveolar allophones are most often investigated by recording word pairs like *kitty* ∼ *kiddie* or *putting* ∼ *pudding* and measuring acoustic and articulatory parameters (Fox and Terbeek 1977; Herd et al. 2010; De Jong 1998; Malécot and Lloyd 1968; Zue and Laferriere 1979). Experiments on whether people reliably perceive or use the sonorant length cues to underlying /t/∼/d/ contrasts have mixed results; however, relatively few experimental paradigms have so far been used to address this question, mainly presenting isolated words in artificial laboratory situations like a forced-choice task (Braver 2014; Riehl 2003; Sharf 1960; Warner and Tucker 2011). The degree of pre-flap sonorant duration difference also varies across speakers and dialects, and in communities with Canadian Raising it can be especially robust in the /aɪ/ vowel where consonant voicing information is most essential for systematically predicting the vowel's quality (raised to [ʌɪ] before underlying /t/ but not /d/; Davis and Van Summers 1989; Davis et al. 2020; Fox and Terbeek 1977; De Jong 2011).

4.1.2 The course of flap acquisition

Acquiring the system of North American English alveolar stop allophones is challenging, as it involves several conditioned allophones, at least two phonemes, and a rapid gesture to produce the flap (Klein et al. 1998). This section reviews studies of children's productions in alveolar flap contexts, from the initial emergence of articulatory ability through schoolchildren's maturing control of alternations in context. Some patterns of error and their implications for children's knowledge are also discussed.

Klein et al. (1998) explore how data for children's articulations should best be coded, and directly compare the value of perceptual and acoustic coding for English intervocalic flap. They argue that while perceptual judgements can vary by person and be subject to perceptual bias, perceptual transcription for alveolar variants is still superior to categorical coding by acoustic/spectral measures: the 'objective' acoustic measures are selected circularly for their tendency to correlate with perceptual judgement, while any basic algorithmic combination of them fails to categorise alveolar stop allophones as well as natural human perception does (Klein et al. 1998). Perceptual coding of flaps and related allophones, which can be aided but not superseded by e.g. viewing spectrograms, remains a standard for both child data and adult speakers (Zue and Seneff 1996; Eddington and Channer 2010; Song et al. 2015). Individual annotators appear consistent in their own perceptual boundaries, and satisfactory inter-annotator agreement can also be reached with training and adjudication (Burrows et al. 2019; Yun et al. 2020).

Song et al. (2015) study the earliest stages of production of monosyllables ending in /t/ or /d/ in several contexts including intervocalic, at 18-30 months, and find no word-final flaps until near the upper age bound. This study constructs a controlled sample from naturalistic speech (Providence corpus, Demuth et al. 2006), using a restricted subset of monosyllabic /t,d/-final words in specific phonological contexts to maximise the ability to compare data across individuals. Still, this sample ends up unbalanced lexically; although the study includes 11 words (6 -/t/ and 5 -/d/), the word 'good' alone accounts for 26% of children's tokens and 59% of mothers' tokens in the analysis, potentially limiting how far the results generalise beyond the study sample. Additionally, significant changes in maturing children's use of variants did not appear in this data (possibly due to small sample sizes) and all child tokens from 18, 24, and 30 months were therefore combined for most analyses, so change over time is not tracked. Inferring a child's knowledge about phonological representation of flaps from their productions in the earliest samples is also methodologically challenging due to children's lack of control over articulatory targets. Klein and Altman (2002) elicit productions of two-syllable trochaic words with typical word-medial flapping environments, for four children ages 21-60 months. Perceptual coding tracks 15 variants of alveolar stop realisation, some of which are not adult target allophones for any alveolar stop but occur in the child data, and development is examined after combining data from all of the children together in three age-based groups (<36, 36-47, 48-60 months). Rates of producing both target [ɾ] and nontarget [t] for medial /t/ increase across the course of the study period, while other nontarget productions, such as syllable deletion, become less frequent; overall, by age 5 none of these children yet achieve adult-like medial flap usage. However, the data presented below in §4.3 indicates nonlinear inflection points in medial flap development, which different children appear to reach on either side of 36 months. This limits interpretability of the data from the <36 months and 36-47 months groups, as each group might contain a mixture of children who have and have not passed a developmental inflection point. Klein and Altman (2002) do not report longitudinal data of individual children, only each individual's total productions for the entire study (21-60 months).

Burrows et al. (2019) find that English monolingual kindergarteners (mean age 68.4 months) accurately repeat only 75% of medial flaps in an isolated word repetition task, while they delete this segment in 5-10% of productions and substitute a stop [t,d] in 14% of productions. The stop substitutions indicate awareness of a connection between [t,d] and [ɾ]. Eddington (2007) suggests that children's ungrammatical flap-context productions continue at least until early literacy. In speech varieties with a complex socially/pragmatically mediated system of flap variation, gradual progress may continue for several years on the way to full adult control (Labov 2011). Treiman et al. (1994) show evidence of ongoing development with a task in which schoolchildren repeat medial flap words extremely slowly, forcing a choice of [t] or [d]. Children starting kindergarten (mean age 65 months) say [t] or [d] at chance for both /t/ and /d/ words. Later in their kindergarten year, children develop a bias towards [d] for all words, resulting in lower-than-chance performance for /t/ words like *city* (although the [d]-bias is still stronger for /d/ than for /t/, with higher rates of choosing [d] for /d/ words than for /t/ words indicating some degree of sensitivity to the underlying contrast). This is reflected in spelling errors as well, where children entering kindergarten select 't' or 'd' at chance but become biased towards 'd' over the next several months; spellings then become more adultlike during first and second grades, particularly in words like *cuter* where the flap is morphologically conditioned (Treiman et al. 1994).

Children express the voicing distinction between flapped /t/ and flapped /d/ well before they fully control the conditioned flapping process. For word pairs like *putting/pudding*, Rimac & Smith (1984) find that schoolchildren have longer flap closure durations than adults, but the two age groups have a parallel relative difference of longer vowels before /d/ flaps than /t/ flaps. Hanson and Shattuck-Hufnagel (2011) find this effect of underlying voicing on preceding vowel length already present at 30-39 months, while Klein et al. (1998) confirm that productions of some children 30-48 months are consistently perceived as flaps despite having long closure durations relative to adult flaps. Therefore, even at a stage when children differ from adults in some fine details of flap articulation, they are already systematically sensitive to the effects of underlying voicing.

Taken together, the studies of English flap development reviewed so far in this section lack data for certain word contexts and ages. Quantitative analyses often group data in intervals of at least 6-12 months, or combine multiple children's data by age for longitudinal analyses. If non-linear trajectories exist, they could easily be obscured by this type of grouped analysis, as could any other developmental changes that occur over a relatively short interval or at different ages for different children (Tessier 2019). As an initial effort to gather detailed longitudinal data for young children's flap productions, Richter (2017) observes one child's medial flap productions at 4-month intervals from 1-4 years (Providence/Lily). 25 high-frequency medial-flap words are perceptually coded as correct (flap) or error (something else), reproduced as Figure 4.1, which strikingly illustrates a non-monotonic acquisition course. Coding is based on perception of a single judge, a native US English speaker from Philadelphia, who had training in phonetics but was not aware of the specific interest in developmental trajectory at the time of coding in order to avoid unintended bias.[1] Flaps are produced with moderate accuracy in Lily's earliest recordings when articulation is still challenging, and then produced with high accuracy, before abruptly dropping to only 51% accuracy at age 3 (with the non-target productions after 2;8 being mainly [t] or null). The marked drop, that has not been fully recovered from by age 4;4, signals that this child has learned an allophonic relationship involving the flap. By the time Lily learns this, she may also already be sensitive to the voicing contrast of /t,d/ flaps, because even before age 2 she was systematically producing an adultlike pattern of relatively longer vowels before voiced consonants (Ko 2007). Motivated by the evidence of Lily's non-monotonic development, §4.2-4.3 will extend this observational study to include three more children and a larger vocabulary sample.

Children also produce occasional innovative errors that are especially diagnostic of allophone learning, such as (4.3) *ready, sodas* with [t]; or similarly *Florida* with the normally flapped 'd' realised as [t] (De Jong 2011: 2725); and [t,d] substitutions for medial flaps in word repetition tasks (Burrows et al. 2019). (4.3a) resembles the slow pronunciation task of

[1] My thanks to Kara Dang (University of Pennsylvania 2019, Cognitive Science) for coding, analysing, and reporting on this data.

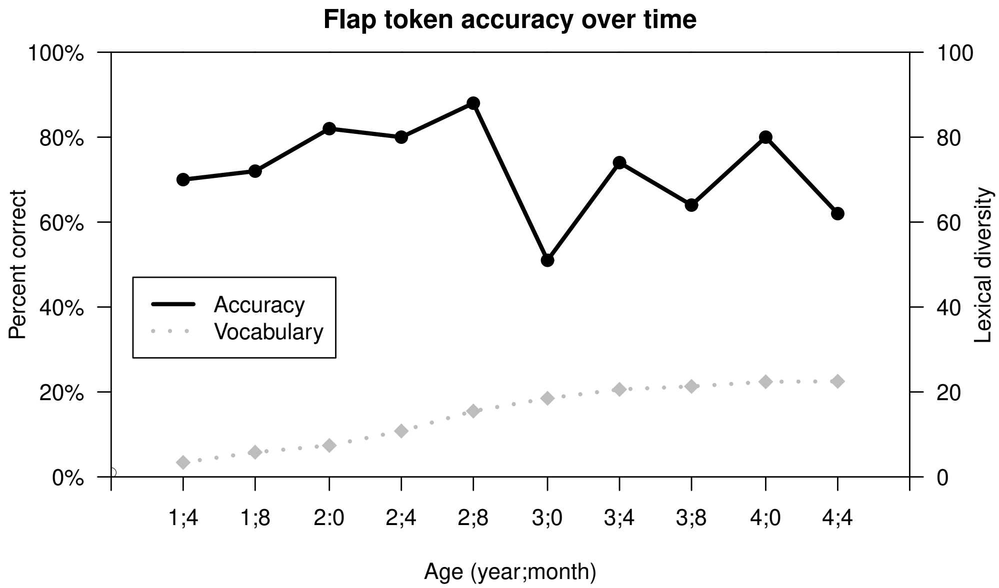


Figure 4.1: One child's productions in obligatory alveolar flap contexts (Providence/Lily). Labelled ages are upper bounds of four-month intervals.

Treiman et al. (1994), although it is spontaneous natural speech; a pause between syllables forces the child to produce something other than a flap as the onset of an isolated syllable. (4.4) is similar, as deleting the first syllable of *battery* makes the flap segment utterance-initial. However, (4.3b) is spoken fluently without word-internal breaks in a normal-speed word repetition task.

(4.3) a. *ready... jump!*
[wɛ ti] (Providence/Lily, 2;6.12)

b. *sodas*
[ˈsotəz] (CHILDES: PaidoEnglish, two children at 3;9 and 4;9)

(4.4) battery
[ˈtuːri] (Providence/Ethan, 2;5.26)

4.2 Methods

4.2.1 Speakers

Medial flap data includes productions of four children in the Providence corpus: Ethan (11–35 months), Lily (13–48 months), Naima (11–46 months), and William (16–40 months). They are monolingual English learners in the area of New England, considered to have typical language development (Song et al. 2015). A secondary analysis of word-final intervocalic /t/ includes only Lily and Naima, who have the most available data in this context.

These four children are included because together their data covers a large range of ages and stages of overall language acquisition, while within this as individuals they represent a variety of different trajectories and rates of development. Overall language development is estimated by Mean Length of Utterance (MLU) and moving-average Type-Token Ratio (TTR). MLU may be measured in either words or morphemes per utterance; morpheme-based MLU is usually considered more reliable, although the word-based version may provide complementary information on development of the lexicon (Scarborough et al. 1991, Wieczorek 2010). Up to MLU of about 4, it has a fairly reliable relationship with children's stage of morphological development in English, and may also describe syntactic development somewhat (Brown 1973, Fenson et al. 1994). MLU can be used as a longitudinal measure to index a child's development over time (Rollins et al. 1996). In this study MLU is measured in morphemes per utterance (Scarborough et al. 1991; Wieczorek 2010), calculated with the *mlu* program in CLAN using morpheme-level annotations provided with the Providence corpus. TTR is a measure of lexical diversity that can describe differences in language use of various adult populations as well as developing children (Wu et al. 2019). Children whose speech often repeats elements from a small inventory have lower TTR, while children who readily command a larger vocabulary may have higher TTR (Covington and McFall 2010). Moving-average TTR is calculated using the *freq* program in CLAN, within a moving window of ten words to find an accurate index of lexical diversity that is not artificially affected by transcripts' total lengths (Covington and McFall 2010).

4.2.2 Words

The sample includes common words (produced by multiple children) and any words that individuals produce multiple times several months apart, to capture an incremental representation of children's language. The medial-/t/ analysis consists of words with intervocalic /t/ preceding an unstressed syllable, optionally with /r/ directly preceding /t/. There is a core set of ten shared word types, which are each produced at least once by three or four of the children (*better, bottom, dirty, getting, gotta, kitty, little, potty, turtle, water*). In addition to this, each child's medial-/t/-flap vocabulary also includes any words they produce in two or more sessions several months apart. Therefore, the vocabulary included in the analysis is slightly different for each child. This strategy for constructing the sample aims to gather some data that is comparable across children while also maximising the amount of individually useful longitudinal data for each child. The final-/t/ sample (for two children) contains words ending in /t/ preceded by a vowel, with optional /r/ directly before /t/. For each child, all such words that they produce at least twice several months apart are included, except for exclusion of extremely high-frequency /t/-final functional morphemes (*that, what*) as it is often hard to know whether instances of these actually contain final /t/ according to a child's word segmentation. For words where flapping varies by speaker or dialect, such as *artist*, inclusion for each child is based on whether their parents' pronunciations use an alveolar flap.

When coding pronunciations for a child, by default all tokens of the medial-flap vocabulary items are included. For the two children whose final /t/ is also studied, pronunciation coding only includes tokens where the final /t/ is intervocalic (followed by a vowel-initial word in the same utterance, rather than in utterance-final position or followed by a consonant-initial word), since this is the only flap-permitting context for final /t/. In all cases, tokens are excluded if the audio quality is poor, such as two people talking at once or excessive background noise. Tokens are also excluded if the child's pronunciation is affected by acts like crying, laughing, etc. Furthermore, to reduce oversampling, if more than three codable tokens of the same word appear in a single recording session, only the first three occurrences are used.

Coding primarily aims to track phonological development involving [ɾ] as shown by production of non-/adultlike targets. Although the Providence corpus is distributed with IPA transcriptions for child utterances, these are made by different transcribers for different ages, and inconsistencies in transcription of /t/ variants make them unsuitable for detailed longitudinal analysis. Instead, pronunciations are labelled by perceptual coding of a single native English speaker,[2] which previous literature has established as a reliable way to track developmental trajectories of alveolar flaps (Klein et al. 1998; Eddington and Channer 2010; see §4.1.2). The categories annotated for word-medial /t/ are: [ɾ], [d], [t], [ʔ], deletion, and other (non-alveolar consonant, etc.). [d] and [t] are distinguished by voicing or aspiration, while [ɾ] is distinguished from [d] and [t] principally by manner of release and duration. Deletions may affect either an entire syllable, or only the /t/ segment; deletion is coded conservatively, only when there is no indication of glottalisation [ʔ], no intervocalic glide such as [j], or any other segment corresponding to /t/. Annotations for intervocalic word-final /t/ are: [ɾ], [tʰ], [d], [ʔ, t̚], deletion, other. Most of these are distinguished like the medial categories. The additional segment [t̚] which does not appear word-medially is coded together with [ʔ] because this difference is not clear in noisy recordings, not critical for this study since either segment is grammatical, and often an artificial distinction as articulatory gestures for glottalisation and alveolar closure may overlap in the same token (Song et al. 2015; Tollfree 2001). Supplementary audio examples showing how the coding scheme is applied are available at https://www.ling.upenn.edu/~ricca/dissertation/files/coding.html.

This corpus study focuses on /t/ pronunciations, and not /d/, for several reasons. More studies on child and adult /t/ variants provide context for interpreting the results, while the studies that have included both /t/ and /d/ flaps do not suggest they are unconnected processes that would need to be examined independently (e.g. Herd et al. 2010; Klein and Altman 2002; Song et al. 2015). The [d] ∼ [ɾ] contrast is perceptually less distinct than

[2]The coder is also a heritage Latin American Spanish speaker; bilingual Spanish/English listeners may discriminate [ɾ] ∼ [d] nonword stimuli better than monolingual English listeners, as these segments are phonemically contrastive in Spanish (Boomershine et al. 2008).

[t] $\sim$ [ɾ], and the suitability of perceptual coding for English /d/ variants has not been evaluated as extensively as /t/ coding (§4.1.2). Finally, some children in the Providence corpus have much less data for /d/ flap-contexts than /t/ flap-contexts. Models for these children in Chapter 5, which predict flap allophone acquisition trajectories from combined /t/ and /d/ vocabulary, can correctly account for the /t/ productions observed here in 4.3.

4.3 Results

Due to inherent drawbacks in the nature of the child speech corpus data, e.g. relative oversampling at older ages compared to sparse data at early ages, and extremely uneven distribution of vocabulary items across recording sessions, the results for children's flap productions are reported here in two complementary styles.

First, §4.3.1 presents a detailed view of only flap-target and deletion rates across time. Segmented linear regression models are fit for each child, using the Segmented R package (Muggeo et al. 2008), to describe whether and when there appears to be a change in their rates of producing target variants. This analysis treats flap-target production rate as a piecewise linear function of a child's age. Unsegmented linear regression models describe whatever dataset they are fit to as if the observations all result continuously from the same process, while regression models with two or more segments can describe the data as the output of a different process for each segment. In this study, acquiring an allophonic grammar is understood as a qualitative change in the process that gives rise to a child's productions, which may produce a different distribution of productions compared to the child's initial non-allophonic representation. While there is no reason to think that children's rates of target production actually rise or fall in a specifically linear manner, fitting linear models requires only one parameter (the slope) for each segment and is therefore a better choice for these small and irregularly-distributed datasets. The precise values of the segment slopes are fairly irrelevant in this analysis; instead, the important outcome is whether there is a breakpoint, and if so, when the breakpoint appears. To test this, both one-segment and two-segment regression models are fit for each child's rate of flap target production (weighted by the total number of flap-context productions in each session) as a function of their age; while it is also possible to test models with more breakpoints, overfitting becomes an increasing concern for this child data. In the two-segment models, the location of the breakpoint as well as both segment slopes are automatically chosen to maximise the model's fit to the whole dataset (Muggeo et al. 2008). Then, the one-segment and two-segment regression models are compared using the Bayesian Information Criterion for model selection. If a single-segment model is selected, this may mean there was no qualitative change in the child's flap grammar during the time period of data collection, while if a model with two segments is selected, the data may be generated by different processes before and after the breakpoint (even if, as stated, a linear model is probably not a particularly correct model of either process). Since the nature of this data set makes it difficult to perform a single meaningful statistical analysis, the approach taken in this chapter is instead to present a comprehensive set of various analyses, which collectively point to the same conclusion; this indicates that the conclusion is robust and not contingent on irrelevant particulars of a statistical test, even if no one test is especially informative in isolation. Highly reliable findings that there are breakpoints where children's productions start to become less adultlike provide firm evidence for the existence of U-shaped learning during alveolar flap acquisition.

Alongside this, §4.3.2 is intended to give a higher-level complete picture of children's usage for all flap-context /t/ variants across time. In this section, data is grouped into longer time intervals, constructed so as to be comparable from child to child as well as across time intervals for each child, although the relatively coarse grouping of data points required for a robust description here does not permit the level of detail shown in §4.3.1 for all of the other pronunciation variants in §4.3.2.

Finally, since a noteworthy aspect of these results was a period of increase in tokens coded as medial flap segment deletion, §4.3.3 examines the nature of those productions in more detail.

4.3.1 Detailed results

This section takes a closer look at U-shaped learning, by tracing each child's frequency of producing both flaps and deletions for word-medial /t/ flaps. This analysis uses a fairly strict or conservative approach, investigating specifically the development of flap productions and not automatically treating [d] as equivalent. In this section, [d] is grouped together with [ɾ] for only one child (Ethan), because due to the relationship between his articulatory and morphosyntactic development (as illustrated and justified further in §4.3.2) it seems indefensible to interpret a complete lack of perceptible flaps in most of his speech as a complete lack of awareness about the flap target as distinct from the voiced alveolar stop. Likewise, deletions are a focus for detailed analyis in this section because the complete results reported in §4.3.2 reveal that, although they are errors in the medial flap context, at some point children's frequency of producing them increases.

Figure 4.2 shows the session-by-session rates of flap (or combined flap and d) productions, and deletions, for each child. This data is quite noisy, as there are sometimes only a few tokens in a session; while the images in Figure 4.2 do not explicitly show how many tokens are in each session, they do serve to display the changing frequency of child recordings at different ages. Figure 4.3 shows the same data more clearly with smoothing applied, using a moving window of size 5 weighted by the number of tokens in each session. In any case, these results show a clear and sudden drop-off in flap rate for three of the children, accompanied by an apparent tradeoff of flap production rate with an increased deletion rate. This is especially visible for Naima and Lily, who generally have the most tokens per session/least sparse data. It seems likely that a low point in the rate of flap production occurs around the same time as a peak in deletion rate.

All of the children start with some period of few or no flaps; although Ethan's graphs in Figures 4.2-4.3 also include [d], Figure 4.12 below indicates that he too has zero flaps for a long initial period. The sharp change from no flaps to a rapidly increasing rate of flap production (approximate times: Lily: 20 months; Naima: 17 months, Ethan: possibly 32 months; William: 25 months) may not be a qualitative phonological change, but could

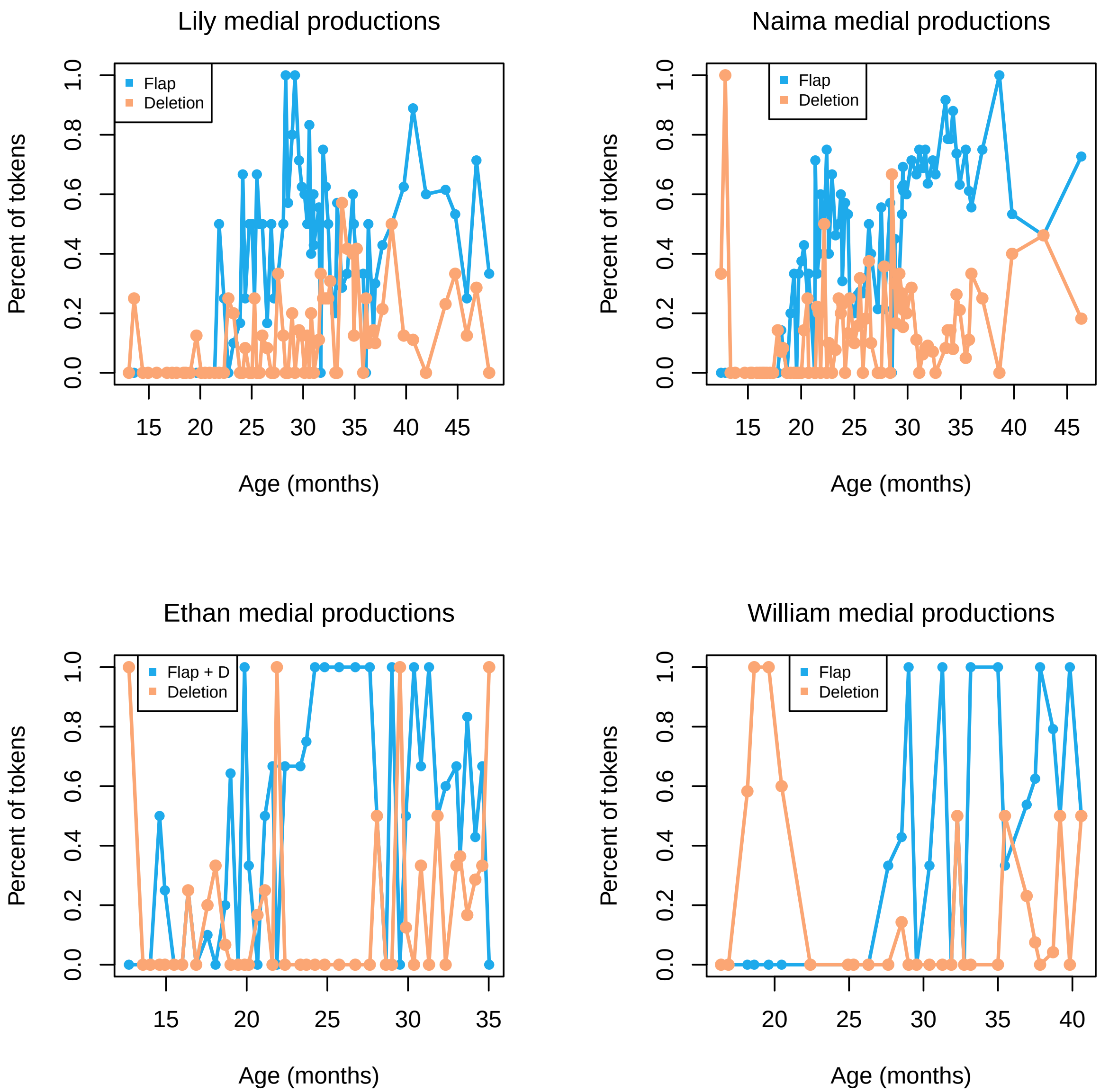


Figure 4.2: Proportions of medial flap-context /t/ realised as a flap, and deleted, for every session of four children in the Providence corpus. Data for Ethan reflects a combination of [d] productions with flap productions.

instead signal when children gain the articulatory skill to produce flaps.

The later stages of development, following the maximum extent of children's regression (the greatest dip in rates of flap production, along with peak in deletion frequency), are rather

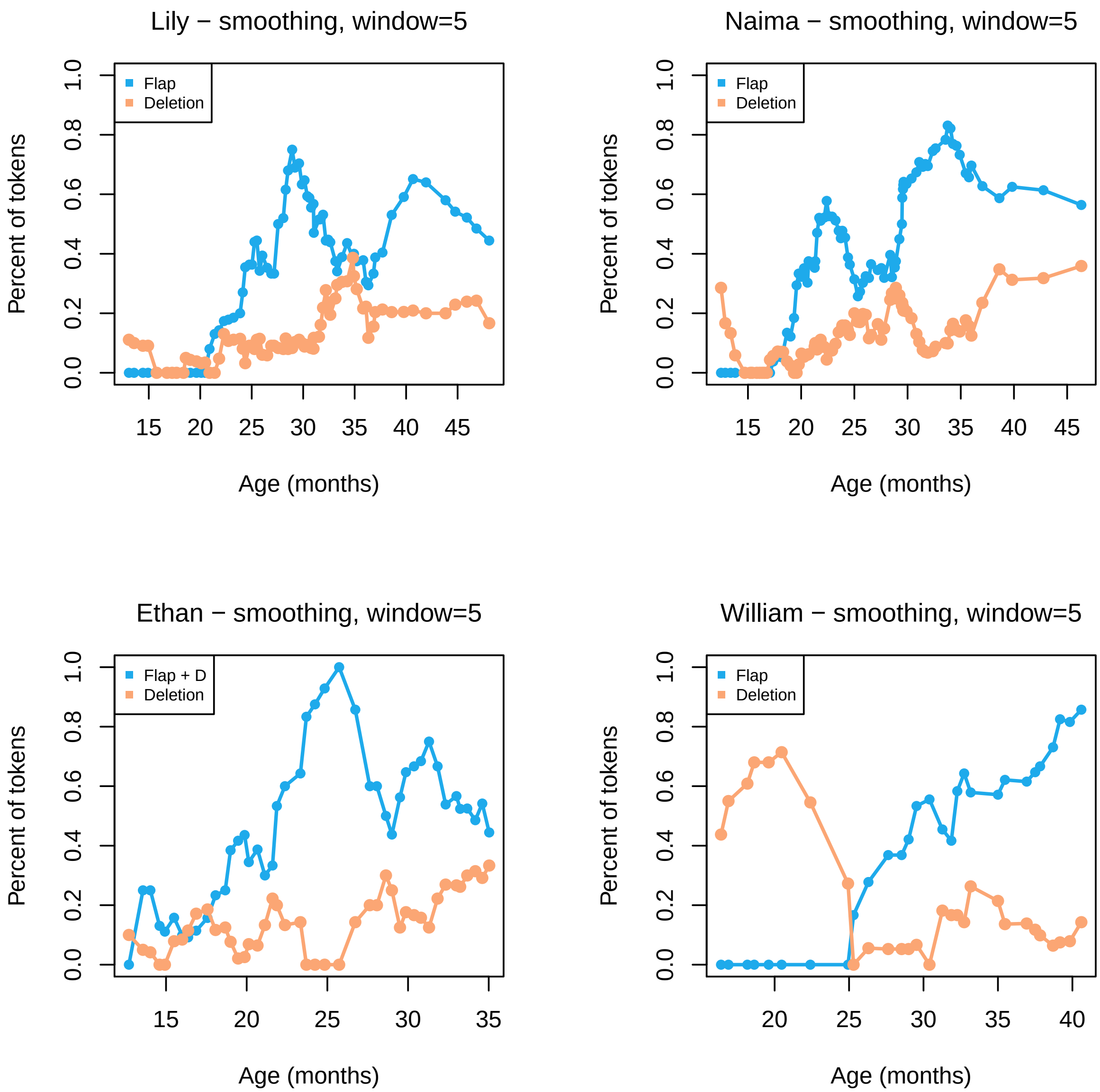


Figure 4.3: Proportions of different medial flap-context /t/ realisations over time, smoothed with a moving window of size 5.

less clear in the present data. For Naima, the overall rate of deletions remains elevated compared to what it was before the sharp turnaround in her rate of flap production that occurred around 23 months, but after 30 months her rate of flap production also clearly increases again (this is possible because her rate of d-production declines in this period). Lily also has some recovery of flap productions after reaching a minimum rate of flap use around 35 months, although she never approaches her previous high frequency; her rate of deletions always remains higher than it had been before 30 months as well. It is not quite possible to tell whether Ethan has reached his maximum deletion frequency or minimum flap rate by the end of recording sessions, and William's data also give no information on this part of development.

Due to the span of developmental stages represented in Figure 4.3, from an initial inability to even articulate flaps all the way to Naima's return to high-frequency flap production (with MLU around 5), it would appear useful to focus analysis on the portion most directly relevant to APL, that is, the 'regression' away from adultlike language. This seems to include the turning point in a child's rate of flap productions, preceded by an increase in appropriate flap frequency and followed by a sharp drop, as well as a low point of minimum flap production following this drop, and the point where the rate of deletions is highest. Although these events generally occur around a similar MLU or TTR for different children, it is not quite possible to use either MLU or TTR as criteria to define a lower and upper age bounding the period of interest. As shown in Figure 4.4, TTR as measured on this corpus does not increase much more after reaching about 0.8, and therefore TTR cannot be used to identify the upper age limit of the relevant interval. MLU in Figure 4.5 is a very noisy measure at the level of individual recording sessions. MLU (along with TTR) is intended as a general index of a child's overall language development, which is difficult to define but assumed to strictly increase over time as a language is acquired; however, even if fairly heavy-handed smoothing is applied, MLU does not behave as a monotonically increasing quantity in this dataset, and therefore MLU thresholds also cannot be used to identify a single continuous age range of interest.

It is still possible to use both MLU and TTR as criteria to identify subsets of 'relevant' sessions that match the idealised thresholds. This will construct a rather artificial dataset for studying child development, as the included recordings will not form a continuous block

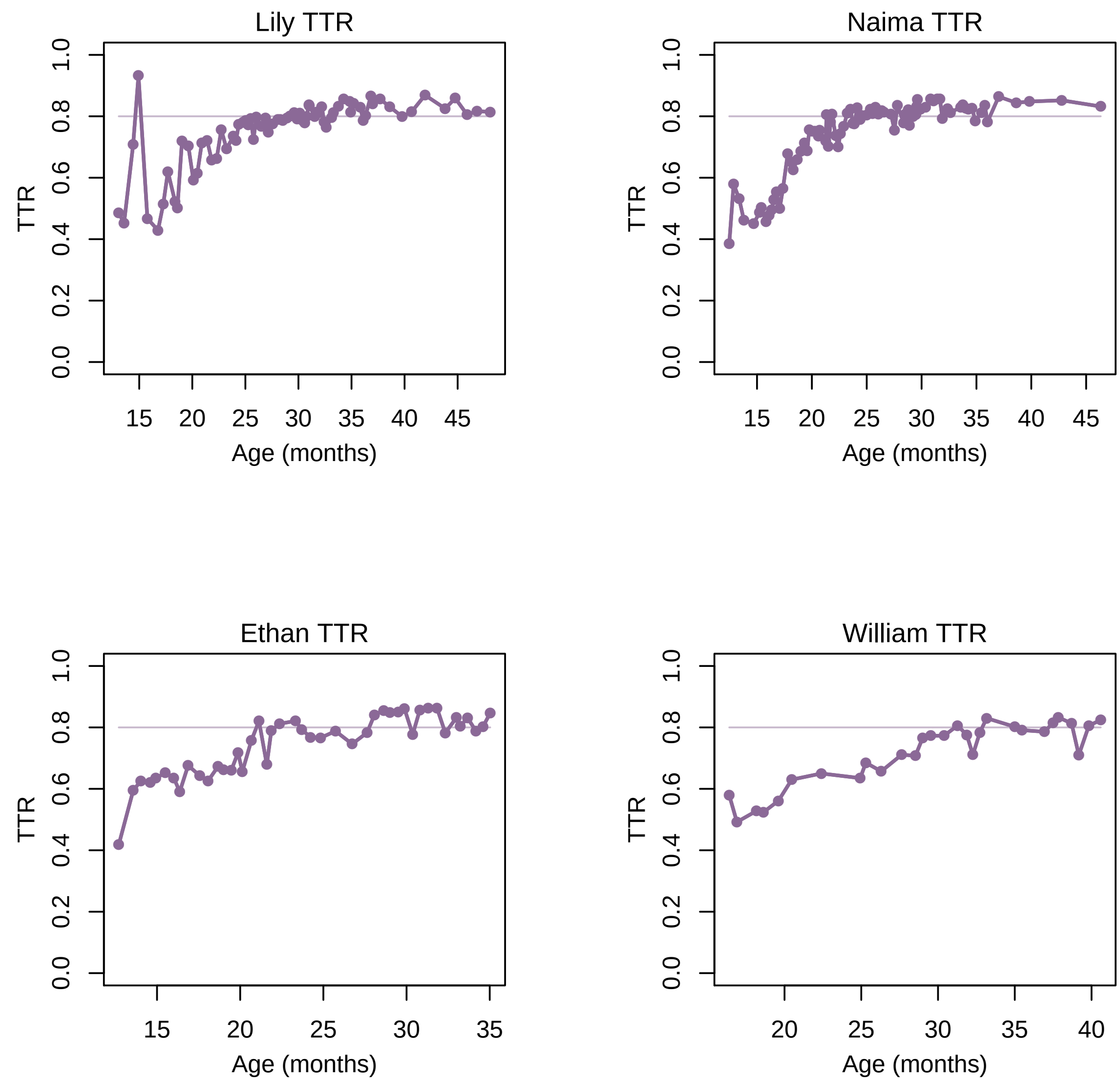


Figure 4.4: Type-token ratio for each recorded session, calculated as stated in §4.2.1. A line is shown at TTR = 0.80 for reference.

of time when TTR or MLU values fluctuate in and out of the defined range. However, it could be seen as a more objective way of processing the data, or an illustration of a method which could be more meaningful on a larger dataset with less random fluctuation. Figure 4.6

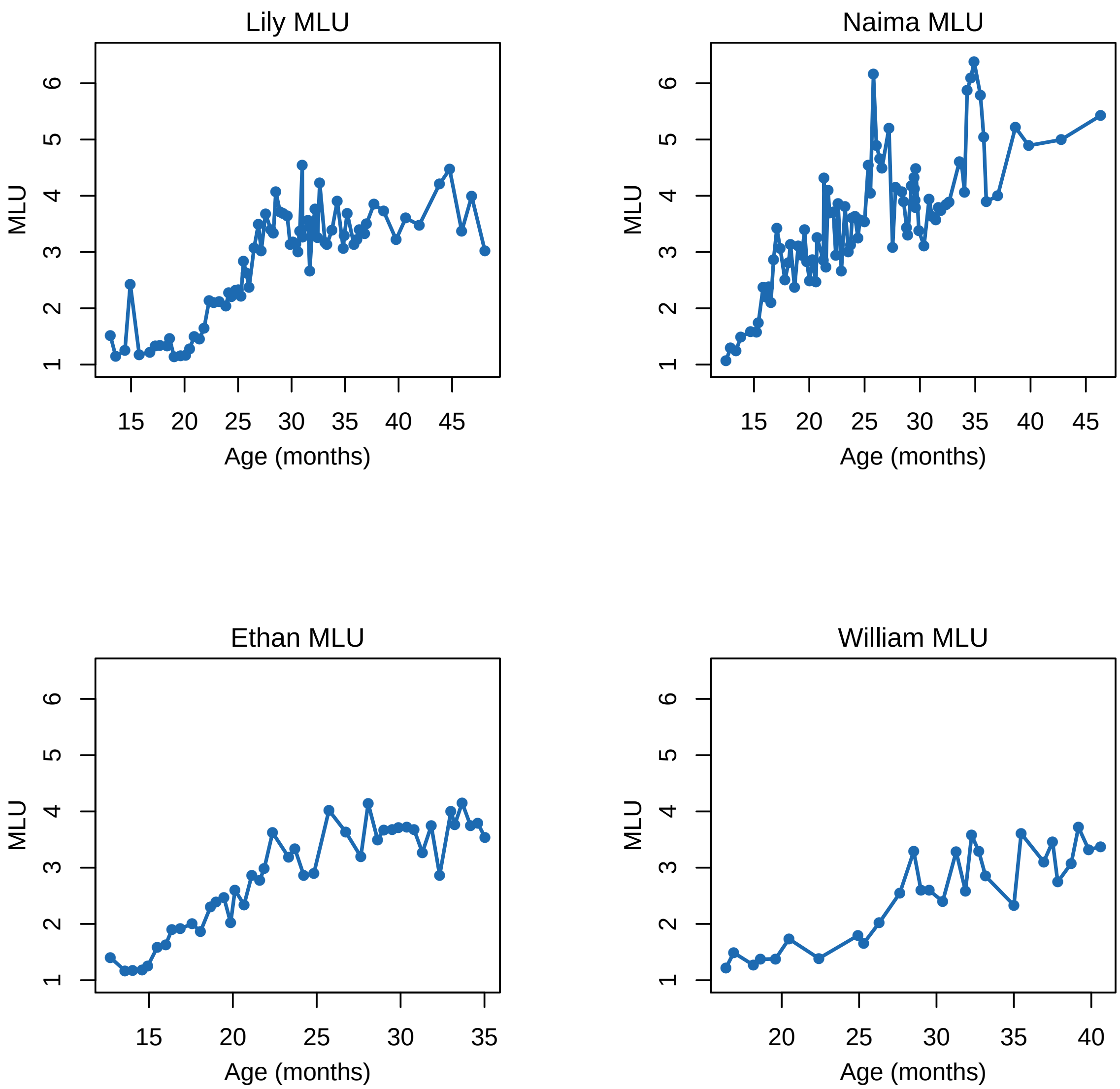


Figure 4.5: Mean length of utterance for each recorded session, calculated as stated in §4.2.1.

uses a TTR range of 0.6-0.8 to try to identify the most relevant sessions. Segmented linear regression models are also fitted for the rates of flap(+d) productions, to help establish whether there is a non-monotonic acquisition course. The Bayesian information criterion (BIC) is used for model selection, preferring lower values, to decide between either a simple

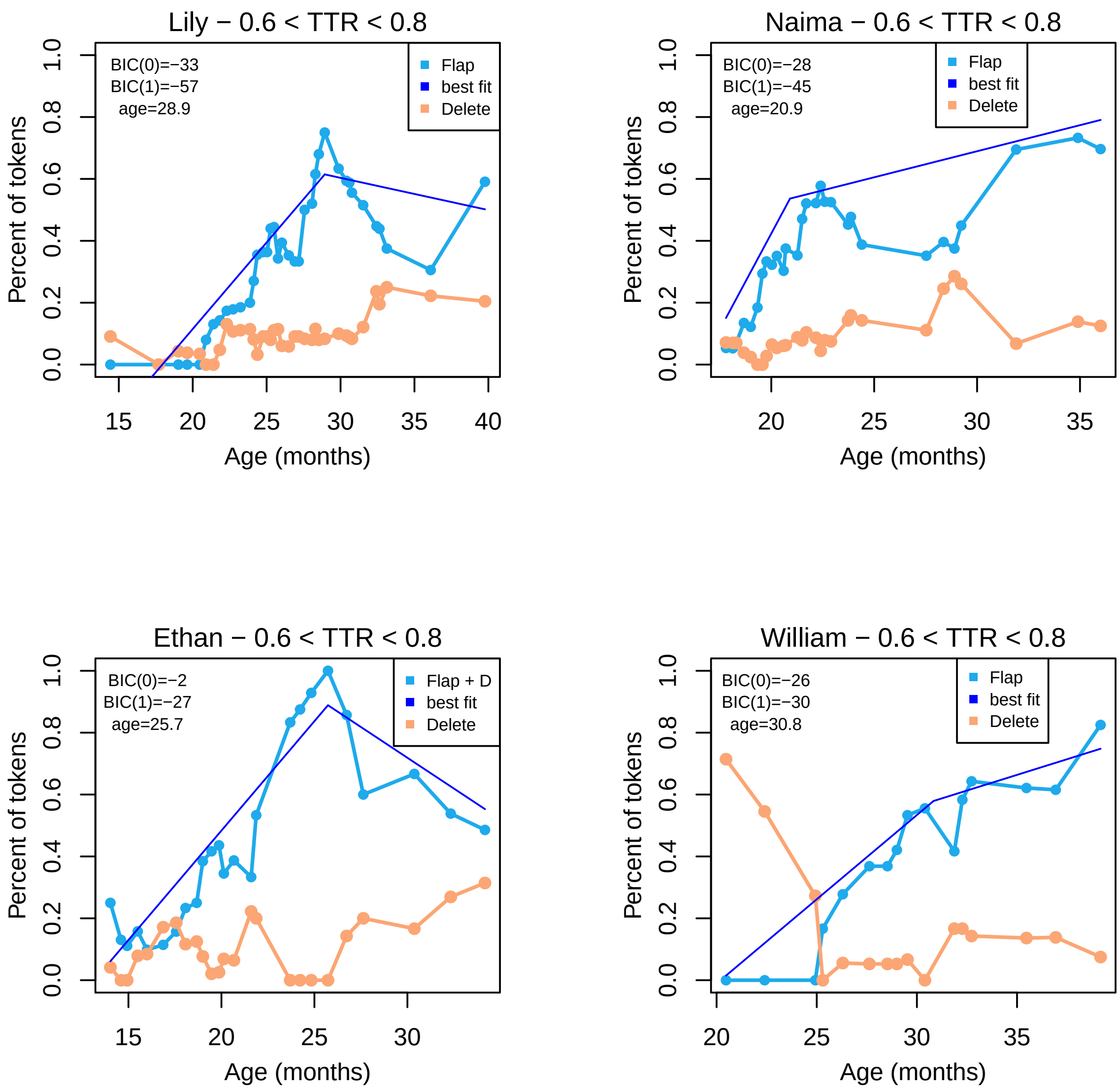


Figure 4.6: Medial /t/ realisations during the developmental stage relevant to U-shaped flap learning as identified by TTR range 0.6-0.8. Production rates are smoothed by moving average with a window size of 5. A segmented linear model is also shown for the flap productions.

linear model or two segments with distinct slopes. BIC reflects a combination of how well the model fits the data and how many predictors are in the model (where fewer predictors,

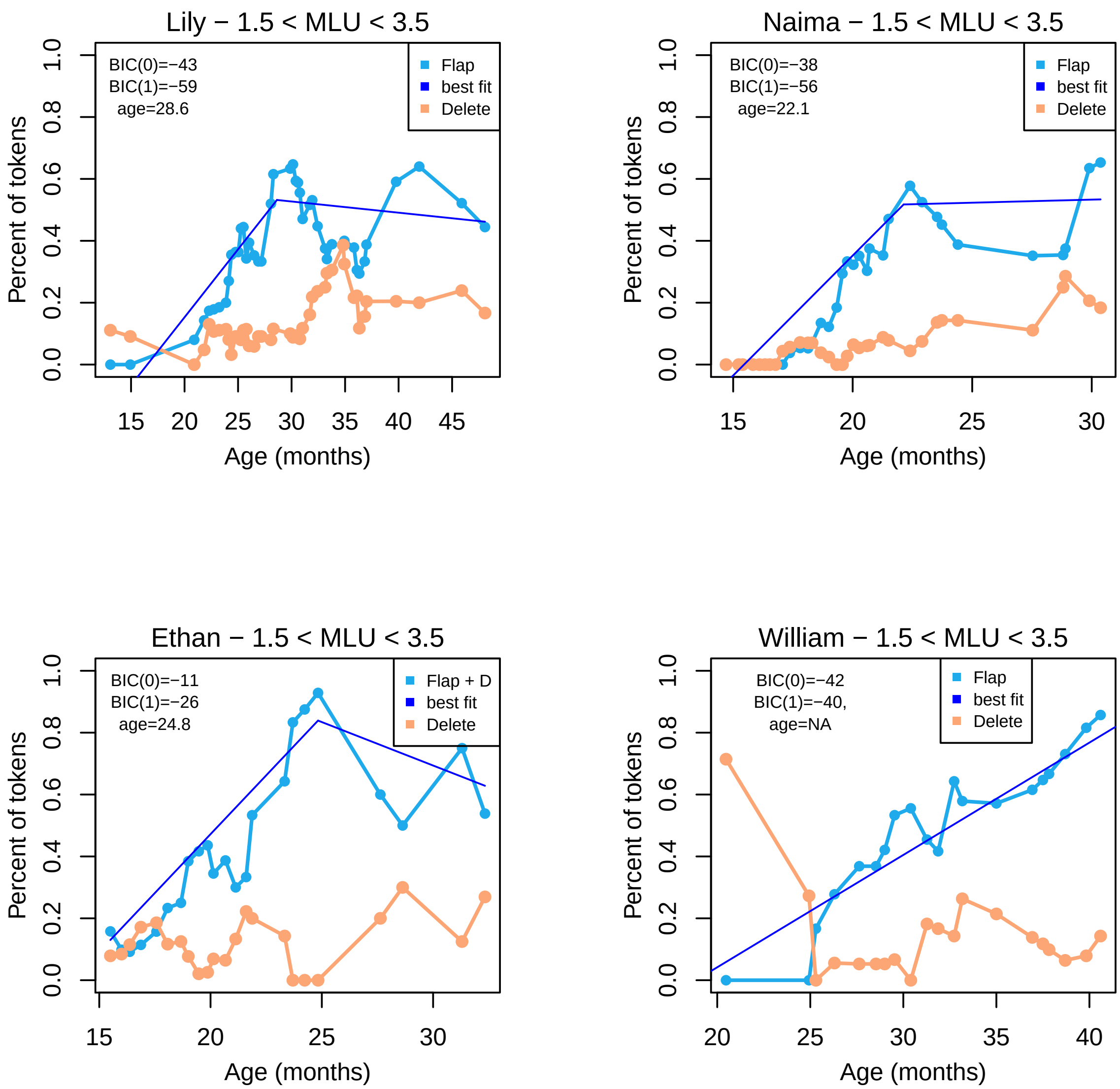


Figure 4.7: Medial /t/ realisations during the developmental stage relevant to U-shaped flap learning as identified by MLU range 1.5-3.5. Production rates are smoothed by moving average with a window size of 5. A segmented linear model is also shown for the flap productions.

i.e. one line segment rather than two, are preferred, so adding a predictor is only acceptable if it makes the model fit considerably better). Two-segment models are selected for all

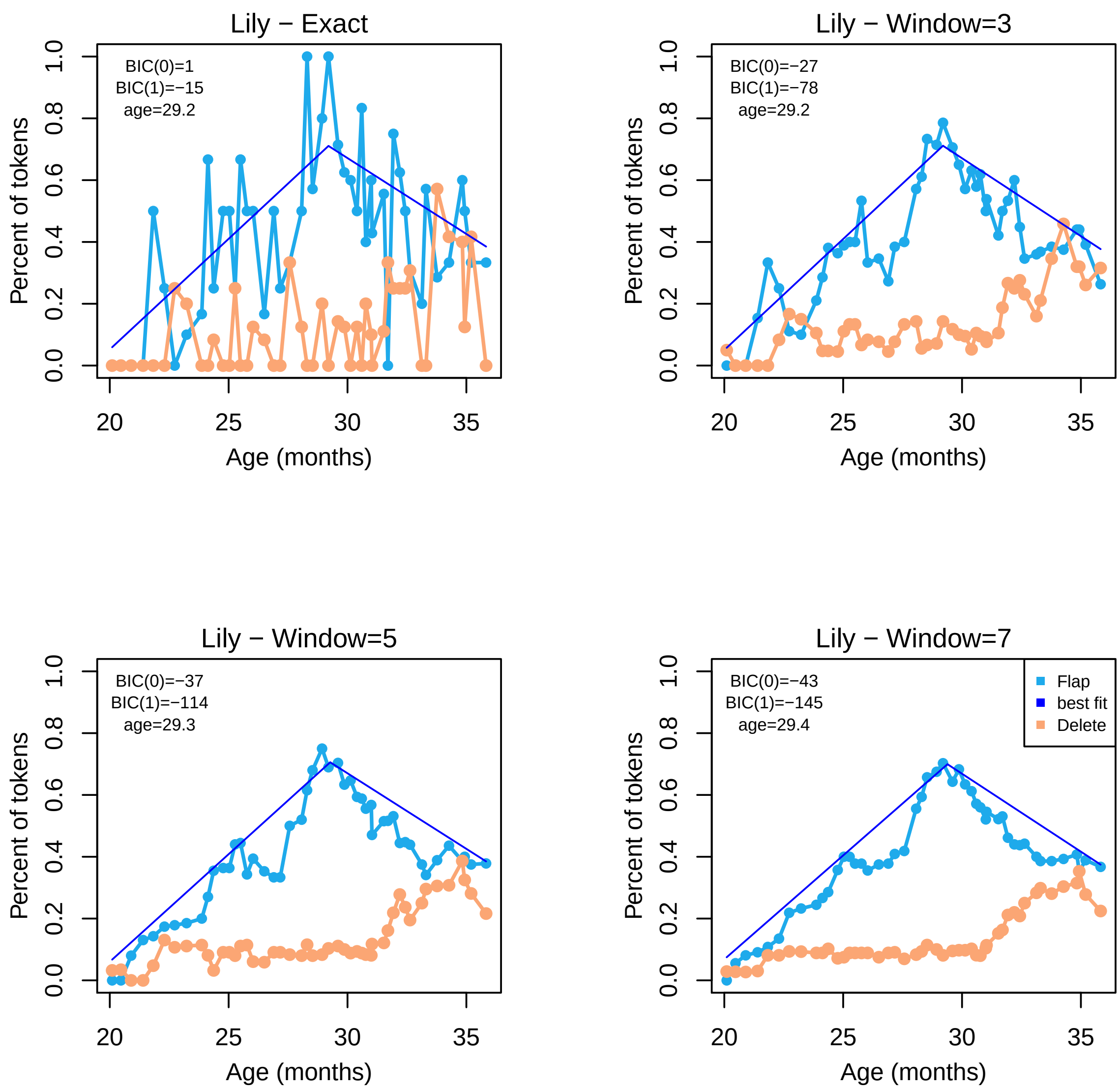


Figure 4.8: Lily: medial /t/ realisations during a selected period of time, showing best-fit segmented linear regression models, using various degrees of smoothing.

children in Figure 4.6, with breakpoints at 21 months (Naima; both slopes are positive but the second segment is much less steep), 26 months (Ethan), and 29 months (Lily); the difference in slopes between William's two segments is not significant. Figure 4.7 is very

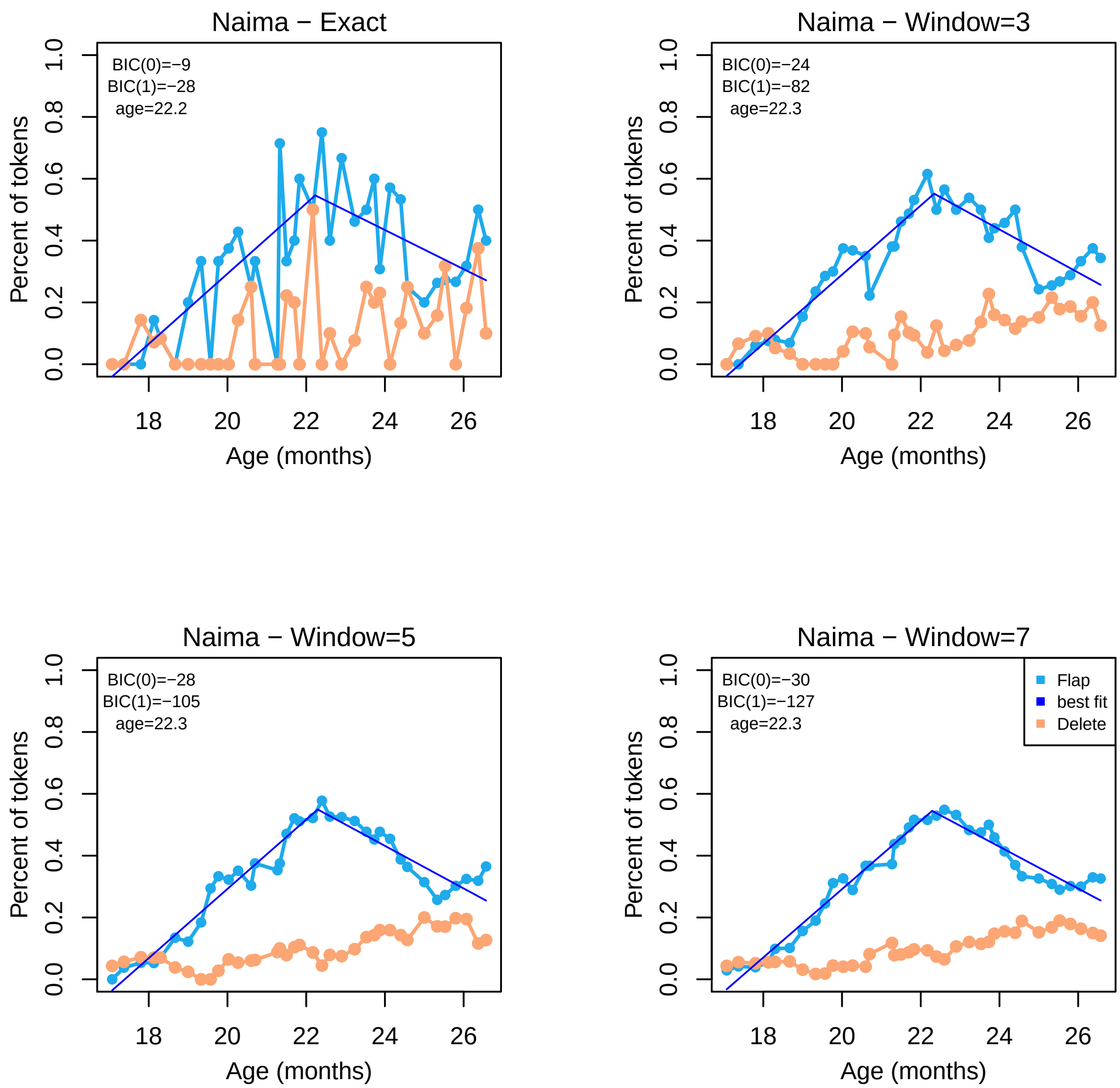


Figure 4.9: Naima: medial /t/ realisations during a selected period of time, showing best-fit segmented linear regression models, using various degrees of smoothing.

similar, but using MLU of 1.5-3.5 to select recordings. In this case BIC determines a single segment linear model for William, indicating monotonic increase of appropriate medial flap production, while there are breakpoints found for the other three children at 22 months

(Naima), 25 months (Ethan), and 29 month (Lily).

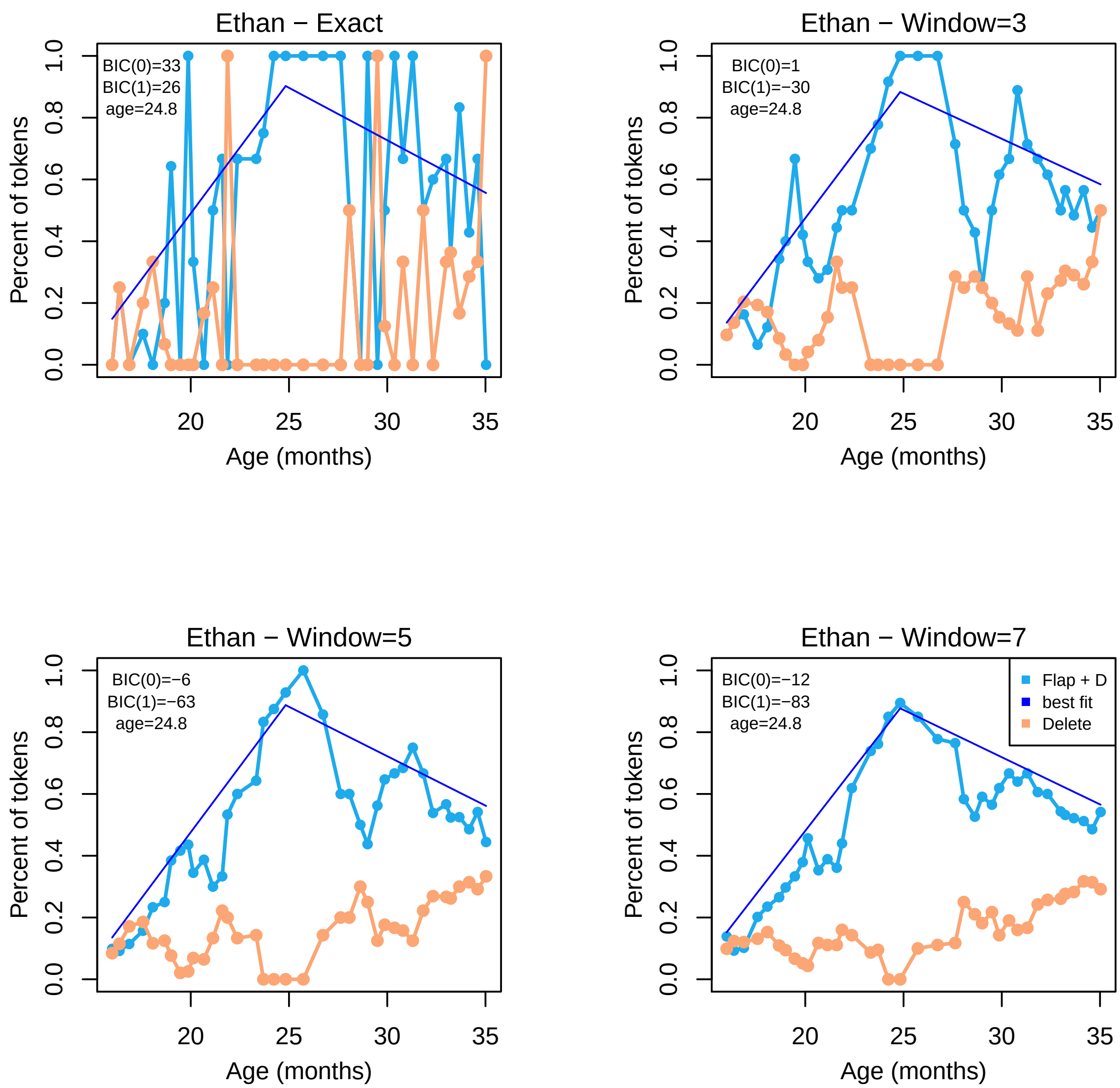


Figure 4.10: Ethan: medial /t/ realisations during a selected period of time, showing best-fit segmented linear regression models, using various degrees of smoothing.

Finally, Figures 4.8-4.11 repeat the segmented linear regression analysis over intervals defined by hand to precisely contain the typical U-shaped learning trajectory as far as possible.

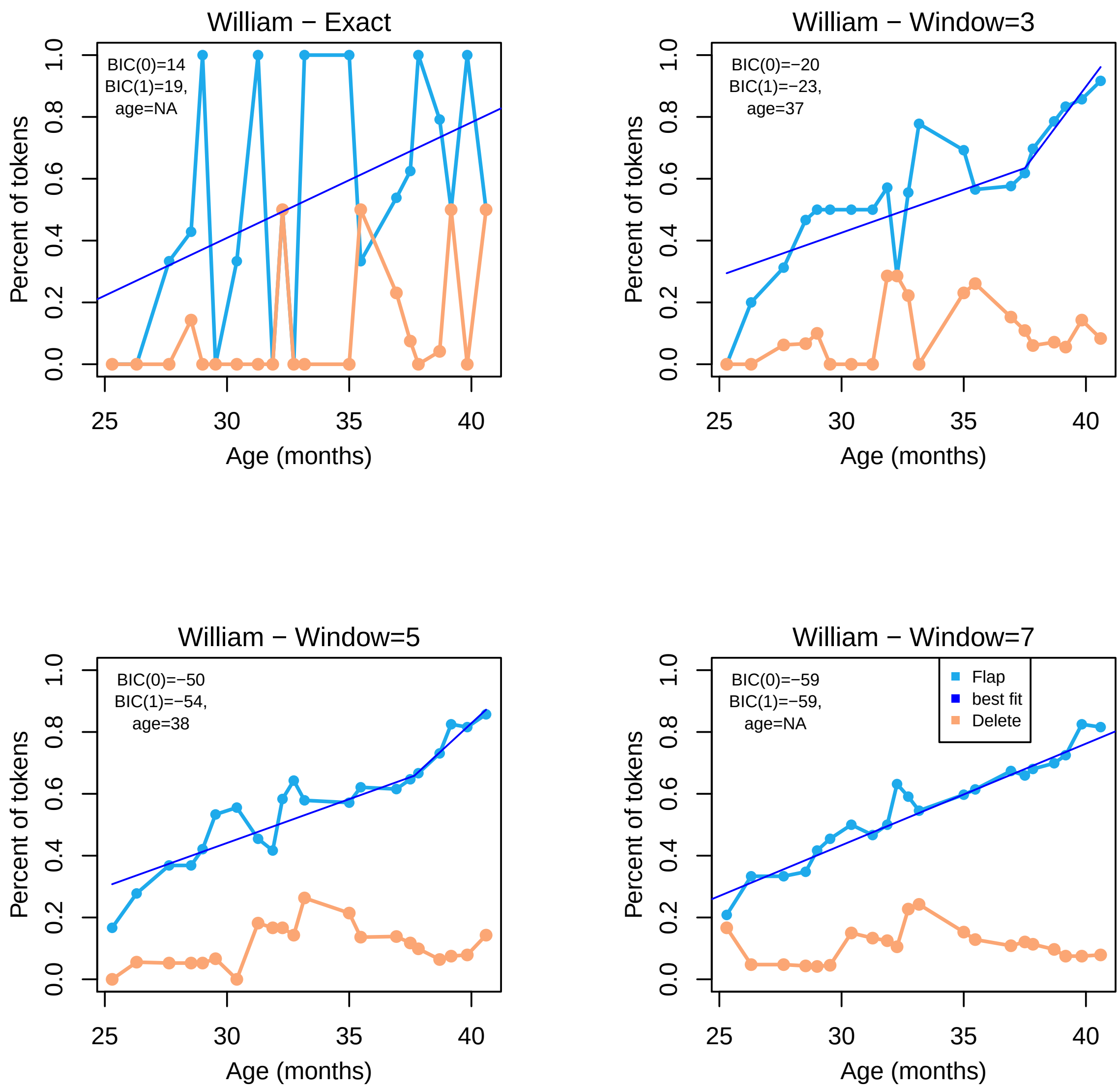


Figure 4.11: William: medial /t/ realisations during a selected period of time, showing best-fit segmented linear regression models, using various degrees of smoothing.

Lower and upper age limits for the relevant interval are chosen, and all recordings within these limits are included. Furthermore, segmented linear models are applied to data that is smoothed to various different degrees, to show that the results are not artificially dependent on certain smoothing parameters. The breakpoint analyses end up having highly consistent results throughout, which altogether provides strong evidence for the occurrence of regression or U-shaped learning. In Figures 4.8-4.11 the estimated ages for the turning point when children's rate of flapping begin to decrease are 22 months (Naima), 25 months (Ethan), and 29 months (Lily). In two cases BIC model selection fits a two-segment model for William's production, when the window for smoothing is 3 or 5; however, in this case the second slope is more positive than the first (rate of flapping starts increasing faster) which would not be a mark of U-shaped learning at this point in his development. In addition, there is minimal difference between how well a single-line model fits William's data here (e.g. BIC value -50 for one segment and -54 for two, often not considered meaningful in BIC comparisons as the difference in values is below 10), whereas the difference in BIC values from adding a breakpoint for the other children's linear model is often at least 20 and sometimes exceeds 100, indicating that the two-segment models provide significantly better fits for their data.

4.3.2 General results

Figure 4.12 shows all medial /t/ variants produced by Ethan (11-35 months, 275 tokens), William (16-40 months, 168 tokens), Lily (13-48 months, 464 tokens), and Naima (11-46 months, 888 tokens). Data is now divided into only four stages per child, aiming to balance even distributions of each child's tokens and time intervals across the bins while also providing some comparable divisions of similar TTR or MLU across all four children where practicable.

Of the medial /t/ categories [ɾ], [d], [t], [ʔ], deletion, and other, as described in §4.2.2, the [ɾ] and [d] realisations are considered to indicate adultlike flap targets in Figure 4.12 (unshaded bars), while [t], [ʔ], deletion, and other segments are labelled as errors or not adultlike targets (diagonal shading). The reasons for grouping [d] with perceptually-similar [ɾ] here are to maximise reliability of the overall adultlike/non-adultlike contrast in coding, as well as to incorporate the fact that adults generally would not perceive these child tokens as errors (in contrast to the clear perception of error in productions of [d] for obligatory

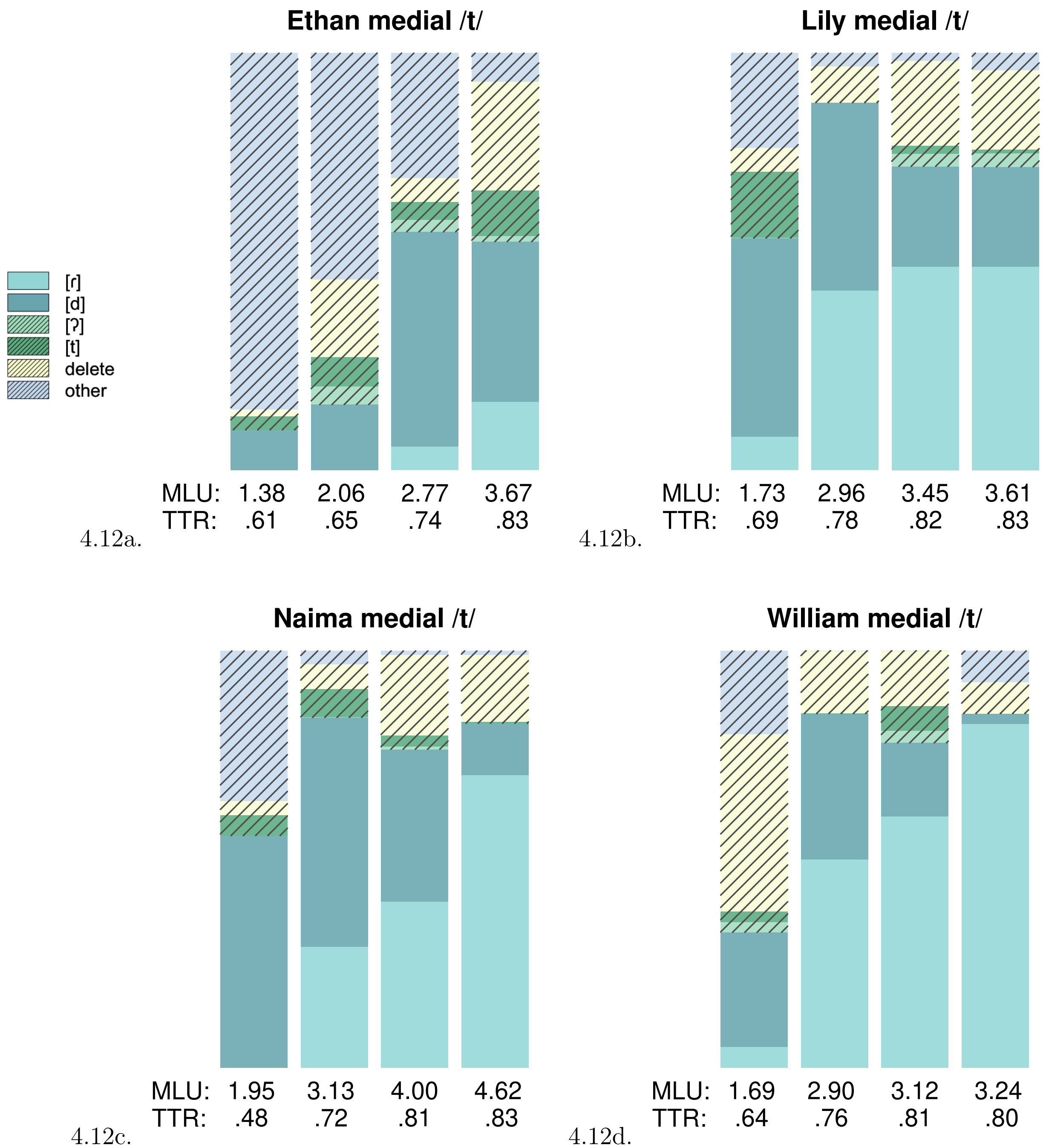


Figure 4.12: Medial /t/ flap-context productions for Ethan, Lily, Naima, and William. MLU counts morphemes per utterance, and TTR is averaged over a 10-word moving window. Shaded segments are errors, or do not resemble adultlike speech.

[t] in contexts like *table*); [d]-coded tokens can be flap-target productions where immature articulatory abilities do not quite achieve the flap percept. This especially seems to be an accurate interpretation for Ethan, who consistently uses [d] in the flap context, but produces no flaps at all until relatively late in the available corpus sample and even then produces them at only a low rate; this is part of a general pattern in his data to reach fairly advanced lexical and morphosyntactic development before mastering articulatory control. In contrast to considering [d] productions as likely flap targets, the categories [t], [ʔ], deletion, and other sounds are always counted as non-adultlike (shaded bars in Figure 4.12) because these are perceptually distinct from input forms and unlikely to represent poorly articulated flap targets. Deletions in the early stages are mainly deletions of whole syllables (4.5a) especially in words of 3 or more syllables, while later deletions omit only the /t/ segment (4.5b). Pronunciations in the 'other' category generally show interference from another sound in the word, either to satisfy the youngest children's apparent requirements for syllable/word structure (4.6a) or in particularly challenging words (4.6b).

(4.5) a. motorcycle man uh-oh
[mozaɪkə] (Providence/Naima 01;05.24)

b. here's the water slide
[wɑɝɹ] (Providence/Lily 03;08.23)

(4.6) a. bottle
[bɑbo] (Providence/Ethan 01;03.15)

b. watermelon
[wɑləmɛlɪ] (Providence/William 03;02.21)

Overall, Figure 4.12 shows that when MLU reaches about 3.5 morphemes per utterance or TTR is above 0.8, /t/ segment deletion increases considerably. Up to this point, medial [ɾ,d] frequency increases, until it becomes the majority of productions for all four children. Therefore, their ability to fluently articulate an acceptable target segment in this context is firmly established before they start omitting it, for the three who eventually do. One child, William, maintains a high rate of [ɾ,d] in his final recordings, and a consistently decreasing rate of deletions. His MLU is also below 3.5, and TTR remains slightly lower than the other children's, meaning that this limited data fits into the larger pattern established by the other three children. At all time points, children's realisations of any given word often vary from one token to the next; the different frequencies of each /t/ realisation are not simply an artefact of using one consistent form per word but shifting in spoken vocabulary composition.

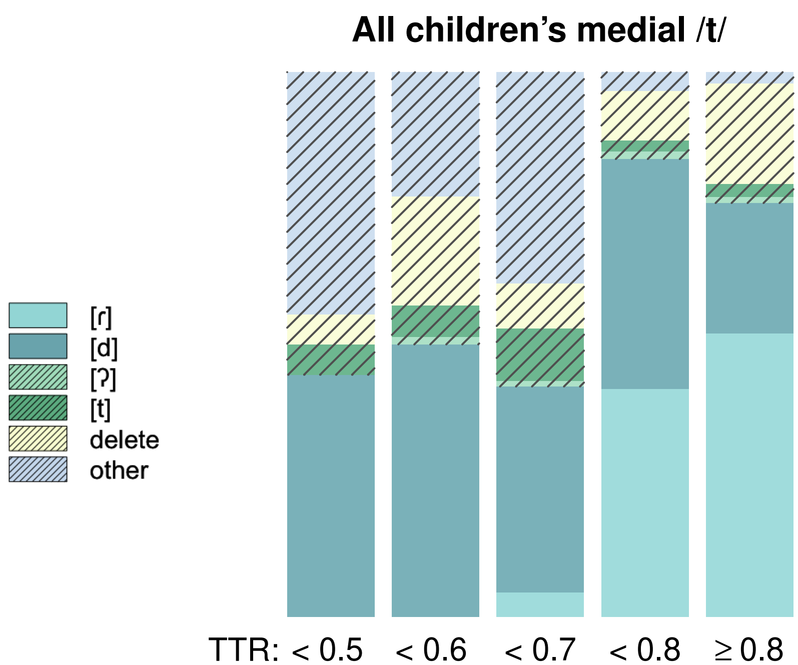
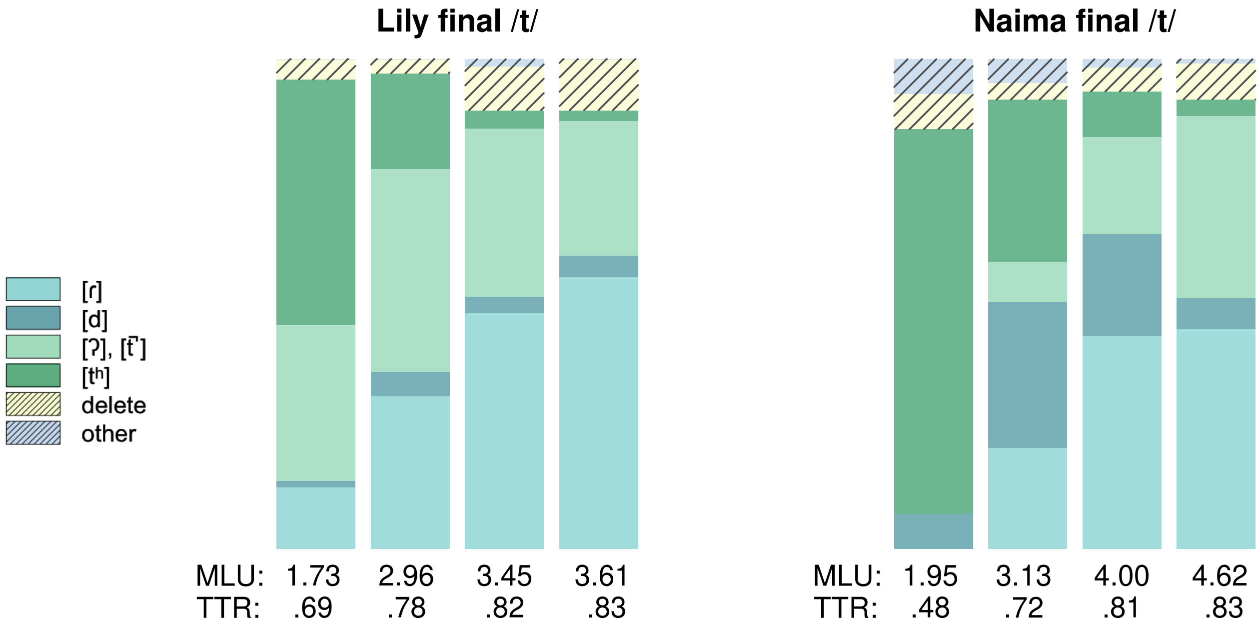


Figure 4.13: Medial flaps of all four children together, binned according to TTR.

Figure 4.13 compiles all data from Figure 4.12 to bring out this common pattern as it appears (at least in part) in all four children's data. Here the tokens are binned according to TTR, using thresholds of TTR < 0.5, <0.6, <0.7, <0.8, and ≥ 0.8. The actual average TTR for the samples in each division is 0.49, 0.53, 0.66, 0.77, and 0.82, while the average MLU is respectively 1.83, 1.93, 1.96, 3.04, and 3.82 morphemes per utterance. The expected increase in deletions is evident around TTR$\geq$0.8, while just before this when $0.7 <$ TTR $<$

0.8, children are highly consistent in producing flap-like segments.



4.14a. 4.14b.

Figure 4.14: Word-final /t/ of Lily and Naima, using the same divisions as Figure 4.12.

Figure 4.14 shows productions of intervocalic word-final /t/, for the two children with sufficient data in this context. The categories coded in this position are [ɾ], [d], [ʔ, t̚], [t$^{\text{h}}$], deletion, and other, as described in §4.2.2. Here, only deletion and ‘other’ segments are not grammatical in adult speech. Early high frequency of [t$^{\text{h}}$] is a notable feature, particularly for Naima, but it becomes much less frequent as children develop. Unlike the medial stop data, here there is never a substantial increase in new errors. Instead, children seem to steadily approach adultlike use of variants.

4.3.3 Deletion

It is not entirely clear why medial /t/ is deleted in children’s productions. Deletion in this context is not part of their input; it does not appear in several detailed studies of adult’s medial /t/ flap-context pronunciations (e.g. Eddington and Channer 2010; Patterson and Connine 2001; Zue and Laferriere 1979). Therefore, if deletion is a child’s target, this is an innovated error. Generally, omission of elements might result from the complexity of having to select a variant (allophones, allomorphs, etc.; Gerken and McIntosh 1993; Kehoe 2015; Tomas et al. 2012), though it is not yet clear precisely what cognitive process links this type of grammatical complexity to children's null productions. In any case, the results shown above cannot indicate if children have innovated deletion as a target for medial /t/, because the acoustic impression of deletion could also come from reduced or immature articulation of some non-null target.

If the target of 'deletions' is [ɾ], the apparent U-shaped acquisition might reflect phonetic or articulatory changes, not phonological development. However, poorly articulated or extremely reduced [ɾ] productions seem unlikely to be responsible for the increase in deletions at MLU$\geq$3.5. By this time, children may have competently articulated [ɾ] for many months, including in the same words where deletions later appear; they may use appropriate medial flap-context segments in 80-90% of productions, which suggests relatively mature articulatory control (§4.3.2). Once deletions increase, [ɾ] continues to be produced in fast as well as slow speech, while deletions also occur in both fast speech and slow or careful speech (audio in supplementary materials). This distribution would not be expected if deletions were reduced instances of [ɾ]. Potentially, a distribution like this could occur if children have some global loss of articulatory skill for [ɾ], but this is implausible: in contrast to U-shaped grammar learning, purely motor 'regression' is not commonly attested in child language acquisition, and there is no evident motivation for it (e.g. Klein and Altman 2002; Tessier 2013; McAllister Byun and Tessier 2016; Vihman 2014, 2019). Furthermore, the word-final data in Figure 4.14 show no such regression; instead, flap production continues increasing.

Children's 'deletions' could derive from articulation of some target that is neither [ɾ] nor null. For example, adult word-final /t,d/-'deletion' is perceived as categorical segment deletion but is still produced with residual [t,d] gestures (e.g. Purse 2019). Due to the complex relationship between articulation and perception, a similar possibility must be considered for children's medial data unless articulatory experiments show otherwise. Plausible targets are word-final /t/ variants like [ʔ, t̚, t^{h}], any of which would be an innovated error since North

American English does not generally permit them in medial flap-contexts.[3] In contrast to possible innovated deletion, these targets have a readily identifiable source. Of them, [ʔ] seems most likely to be articulated imperceptibly in the medial intervocalic context. Although this intuition cannot be evaluated with the present data, reduced or mistimed gestures easily obscure acoustic signatures of glottal stops, and these children have no adult models of [ʔ] to learn from in this context. Also, two children (Lily, Naima; Figure 4.12) begin to occasionally produce perceptible medial glottal stops around the same time as their rate of deletions increases (while Ethan has always produced some glottal stops in this context, and continues to). This data is highly compatible with the possibility that children's increased deletions reflect medial [ʔ] targets.

4.4 Discussion

Longitudinal examination of individuals' flap-context productions reveals new patterns not shown by previous methods of studying flaps or /t/-allophone development. In the context of existing research, these results are generally consistent with previous data to the extent this can be assessed, but also highlight some gaps in knowledge that still need to be addressed. Medial /t/ data in §4.3 shows U-shaped learning, principally signalled by an abrupt phase of less frequent /t/-target productions followed by increased /t/-deletions as observed for three children around MLU of 3.5 morphemes and TTR over 0.8. The exact MLU and TTR figures are not especially meaningful; the relevant aspect is their consistency across children, showing that a point of non-linearity in children's medial /t/-flap productions has a predictable relationship to overall morphosyntactic and lexical development. This linguistic relationship remains fixed while being expressed at a variety of ages.

Development of word-final intervocalic /t/ (Figure 4.14) is distinct from the medial context, strengthening the interpretation that the changes observed word-medially involve phonological learning, not just articulatory skill. The differences across contexts also show the value

[3] Careful speech can permit alveolar stops in medial flap contexts, but these are rare in parents' speech. More to the point, stylistically permissible stops are not similar to the distribution and frequency of children's deletions, nor is there any clear reason why a [t] target would be pronounced imperceptibly so often.

of this study's methodology, as a complement to studying smaller samples of productions or conducting distributional learning experiments, to investigate the relationship between word segmentation and allophone learning (e.g. Martin et al. 2013; Mattys and Jusczyk 2001; White et al. 2008). In Chapter 5, APL is used to model only the word-medial flap context, due to the distinct acquisition courses and because learners may not necessarily process or learn from alternations at word edges in a qualitatively similar way to medial alternations involving productive morphology (Peperkamp et al. 2003; Pierrehumbert 2003; Vihman and Croft 2007).

Medial flap deletion has been previously observed in about 10% of kindergarten students' tokens during a word repetition task (Burrows et al. 2019). Lily's medial flap data in §4.3.1 also replicates the pattern first seen in a smaller sample of her tokens coded by a different single annotator (Richter 2017). Word-final data (Figure 4.14) can be compared to results of Song et al. (2015), who sample pronunciations of six /t/-final words for children 18-30 months in the Providence corpus. For this comparison, when averaging Lily and Naima's productions of these six words from all sessions between 18-30 months, their productions include 23% unreleased stops, 9% flaps, 11% glottal stops, and 47% released stop, which is similar to the average rates Song et al. report for all six children studied (including these two). However, Song et al. (2015: 160) do not observe flaps until close to 30 months, and suggest 'the possibility that children are able to produce flaps in an adult-like manner only by 2;6 or later'. The coding for this dissertation identified no earlier instances of flaps within the same 6 words and 3 time-points sampled by Song et al., indicating consensus on annotation, but as shown in Figure 4.3, several children produce medial or final flaps elsewhere well before age 2;6. Furthermore, Song et al. (2015: 162) theorise that children 'start out by producing a canonical form of the alveolar stop and only later produce more contextually appropriate phonetic variants'. This accurately describes the overall word-final development of Naima (Figure 4.14), who produced 47% of the total child tokens analysed by Song et al., but the present broader sample of Lily's word-final data in the 18-30 months period shows individual differences that make this description less complete. Although Lily initially overproduces the canonical (released) form compared to adults' frequency (Eddington and Brown 2020), she also often produces the 'contextually appropriate' glottal stop variant, and even some flaps. Additionally, Figure 4.12 shows that a developmental course starting from canonical released alveolar stops has little relevance to word-medial flap acquisition. Overall, the present study's results within the overlapping portion of the data successfully replicate Song et al. (2015), but additional data questions some generalisations drawn from their controlled sample.

U-shaped development is perhaps best known for overregularised verb morphology (Cazden 1968; Clahsen et al. 2001b; Ervin and Miller 1963), but it also occurs in early word learning (Ferguson and Farwell 1975; Fikkert and Levelt 2008; Menn 1971; Tessier 2013; Vihman 2019). Children's earliest words have relatively accurate forms (Menn and Vihman 2011; Fikkert and Levelt 2008), but around 20-100 words, forms may become 'radically modified' to conform to phonological patterns (Vihman 2019: 265). For example, Vihman and Velleman (1989) describe a one-year-old child who initially pronounced *button*, *balloon*, *banana*, and *bunny*, each distinctly until developing a CVNV pattern for words with nasals, after which all four became indistinguishably [bʌnːə] and [banːə]. Oliveira-Guimarães (2013) observes a child acquiring Brazilian Portuguese who began replacing initially-accurate codas with innovated [m], which is not a possible coda in the adult language. Such nonlinear development, or regression, occurs as children first learn a few individual items, but then accumulate more information that forms the basis for the 'emergent systematicity that underlies the child's temporary retreat from accuracy' as new generalisations are overapplied (Vihman 2019: 3). Early generalisations like consonant harmony and reduplication often relate to syllable or whole-word structure (Fikkert and Levelt 2008; Kehoe 2015). Children tend to abandon these, in favour of forms that correspond more directly with adult segmental representations, when they start combining words and controlling some morphology (Priestly 1977; Vihman 2019). Tessier (2013, 2019) describes additional instances of U-shaped phonological learning, with attention to how well an error-driven learner in a constraint grammar framework accounts for them. Although the decrease in rates of medial flap production and increase in deletions happens later in development than most of these word-learning regressions, Dinnsen and Gierut (2008) observe phonological overgeneralisations about initial fricatives from an older child (4;2-4;5), and nonlinear allophone representation learning broadly parallels the 'temporary regressions as an effect of overgeneralization or reorganization' reviewed by Vihman (2019: 210). However, early word-structure regressions are unpredictable, variable from child to child, and can be extremely unlike adult language (McAllister Byun and Tessier 2016; see also discussion in Vihman 2019: 270-274). Medial flap regression seems to be initially expressed the same way by several children, likely involving a generalisation about surface alternations in the adult language; these aspects instead resemble U-shaped morphology learning. Although the interpretation of children's medial flap deletions remains uncertain, some points are clear. There is an identifiable event in English flap acquisition, with a predictable connection to general morphosyntactic and lexical development indexed by MLU and TTR. This event probably includes a change in phonological knowledge, since the most reasonable explanations for child production data involve targets that differ from children's input. Given the uncertainty about the nature of the 'deletions', it could be particularly interesting to obtain acquisition data from an English variety that allows medial intervocalic glottal stops. Medial intervocalic flapping rates are increasing in New Zealand English, and glottal stops are becoming more frequent in many global English dialects, providing additional opportunities to compare development in various input distributions (Eddington and Brown 2020; Hay and Foulkes 2016; Klánová 2016; Marshall 2003).

This corpus study will provide a reference for validating a cognitive model of allophone learning in the next chapter. Iterative evaluation of representations in APL, as a child's vocabulary expands, describes how gradual vocabulary growth leads to significant qualitative transitions in the learner's grammar. The existing distributional models for allophone learning (Chapter 2) have not been implemented to track progressive trajectories of allophone learning, only as models of what is ultimately learnable from 'total' input (Peperkamp et al. 2006; Martin et al. 2013). APL identifies the critical downwards inflection in a U-shaped trajectory, the point at which alternations in the input first motivate an abstract allophonic generalisation, which is reflected in clear landmarks visible in the the longitudinal production data in §4.3.

Chapter 5

Longitudinal model of alveolar flap development

This chapter shows in detail how APL models the way that input shapes children's acquisition of abstract representations. APL traces acquisition trajectories by grounding iterative application of the Tolerance Principle in a specific early phonological state, and this is validated by modelling the course of [ɾ] allophone development. Lexical, semantic, and morphological development, estimated as described in §5.3, underlies generalisations about systematic correspondences between surface-distinct sound segments (Brown 1973; Bergelson and Swingley 2015; Stoel-Gammon 2011). For this application, APL initialises with a default concrete H0 that [ɾ] is just exactly /ɾ/, and evaluates this hypothesis as items with multiple surface forms accumulate in the learner's vocabulary, potentially rejecting it in favour of some abstract generalisation if surface alternations are too prevalent. This change in representation is expected to coincide with children's 'regression' in [ɾ] productions observed in Chapter 4, similar to other instances of U-shaped language acquisition.

In the modelling results of this chapter, APL predictions based on aggregated child-directed input (§5.4.1) appear compatible with the general developmental course indicated by group-level production data (Figure 4.13), while predictions derived from single children's vocabularies (§5.4.2) correspond reliably with the individual variation observed in these children's development (§4.3.1). This validates APL as a cognitive model at the level of individual learners.

§5.1 formally describes alveolar flap acquisition in the APL framework. Phonological knowledge is a function of observable surface alternations in a child's lexicon, so applying APL requires an estimate of developing children's vocabularies, which may be obtained from various sources. Since learners' input determines the language they acquire, §5.2 approximates a generic learner's vocabulary from child-directed speech input. Alternatively, children's productions are a signal about the inventory of words they know, so in §5.3 individual children's productions are used to model development for each of the four children whose data is shown in §4.3.1. These individual implementations can be compared across children to examine age, vocabulary composition, or other aspects of language development throughout flap allophone acquisition, and identify shared developmental patterns. The input-based vocabulary and production-based vocabularies give overall convergent results for allophone acquisition, as presented in §5.4. The implementations based on children's individual data allow for the most precise modelling, while the generalised estimate from a corpus of child-directed speech shows that APL can also give interpretable results using more accessible forms of data.

5.1 Formal model for alveolar flap acquisition

APL identifies whether a child's current vocabulary motivates acquisition of an allophonic relationship between a distinct surface segment [ɾ] and other alveolar stop allophones. The intervocalic flapping process of adult North American English as exhibited by the words in (4.1-4.2) can be stated by the generalisation (5.1), which a child has no direct access to, but which describes the form of their input.

(5.1) If medial /t,d/ precedes an unstressed syllable, and is intervocalic, with /r/ optionally directly preceding /d,t/, then it is realised as [ɾ].

As usual, this generative rule formulation is not a modelling requirement or theoretical commitment of APL; the notation describing alveolar flapping and identifying particular phonemes and allophones is chosen only for convenience of concrete illustration, and alternative descriptions that specify the same surface forms would also be compatible. For the case of alveolar flapping, projection of surface segments to underlying forms (§3.2) indicates the following initial grammatical hypothesis:

(5.2) **H0:** If *surface flap* then *underlying flap*

or If [ɾ] → /ɾ/

This hypothesis applies from the perspective of a listener/comprehender; that is, the learner whose grammar contains H0 considers every [ɾ] they hear to be a candidate subject to H0. The Tolerance Principle (3.2) calculates when learners abandon their H0, as exceptions e becomes too great relative to the number of potential applications N. Therefore, to model the evaluation of whether (5.2) H0 for [ɾ] is tolerable to a learner at a certain point in their development, each word in that child's current linguistic knowledge is assessed and added as appropriate to counts for:

N: words that H0 is applicable to, that is, words containing surface flap

e: exceptions, or sources of cost, where H0 is applicable but [ɾ] seemingly corresponds to something else in other forms of the same lexical item

As discussed in §3.3, these 'exceptions' are word forms belonging to alternating lexical items like wri[t]e ∼ wri[ɾ]ing, which need multiple stem forms memorised in the lexicon unless a productive process relates the pronunciations. Coding costs e is based on the child's entire inventory of relevant vocabulary items at a given time. For example, while the word dirty [dɝɾi] may be present among the candidates N for a very young child, it is not associated with cognitive cost e until later when another surface variant dirt [dɝt] also enters the vocabulary. Excessive cost will motivate the learner to abandon H0 in favour of some allophonic grammar, and the model instantiated at this critical point, if it is reached, can be interpreted in terms of the composition of the learner's vocabulary and their lexical items N that contain flaps.

5.2 Data source: child-directed speech

5.2.1 CDS: Vocabulary

To model flap acquisition based on realistic language input, learners' vocabulary is drawn from a large amount of child-directed speech, with about 3 million English words of the CHILDES corpus providing an adequate sample for a generalised estimate (MacWhinney 2000). Progressive vocabulary development is approximated according to word frequency in this sample, with more frequent words considered likely to be acquired earlier, and excluding heavily oversampled names (Adam, Peter). Nine 'stages' of development are modelled by incrementally expanding vocabulary from the 50 most frequent words up to 3000 words. Open-class lexical items are the main source of flaps and alternations in children's vocabularies; input frequency is a good predictor of the average age that these items are acquired within a large population of learners, and using word input frequency to approximate an order of word acquisition is therefore a standard and frequently successful method in group-level language acquisition modelling (Braginsky et al. 2019; Cartmill et al. 2013; Goodman et al. 2008). Naturally this is a rough approximation, and each of the nine arbitrary 'stages' does not represent any particular acquisition benchmark, although they may be loosely grounded in estimates of vocabulary development at different ages (e.g. Hart and Risley 2003). Relative comparisons between stages are more justifiable, as certainly the 500-word stage represents a child further along in their language acquisition than the 100-word stage.

5.2.2 CDS: N

In each of the progressive acquisition stages, a word in the learner's vocabulary is counted as one of the candidates N if the word's pronunciation contains a medial alveolar flap, unless the word is another form of a lexical item whose flap pronunciation has already been accounted for in N (e.g. *puddle* and *puddles* are not counted separately). The CMU Pronouncing Dictionary provides phonetic input to determine which words contain flaps. CMU pronunciations are specified in the 39-phoneme ARPAbet, which has one T covering all

/t/ allophones including the flap, and distinguishes three levels of lexical stress. While the North American English medial flapping process is most often described with reference only to 'stressed' and 'unstressed' syllables, treating CMU secondary stress as unstressed for purposes of computing the flapping process (5.1) yields accurate pronunciations corresponding to native speaker judgement for the most frequent child-directed words. CMU stress annotation is not entirely consistent for this description of flapping, e.g. *ability* and *altitude* are both assigned primary-unstressed-secondary stress in their rightmost three syllables; therefore, in *ability*, secondary stress must be 'unstressed' to predict flapping, while in *altitude* it must be 'stressed' to derive [t], but issues like this do not arise in the limited early-childhood vocabulary.

5.2.3 CDS: *e*

Within each simulated acquisition stage, any word in N is also counted towards *e* if the current vocabulary contains another alternating form of the same lexical item, provided that the learner's overall language development is sufficient to understand how the two differing surface segments correspond. Assessment of whether this correspondence is interpretable is based on the usual course of morphological acquisition (e.g. Brown 1973; Haugen 1938), along with the prevalence of other instances of the same morphological process in the learner's current vocabulary.

5.3 Data source: individual productions

5.3.1 Individual: Vocabulary

To model flap acquisition using production data in the Providence corpus, children's flap-related vocabularies are estimated based on transcriptions of their utterances. Since children do not control all the words that can appear in transcripts of their speech, the words relevant to the model (containing flaps, or alternating with flap words) are validated according to the following set of criteria to determine when or if the child's knowledge of the word can be established. Overall, words are generally considered in-vocabulary when a child produces them before an adult has used them in a recorded session, or when the child uses them in their own construction, but not if they occur only in a memorised or echoed phrase. Specifically, a word is considered in-vocabulary for a child's lexicon when either condition (5.3) or condition (5.4) is met, as long as none of the exclusion criteria (5.5) apply.

(5.3) A word is **in-vocabulary** if it appears in the child's transcript and **at least one of (5.3a)-(5.3e)**:

- **(a)** The child produces the word before any adult does, in some recording session.
- **(b)** They use the word while expressing their own thoughts.
- **(c)** They use the word, or related forms of a lexical item, in several different utterances within the same session, or in paraphrases that show comprehension of a parent's speech.
- **(d)** They appear to be actively learning the word during the session.
- **(e)** A parent provides additional comments confirming what the child knows.

A child's articulation, apparent pronunciation target, or syntax do not need to be adultlike to match the criteria (5.3a-e).

(5.4) A word can still be **in-vocabulary** even if it might not appear in a child's transcript, if **all of (5.4a)-(5.4c)**:

- **(a)** There is another form that the child knows according to criteria (5.3), that is the 'missing' word plus an affix.
- **(b)** The child is assumed to know the inflectional or derivational process that relates the simpler absent form to the more complex present form.
- **(c)** The absent form is not unusual or marked in child-directed speech, as compared to the present form.

- spread.V is considered part of the vocabulary even if a child only produces the form *spreading*
- kid.V is not considered in-vocabulary if children produce only *kidding* (5.4a), as this does not satisfy (5.4c) and the processes relating the forms (5.4b) may be less transparent.

(5.5) Words that appear in a child's transcript may still be considered **out-of-vocabulary** for a child's lexicon in certain situations, if **at least one of (5.5a)-(5.5f)**:

(a) The word is only produced as part of a proper name, book title, restaurant name, etc.; and there are no utterances showing whether the child can use, understand, or even segment the word outside of quoting the title.

(b) The word only appears in a memorised text or song, often distinct from the child's own style, or archaic; and there is not additional conversation showing whether the child understands the word outside of the memorised passage.

(c) The 'word' is sound-based, not an ordinary lexical item, e.g. $giddy - up$, $pitter - patter$.

(d) The child only produces the word directly after a parent has, without evidence of using the word to express their own different ideas.

(e) The production is uncertain, or notably different from the child's usual patterns of divergence from adult targets, implying they may not know the target form; or it is impossible to tell what the target word was even with context.

(f) A parent comments with doubt of the child's linguistic knowledge.

Sometimes there is not sufficient evidence to tell if a child knows the word or is just echoing it. They may repeat a parent's single word or phrasing in a way that is not necessarily situationally inappropriate, or reply to a familiar form of question by copying potentially unfamiliar words contained in the question. Since there is not clear evidence that the child knows the word, it is not considered in-vocabulary.

Examples of child speech showing how the criteria (5.3-5.5) are applied are available at https://www.ling.upenn.edu/~ricca/dissertation/files/vocabulary.html.

5.3.2 Individual: *N*

A child's N at a particular age consists of all lexical items in their vocabulary that have medial flap targets. The parents' pronunciations are used as a guide for any words where this varies by speaker or dialect (e.g. *artist*). To determine whether potentially-related words (*bottle* $\sim$ *bottles*, *little* $\sim$ *littlest*) are considered the same lexical item or two distinct items, each child's morphological knowledge is assessed based on their entire corpus. (5.6-5.9) reviews the major morphological elements found in the children's vocabularies, and outlines how decisions were reached on their knowledge of morphology, combining evidence from established literature about the age and order of English morpheme acquisition, specific ways of using morphology in spontaneous speech, and applying the Tolerance Principle as a consistent and unbiased way to infer what processes may or may not be productive given their prevalence in child's entire apparent vocabulary. If the child seems to understand the morphological relationship according to these criteria, then the words do not count as separate lexical items in N. Examples of child speech showing in more detail how the criteria (5.6-5.9) are applied together are available at
https://www.ling.upenn.edu/~ricca/dissertation/files/morphology.html.

(5.6) **Verbs.**

All four children are considered to understand **progressive -/ing/, present -/s/, past -/ed/,** and **negation -n't**.

(a) Literature on acquisition, e.g. (Brown 1973), indicates when typically developing children acquire these, and these elements are used very often throughout the children's speech as expected.

(b) New word forms show productivity.

(c) Children's utterances indicate they are working out morphological relationships.

(d) Longitudinal studies suggest that children who produce verb forms with contracted negation ($didn't$, $can't$) understand the morphological and syntactic composition of these forms, at least for children who have started using the present form $doesn't$, often around 2-3 years; a productive form $amn't$ is also occasionally created by children whose input dialects do not contain this form (Thornton and Tesan 2013; Thornton and Rombough 2015; Capdevila i Batet and Llinàs i Grau 1995; Major 1974; de Villiers and de Villiers 1973; Yang 2017; CHILDES:Forrester 03;05.08).

- Ethan, Lily, Naima, and William all produce $doesn't$, so their production-based models include the *did* $didn't$ relationship.
- Table 5.1 shows two versions of models estimating a vocabulary from input; one version incorporates knowledge of this relationship and the other does not, since either is possible for children under 3.5 years.

(5.7) **Nouns**.

All four children know the **plural -/s/**, at least as -[z].

(a) This is also expected of typically developing children, and used frequently throughout children's speech in the Providence corpus.

(b) In the words relevant to this study, only the -[z] allomorph occurs, other than one instance of -[s] (Naima).

All four children know **diminutive -/y/**. Ethan, William, and Naima also know **agentive Verb+/er/** (which is not relevant to Lily's /t,d/ vocabulary).

(a) Semantically appropriate use.

(b) New word forms show productivity.

(c) Children's utterances indicate they are working out morphological relationships.

(d) Likely productive according to Tolerance Principle/Sufficiency Principle (Yang 2016), when considering child's entire vocabulary:

- William knows at least 16 -/y/ diminutives, 11 of which are likely to be interpretable (*horsie*, *piggy*).

  $\theta_N = 5.8$ for $N = 16, e = 5$; since $\theta_N < e$, diminutive -/y/ would be productive in this vocabulary.

(5.8) **Adjectives**.

All four children know **comparative -/er/**.

(a) Predicted according to TP:

- Ethan knows about 11 comparatives ending in -/er/. 9 are completely regular (*bigger*, *easier*, *flatter*) and he also produces the base forms for all 9 of them (*big*, *easy*, *flat*).

  $\theta_N = 4.6$ for $N = 11, e = 2$; since $\theta_N < e$, this process is productive.

- William has about 8 -/er/ comparatives; 7 are regular and he also produces their stem forms.

  $\theta_N = 3.8$, so this is productive.

- Lily has about 10 -/er/ comparatives; 8 are regular and she also produces their base forms.

  $\theta_N = 4.3$, so this is productive.

- Naima has about 19 -/er/ comparatives; 16 are regular and she produces the stem forms of 15.

  $\theta_N = 6.5$, so this is productive.

Lily, Ethan, and Naima have productive **Noun+/y/ adjective** formation, and William does not.

(a) New word forms show productivity.

(b) Predicted according to TP:

- William has about 27 adjectives ending in -/y/, 2 paired with verbs and up to 4 paired with nouns.

  $\theta_N = 8.1$, which indicates no productive generalisation with this vocabulary.

Lily and Naima know **superlative -/est/**.

(a) Predicted according to TP:

- Lily has 7 -/est/ superlatives. 6 are regular and she also produces their base forms.

  $\theta_N = 3.6$, so this is productive.

(5.9) **Non-related items**.

Relationships involving **complex lexical items** are not considered transparent to learners without evidence of understanding the internal structure.

(a) Children who know *outer* only in the context of *outerspace* may not perceive any alternation with *out* even if they understand comparative -/er/.

(b) *Kiddiepool* is not considered related to *kid* if there is no evidence that a child composes its meaning out of these parts, rather than *kittypool* etc.

Zero derivations are not considered productive, as there are not many of these in the young children's vocabularies and they do not have a systematic predictable meaning.

(a) *record.V, recording.N; meet.V, meeting.N; garden.N, gardening.V; water.N, watered.V* are modelled as distinct lexical items.

5.3.3 Individual: e

Items in N are counted as exceptions e when a child's vocabulary contains a non-flapping form of an item that observably alternates with a corresponding flap. The same criteria for morphological knowledge (5.6-5.9) determine which alternations (e.g. $eat \sim eating, flat \sim flatter, did \sim didn't$) provide interpretable evidence to each child. As noted above, some affixes, such as verbal -*ing*, -*ed* and adjectival comparative -*er*, seem productive for all four children by the end of recording. Other processes, like formation of adjectives from noun+/y/, vary from child to child according to the composition of their vocabularies.

5.4 Results

Tables 5.1-5.5 model tolerability of the default rule H0 as vocabulary grows and exceptions accumulate. Table 5.1 shows a general expected learning trajectory based on data aggregated from several parent-child pairs, as described in §5.2, while the tables in §5.4.2 each track the individual development of one child. The predicted times of allophone learning correspond to the nonlinear development of children's pronunciations observed in Chapter 4, and this consistently occurs at a similar point during children's general acquisition of English.

5.4.1 Estimating vocabulary from input

In Table 5.1, the default hypothesis H0 appears tolerable early on; at first, the learner knows very few words with flaps, and then for a time the prevalence of alternating pairs remains relatively low. However, by 750 words the learner has crossed a tipping point where H0 is rejected. The relevant words N for this stage are given in (5.10), with those that may contribute to the lexicon's count of exceptions e underlined.

Vocabulary	Flap words, N	Exceptions to H0, e		Tolerable e	H0 tolerable?	
50	0	0		NA	NA	
100	1	0		0	yes	
200	7	1	[0]	3	yes	
300	11	3	[2]	4	yes	
500	21	6	[4]	6	yes	
750	29	10	[7]	8	NO	[*]
1000	42	14	[11]	11	NO	[*]
1500	67	22	[16]	15	NO	
3000	122	38	[28]	25	NO	

Table 5.1: **Tolerability of default H0 <[ɾ] IS UNDERLYING FLAP>** through vocabulary stages, as estimated from child-directed speech. Shaded cells mark model stages where H0 is rejected and flap allophony is acquired, as e exceeds the tolerable threshold θ_N. Alternatives in brackets compare a learner who does not decompose negation contractions (e.g. *would* ∼ *wouldn't*), and starred entries in the final column indicate differences in when H0 is still tolerable for this learner as its e remains below θ_N.

(5.10) Vocabulary ∼750 words, N=29: beautiful, better, bottle, bottom, eating, getting, gotta, kitty, later, little, matter, pretty, putting, sitting, sweetie, water, what'd, what're, anybody, everybody, somebody, already, [*]couldn't, daddy, daddy's, [*]didn't, lady, ready, [*]wouldn't

From the 750-word stage onwards, cost of H0 only increases further; in the end the English vocabulary strongly requires recognising the allophonic status of [ɾ]. The vocabulary stages used here are modelling abstractions such that, for example, the 750-word stage should not really be interpreted as directly implementing a model of a child with a 750-word vocabulary. However, to whatever extent the divisions could be linked to age, a 750-word stage may be expected around or slightly before 3 years old depending on factors such as socio-economic background (Hart and Risley 2003). This is only a provisional interpretation of the stages but learning the allophonic nature of the flap around this point would be consistent with observations of (Klein and Altman 2002), as well as the children in the Providence corpus whose productions are studied in Chapter 4 and modelled individually in § 5.4.2.

Principal contributors to flap allophone discovery seem to be the verbal lexicon and morphology, e.g. *-ing*, $-n't$. Judgements informed by the criteria of (5.6-5.9) indicate that children by this stage of vocabulary development are probably aware of the relevant morphological segmentation, and therefore it is accurate for a model to treat them as able to observe alternations in the stems as shown in 5.10. However, an alternative model for a learner who finds contracted negation opaque has been offered for comparison in Table 5.1. This could reflect some individuals' development if -$n't$ segmentation is acquired later. More generally, it illustrates how fine details of APL predictions can vary based on a child's particular knowledge (flap allophony is predicted later without evidence from -$n't$) while retaining the same essential qualitative prediction for non-linear acquisition.

5.4.2 Estimating individual vocabulary from productions

Tables 5.2-5.5 each illustrate development of one child in the Providence corpus. Ages are listed as the upper bound of 3-month intervals, covering each child's complete set of recordings. N, e, and tolerability are calculated as described above. MLU and TTR provide general indications of language development, as stated in §4.2. Samples of each child's vocabulary at a critical stage in the learning model are provided in (5.11-5.14), showing which lexical items and alternations motivate their allophone acquisition.

APL predicts that Lily probably learns flap allophony by 36 months (Table 5.2). A period between age 2-3 years is near the threshold to reject H0, and a strict interpretation (rejection between 33 and 36 months) is not appropriate due to the noise and error inherent to this method of estimating a vocabulary. However, 36 months is a clear upper bound, because the trajectory continuing from 3 to 4 years has a larger margin predicting that H0 is not sustainable. Overall, these predictions have a clear relationship with Lily's pronunciations in Chapter 4, where §4.3.1 shows a sharp decline in her rate of medial flap production that starts around 30 months and reaches the lowest point by 37 months, with an increase in the rate of segment deletions that peaks around 35 months. Although the exact relationship between allophone acquisition and any one of these behavioural landmarks is not yet known, as a whole this production data validates the model's prediction that a major qualitative change in the child's grammar occurred by or before 36 months.

LILY Age (months)	MLU	TTR	Flap words, N	Exceptions to H0, e	Tolerable e
15	1.33	0.55	1		
18	1.28	0.53	2		2
21	1.29	0.64	6	1	3
24	1.96	0.71	14	4	5
27	2.58	0.78	44	11	11
30	3.50	0.79	52	13	13
33	3.51	0.81	71	16	16
36*	3.37	0.83	85	20	19
39	3.55	0.83	90	21	20
42	3.44	0.83	94	21	20
45	4.33	0.84	101	23	21
48	3.59	0.80	106	26	22
51	3.02	0.81	107	27	22

Table 5.2: **Tolerability for Lily of default H0 <[ɾ] IS UNDERLYING FLAP>**, as estimated from her productions. Shaded rows indicate when $e > \theta_N$. Vocabulary for N and e at the starred age (36 months) is listed in (5.11), where different forms of one lexical item are connected with tildes and alternations e are starred.

(5.11) **Lily:** beautiful, beauty, beetle∼beetles, better, bite∼biting*, boot∼boots∼bootie∼booties*, bottle∼bottles, bottom, butter, butterflies∼butterfly, calculator, caterpillars, city, computer, cut∼cuts∼cutting*, defecating∼defecate*, dirt∼dirty*, eat∼eating*, escalator, firefighter, get∼gets∼getting*, gotta, havarti, hospital, hurt∼hurting∼hurts*, jupiter, kitty, later, letter∼letters, lettuce, little∼littlest, otter, outer, party, petal∼petals, photograph, pluto, potty, pretty, put∼putting*, quarters, refrigerator∼refrigerators, saturn, scooter, scuttle, sit∼sitting*, split∼splitting*, starting, sweater, tomato∼tomatoes, turtle∼turtles, visit∼visiting*, water, watered∼watering, watermelon, already, bird∼birdie∼birds*, body, colorado, corduroy, dad∼daddies∼daddy*, didn't∼did*, dreidel, eddie, everybody, florida, garden, hard∼harder*, hide∼hiding*, kidding, ladder, ladybug∼ladybugs, medicine, need∼needed∼needs*, nevada, nobody, paddington, pudding, puddle, read∼reading*, ready, ride∼riding*, soda, spider, teddy

ETHAN Age (months)	MLU	TTR	Flap words, N	Exceptions to H0, e	Tolerable e
12	1.17	0.48	1		
15	1.23	0.58	4		2
18	1.78	0.64	12	2	4
21	2.29	0.68	19	6	6
24	3.16	0.80	32	9	9
27	3.38	0.77	35	9	9
30*	3.65	0.84	47	14	12
33	3.39	0.83	54	17	13
36	3.85	0.82	64	19	15

Table 5.3: **Tolerability for Ethan of default H0 <[ɾ] IS UNDERLYING FLAP>**, as estimated from his productions. Shaded rows indicate when $e > \theta_N$. Vocabulary for N and e at the starred age (30 months) is listed in (5.12).

(5.12) **Ethan:** alligator, batteries∼battery, better, bite∼biting*, bottle∼bottles, bottom, butterfly, caterpillar, dirt∼dirty*, eat∼eating*, excited, firefighter, get∼gets∼getting*, gotta, kitty, late∼later*, letters, little, matter, naughty, potatoes, potty, put∼putting*, sit∼sitting*, water, watermelon, bird∼birds∼birdie*, body, corduroy, dad∼daddy*, did∼didn't*, gordon, hide∼hiding*, ladder, lady, load∼loading*, middle, nobody, noodle∼noodles, pleaded, read∼reading*, ready, riding∼ride*, shady, somebody, speeding, video

The implementation for Ethan (Table 5.3) has similar results: again an extended period is close to the threshold to reject H0, but after 30 months the threshold is clearly and permanently passed. This is also essentially in line with Ethan's pronunciation data in §4.3.1, where the rate of medial /t/ realised appropriately as either flap or d begins to sharply decline around 26 months. After this Ethan's production data is somewhat indeterminate; the rate of target productions may reach a low point by 29 months along with a high rate of deletions at this same time, but it is also possible that after the end of data collection at 35 months the rate of target productions fell further or that the rate of deletions continued increasing.

Still, at minimum it is clear that Ethan's productions show a characteristic 'regression' of U-shaped learning, during the months directly leading to the APL upper-bound prediction.

Naima Age (months)	MLU	TTR	Flap words, N	Exceptions to H0, e	Tolerable e
12	1.69	0.26	1		
15	1.41	0.48	2	1	2
18	2.34	0.52	24	6	7
21	2.91	0.71	48	14	12
24	3.43	0.77	63	14	15
27*	4.29	0.81	78	19	17
30	3.92	0.82	111	28	23
33	3.68	0.84	123	34	25
36	5.26	0.82	133	36	27
39	4.43	0.83	134	36	27
42	4.89	0.85	135	37	27
45	5.00	0.85	140	38	28
48	4.10	0.83	145	40	29

Table 5.4: **Tolerability for Naima of default H0 $<$[ɾ] IS UNDERLYING FLAP$>$**, as estimated from her productions. Shaded rows indicate when $e > \theta_N$. Vocabulary for N and e at the starred age (27 months) is listed in (5.13).

(5.13) **Naima:** accident, alligator, artist, avocado∼avocados, babysitters, better, bird∼birds∼birdie∼birdies*, body, boot∼booties*, bottle, bottom, butter, butterfly∼butterflies, cardinal∼cardinals, carter, caterpillar, computer, corduroy, cottage, cut∼cutting*, dad∼daddy*, daughter, did∼didn't*, dirt∼dirty*, eat∼eated∼eats∼eating*, electricity, escalator, everybody, feeder, feed∼feeding*, firefighter, garden, gardening, get∼gets∼getting*, glitter, grate∼grating∼grater*, havarti, hide∼hiding*, hospital, judy, kitty, lady, ladybug, late∼later*, letter, little∼littlest, medium, motorcycle∼motorcycles, myrtle, need∼needs∼needed*, nobody, party, patting, peter, photograph∼photographs, portishead, potato, potty, pretty, put∼puts∼putting*, radio, read∼reads∼reading*, ready, serengeti, sit∼sits∼sitting*, slide∼sliding*, soda, somebody, spaghetti, spider∼spiders, spit∼spitting*, tomato∼tomatoes, turtle∼turtles, visiting, water, waterfall∼waterfalls, wooden, write∼writing*

Naima's implementation in Table 5.4 places allophone acquisition within a few months on either side of 2 years, but not later than 27 months. This is considerably younger than the previous two children, but it matches Naima's data in §4.3.1, where the rate of medial flap productions has begun to suddenly decline already by 23 months, reaching a lowest point after 25 months, while the rate of segment deletions in this context is especially high around 29 months. Tables 5.2-5.4 show that regardless of variation in age, the relevant vocabulary size (N) as well as MLU and TTR are comparable for Naima, Lily, and Ethan at the critical point predicted by APL. Naima's data after a period of somewhat ambiguous model predictions also shares the pattern of exceptions e then accumulating far above a tolerable threshold.

Predictions for William in Table 5.5 diverge from the others, because rejection of H0 does not appear by the upper age limit of his recorded data in the corpus. This is consistent with his pronunciations in §4.3.1, where no nonlinearity appears. However, Table 5.5 shows a long period close to the threshold for allophone learning, and the MLU, TTR, and flap vocabulary size N for William's final recordings are also reaching the values at which other children seemed to reject H0. This suggests that William might have shown evidence of flap allophone acquisition quite soon after the end of data collection.

(5.14) **William:** batteries∼battery, beautiful, better, bite∼bited∼biting*, bottle, bottom, butterfly, cheetah∼cheetahs, computer, dirty, eat∼eating*, firefighter, get∼getting*, gotta, hippopotamus, hit∼hitted*, hurt∼hurting∼hurts*, katie, kitty, late∼later*, letter∼letters, little, motorhome, motorcycle, party, potato, snotty, spaghetti, turtle, visited∼visit*, vitamin, water, watermelon, aidan, already, anybody, audubon,

WILLIAM Age (months)	MLU	TTR	Flap words, N	Exceptions to H0, e	Tolerable e
18	1.33	0.55	1		
21	1.46	0.57	3		2
24	1.52	0.64	5		3
27	1.81	0.66	7	1	3
30	2.71	0.74	21	5	6
33	3.03	0.77	33	10	9
36	2.95	0.81	39	10	10
39	3.13	0.81	53	11	13
42*	3.47	0.78	58	14	14

Table 5.5: **Tolerability for William of default H0 <[ɾ] IS UNDERLYING FLAP>**, as estimated from his productions. Shaded rows indicate when $e > \theta_N$. Vocabulary for N and e at the starred age (42 months) is listed in (5.14).

bird∼birdie∼birds*, buddy, crocodile, dad∼daddy*, did∼didn't*, everybody, garden, hide∼hiding*, ladder, load∼loader*, louder, medicine, nobody, pedal, read∼reading*, ready, ride∼riding*, somebody, spider, spiderman, thursday

5.5 Alveolar flap acquisition model discussion

5.5.1 Consistent patterns in development

The four children whose alveolar flap learning is modelled in §5.4.2 differ in individual timing and details of language acquisition, yet the fit of APL to their developmental data is quite similar. Considering Tables 5.2-5.5 together, children tend to have several months with ambiguous model predictions near the threshold for allophone acquisition, but eventually in Tables 5.2-5.4 there is a point of no return where the prevalence of alternations unambiguously and consistently requires knowledge of allophones. A sudden drop in the rate of appropriate medial flap target productions, along with an increase in medial flap deletions, provide a consistent behavioural signal in the corpus study of Chapter 4. This signal appears in the productions of three children, when MLU is around 3.5 morphemes per utterance and

TTR is just over 0.8; the fourth child matches the trajectory leading up to this point, with MLU and TTR below or barely reaching these levels in his final recorded sessions. For the three children with evidence of U-shaped learning in Chapter 4, APL predicts an upper bound on the time of allophone acquisition 4-6 months after the rate of medial flap target production begins to drop, keeping in mind that the model reports a child's status only at intervals of every three months while §4.3.1 shows detail from several recordings per month. It is less straightforward to identify a reliable relationship between the model's predictions and the time of the lowest rate of flap target production after this drop, or the time when the rate of deletions becomes highest. It is impossible to be sure whether either of these two events have occurred in Ethan's data, and Lily's highest rate of deletion occurs before her lowest rate of flap accuracy while Naima's highest rate of deletion occurs a few months after her lowest rate of flap accuracy; therefore, there is no clear expectation for how a model should correspond to these events and no way to validate whether a model has or has not made correct predictions about them. The earliest sign of nonlinearity in children's productions is the abrupt change from increasing to decreasing rates of medial flap target productions, and APL predicts an upper bound on the time of allophone acquisition at a fairly constant time after this sign. Overall, even though various aspects of the production data itself are not yet fully explained, this longitudinal study has therefore shown that alveolar flap development is an integral part of general English acquisition and APL seems appropriately sensitive to how it is learned.

Similar morphology, including verbal -*ing* and diminutive -*y*, is involved in the critical alternations for all four individual children as well as the implementation of §5.4.1. For many alternating lexemes, the flap variant is conditioned by some of the earliest inflectional morphology that English-learning children acquire (e.g. verbal -*ing*, -*ed*; Brown 1973; Haugen 1938), so the modelling assumption that learners have cognitive access to the alternation is well justified.

Although APL does not constrain exactly what allophonic grammar a child may learn when they do acquire such a representation, in order to demonstrate how this model fits into the entire language acquisition process it may be helpful to establish that a reasonable hypothesis is likely to be available at the predicted time. As one example, the inductive learning model of Albright and Hayes (2003) can be applied to identify patterns of realising medial /t/ and /d/ as flaps, given the same input data that APL predicts this alveolar flapping allophony is learned from. This has been implemented for the 750-word vocabulary based on child-directed speech that was identified as the source of flap allophone knowledge in §5.4.1. To reflect the APL acquisition scenario, the Minimal Generalization Learner[1] is applied from the perspective of a comprehender who hears a word containing some alveolar stop allophone, and resolves its underlying form. Transforming surface to underlying forms is an unconventional use of Albright & Hayes' learner, but this implementation establishes that input is sufficient for a Minimal Generalisation strategy to generate useful hypotheses about flapping. Given all words in the 750-word child vocabulary that have a stop or flap allophone of /t/ or /d/, and representing the vocabulary with a basic set of (binary and multiply valued) phonological features, Albright & Hayes' learning model proposes the following conditioning for allophones:

(5.15) [ɾ] → /t/, [ɾ] → /d/ /
[+syllabic, [5]sonority, +stress, [2]aperture, +voice, -post-glide] ___
[+syllabic, [5]sonority, -stress, [2]aperture, +voice, [1-2]high, -post-glide]

This correspondence between /t,d/ and [ɾ], following a stressed vowel and preceding an unstressed non-low vowel, accurately describes this limited vocabulary subset which happens to not require a broader generalisation. Therefore, in this case the input is shown to contain a reasonable basis for some generalisation about allophones by the time APL predicts the learner must acquire one, although children's early allophone systems may differ from adults' and APL itself does not specify how the potential alternative grammars become active in a child's hypothesis space. The APL grammar selection/rejection model should be compatible with many models for grammar discovery or induction with known computational proper-

[1]Version from http://www.mit.edu/ albright/mgl/ with phonological features based on the CELEX set, slightly adapted for CMU pronouncing dictionary

ties, like maximum entropy or minimum description length (Berger et al. 1996; Hayes and Wilson 2008; Solomonoff 1964), with the potential to join into a larger cohesive picture of phonological development.

5.5.2 Individual variation

The Providence corpus includes data from 11 to 48 months, and children of quite different ages happen to reach simultaneously similar stages for both general language development (estimated by MLU or TTR) and flap-specific measures such as N or allophone learning status. While age-based variation in measures like MLU and TTR is expected in any sample of children and not remarkable in itself, the relevant point for APL is how closely alveolar flap allophone development is tied to global morphosyntactic abilities rather than to age (Cazden 1968).

Allophone learning seems to be predicted (and reflected in behaviour; §4.3.1) by the time when children know roughly 15-20 relevant alternations e and approximately 50-80 total words containing flaps N. Although the quantities N and e are more variable across children than MLU and TTR, they do not vary nearly as much as the model itself would permit. The condition for allophone learning could be satisfied with a vocabulary of 5 flap-containing words 4 of which alternate, or alternatively not until a child learns 275 flap-containing words including 50 that alternate, but neither of these situations emerges from the distribution of English flaps.

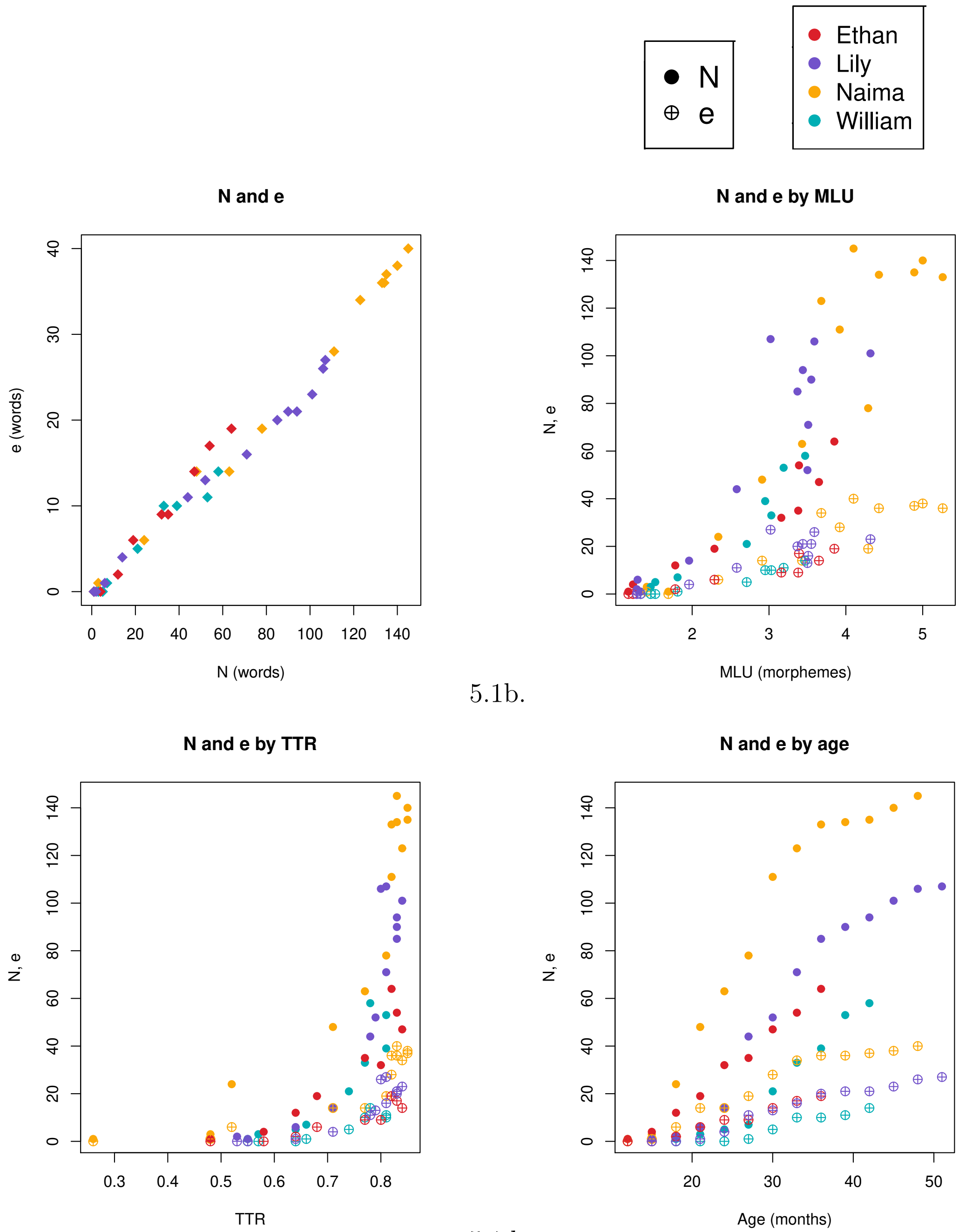
As mentioned above, evidence from children's medial flap pronunciations, MLU, TTR, and model terms (including N, e, and prediction of allophone acquisition) all point towards about the same time when each child seems to learn something significant about flaps. Based on §4.3.1 and §5.4.2, this linguistically convergent point happens around 30-35 months (Lily), 25+ months (Ethan), 23-28 months (Naima), or no sooner than 42 months (William). Examining the five variables MLU, TTR, age, N, and e, for all four children shows that N and e have high linear correlation ($r(42) = 0.99$, $p < 0.001$, Figure 5.1a), while relationships between other variables are nonlinear (Figures 5.1b-d, Table 5.6). The linguistic factors are closely related to each other. For example, a linear model predicting the logarithmic transform of N from MLU, TTR, and age ($R^2 = 0.91, F(3,40) = 132.5, p < 0.001$) finds a significant relationship of N with both MLU ($p < 0.001$) and TTR ($p < 0.001$) but not age ($p = 0.215$). A reduced model including only MLU and TTR as predictors ($R^2 = 0.91, F(2,41) = 195.2, p < 0.001$) has essentially the same predictive value as the full model, indicating that N is a product of language ability independent of age.

	MLU (morphemes)	TTR	N (words)
Age (months)	0.78	0.81	0.82
MLU (morphemes)		0.82	0.91
TTR			0.91

Table 5.6: **Correlations between age, MLU, TTR, and log-transformed N.** e grows strictly in parallel with N (correlation $r(42) = 0.99, p < 0.001$) and is omitted.

A further individual difference between children is the relationship between articulatory or motor development and systematic language knowledge. For example, in the production data of Figure 4.12, Ethan frequently produces ‘other’ segments in flap contexts up until he shows evidence of flap allophone learning, and even after that point he produces few flap-like realisations and more d-like realisations, whose articulation requires less speed. In contrast, Lily in Figure 4.12 reliably produces identifiable flaps long before evidence of allophone learning. This reinforces that ‘language acquisition’ is not one uniform process, even though some components are heavily interdependent.

A final notable area of variation is the composition of each child’s flap-related vocabulary. §5.4.2 includes each vocabulary at the point where identifiable allophone-learning status, MLU, and TTR seem most similar (starred in Tables 5.2-5.5). Despite the similar learning outcomes, the Jaccard similarity of these vocabularies for any two children ranges from only 0.25 (Naima and William; Naima’s flap-related vocabulary is about 1.5 times the size of William’s but about half the words in William’s vocabulary are not in Naima’s) up to 0.37 (Ethan and William, matched by 0.36 for Lily and Naima). The role of morphology also varies somewhat. Some frequent forms are informative for all children, but adjectival -*y*



5.1a. 5.1b. 5.1c. 5.1d.

Figure 5.1: Relationships between N, e, MLU, TTR, and age

derivation does not seem productive for one child, whose vocabulary does not yet include many pairs like *dirt* ∼ *dirty* to motivate this generalisation. Likewise, only two children's flap inventories include superlative *-est* as in *biggest*, though it seems productive for both of them. Finally, results of Klein and Altman (2002) further support the APL proposal that allophone learning is sensitive to morphological surface alternations: error production patterns in children up to 5 years show that the context in which they most often fail to produce target forms is the morphological alternation context -*ding*, while flaps are most reliably present in the rarely-alternating context -*ty* (*Betty*, *dirty*, *kitty*), consistent with surface-based memorisation of non-alternating forms (Kiparsky 1982).

Chapter 6

Secondary split

This chapter uses APL to examine the circumstances and progressions of a recent phonemic split in Menominee, the development of ū from an environmentally conditioned allophone of ō to a phoneme in its own right. The present-day phonemes /o/ and /u/ are descended from the historical segments [oː] and [uː], normally written ō and ū in Menominee orthography; changes in short vowel pronunciation obscured the distinction in these segments' environments to the point that ō ∼ ū vowel quality is no longer predictable from context and instead has to be specified in the lexicon (Bloomfield 1962). This split is somewhat parallel to an older completed split in the front vowels, which had originated in allophonic height alternations of ē and ī, although in the front vowels /e/ resulted from lowering (Bloomfield 1962; Miner 1979a; Oxford 2015).

Recalling §3.4, in the APL framework a phonemic split in a language is the result of learners' failure to merge surface-alternating phonological outputs, like the segments [oː] and [uː], into an equivalence class where some predictable phonological process can generate both surface forms as appropriate from the same underlying representation. Persistence of the default surface-based grammar as indicated by Invarient Transparency occurs when learners do not encounter enough lexical items with alternating outputs of a phonological process, so when changes in a language eventually reduce the prevalence of such alternations then a child's initial 'what you hear is what you get' grammar may remain tolerable at a large vocabulary size.

In this chapter, §6.1 gives background describing the Menominee vowel system, relevant allophonic processes, and progression of the high back vowel split. Menominee is an Algonquian language that is currently spoken in Wisconsin; as of 2019 the number of adult L1 speakers was estimated as three to five (Cudworth 2019). However, the language has been consistently spoken by a larger number of L2 speakers, and in addition to L2 language immersion schools for older speakers, more early-childhood language learning initiatives have been developed within the past few years including language immersion daycare (Wisconsin Department of Public Instruction, dpi.wi.gov). Throughout the history of the language, short vowel qualities gradually came to share several of one another's allophones and underwent reduction, so a present-day short vowel pronunciation is not very informative about underlying vowel quality, either to specify a conditioning environment of nearby ō ∼ ū or as a shortened counterpart of ō ∼ ū in another inflected form of the same lexical item.

Following the section of language background, §6.2 explains how acquisition of ō ∼ ū can be modelled with APL. For Menominee, there is not adequate corpus data available to estimate children's vocabularies throughout early development as was done in §5.2 or §5.3, so instead a dictionary is used to model the ultimate outcome of acquisition, analogous to the final rows in Tables 5.1-5.5. In §6.3, the model is applied at three different stages in the history of Menominee short vowel quality reduction. These implementations illustrate how the relationship between data and acquisition works during the process of secondary split. An allophonic relationship is shown to be learnable from the presumed input at a historical stage, when the alternation is observable in both short and long vowels, but not learnable in the modern language, when the alternation and its environment are only reliable in long vowels with a reduced number of obvious surface alternations. Finally, §6.4 interprets the modelling results with attention to new evidence about the acoustic realisations of post-split phonemic ō ∼ ū, and reviews the overall contributions of this model for secondary split.

6.1 Menominee ū

This section describes the origin of phonemic ū in Menominee, along with background on certain elements of the general vowel system that have interacted with this development, including phonologically-conditioned vowel length alternations and change over time of the vowel qualities produced in short vowels. Cudworth (2019: Ch. 2) gives a broader overall review of the Menominee vowel system, summarising the most important patterns, while the full details that these (and most other authors') descriptions are largely based on can be found in Bloomfield 1962.

Orthographic note.

Following the present-day conventions for writing Menominee, as used in the Menominee Dictionary of Macaulay (2012) and approved by the Menominee tribal legislature, long vowels are written with a macron (*ū, ē, ī*, etc.) and short vowels without (*u*, *e*, *i*). In this system, *ae* is a single short vowel whose long counterpart is *$\overline{ae}$*. There are also two long diphthongs *ia*, *ua* whose short counterparts are spelled *ya*, *wa*; the conventional analysis is that the long diphthongs are a glide followed by a long vowel while the short diphthongs are a glide with a short vowel (Bloomfield 1962).

The orthography for most consonants is fairly intuitive for readers familiar with basic IPA, and English-speaking readers. Glottal stops are spelled with *q*. Besides the short diphthongs spelled *ya* and *wa*, the glides *w* and *y* can function as consonants, including insertion of *y* between two otherwise-adjacent vowels; vowel clusters are not permitted, so when vowel sequences occur at morpheme boundaries either a segment will be inserted between them or one of the vowel qualities will be lost (Bloomfield 1962; Cudworth 2019; Macaulay 2012). Given these restrictions, it is generally unambiguous whether the orthography of a word indicates short vowels/diphthongs, long vowels, or consonantal glides.

When not otherwise specified, all Menominee wordforms and definitions in this chapter are sourced from the dictionary published by Macaulay and Menominee Tribe of Wisconsin (2012).

6.1.1 ō-raising

The Menominee ū entered surface forms only as the allophonic output of a raising process ō → ū (6.1).

(6.1) a. watōp 'alder'
watūpyak 'alders'

b. maskūcīhsepak 'bean leaf'

This raising was conditioned on the occurrence later in the same word of either long or short high vowels *u* and *i* (but not non-high *a*, *ae*, *o*, *e*) or postconsonantal semivowels *w* and *y* (Bloomfield 1962: §4.66). Back vowel raising is conventionally described as a form of vowel height harmony, because the front mid and high vowels ē and ī also alternate according to similar conditioning environments, although in the case of the front vowels either can also be an underlying vowel quality regardless of length (Cudworth 2019; Bloomfield 1962; Miner 1979b). Some literature explores the possibility that Menominee should be analysed as having ATR harmony rather than height harmony (Archangeli and Pulleyblank 1994; Milligan 2000; Walker 2018). However, as Kimper (2011: 111) notes, these 'justify their reanalyses primarily on analytical rather than empirical grounds' in a way that 'crucially depends on the specific set of theoretical assumptions each author works under, and poses no necessary or inherent problem'. Oxford (2015: 331) agrees that the ATR harmony analyses are driven by 'analytical convenience', and also finds theory-internal grounds to prefer a height harmony analysis as originally described by Bloomfield (1962). While it is not critical for APL which feature drives raising of the back vowels, as long as there is some discernible contrast in learners' input, this chapter therefore treats height as the relevant dimension.

Due to the phonologically predictable complimentary distribution of ō and ū, in a description or analysis of the language when raising first appeared there would be no need to posit an underlying ū. Furthermore, there was presumably no short u in the language, only short o, since this raising process applied only to long vowels (Bloomfield 1962, but see §6.4).

A recent description for Menominee vowel harmony given by Cudworth (2019: 33), based on review of language data along with previous analyses (Bloomfield 1962; Macaulay and Menominee Tribe of Wisconsin 2012), is quoted in (6.2):

(6.2) "Menominee Vowel Harmony raises long mid vowels to long high vowels and pre-glottal short *o* /o/ to short *u* /u/ when followed later in the word by a high vowel or diphthong, unless *ae* /æ/ or *$\overline{ae}$* /æ:/ occurs between the target (ē /e:/, ō /o:/, or oq /oʔ/) and the trigger (i /i/, ī /i:/, u /u/, ū /u:/, Cy, or Cw). a /a/ is transparent to the Vowel Harmony: it does not raise, it does not block, and it does not trigger the rule."

This formulation is a useful description that will be adopted for this chapter, as it specifies what is found in the present-day language without having to build in assumptions about the relationship of glottal stops and vowel length (e.g. a theory-dependent assumption revealing how in some way the glottal stops in *oq*/*uq* make the vowels function as long vowels for the raising process, or assuming that raised *uq* is a quite recent development that must have been *oq* some time earlier).

Still, even in a fully allophonic historical variant of the language, where all surface ū is generated from underlying ō and the realisations are in perfect complementary distribution, many lexemes would contain a non-alternating ū. For example, no inflected form of the noun *maskūcīhsepak* 'bean leaf' (6.1b) will ever have surface ō instead of ū. This is because the ū in this stem is conditioned by ī that is also part of the stem rather than an occasional affix, in the stem syllable following ū. The non-alternating lexical items like this open the possibility of learners acquiring a phonemically split grammar with underlying ū, because regardless of whether a learner believes a non-alternating word has underlying ū or underlying ō, only one stem form must be stored for it in the lexicon.

Furthermore, subsequent developments in Menominee have nevertheless made it clear that 'a theoretically superfluous and covert underlying /u:/' (Ringe and Eska 2013: 132; notation *u:* equivalent to *ū*) must have somehow been learned at some point in the language's history.

6.1.2 Present phonemic status of ū

This section presents evidence that long ū does now have some kind of phonemic, or underlyingly independent, status in Menominee (Bloomfield 1962; Milligan 2005; Miner 1979b). Overall, this changed status is clear because words can now form minimal pairs based on this vowel quality contrast, and because some speakers inherit distinct ō and ū surface pronunciations despite these speakers' merger of the raising trigger /i/ with non-raising /e/. Word borrowings into Menominee recorded by Bloomfield (1962: §1.16) offer a clean illustration of the phonemic contrast between the high back vowels, with a minimal pair *cōh* 'Joe' and *cūh* 'Jew'. This the same type of occurrence as the recent Zoque loanwords discussed in §3.2.5, where foreign vocabulary was borrowed into the language using surface segments from its existing inventory in environments where these segments should be impossible according to regular allophonic conditioning.

A handful of native Menominee lexemes also have surface forms that violate the conditioned raising process and therefore require vowel height to be specified in the lexicon. For example, the interjections *capūq* 'splash!' and *kūh* 'stop it!', require the raised ū pronunciation despite having no conditioning triggers (Bloomfield 1962: §1.16), as do *pūceman* 'boots', *atūtapekan* 'chair', *yūh* 'never mind' (Macaulay 2012). Alternatively, the onomatopoeic *ōhōpīwāēkat* 'it makes a whooping sound, there is a sound of whooping' always has ō where shown, despite the presence of ī in the third syllable which should ordinarily trigger the raising rule (Macaulay 2012).

In addition, a short u segment has entered the language in a small number of words, notably the verbal element *uw*. This appears without any following raising trigger, as in (6.3a) where it is immediately followed by *ae* which would block any following raising triggers, or the past tense of the same word (6.3b) which has no blocking but still no raising triggers. Historically, long ū would be expected to alternate with short o, because only *o* occurred as an underlying vowel quality and the raising rule affected only long vowels (or those followed by a glottal stop, whatever the analysis); in theory the short u could arise either if ū is thought to have a distinct height-specified underlying quality in some lexical items, or if the raising process expanded to apply to short as well as long vowels in a selection of items (Bloomfield 1962; Macaulay 2012).

(6.3) a. osīhtuwaew 'he or she makes, gets it ready for him or her'

b. osīhtuwacen PST

There is not a clear record of when all these pronunciations arose in the language, but they are language-internal developments rather than foreign loans.

However, the most convincing line of argumentation for the underlying independent status of ū comes from considering the Menominee vowel system and variation as a whole. In spoken Menominee, Bloomfield (1962: §1.8) observes that the ō ~ ū alternation 'is entirely living and automatic' (productive), 'but there is much fluctuation' in both directions, as 'the lower vowel is sometimes spoken when the higher one would be expected, and vice versa'. Importantly for the idea of an independent underlying ū, Bloomfield (1962: §1.16) observes that 'moreover, the difference of o: and u: is maintained by persons in whose speech this alternation has lost its regularity' with reference to environmental conditioning . Since the surface distinction of ō ~ ū could be reliably produced in the absence of systematic conditioning that would be able to cue each variant where-ever needed, the different surface vowel qualities must have their height distinction encoded in the forms in these speakers' lexicons.

Finally, some speakers of Bloomfield's time already manifested a quite advanced classic secondary split of ō ~ ū from conditioned allophones to distinct phonemes amid the dissolution of the original conditioning environment: 'Some younger speakers have no *i* [which triggers ō-raising], replacing it everywhere by *e* [which is not a trigger]; but these speakers maintain the alternation as regularly as do older ones' (Bloomfield 1962: §1.8). Beyond this specific merger of *i* with *e*, the Menominee short vowels in general now contain massive variation and neutralisation of vowel quality (to be discussed in §6.1.4), even further disrupting any possible conditioning environment for allophonic raising (Cudworth 2019). §6.4 will return to more detailed evidence from speech.

It is necessary to mention that there are analyses of present-day Menominee that do not consider the language to contain an underlying high back vowel quality, which contradicts the approach of this chapter. Notably, Cudworth (2019: 202) concludes that 'high back vowels can be derived in nearly all instances, and as a result do not occupy a node in the phonological hierarchy; there is no phonological u /u/ or ū /u/'. This analysis of the language seems oriented to correctly specifying the behaviour of morphemes and surface forms (to the extent these are known; for example, the words *pūceman* 'boots' and *atūtapekan* 'chair' as mentioned above are not present in Cudworth's (2019) data, presumably because no audio recording of them is available) in a way that is useful to linguists studying the language's historical development, but it does not explicitly consider what kind of system could plausibly be learnable for speakers now or in the recent past. Therefore, it may be reasonable to interpret Cudworth's statement to mean that for certain purposes there is good reason for linguists to use a structural description of the language in which the high back vowels are allophones of mid back vowels. As stated earlier in §3.2.1, this does not necessarily mean that people learn these segments as allophones in the way that results in the kind of diagnostic behaviours presented in §2.1, which is the context of this chapter's assertion that high and mid back vowels are phonemically contrastive. APL explicitly rejects the idea that being able to derive certain surface segments in a formal description is sufficient evidence that those segments are synchronically derived that way for users of the language. The case study of German in Chapter 7 will include an even stronger contrast between formal descriptions and learners' representations.

6.1.3 Vowel length alternation

Menominee vowels in open-class words undergo automatic length/quantity alternation between long and short. Diphthongs are included in this system, with the short diphthongs acting as short vowels (Bloomfield 1962). These length alternations arise from strictly phonological processes whose outputs, although intricate and 'famously complex' (Cudworth 2019:

37) to describe (see Bloomfield 1962: §4.45-4.58), are determined purely by the shape of a word, influenced by both the content of the word's individual sound segments and its syllabic structure. The length alternations are robust and reliable, confirmed through multiple researchers' acoustic analyses of their durations showing that surface short vowels are indeed shorter than surface long vowels (Cudworth 2019; Milligan 2005).

Furthermore, Menominee is consistently described as having both underlyingly short and underlyingly long monophthongs and diphthongs, and the surface length alternations involve both shortening of long vowels and lengthening of short vowels. Cudworth (2019) shows that although surface long vowels have similar lengths regardless of whether they are underlyingly long or have been lengthened from an underlyingly short vowel, surface short vowels that are underlyingly long are realised with durations about 1.4 times as long as surface short vowels that are also underlyingly short (but both types of surface-short vowels have shorter acoustic durations than surface-long vowels). This evidence from speech supports the assertions that Menominee vowels can be underlyingly either short or long, and that length alternations are part of the synchronic grammar since shortened long vowels are distinct from underlyingly short vowels.

Some of the length alternation patterns are typologically unexpected: long vowels can be shortened in open syllables while short vowels can be lengthened in closed syllables, the opposite of the usual relationship (Bloomfield 1962; Buckley 2000; Cudworth 2019; Milligan 2005; Miner 1975; Hayes 1995). Therefore, while acoustic studies have confirmed surface alternation patterns pretty much as Bloomfield (1962) describes, there has been some debate on precisely what kind of system these patterns result from. For example, Miner (1975) suggests that the surface duration differences realise syllable stress contrasts rather than length contrasts. Milligan (2005) provides a more comprehensive review of the various proposals about Menominee vowel length alternations, pointing out flaws and frequent misunderstandings of the primary data, and does not identify any entirely satisfactory account for the complete language. Following a consensus that has emerged on the basic nature of the system if not all details, this chapter is written as though Menominee has vowel length alternations, but the essential assumptions for this application of APL are just that there is some kind of learnable, synchronically active, alternation behind the surface vowel duration differences.

The vowel length alternations are an entirely separate process from the ō ~ ū vowel quality alternation described in §6.1.1 above, but they are necessary to mention because both processes can be active in the same lexeme, and their interaction is important throughout the history of the ō ~ ū phoneme split. Specifically, because the Menominee raising rule is active only on long vowels, different environments with correspondingly different length conditions can cause different forms of a lexeme to exhibit vowel quality alternation between ū and short o (6.4).

(6.4)	kockuan	'hook'
	ne-kūckwan	1SG.POSS.SG

To model the acquisition of vowel quality in the face of length alternations in the rest of this chapter, Menominee length rules will be treated as already transparent/understandable to the learner. This means introducing the modelling assumption that the learner is well aware of the underlying connection between phonologically conditioned length-varying surface forms, such as the *ae* and *$\overline{ae}$* in *kāsakaeh* 'cat' and *kāsak$\overline{ae}$hsaeh* 'kitten' (literally, 'cat' + diminutive suffix).

This assumption has two parts, supported independently: (1) that Menominee vowel length alternations have been, and are, controlled by regular phonological processes, active in speakers' productive grammars and acquired by children learning the language, and (2) that the status of ū in the phonology is determined relatively late in acquisition, or at least at some point after the acquisition of the length alternation system, for the generation(s) of language learners around the critical time at which ū shifted from allophonic to phonemic status. (1) is supported by the regular system of length alternations described by Bloomfield (1962) and recent acoustic data confirming this system (with a slightly revised, but still regular and productive, phonological analysis; Milligan 2005; Cudworth 2019). (2) is entailed by a general principle of language change through learner error, where ‘error’ refers to the acquisition of a grammar that differs from, rather than replicates, the input grammar: ‘only learner errors from late stages of acquisition, in which the adult grammar is more or less in place, survive into adult speech’ (Ringe and Eska 2013: 112). Essentially, (2) claims that for the last generations to learn allophonic ū, this allophone was probably determined rather late in acquisition (maybe even late enough for literacy to play a role, as has been claimed for alveolar flap allophony; §4.1.2) as the language had only barely enough surface alternations to require allophone learning, until eventually there were not enough alternations and speakers could reach a stable adult grammar without their default contrastive grammar ever reaching the threshold of intolerability. With the (potential) allophonic status of ū being acquired so late, it is reasonable to assume that learners would already control the language’s system of length alternations. This somewhat resembles the assumption in Chapter 5 that since English alveolar flap allophony is acquired relatively late, it is learned against a background of a mostly adult-like English phoneme inventory.

6.1.4 Short vowel quality neutralisation

The Menominee surface quality distinctions of long vowels are represented by ā, a̅e̅, ē, ī, ō and ū. The Menominee short vowels are ostensibly the short counterparts of each long vowel. In the present-day language this includes short u, though as mentioned in the preceding sections it is found in only a highly restricted distribution, mostly either as a syllabic allophone or orthographic variant of *w* (Macaulay 2012). Short u is by far the least frequent vowel in regular production; Milligan (2005) encountered no tokens of it in 1011 words of narrative speech, while Cudworth (2019) found only two instances of it in 1160 words, with both instances being realised with very fronted pronunciations.

In contrast to the reasonably distinct pronunciations of long vowel qualities, Menominee short vowels as a group have become greatly neutralised in contemporary speech. Bloomfield emphasises that ‘The determination of the short vowels in the normal form is the greatest difficulty in Menominee phonetics’ (and difficult for native speakers as well as his foreign ear; 1962: §1.17), and 'What with the slurring and surface fluctuation of short vowels, one is likely to err as to the [...] normal phonemic form' (1962: §1.25). Acoustic analysis of speakers born in 1915-29 shows patterns of variation that Cudworth (2019) interprets as exhibiting both predictable allophony of some vowels/speakers and 'idiosyncratic allophony most likely in free variation' of others.

Although the range of possible variants differs across speakers and over time, and the exact set of surface forms described for each short vowel depends on the particular speaker population whose pronunciations are described (e.g. Bloomfield 1962; Cudworth 2019), a sampling of the vowel quality alterations potentially affecting Menominee short vowels includes the following list (6.5):

(6.5)

- initial /a/ ‘indistinguishable’ from initial /o/; ‘Younger native speakers make nonce-formations which show uncertainty’ (by pronouncing obvious long [ō] in place of long [ā], for long-vowel forms of lexemes where this vowel occurs usually in a short and therefore ambiguously pronounced position) (Bloomfield 1962: §1.19)
- ‘the /a/ is occasionally fronted so far as to coincide with [e]’ (Bloomfield 1962: §1.19)
- ‘the higher variants {of /ae/} which coincide with [e] seem to be favored in rapid speech’; ‘this partial coincidence makes it sometimes hard to determine where /ae/ rather than /e/ is to be set up in normal forms [...] even a native speaker would probably find it difficult to decide’ (Bloomfield 1962: §1.20)

e ‘is the favorite short front vowel, and [...] widely replaces normal [æ] and [i] and in certain positions even normal [a]’ (Bloomfield 1962: §1.21)

- ‘some speakers are fairly consistent in giving /i/ its typical coloring, while others, especially younger speakers, apparently replace [i] by [e] in all instances’ (Bloomfield 1962: §1.22)

Figure 6.1 additionally reproduces the pronunciation guidance for short vowels given by Macaulay (2012), which was compiled some generations after Bloomfield’s data, highlighting how ‘Many of the short vowels end up sounding a lot alike, in fact’ (Macaulay 2012: x).

This section has aimed to show that short vowel pronunciations in present/recent Menominee speech are often ambiguous, as they underdetermine which vowel quality they would correspond to in a lengthened phonological environment. Putting this another way, they are compatible with more than one hypothesis about underlying vowel quality. §6.2 considers a few different modelling treatments of short vowels, to reflect the historical decrease in

The short vowels are all shorter in length than the long vowels are, and their pronunciation varies more than the pronunciation of the long vowels does. When they're not stressed or accented, they often sound like:

- the final sound in the word *sofa* (which is called a 'schwa'),
- the *i* in the word *bit*, or
- the *e* in the word *bet*.

Many of the short vowels end up sounding a lot alike, in fact. In the list below, the English words show the many different ways each short vowel and short diphthong can be pronounced.

Symbol	Sounds like	Found in Menominee Word
a	but, sofa, father	kōhkōhsak, pakāhcekan, sōmenapoh
ae	bit, bet, bat	asāqcekaew, maehkāēnāhkok, mīhekanaehkaew
e	bit, bet, beet, sofa	māhkesen, nenēkehekok, kanew, āēhsepanak
i	bit, beet	kenūpik, onāwanikok
o	put, sofa, boat	apāēhsosok, omāhkahkowak, oqsāskok
u	put	
ya	yuck	ohpāēnyak
wa	won	mīcwah, pākaqāhkwan

Figure 6.1: Present-day short vowel pronunciations (Macaulay and Menominee Tribe of Wisconsin 2012).

information about underlying vowel quality that short vowel pronunciations can give to a listener.

6.2 Methods

6.2.1 Formal Model

For the case of Menominee, APL can be interpreted to mean that the learner's default hypothesis about ū, when their grammar projects surface segments to underlying forms, is as follows:

(6.6) H0: If *surface ū* then *underlying ū*

Since this rule is modelled from the perspective of a listener/comprehender, the learner considers every ū they hear to be a candidate subject to H0, and a violation of this rule is counted when the learner has morphological and semantic evidence to want to treat a candidate as underlyingly not ū. A child whose vocabulary includes both *watōp* 'alder (sg.)' and *watūpyak* 'alders (pl.)', and knows how they are related by the plural suffix *-yak*, must memorise the surface [ō] and [ū] alternation as an exception (surface ū in *watūpyak* = underlying ō in *watōp*) to their generalisation H0.

If the learner hears too many exceptions like this, as defined for their input by the Tolerance Principle, they must reject H0 and learn that there is an underlying relationship between surface segments. However, since all non-alternating ū lexemes (like *sūhkwaqsaen* ∼ *sūhkwaqsaenyak* 'whetstone/s') are also candidates subject to H0 and do not violate it, if the surface alternations are not too prevalent among these then ū and ō must be acquired as distinct phonemes. In §6.3 below these outcomes are both modelled, using data sets representing the language environment of learners at different historical stages.

6.2.2 Data

The Menominee data (lexemes) to use in these models come from the Menominee Dictionary (Macaulay and Menominee Tribe of Wisconsin 2012). This dictionary's words are from Menominee elders consulted by the dictionary's compiler, as well as from Bloomfield's (1975) Menominee Lexicon. It is marked as an Intermediate-level dictionary for heritage speakers/learners, and contains English-Menominee lookups as well as about 7400 Menominee-English definitions.

While many entries have a single given word form in a standardised inflection (like 3rd person singular independent for verbs), some entries provide an additional variant form for the headword. Although the dictionary does not explicitly state the criteria for including a variant form, based on comparison with Bloomfield (1962) it appears that variants are included exactly when this inclusion shows common forms that are not straightforwardly predictable by the learner. For the compiler of the modern dictionary, ū/ō vowel quality alternations are 'irregular' enough to print, while the quantity alternations (§6.1.3) are predictable enough to not be printed as alternate forms.

The interpretation that listed variant forms are common unpredictable forms makes sense given the dictionary's intended intermediate-level language users. This also permits the modelling assumption that the dictionary's entries and variant forms contain approximately the kind of data from which native speakers would acquire their phonemes: several thousand basic lexemes and their common inflected forms, including memorised exceptional stems. This crucially allows the model to assume that when a lexeme occurs either as only ū or only ō forms in this dictionary, then it is non-alternating, while if a lexeme is present with both ō and ū forms, then it does alternate in a common basic inflectional form, and it therefore counts as a violation of H0.

The text dictionary used by the present work was developed in conjunction with a multimedia Menominee Database, which includes the same information augmented by audio examples of at least some form of the word for many entries, as well as morphological analysis specifying underlying forms and composition for the majority of lexical items. However, the database is actively under construction, and many lexical items do not yet have some or all of this additional data attached. Since APL relies on fairly complete vocabulary lists, it is necessary to use a data source that enables all relevant words to be included in the model. In addition, the text dictionary has been released to the public, which facilitates evaluation and extensions of the present research. Outside of the Menominee Database, there are some audio recordings of the language freely available at http://www.menomineelanguage.com/language-music-and-more; these are not provided with transcriptions or analyses of morphology/phonology, and therefore also have not been incorporated in the present study, but they are an easily available source of pronunciations to listen to.

6.2.3 Procedure

To model whether (6.6) H0 for ū is tolerable to a learner, based on the total number of candidate lexemes subject to (6.6) and the number of violations, each ū-containing dictionary entry was processed and coded as one of the following five entry types, which together describe the full range of possibilities seen for ū in Menominee lexemes:

(6.7) **Ū**: Lexeme occurs only with long ū

kenūpik 'snake'

kenūpikok 'snakes'

ŪŌ: Long ū alternates with long ō in this lexeme, phonologically conditioned

mōhkotāqsow 'do carpentry (3rd pers.)'

nemūhkotāqsim 'do carpentry (1st pers.)'

Ūu: Long ū alternates with short u, and not with any form of o

kuqsyatam 'is shy/bashful (3rd pers.)'

nekūqsyatan 'is shy/bashful (1st pers.)'

Ūo: Long ū alternates in this lexeme with short, but not long, o

kockuan 'hook'

nekūckwan 'hook (1st pers. possessor)'

DUPL: duplicate entry repeating a stem that is already counted (for example, several verbs have entries once with animate and once with inanimate object), or a transparent compound of lexemes that are already counted, so this entry should not be counted independently in lexeme-based modelling.

The Menominee Dictionary is distributed as a PDF, so when converting[1] it into organised data (headword, part of speech, variant forms, etc.) its entries were parsed heuristically. No patterns of error were identified with this entry parsing, but there is some chance that the dictionary as used for modelling does not perfectly reflect the information in the dictionary

[1]using PDFMiner followed by a purpose-written Python script. No OCR was involved as the PDF already contained machine-readable characters.

as published for human readers.

The coding of each entry was based on the entire dictionary, so an entry that itself only contained ū could still be marked Ūu, ŪŌ, or Ūo if another entry with the same stem contained the relevant variant. A few lexemes occurred with both o and ō in addition to ū; these were coded ŪŌ reflecting the cost that such a paradigm would always incur. Duplicate entries were marked by hand, referring to productive morphology (standard diminutive suffix, verbal agreement with animate vs. inanimate objects), transparent compounding, etc. to determine the final counting of lexemes.

Around 3100 of the Menominee Dictionary's 7400 entries give a variant form alongside the entry headword. 337 entries, from this presumably fairly balanced sample of the language, contained at least one form with ū; 106 of these were discarded as duplicates of other lexemes, while the following table shows the breakdown by coding of the 231 principal ū lexemes:

Table 6.1: Types of paradigms in which ū appears.

Ū	ŪŌ	Ūu	Ūo
138	31	5	57

6.3 Implementations and results

The counts Ū, Ūu, ŪŌ, and Ūo can be used in the APL model to derive the number of tolerable and actual exceptions to H0 for a given set of language data. The following three model applications vary in the way that the learner's exceptions for short vowels are assessed, based on different assumptions about their possible input data.

The Pre-split version represents a historical stage in which vowel quality of short vowels as well as long vowels was informative. In this case, the model predicts that H0 for ū must be abandoned, that is, ū does not have to be learned as underlying.

The Split model version represents a current or idealised near-future state of the language in which short vowels are completely uninformative about vowel quality, while long vowels remain informative; in this model, H0 (6.6) is not rejected because there are not enough input alternations incompatible with it, and so ū is learned as underlyingly distinct from ō. The *Long-vowel version is an abstraction away from the disorder of short vowels, and strengthens the overall reliability of modelling the influence of Invariant Transparency in Menominee, by showing that despite anything potentially going on with the short vowel system, the long vowels alone are (just) sufficient to drive the vital secondary split outcome of learning underlying ū.

6.3.1 Pre-split

In Pre-split, N (the total number of candidate lexemes that could be subject to (6.6)) is comprised of all lexemes that contain ū. Short vowels are considered informative about vowel quality for this model; that is, a lexeme in which short u alternates with long ū is compatible with (6.6), while lexemes in which either short o or long ō alternate with long ū make exceptions e to (6.6) due to the differing vowel qualities (both [ō] and [o] are understood to normally have the underlying quality /o/, just as [ū] is understood to normally have the underlying quality /u/).

(6.8) $N = \bar{\text{U}} + \bar{\text{U}}\text{u} + \bar{\text{U}}\bar{\text{O}} + \bar{\text{U}}\text{o} = 138 + 5 + 31 + 57 = 231$

$e = \bar{\text{U}}\bar{\text{O}} + \bar{\text{U}}\text{o} = 31 + 57 = 88$

$\theta_N = 231/\ln(231) \approx 42.4 << 88$

In this model, the learner is faced with 88 exceptions to the maximally contrastive H0 (6.6), but H0 is only able to tolerate 42 exceptions. Therefore, this learner has found sufficient motivation to depart from the early grammar, so they must learn an allophonic relationship and the are not required to maintain ū underlyingly distinct from ō. Pre-split may reflect a historical stage in Menominee, at which vowel quality in short vowels was more often informative, and when ū was truly an allophone of ō.

The handling of Ūu lexemes in this model may be dubious: would there have been any such lexemes at the historical time when Menominee short vowels were informative as to vowel quality? Rather than definitively answering this question, it is enough to observe that the

5 Ūu lexemes, all of which have a glottal stop following the short u, can be either totally removed from the model (N=226, e =88, $\theta_N \approx 41.7 << 88$) or assumed to have actually been Ūo at an earlier time (N = 231, e = 93, $\theta_N \approx 42.4 << 93$) without affecting the robust general outcome that $\theta_N \ll e$.

6.3.2 *Long-vowel

*Long-vowel is a maximally simplified model, in which the learner exclusively pays attention to long vowels and does not (as far as the evaluation of H0 is concerned) even bother to acknowledge the existence of short vowel alternations.

(6.9) $N = \bar{\text{U}} + \bar{\text{U}}\bar{\text{O}} = 138 + 31 = 169$

$e = \bar{\text{U}}\bar{\text{O}} = 31$

$\theta_N = 169/\ln(169) \approx 32.9 \gtrsim 31$

In this model, the number of actual exceptions is just barely below the number of tolerable exceptions; this means that the model predicts a phonemic split, because by the end of acquisition there are not quite enough surface alternations to discard H0.

One interpretation of *Long-vowel could be that it represents acquisition for a learner who considers ū-lexemes with short vowel alternations to be too irrelevant, baffling, variable, etc. to reliably count as either compatible or exceptions to a hypothesis H0 about the long vowel (at least while they are young enough to still be shaping their phonological inventory). Another possible interpretation treats *Long-vowel as a theoretical abstraction away from the complications of short vowels (whatever they may be doing at some historical stage, likely with some vowel qualities compatible and others violating H0); *Long-vowel then reflects something like how much power the long vowels alone could have to drive historical change, if long vowels were the only definitively reliable information about vowel quality that learners are certain to hear. On this interpretation, the model does not itself describe a learning scenario that actually happened, but it does allow a bound to be placed on what could have happened even though some relevant details are impossible to know. The phonemic split in Menominee is robustly predicted to occur, without this prediction hingeing critically on any specific course of changes in the short vowel system.

6.3.3 Split

Split is like Pre-split in that H0 is evaluated with all ū-lexemes; however, in contrast to Pre-split, Split treats short vowels as completely uninformative about vowel quality, such that any short vowel is compatible with any hypothesis about the underlying vowel quality of the lexeme's corresponding long vowel.

(6.10) $N = \bar{\text{U}} + \bar{\text{U}}\text{u} + \bar{\text{U}}\bar{\text{O}} + \bar{\text{U}}\text{o} = 138 + 5 + 31 + 57 = 231$

$e = \bar{\text{U}}\bar{\text{O}} = 31$

$\theta_N = 231/\ln(231) \approx 42.4 > 31$

In this model, the number of exceptions $e = 31$ is less than the number of tolerable exceptions for H0, and so the default grammar projecting surface contrasts to underlying contrasts will be maintained, without acquisition of allophony. Therefore, the model predicts phonemically distinct ū and ō in present day Menominee.

Split is positioned as a current or near-future (possibly idealised) state of Menominee, in which short vowels are maximally ambiguous and therefore do not raise exceptions to (6.6) regardless of how they are pronounced. Bloomfield's (1962) observation of speakers producing varying long vowels for historical /ā/, in lexemes where learners usually encounter this vowel in a short form whose quality can evidently be interpreted as something other than /a/, shows a step towards this state. Essentially, Split is the final endpoint of a continuum that begins at Pre-split. On this continuum, the Menominee short vowel system gradually disintegrates over time, as the variation in pronunciations per vowel increases while the number of total surface contrasts in short vowels decreases. As the short vowel system degrades, the learnability of ū shifts gradually from definitely an allophone of ō to definitely independent underlying ū. This work therefore accords with a natural qualitative explanation for how ū might attain phonemic status, while adding a quantitative account of precisely how certain input requires learners to acquire a particular phoneme system.

6.3.4 Summary of results

Table 6.2 compiles key information on all three model variants described above, including how their input was coded as described in 6.2.3 and their main outcomes.

	U	UO	Uu	Uo	N	e		θ_N
Pre-split	✓	X	✓	X	231	88	$>$	42
*Long-vowel	✓	X	n/a	n/a	169	31	$\leq$	32
Split	✓	X	✓	✓	231	31	$<$	42

Table 6.2: *TP evaluation of non-allophonic H0 (6.6) at* Pre-split, *Long-vowel, *and* Split *stages. X and* ✓ *designate that a word type $\bar{U}\bar{U}$, $\bar{U}$u, etc. is in/compatible with the default grammar H0 at the specified stage. N, e, and θ_N follow the TP.*

6.4 Discussion

The phonemic split of conditioned allophones into contrastive phonemes, as illustrated here by Menominee back vowels, is a primary mechanism for languages to gain new phonemes. Although it has generally been difficult to point to the moment of a split itself, there is eventually clear after-the-fact evidence where splits occurred, as described in §6.1.2. In this case, restructuring of lexical items to specify an underlying distinction between ō and ū, when these had previously been predictably conditioned allophones of the same phoneme, silently innovated a new phonemic contrast in the language which later allowed retention of the surface distinction even in the eventual absence of distinct conditioning environments. If the data do not contain enough alternations among some number of non-alternating lexemes, then ū will be learned as underlyingly independent from ō, regardless of a complimentary distribution of surface ū with surface ō. Using APL to model this language change gives new insight on Invariant Transparency and the process of phonemic split, by offering a quantitative answer to how many surface ū/ō alternations would be 'enough' for Menominee learners to depart from a concrete default grammar and learn ū as a surface allophone of ō. The current model implementations use a dictionary to stand in for a child's lexicon. Therefore, this is a more abstract and approximate use of the model than the individual implementations in §5.4.2 that provided strong validation of APL. Still, the success of a group-level model in §5.4.1 suggests that this modelling framework can still give useful results even with less ideal data as used here by necessity.

In addition, these dictionary-based implementations incorporate only very high-level and broad generalisations about the nature of input children learn from. The models use minimal phonetic detail, instead focusing on a binary notion of non-/contrast between segments, and assuming that the language's short vowels can simply be either always completely informative about quality or never at all informative about quality. Recently published data on Menominee vowels as they occur in running speech can therefore add nuance to the interpretation of these results.

Milligan (2005) and Cudworth (2019) report frequency counts of all long vowels, short vowels, and diphthongs in small samples of narrative speech. As discussed throughout this chapter, the extreme rarity of short u in these sources, even relative to its already scarce occurrence in dictionary entries, is relevant to any modelling that might consider short u as a source of vowel quality information to learners. In Milligan's data the long ū is also notably infrequent, though in Cudworth's data it is not so uncommon, but still lowest-frequency long vowel, accounting for about 5% of all long vowels.

Cudworth also characterises the formant values that each vowel is produced with, and this data makes it possible to see which pronunciations are actually ambiguous about their underlying form and what the possible forms are. The acoustic values of short vowels studied by Cudworth (2019) overlap at least as much as expected from previous descriptions of the language. For example, e and i are barely distinct in the F1/F2 dimensions. As for the long vowels, overall ū and ō occupy different enough ranges of acoustic F1/F2 space that there can be no question that these are two separate surface-distinct segments, although one particular speaker has enough overlap for their targets that many tokens could be fairly ambiguous (Cudworth 2019). For the whole population of speakers, there is actually somewhat greater (although still by no means complete) overlap between the range of ō and the long low back vowel ā. The fact that there is quite a lot of acoustic overlap between the possible targets of these different vowels, but still some distinction between how likely it is for an ambiguous acoustic target to belong to each vowel, affirms that there is a real synchronic distinction between ū and ō (despite the lack of discernible conditioning environments for these segments) but also shows that learners are faced with quite a lot of variation and uncertainty (Cudworth 2019). Based on the F1/F2 measurements, ū and ī have similar heights but ū is considerably further back, as expected. However, it is particularly interesting that some of the long vowels, including both back ū∼ō and the front ī that historically conditions their raising, exhibit previously undescribed variation in vowel quality. Notably, there has been a pattern of increasing long front vowel acoustic overlap, and this involves the long ē, which did not condition back vowel raising, becoming indistinct from the long ī, which did condition back vowel raising (Cudworth 2019). As long vowels, ī and ē were expected to remain quite distinct in production (Bloomfield 1962; Milligan 2005) and serve as a main source of evidence about different environments for ō and ū when reduced short vowels are unable to provide much information; therefore, the fact that these long front vowels actually may not be able to signal much about a trigger for back-vowel raising further reinforces the idea that ō and ū have become distinct phonemes in the present-day language.

Together, the relative vowel frequencies and acoustic data that have been gathered from small speech corpora help ground APL implementations better in the actual distribution of surface alternations in Menominee. Out of the 143 non-alternating ū lexemes in the Menominee Dictionary (Macaulay 2012), for 88 items the apparent raising trigger is orthographic i or ī, whose pronunciation now is not generally distinct from the mid front vowel that is not a raising trigger. In addition, of the 57 items where ū alternates with short o, 25 have i or ī following the short o and 22 more instead have a diphthong following the short o; therefore, these are considered unraised only because of their length and not for absence of a raising trigger. For the vast majority of these 47 words, the long ū is in fact conditioned by the same trigger present in the short o form of the word, including many instances of ambiguous i or ī. Finally, for the 31 words where ū and ō alternate, the ō generally precedes mid or low vowels o, e, and a, or else is the final vowel of the wordform; meanwhile, in the ū forms of these 31 alternating words, i or ī is the raising trigger for 23 words. Therefore, it seems extremely unlikely for a learner to discover a set of distinct environments for ū and ō. However, there is a caveat that Cudworth (2019) studied a sample of only 5 speakers, all female and born 1915-29, including more detailed studies of the individual systems of 3 of these speakers. Some vowels had fewer than 10 tokens total (while others had hundreds), and besides this the format measurement gives a very incomplete characterisation of vowel realisation relative to other speech representations like MFCCs, so it is not possible to unambiguously understand what the various vowels' distribution overlaps mean.

Finally, some analyses (e.g. Kimper 2011; Milligan 2000) question the specification of only long vowels as targets of harmony, at least in the present-day language rather than original appearance of the high back vowel. For example, Milligan (2000: 251) notes that 'it is impossible to hear if vowel harmony affects short vowels since there is no phonetic difference between an /e/ and /i/'. This is relevant for back vowels as well since there is virtually no acoustic evidence about the realisation of segments conventionally described as short u, so it cannot be determined how similar or different they are from the range of short o realisations (see §6.1.4). However, in acoustic data of Cudworth (2019) the distribution of short o is in general about the same as long ō, and likewise distinct from long ū, without any strong suggestion of bimodality which would be the clearest signal for a subset of short o tokens being raised; furthermore, the data includes a relatively generous 113 tokens of o compared to 43 ū and 49 ō (and 2 u), so there is a fair chance for such bimodality to become apparent if it were there. As stated above, due to the nature of this F1/F2 acoustic evidence it still cannot firmly falsify the idea that short o could be a raising target, but as far as possible, the acoustic data supports this chapter's decision to proceed with the dominant assumption that short vowels are not a raising target.

Chapter 7

Differentiating learnability of extensionally equivalent grammars

This chapter uses APL to distinguish between some of the infinite set of abstract grammars that are descriptively compatible with an input language, and the subset of these grammars that are reachable through a child's acquisition process. This application of the model is illustrated by assessing the phonemic status of velar nasals in German. The segment [ŋ] originally entered the language as a nasal allophone, through place assimilation /ng/ $\rightarrow$ [ŋg], /nk/ $\rightarrow$ [ŋk].

Historically, [ŋ] has occurred with velar stop presence/absence alternations like *lang* [laŋk] ~ *lange* [laŋə]. These alternations motivated acquisition of an underlying two-segment nasal + velar stop sequence with conditioned deletion of the stop. Given the presence of underlying velar stops, it is not descriptively necessary to specify velar place on the nasal as well, since this can instead be received through the assimilation process. However, subsequent changes in the language have eliminated such velar stop presence/absence alternations from modern Standard High German (SG), and many lexical items with a velar nasal now have no forms where a following velar stop is pronounced. For this present-day language there are still descriptively adequate assimilation-based grammars with no velar-specified nasal phoneme /ŋ/, but the same SG data can also be accounted for by grammars with a phonemic /ŋ/ which need not always be followed by a deletable underlying velar stop (Vennemann 1970). Both possible representations are tolerable grammars, capable of describing the language's forms without too many exceptions. However, without actual alternations, children's input seems to provide no especially identifiable reason to acquire a grammar of synchronic assimilation where all [ŋ] are associated with a subsequent underlying velar stop (Kiparsky 1982).

The present-day assimilation analysis for [ŋ] (§7.1.1) implies rule reordering as a mechanism of language change, while the alternative phonemic analysis (§7.1.2) reflects inventory restructuring, silently gaining specification of velar place on the nasal segment itself at some point when 'there is a quite different grammar which accounts for the same language' after overt velar stops have completely disappeared from some paradigms (Kiparsky 1982: 6). Rule reordering in general is not particularly well understood, with questions around the conditions that could trigger it, factors or pressures that would motivate it, and even whether it truly occurs (Kiparsky 1982). On the other hand, the exact linguistic conditions that might be able to trigger restructuring will be difficult to detect, because this process inherently requires there to be a period of descriptive ambiguity. APL contributes to this inquiry by providing independently motivated tests for the occurrence of restructuring, without relying on case-specific subjective judgements.

The basic language data and an outline of the two major analyses for the German velar nasal (GŊ) are presented in §7.1. Then in §7.2, APL is applied to model the acquisition of each type of descriptively adequate grammar for modern SG, with a clear finding of acquiring underlying velar-specified /ŋ/. This application addresses uncertainties in the relationship between formal description and psychological representation, and implies language change through restructuring rather than rule reordering in the case of GŊ. In §7.3, APL is applied to a corpus of Middle High German (MHG), where relevant velar stop presence/absence alternations are prevalent. This modelling scenario helps reach a more complete understanding of the language's historical development. It also validates that, besides the prediction of a one-segment representation for present-day data in §7.2, with suitable input APL is also capable of predicting the kind of two-segment representation that historically preceded either rule reordering or restructuring and would have still existed after hypothetical reordering. Finally, §7.4 returns to the discussion of rule reordering and restructuring as processes of language change, and reviews how APL provides a general unbiased tool to investigate when they may occur.

7.1 German velar nasal

The velar nasal in present-day German indisputably has its historical origin in the assimilation of an alveolar or place-unspecified nasal followed by a velar stop (Issatschenko 1963; Hall 1989; Kiparsky 1982; Penzl 1968; Steinberg et al. 2016; Vennemann 1970). This origin is visible in orthographic *'ng', 'nk'*, and the standard language can still be described by a grammar of this nature, in which all surface velar nasals are synchronically generated through assimilation without any underlying place-specified velar nasal. However, the language can be about equally well described by referring to an independent velar nasal phoneme whose pronunciation is not synchronically dependant on an adjacent velar stop. This is because Modern Standard High German (SG) lacks lexemes where [ŋ] alternates with any of [n], [ŋg], or [ŋk]; instead, all forms of a lexeme nearly always have the same velar nasal variant, either [ŋ] or [ŋk] (7.1b). In contrast, in Middle High German (MHG) the association between velar nasals and velar stops was easily observable (7.1a): velar stops following nasals were often pronounced devoiced in syllable codas, while elsewhere they were not devoiced but could be deleted, and therefore words often had predictably conditioned alternating pronunciations that could reliably indicate the presence of an underlying velar stop following [ŋ] in a lexical item (Ringe and Eska 2013; Gress-Wright 2010; Penzl 1968; Wright 1907).

(7.1) a. Middle High German:

lanc	lange	'long'
-[ŋk]	-[ŋg]- or -[ŋ]-	

b. Modern Standard High German:

lang	lange	
[laŋ]	[laŋə]	(Wright 1907: §226,250)

The phonemic status of SG velar nasal has been of interest to phonologists due to tension between its obvious historical origins in /ng/ contrasted with the present questionable discoverability of a rather inert assimilation process, or a never-pronounced underlying stop in some lexical items (Ringe and Eska 2013). Most of this discussion centres around two interdependent issues: the representation of velar nasal segments in the phonology of dialects similar to modern SG, and the historical processes that could have led to this standard language as well as a variety of other dialects. Resolving either one of these issues goes some way towards clarifying the other, as knowing what kinds of language changes are possible constrains their potential outcomes, while knowing the outcome of a change helps identify how it happened.

In the case of GŊ, two different representations are descriptively adequate, and each representation would be the outcome of a different mechanism of language change whose conditions for operation have both been far from clear. Because of this, most previous argumentation on the velar nasal's present phonemic status has been based in linguists' subjective qualitative judgements about the nature of the two potential grammars. These arguments take into account factors like whether ordered rules are permissible in a theory of phonology (Vennemann 1970; Kohrt 1980), whether learners seek to minimise phonological inventory size and/or complexity of transformation between underlying and surface forms (Kiparsky 1982; Penzl 1968), and what kinds of small differences may be expected either in the grammars of similar neighbouring dialects or from generation to generation within a dialect (Dressler 1979; Hall 1989; Kiparsky 1982; Kohrt 1980; Wiese 1986).

§7.1.1 and 7.1.2 present data on the distribution of [ŋ] and the two major analyses for it in SG, including discussion of the motivation and qualitative evaluation for each grammar. Following this, §7.1.3 describes a common variant of the modern language, in which post-nasal stop epethensis causes an apparent velar stop presence/absence alternation within a subset of lexical items. APL will be applied to both SG and the epenthesis variant in §7.2, illustrating how two different grammars may describe the same language equally well without both being equally learnable.

7.1.1 Synchronic assimilation account: Two-segment velar nasal

The conventional treatment of SG velar nasal considers the present-day forms (7.1b) to result from a re-ordering of the rules that generated historical forms (7.1a). The historical forms, whose alternation patterns may be found also in modern Northern dialects (Paul 1916; Gress-Wright 2010; Wiese 1986; Wright 1907), can be accounted for by assuming three ordered processes (7.2, 7.3). The modern standard forms could be derived with a different ordering of these same three rules (7.4, 7.5), placing g-loss before final devoicing and therefore uniformly deleting all underlying /g/ from /ng/ sequences (Kiparsky 1982; Vennemann 1970; Kohrt 1980). There is some question whether the 'final devoicing' processing precisely consists of *devoicing*, as it has also been described as a tense/lax or fortis/lenis alternation; it will be referred to as devoicing for the rest of this chapter because what is essential for the arguments developed here is just that there is some alternation between different kinds of velar stop (Gress-Wright 2010; van Lessen Kloeke 1982).

(7.2) *Historical rule order leading to alternations:*

VA Velar place assimilation: a non-velar nasal assimilates to a following velar stop.

FD Final obstruent devoicing: in morpheme-final syllable codas, g (along with several other segments) is devoiced.

GL G-loss: g is deleted directly after ŋ; this does not delete either underlying /k/ or underlying /g/ that has already undergone final devoicing.

(7.3)

	/lang/	/lang+e/	'long'
Assimilation	laŋg	laŋge	
Devoicing	laŋk	–	
g-loss	–	laŋe	

(7.4) *Rule order not producing alternations:*

VA Velar place assimilation: a non-velar nasal assimilates to a following velar stop.

GL G-loss: g is deleted following ŋ.

FD Final obstruent devoicing: in morpheme-final syllable codas, any g or other voiced obstruent would be devoiced; however, no -ŋg codas exist to be transformed by this process.

(7.5)

	/lang/	/lang+e/	'long'
Assimilation	laŋg	laŋge	
g-loss	laŋ	laŋe	
Devoicing	–	–	

The ordered rules of (7.4) neatly account for the distribution of [ŋ] in SG, which has salient restrictions. It occurs often word-finally, before the reduced vowel [ə], and before a variety of stops, especially the voiceless velar stop [k], but nearly never before the voiced [g]; furthermore, the alveolar nasal [n] is not found before the velar stops [k,g] (Hall 1989; van Lessen Kloeke 1982; Paul 1916; Vennemann 1970).

A small digression is necessary in this description, as sequences of [ŋg] as well as [ng] and [nk] can occur when a major boundary separates the nasal and the stop. These are either word boundaries or some sufficiently strong word-internal morpheme boundaries; essentially, the domain of processes associated with the velar nasal does not necessarily align with an easily definable concept of a 'word'. A typical example of a lexical item with internal [ng] pronunciation is the verb *hingehen* [hɪngeːən] 'go elsewhere', from *gehen* 'go' and the separable prefix *hin* 'to there'. Other forms of this verb include *gehe hin, hingehe, gehst hin, hingehst, hingegangen, hinzugehen*, etc.; in all cases *hin* is realised with [n] rather than [ŋ]. Overall German has fairly perceptually clear acoustic cues to the relevant strong morpheme boundaries, and quite a lot of transparent productive processes that are sensitive to these same boundaries (Gress-Wright 2010; Hall 2010; Moulton 1947; Penzl 1968). Therefore, granting that all learners of German must learn appropriate sensitivity to these boundaries in general, e.g. to arrange the various forms of *hingehen* as shown, it is reasonable for any account of velar nasals to depend on these boundaries just as other phonological descriptions can depend on referring to more clear-cut word boundaries. Furthermore, in the estimated child vocabulary used to model GŊ acquisition in §7.2, nearly two-thirds of the instances of ostensibly word-internal [ng] occurred due to a single morphological element, the prefix *ge*− as in past participles *hingefallen*, *angehabt*, *angemacht*; other forms of these verbs (e.g. *hinfallen, hinfallt, fallt hin* 'fall down', *anhaben, anhast, hast an* 'have on/wear', *anmachen, anzumachen, mache an* 'switch on') make the major [ng] boundary entirely clear. Therefore, for simplicity this dissertation will generally discuss the velar nasal as if its conditioning were simply word-internal, except when it is specifically relevant to be more precise.

The most serious complications for analyses along the lines of (7.4) are a very small number of words pronounced with morpheme-internal [ŋg]. Going with the simplest interpretation of every surface [ŋ] as underlyingly /ng/, the implied underlying /ngg/ sequence for these [ŋg] words would be unanticipated given the rest of the language, and could require commitment to a theory of phonology in which g-deletion can apply exactly once (Penzl 1968). However, it is notable that the words with morpheme-internal [ŋg] tend to be obvious loanwords not obeying native phonology, like *flamingo* and *känguruh* 'kangaroo'. The very few loanwords that actually have [ŋ]∼[ŋg] stem alternations are unlikely to be in the vocabulary of a young child acquiring their phonology, such as the technical term *Diphthong* [ŋ] ∼ *Diphthongierung* [ŋg] 'diphthong/isation', or the river *Ganges*, nominative Ga[ŋg]es and genitive Ga[ŋ]es (van Lessen Kloeke 1982: 119, Penzl 1968: 340). Therefore, the existence of this vocabulary would hardly challenge even a simple analysis with underlying /ng/ in a framework that admits some exceptions to linguistic generalisations.

Still, if a more complete description of the language is demanded, systems of rules without unusual assumptions can be devised to generate all SG surface forms including the handful with morpheme-internal [ng] (van Lessen Kloeke 1982; Penzl 1968; Vennemann 1970). For example, Vennemann (1970) describes an account with underlying /ng/ that generates [ŋg] where required by further restricting the environments in which the post-nasal g-loss rule applies, such as deleting *g* before ə but not before unreduced vowels. Hall (1989: 839) mentions additional similar proposals of varying complexity, but rejects such descriptively adequate collections of 'apparently unrelated environments which trigger the [g deletion] rule' for lacking explanatory power. Indeed, all of these /ng/ accounts succeed by dividing native from foreign vocabulary, either overtly by marking words with a +Native/+Foreign diacritic, or in effect by defining rules over some combination of environments that serve to separate borrowed and native vocabulary (Hall 1989; Issatschenko 1963; Vennemann 1970). For example, the history of the language is such that natively any vowel directly preceded by [ŋ] ends up reduced, while borrowed [ŋg] tends to precede a non-reduced vowel in words whose phonology is often abnormal for German in other respects as well (van Lessen Kloeke 1982; Vennemann 1970). Therefore, while the systems of rules proposed to account for occasional morpheme-internal [ŋg] in loanwords may lack desired explanatory power from some perspectives, it is questionable how far a description of core German phonology should need to account for these words anyway; (Vennemann 1970: 66) notably does not consider a small amount of foreign vocabulary worth building an analysis around, and judges that 'the name of an Indian river cannot destroy an otherwise well motivated hypothesis about the phonology of German'. Furthermore, besides this set of words with [ŋg], there is no other serious challenge to describing the standard language with ordered rules as in (7.4).

Finally, Kohrt (1980) examines a number of subtly different dialects, and determines that different orders of the same set of rules do not give a satisfactory account of this variation beyond SG, which can include within-dialect optional stop pronunciation (e.g. both [eŋ] and [eŋk] possible for *eng* 'narrow') and occasional lexical exceptions. However, Kohrt suggests instead that the scope of g-deletion expanded without becoming reordered relative to final devoicing, and dialects could relexicalise various sets of words at different points in the rule extension process, all while maintaining a two-segment /ng/ (or /nk/) representation associated with every [ŋ] along with the same basic processes of assimilation, final devoicing, and g-loss. This analysis also recognises that morphemes borrowed into the language relatively recently may not meet the structural conditions of input to the language's older sound changes, and may therefore have surface forms in violation of what can be derived for native vocabulary. However, while Kohrt (1980) justifies the idea of rule expansion by arguing against the traditional rule reordering account for [ŋ], the possibility of velar-specified phonemic /ŋ/ is not addressed.

Motivation for the two-segment velar nasal representation has extended beyond just its descriptive adequacy. Penzl, who ultimately considers German to have a /ŋ/ phoneme, first concedes a compelling intuition that 'Indeed the entire diachronic development of /ng/ to [ŋ] can be quoted as evidence for an allophone [ŋ] before velars' (1968: 343), although this quite common line of reasoning expresses what Kiparsky (1982: 130) cautions is only 'a very natural, though theoretically unjustified, desire to have synchronic descriptions reflect diachrony to the greatest possible extent' (Dressler (1979: 450) makes a more prudent statement that 'we cannot derive any positive evidence for the abstract analysis from diachrony, but at least this analysis is compatible with a diachronic analysis of the history of [ŋ] and with diachronic theory in general'). The restricted environments of the GŊ segment can also seem to point to an obvious conclusion, e.g. Hall's (1989: 818) statement that 'In this section I will similarly assume that [ŋ] is a derived segment because of its limited distribution, although I will not provide any arguments to this effect.' Vennemann (1970: 75) makes a more concrete argument, based in the ability of the two-segment representation to 'save the generality' of productively deriving certain alternations in strong verb stems. However, Dressler (1979: 446) considers the evidence of these verbs to be 'either diachronic or too abstract for contemporary German', and in light of psycholinguistic results showing that for present-day speakers these strong verbs are mentally represented as irregular paradigms with multiple memorised stems (§3.3.2), it certainly seems more likely that the processes whose productivity Venneman's argument relies on are no longer synchronically active, although [ŋ] was presumably a associated with a nasal+stop sequence when they were. Overall, a two-segment velar nasal representation for SG seems descriptively suitable but not essential.

7.1.2 Phonemic account: Single-segment velar nasal

There is also not much difficulty accounting for SG surface forms with the competing analysis of an underlying /ŋ/ that does not rely on a following stop for its velar place feature. This representation would indicate that at some point restructuring of the language's phoneme inventory and lexicon took place. Assuming the existence of phonemic /ŋ/, surface [ŋk] could be associated with underlying /ŋk/ while surface [ŋ] without a following velar stop would be simply /ŋ/. Describing all attested SG surface forms is therefore fairly trivial. The main issue for a phonemic GŊ proposal is to provide a satisfactory account of the restricted distribution of [ŋ]. This is not strictly necessary, as in principle /ŋg/ sequences might just be absent from native German vocabulary (in the same way that /mb/ sequences do not occur while /mp/ do; Vennemann 1970) as if by coincidence from present-day speakers' points of view. However, if linguists find it important for the grammar itself to eliminate the kind of /ŋg/ sequences that are not in fact found in the language, then the restricted environments in which [ŋ] occurs can also be defined with morpheme structure constraints involving the phoneme /ŋ/, such as prohibiting morpheme-initial /ŋ/ or prohibiting velar nasal + stop clusters after a long vowel or diphthong (Issatschenko 1963; van Lessen Kloeke 1982; Vennemann 1970). As with the accounts presented in §7.1.1, for the most part these descriptions handle morpheme-internal [ŋg] by dividing the lexicon into native and borrowed wordforms without much further explanatory power.

Still, analyses with underlying single-segment /ŋ/ seem less common recently. Van Lessen Kloeke (1982) provides one such description, motivated primarily by minimal contrasts as in *Wamme*, *Wanne*, *Wange*, including a thorough specification of the environments for [ŋ] and morpheme structure constraints concerning /ŋ/. Penzl (1968) also considers [ŋ] to not derive from a synchronic process with /g/, based on the small amount of foreign vocabulary with [ŋ]∼[ŋg] alternations (along with some confusion about surface vs. underlying constraints; claiming, presumably because of coda devoicing, that phoneme sequences like /ngt/ as for *singt* [zɪŋt] are 'structurally impossible'). Finally, Steinberg et al. (2016) examine mismatch negativity in an oddball paradigm EEG study, and determine that a phoneme /ŋ/ specified for velar place is available to German speakers, and it is not automatically associated with a following velar stop when no such stop is pronounced.

Additional support for the one-segment velar-specified /ŋ/ analysis is based in considerations of learnability, and the possible role or limitations of purely distributional evidence. Children's input in this situation is similar to what Kiparsky (1982: 51) describes for acquisition of the present-day English word *knife*, where due to lack of alternations 'the child will have no reason whatever to postulate a basic form with initial [k] and to set up a rule deleting initial [k] before [n]'. With no alternations between presence and absence of velar stops after [ŋ], SG learners have no motivation to hypothesise mandatorily deleted underlying velar stops in the words that never have surface velar stops, as suggested by participants' responses in the EEG study of Steinberg et al. (2016). Vennemann (1970: 69) agrees in general with a default of underlying segments being identical to surface segments, that 'should be overruled only by stringent evidence from phonological rules, not in cases where some abstract representation would only fill a "hole in a pattern"', although as mentioned above he believes that there is a verb stem alternation that provides sufficiently compelling evidence in the case of GŊ. In the end, the accumulated arguments for an underlying nasal+stop sequence may be 'very plausible and descriptively appealing, but they do not make an abstract analysis necessary': the phonemic /ŋ/ analysis is also 'possible and does not lack plausibility' (Vennemann 1970: 75).

7.1.3 K-epenthesis dialects

In addition to the historical and Northern forms like (7.1a), and the Southern/Standard modern forms like (7.1b), there are many present-day dialects whose forms are similar to SG but with [k] additionally present in syllable codas only between [ŋ] and another consonant in the coda, as shown in (7.6a-b).

(7.6)	a.	singt	[zɪŋkt]	‘sings’
		sinkt	[zɪŋkt]	‘drowns’
	b.	hengst	[hɛŋkst]	‘stallion’
	c.	hemd	[hɛmpt]	‘shirt’

The [k] in this restricted distribution has sometimes been assumed to be partial survival of the stop from original underlying /ng/ clusters (as affected by coda devoicing), but it is more likely to be a novel epenthesis process inserting [k] into coda clusters after /ŋ/ (Kohrt 1980; Gress-Wright 2010). This epenthesis is a natural sound change that includes additional homorganic stop epenthesis in the same dialects, like (7.6c). It can be easily accounted for by an epenthesis process relatively late in the derivation whether or not [ŋ] is analysed with an underlying velar stop, as shown in (7.7) for a two-segment velar nasal with rule reordering and (7.8) for a phonemic velar nasal with unchanged rule application.

(7.7)	/hengst/	‘horse’
Assimilation	heŋgst	
g-loss	heŋst	
Devoicing	heŋst	
Epentheis	heŋkst	

(7.8)	/heŋst/	‘horse’
Assimilation	heŋst	
Devoicing	heŋst	
g-loss	heŋst	
Epentheis	heŋkst	

Post-nasal homorganic stop epenthesis in final clusters can result in surface alternations (7.9, [k] and [t] insertion) that individually are indistinguishable from alternations caused by deletion of an underlying stop like (7.3). Therefore, in §7.2 velar nasal acquisition is modelled in both SG and a k-epenthesis dialect. However, the prevalence of velar stop alternations caused by k-epenthesis is low relative to the total occurrences of [ŋ], so this does not change the predicted phonemic status of velar nasals.

(7.9)	a.	springst	[ʃprɪŋkst]	'jump' 2.sg
		springen	[ʃprɪŋən]	'to jump'
	b.	rennst	[rɛntst]	'run' 2.sg
		rennen	[rɛnən]	'to run'

7.2 Acquiring present-day German velar nasal

This section models the predicted phonemic status of velar nasals in two varieties of present-day German. §7.2.1 summarises how APL applies to this situation. §7.2.2 describes the language data serving as the model's input, while §7.2.3 explains how this input is coded to assess tolerability of an early surface-mapping phoneme inventory. §7.2.5-6 present model results and §7.2.7 reviews the evidence that this has provided about speakers' likely representation of this segment.

7.2.1 Formal model

GŊ acquisition is modelled by assuming that, since [ŋ] occurs as a surface segment, learners initially hypothesise that it is also an underlying segment (7.10). The model therefore initialises with a phoneme mapping exactly to [ŋ], and evaluates whether this segment participates in too many alternations; it will remain an underlying segment as long as it does not and the initial grammar is tolerable (3.2).

(7.10) H0: If [ŋ] $\rightarrow$ /ŋ/

In the APL model, surface alternations in the lexicon are what motivate allophone learning, regardless of any patterns of gaps in segments' environments that may also stand out to adult observers of a language. However, there are no direct alternations of [ŋ] with [n] in SG (or for several centuries previously) that could provide the kind of evidence against H0 that was used in Chapters 5-6. The only kind of alternations that might be able to motivate acquisition of some kind of abstract transformation around surface [ŋ] are instead paradigms where [ŋg] or [ŋk] alternate with a plain [ŋ] not followed by surface velar stop. Paradigms with these alternations would still require the learner initially to memorise more than one lexical stem, when they understand the morphological relation between the distinct forms and have no productive process to derive one stem form from the other; therefore, the post-nasal stop presence/absence alternations can be counted as observable evidence indicating that surface [ŋ] might not be always a simple reflex of underlying /ŋ/. With this type of alternation as the source of cognitive cost, APL in this case can be thought of as modelling 'is [ŋ] just exactly /ŋ/, or is it some sequence /AB/', rather than the previous studies modelling 'is [x] just exactly /x/, or is it some /Y/ which has multiple realisations including both [x] and [z] where [x] $\neq$ [z]'. Using the model this way is possible because in either case it is only necessary to explicitly evaluate/reject H0, whose form does not vary (e.g. 7.10). Therefore, this extends the APL model to potentially learn allophony beyond only single-segment conditioned replacement.

As usual, if the total number N of lexical items containing [ŋ] includes too many alternating paradigms e where the [ŋ] occurs in different forms both with and without a following velar stop, the learner's system of phonology after rejecting intolerable H0 (7.10) must recognise a productive relationship between [ŋ] and some underlying representation that is not just /ŋ/. If this happens, the phoneme inventory might no longer require a nasal phoneme specified for velar place, although it is certainly possible that a learner could hypothesise a redundant transformation where e.g. every surface [ŋ] is associated to an underlying sequence /ŋg/ or /ŋk/ with conditional stop deletion. However, with the input modelled in §7.2.5-6, the threshold to reject H0 (7.10) is never reached, so in this case the model interpretation is unambiguous: [ŋ] straightforwardly maps to underlying velar-specified nasal /ŋ/, whose existence as a phoneme distinct from /n/ is therefore required.

7.2.2 Data

Like the input-based English alveolar flap model described in §5.2, children's vocabulary composition is estimated from frequency in corpora of child-directed speech by assuming that the most frequent words are more likely to be learned by many young children. Frequencies of over 50,000 word types were obtained from about 4 million tokens of adult speech (presumed generally child-directed) in a combination of several German CHILDES corpora (Caroline, Leo, Miller, Rigol, Szagun, Wagner), excluding child speech from the transcripts as far as possible (MacWhinney 2000):

Caroline One child, age 0;10-4;3.

Leo One child, age 1;11-4;11.

Miller Two children, age 1-4y.

Rigol Four children, age 0-8y.

Szagun 22 children, age 1;4-2;10.

Wagner 13 children, age 1-15y.

The models of §7.2.5-6 are based on the most frequent 9,000 word types of child-directed speech, which have frequency ≥ 5 in the combined corpus. These word types represent common inflected forms of about 5900 lexical items, which is the estimated vocabulary size of a 6-year-old German child (Segbers and Schroeder 2017). The count of lexical items is derived from lemmas given for each of the 9,000 word forms by CELEX (Baayen et al. 1995). It is probably unnecessary to model GŊ acquisition with such a large vocabulary, because the simplicity of the standard system, lacking enough alternations to reject H0 (7.10), means that this phoneme's acquisition is likely settled well before age 6. However, if a model with a 3-year-old's smaller word knowledge failed to acquire a productive transformation, this could leave concerns that the model's limited vocabulary might be insufficiently informative about the phonemic inventory of SG, only representing a child whose language had not yet developed enough to reach a generalisation that adult speakers have. The choice of implementation with a school-age child's considerably larger vocabulary is intended to help remove this concern. Standard German pronunciations used in §7.2.5 were also taken from CELEX; any items from the child-directed corpus that were not present in CELEX (*www*, *hm*, *'s* etc.) were excluded before sampling the 9,000-word vocabulary. Alternate pronunciations representing a widespread dialect variant with k-epenthesis in coda consonant clusters (§7.1.3, 7.2.6) were simulated by applying the epenthesis process to CELEX pronunciations, as provided by a native speaker of one such variety.[1]

7.2.3 Coding

Model inputs N and e were determined by coding all words that included spelling of 'ng' or 'nk' and/or surface [ŋ], in order to include all velar nasal pronunciations and the possible wordforms they could alternate with. A few loanwords with 'ng' spelling but unrelated pronunciation, like *Melange* [melãʒ] from French, were excluded. Remaining forms are labelled in the following pronunciation categories:

(7.11) **ŋ+** [ŋ] plus velar stop; surface [ŋg] or [ŋk]

ŋ_ [ŋ] not followed by a velar stop; either followed by some other segment or word-final [ŋ]

n+ no surface velar nasal; pronunciation [ng] or [nk] with a significant morpheme boundary between the nasal and the stop

The third category n+ does not in the end play a role in model calculations, because there are no paradigms where any [ŋ] alternates with an [n]. All words in this n+ category have the non-velar [n] pronunciation in every form of their paradigms, because all of these words have major morphological boundaries preventing any velar assimilation in the first place,

[1] My thanks to Beatrice Santorini for providing her pronunciations of all relevant words in the 9000-word child vocabulary sample.

like *angehabt* as described in §7.1.1. This category will not be mentioned further.

APL is applied in implementations over both words and lexemes. The relationship between wordforms and lexical items is determined according to the CELEX assignments of word types to lemmas; therefore, to whatever extent the CELEX lexicon does not reflect a child's knowledge about how the items in the present 9000-word model vocabulary are related to each other (e.g. through inflectional or derivational morphology), this could introduce some error to models of acquisition. Still, the lexeme model seems more cognitively faithful according to §3.3.2 , but convergent predictions from the word-level model may give further assurance of the overall conclusion, by showing that the outcome does not critically depend on particular modelling assumptions being exactly correct.

Additionally, models are implemented either as $-As$, no assimilation required to account for velar nasals, or as $+As$ with synchronic assimilation only for the ŋ+ forms in which the velar nasal is followed by a pronounced velar stop. In the −As models, a straightforward implementation of APL, all surface [ŋ] are counted among N subject to H0 (7.10). The +As model scenario provides an especially stringent or conservative test of whether /ŋ/ must exist as a phoneme, by granting that a learner who is aware of general nasal place assimilation in German might not analyse surface [ŋk] sequences as underlying /ŋk/. For this learner, perhaps only surface [ŋ] not followed by a velar stop is required to be interpreted as /ŋ/ and count towards N under H0 (7.10). Intuitively, if a lexeme includes a velar stop like [ŋk] or [ŋg] in all of its surface forms, learners could assume it has an underlying velar stop as well, and in that case learners who know a general nasal assimilation process (presumably motivated by other evidence, including assimilation at other places of articulation or across word boundaries in some speech styles) might derive [ŋk] from an underlying non-velar nasal + velar stop sequence. The +As model therefore tests whether the instances of [ŋ] without a following surface stop must also still have an underlying stop (two-segment representation), by restricting the model's relevant class of lexemes to those with at least one ŋ_ surface form and evaluating whether these ŋ_ words alone are sufficient to keep simple /ŋ/ in the underlying inventory for learners who do not analyse non-alternating [ŋk] paradigms as having underlying /ŋ/. The +As model is probably less realistic than the −As model; it does not explain or motivate why children would discard the possibility that non-alternating [ŋk] includes an underlying /ŋ/. However, a learner could potentially reach a state corresponding to the +As model if they aimed in some sense to minimise the size of their phoneme inventory, or aimed to specify as few features as possible for items in their lexicon; furthermore, these modelling assumptions provide the most generous chance to reject the single-segment phonemic /ŋ/ H0, since in this scenario a two-segment representation with no velar-specified nasal is already assumed for many instances of surface [ŋ]. Still, in the end, APL always predicts that /ŋ/ must be a phoneme, as the learner is not required to learn a productive generalisation that ties it to unpronounced velar stops. This conclusion for SG is therefore robust to any variant of reasonable assumptions.

7.2.4 Implementations

Table 7.1: **Modern Standard High German illustrative data**

Word	Pronunciation	Frequency rank	LemmaID	GŊ realisation	Alternation
ding	'dIN	585	9169	ŋ_	no
dinge	'dI-N@	1311	9169	ŋ_	no
dings	'dINs	1579	9169	ŋ_	no
dingen	'dI-N@n	1617	9169	ŋ_	no
finger	'fI-N@r	383	13120	ŋ_	no
fingern	'fI-N@rn	1882	13120	ŋ_	no
langsam	'l&N-zam	454	22207	ŋ_	no
langsamer	'l&N-za-m@r	3393	22207	ŋ_	no
zeitung	'=W-tUN	1105	50239	ŋ_	no
denke	'dEN-k@	519	8868	ŋ+	no
denkst	'dENkst	1281	8868	ŋ+	no
denken	'dEN-k@n	1582	8868	ŋ+	no
links	'lINks	1396	23115	ŋ+	no

Table 7.1 shows a selection of data coded as stated in §7.2.3, to facilitate concrete illustrations of how N and e are determined in each implementation described below. In this table, each wordform is listed together with its pronunciation as given in CELEX, its fre-

quency rank in the compiled child-directed speech corpus described in §7.2.2, its lemma ID assigned by CELEX (an arbitrary number that indicates which wordforms belong to the same lemma), the category of the GŊ realisation for the CELEX pronunciation as defined in (7.11), and whether the overall paradigm that the word belongs to contains alternation in this pronunciation. The table shows thirteen words belonging to a total of six lemmas; this set of items is selected only for clarity of illustration and is not meant as a representative sample of a child vocabulary.

Table 7.2: **Present-day nasal cluster epenthesis illustrative data**

Word	Frequency rank	LemmaID	GŊ realisation	Alternation
ding	585	9169	ŋ_	yes
dinge	1311	9169	ŋ_	yes
dings	1579	9169	ŋ+	yes
dingen	1617	9169	ŋ_	yes
finger	383	13120	ŋ_	no
fingern	1882	13120	ŋ_	no
langsam	454	22207	ŋ+	no
langsamer	3393	22207	ŋ+	no
zeitung	1105	50239	ŋ_	no
denke	519	8868	ŋ+	no
denkst	1281	8868	ŋ+	no
denken	1582	8868	ŋ+	no
links	1396	23115	ŋ+	no

Table 7.2 shows the same set of words as they are pronounced in a k-epenthesis language variety. In this table the CELEX pronunciation is not listed, and the the GŊ realisation instead reflects speaker judgements as described in §7.2.2. The alternation status of each paradigm is determined according to these judgements, and this data will also be used below to illustrate the mechanics of APL for k-epenthesis dialects.

For the no-assimilation lexeme-based model (−AsL), the number of items N potentially subject to H0 (7.10) consists of every lemma where at least one form has any surface velar nasal, ŋ+ or ŋ_. A cost e is assessed for each lexeme that has at least one ŋ+ form and at least one ŋ_ form both attested in the 9,000-word child vocabulary (in practise, these are all alternations of [ŋ]∼[ŋk]; there are no [ŋ]∼[ŋg] alternations). Therefore, applying −AsL to the SG data in Table 7.1 yields $N = 6$ since there are six distinct lemmas, and $e = 0$ since there is no alternation of pronunciations within the word forms belonging to any of these lemmas. Using instead the k-epenthesis data in Table 7.2, $N = 6$ still, but there is one lemma whose wordforms contain alternation of k presence/absence, so $e = 1$.

For the lexeme-based model with assimilation in the presence of overt velar stops (+AsL), the number of items N potentially subject to H0 (7.10) consists of every lemma where at least one form has surface ŋ_, a velar nasal not followed by a velar stop. Lexemes that contain only ŋ+ forms (non-alternating [ŋk] or [ŋg] in all forms) are not considered candidates for underlying /ŋ/ under the assumptions of the +As model. Of the lexemes that do have at least one ŋ_ form, a cost e is assessed for each one that also has at least one ŋ+ form. When applying +AsL to the SG data in Table 7.1, $N = 4$ because there are four lemmas (*ding*, finger*, langsam*, zeitung*) with at least one wordform having an ŋ_ pronunciation. In this case $e = 0$ because none of these lemmas have alternations between wordforms with ŋ_ and ŋ+ pronunciations. When applying +AsL to the k-epenthesis data in Table 7.2, now N=3, because here all forms of *langsam** are pronounced with [ŋk]. In addition, $e = 1$, because one of the 3 ŋ_ lemmas also has a form with ŋ+.

For the no-assimilation word-based model (−AsW), the number of items N potentially subject to H0 (7.10) consists of every word with any surface velar nasal, i.e. the total number of ŋ+ and ŋ_ words, and a cost e is counted for each ŋ_ word that is a member of a paradigm that also has at least one ŋ+ member. This measurement of e for individual words reflects the fact that a ŋ+ form guarantees the existence of a corresponding underlying form with a nasal+stop sequence, in accordance with the presumed general straightforward mapping between surface and underlying forms, and therefore if there is also an ŋ_ form of the same lexical item, a learner maintaining H0 (7.10) must memorise a different representation that lacks this stop to account for the ŋ_ form. When applying the −AsW implementation to the SG data in Table 7.1, $N = 13$ since there are 13 total wordforms. To calculate e, Table 7.1 shows 9 ŋ_ forms. When examining the full paradigm for each of these wordforms, none of the 9 ŋ_ belong to alternation-containing paradigms, so $e = 0$. Turning to the k-epenthesis data in Table 7.2, again $N = 13$. However, in this data there are only 6 ŋ_ wordforms, and three of these six (three different forms of *ding**) belong to a lemma that also has at least one ŋ+ form, so $e = 3$. This word-based count may be an unjustifiably high cost assessment, with the lexeme-based models being a better reflection of the true cost; even so, H0 will be found to remain tolerable in the −AsW model.

For the word-based model with assimilation before overt velar stops (+AsW), the number of items N potentially subject to H0 (7.10) consists of every ŋ_ word, and a cost e is counted for each of these words that belongs to a paradigm that also has at least one ŋ+ member; again, this likely overestimates the number of exceptions, and as a +As model it also probably underestimates the number of velar nasal items N in the learner's vocabulary. In any case, applying +AsW to the SG data in Table 7.1 yields $N = 9$ words with ŋ_ realisations, and $e = 0$ since none of them are part of paradigms where they alternate with any ŋ+ form. With the k-epenthesis data in Table 7.2, the count is instead $N = 6$ words realised as ŋ_, and $e = 3$ since three of these words belong to a paradigm whose forms also include ŋ+.

7.2.5 Results: Standard German

In the standard pronunciations of CELEX child vocabulary, there are no alternations involving the velar nasal and stops. All paradigms contain either only ŋ_ forms, only [ŋk] forms, or only [ŋg] forms. Therefore, the learner has no motivation to abandon the hypothesis of underlying /ŋ/, under any of the four sets of modelling assumptions (Table 7.3).

7.2.6 Results: K-epenthesis dialects

When epenthetic [k] appears between [ŋ] and another syllable coda consonant, some paradigms have all of their forms changed from ŋ_ to [nk], which decreases the count of N in +As models but does not introduce any alternations. An example of this is seen with *langsam** in Tables 7.1 and 7.2. Other paradigms, as shown with *ding**, have only some forms trans-

Table 7.3: **H0 for ŋ in Standard German** (tolerable when $e < \theta_N$, $\theta_N = N/ln(N)$)

Model	N	e	θ_N
−AsL	344	0	59
+AsL	240	0	43
−AsW	504	0	80
+AsW	345	0	59

formed from ŋ_ to [nk] due to differences in syllabification and coda content, and this adds obvious surface alternations e.

Table 7.4: **H0 for ŋ with k-epenthesis** (tolerable when $e < \theta_N$, $\theta_N = N/ln(N)$)

Model	N	e	θ_N
−AsL	344	15	59
+AsL	224	15	41
−AsW	504	37	80
+AsW	300	37	52

Although the insertion of [k] in coda clusters creates apparent alternation within paradigms, in the estimated child vocabulary used for this model, these alternations are not prevalent enough relative to the amount of non-alternating [ŋ] to make learners productively generalise about the relationship between nasal and stop, as shown in Table 7.4 for all four modelling scenarios. Instead, it remains tolerable to memorise this number of duplicate stem forms while maintaining a straightforward mapping from surface to underlying segments, including a single-segment phonemic /ŋ/ that is not associated with an underlying velar stop in lexemes where such a stop is never pronounced.

Furthermore, it can be questioned whether the alternations caused by homorganic stop epenthesis should be considered exceptions to H0 (7.10). They are likely obvious insertions rather than underlying segments, as is the [p] in *Hemd* [hɛmpt]. This post-nasal coda epenthesis appears to be a regular productive transformation, whose N is restricted to nasals in a certain set of syllable coda environments, but not restricted to nasals with velar place (7.5, 7.9). Even in accounts that analyse [ŋ] as a nasal+stop sequence with no phonemic /ŋ/, it is uncontroversial that epenthetic [k] is (re-)inserted after [ŋ] has its velar feature, such that this stop receives its place from the nasal rather than the other way around (Kohrt 1980). The purpose of nevertheless counting epenthetic [k] as a source of violations of H0 (7.10) was that if H0 remains tolerable even when these epenthetic velar stops are considered evidence for 'alternating stems', then the prediction of non-productivity (independent phonemic /ŋ/) is especially robust. This also indicates that the model's predicted outcome is not contingent upon any assumptions about when children might learn that epenthetic [k], [p], etc. are insertions, relative to the course of learning about /ŋ/. Acquiring the general post-nasal coda epenthesis process is not a focus of this dissertation, beyond having shown that its effects in the input do not impact the phonemic status of the velar nasal.

7.2.7 Discussion

This analysis of present-day GŊ phonology brings out certain theoretical commitments and implications of APL, illuminating differences between this model and alternatives when accounting for the same data. As reviewed in §7.1, there are present-day varieties of German that seem to have reasonable formal analyses both with and without a velar nasal phoneme, so neither way clearly objectively describes the language data better. Particularly in children's vocabulary, the environment of [ŋ] today remains consistent with synchronic persistence of a historical allophonic process that does not require underlying velar place of the nasal segment but does require it to be followed by a velar stop; on the other hand, there are no active alternations between [ŋ] and a non-velar nasal that would require the active transformations of such a process. The learnability of either type of grammar therefore depends on the extent to which allophony is acquired whenever possible, such as if children look for patterns in input distribution to minimise phoneme inventory size or feature specification of lexical items, or whether it is learnt only when necessary as required to account for alternations. APL examines this question with a quantitatively defined general model of language learning, which predicts definitively that there is a velar nasal phoneme in some major present-day dialect types, because there are not enough alternations involving [ŋ] to require learning a productive association between underlying velar stops and the surface velar quality of the nasal. Specifically, in this model the (passive, or potential) observation that many [ŋ] precede [k] while no [n] precede [g] or [k] does not motivate a child to generalise that all surface [ŋ] are synchronically derived through velar stop assimilation.

APL's ability to discriminate between *tolerable* and *learnable* grammars is a feature of the analysis. Because both possible GŊ representations provide tolerable and practically exceptionless descriptions for the adult language, neither would incur excess cognitive cost for a speaker to use, so the Tolerance Principle alone (outside the context of this learning model) does not determine whether one is more likely than the other to match speakers' mental representations, just as earlier subjective assessments of how well each grammar describes the data (§7.1) did not resolve this question.

Another contribution of this case study is extending APL to model acquisition of representations whose possible alternations involve more complex transformations. The two case studies presented previously have been limited to instances of one-to-one segment substitutions, modelling whether or not two different single surface segments are allophones of one another. In contrast, GŊ alternations could involve a single surface segment deriving from two adjacent underlying alternating segments. While such a two-segment representation was not predicted for modern SG in this section, §7.3 examines Middle High German data and finds that the input during this historical stage does require rejecting the single-segment phonemic /ŋ/ representation.

7.3 Phonemic status of velar nasal in historical data

This section models GŊ learnability during several centuries of Middle High German, establishing that in contrast to modern SG input (§7.2) APL does learn an association between the surface velar nasal segment and alternating velar stops when it is applied to some historical data where this is uncontroversially productive. Modelling learnability during the MHG period will address the relationship between historical sound change and synchronic productivity that has featured in discussion of present-day GŊ.

Changes in GŊ pronunciations and alternations were shaped by a complex interaction of several processes, with subtle differences in their effects across different dialects. For example, the loss of [g] following [ŋ] initially occurred in a restricted medial environment, before gradually expanding to the complete g-loss described in (7.2-3) (Penzl 1968; Wright 1907; Kohrt 1980). The coda devoicing process, besides interacting with g-loss in that devoiced [k] is not deleted, is also affected by the presence of word-final [ə], whose conditioning and productivity (as a phonological and morphophonemic process) likewise change over time (Gress-Wright 2010; Hall 1989; Vennemann 1970). Because of this, as well as the synchronic variation inherent to any language change, it would not be possible to obtain a reasonable simulation of progressive historical stages by systematically applying some well-behaved set of hypothesised transformations backwards from modern child-directed data; see Gress-Wright (2010) for discussion of what can be discerned about the routes from Middle High German to Modern German in various locations. Instead, modelling acquisition of historical language varieties is best done with actual historical corpus data. Data sparsity, and modelling something intended to correspond to a child's vocabulary despite having no source of child-directed input, will be challenges of using historical corpus data rather than reconstruction. However, as in §7.2 and in contrast to some predictions in Chapters 5 & 6, the model's outcome in §7.3.3 is very clear, and appears robust to even a large amount of error or noise. Following the presentation of these results, §7.4 will use them to discuss what APL can contribute to debates of restructuring and GŊ rule reordering.

7.3.1 Data

Historical language data is sourced from the Referenzkorpus mittelhochdeutsch (1050–1350) (ReM), a large corpus of diplomatic transcriptions of Middle High German texts spanning the stated years with uniform searchable orthographic conventions applied. The texts are labelled as belonging to five dialect regions, and annotated with word information including lemma, part of speech, and inflection (case, person, tense, etc.). In total the corpus contains about 2.2 million tokens belonging to about 35,000 lemmas (16,000 of which appear only once), or 75,000 normalised word types (e.g. different inflected forms of the lemmas, like the 13 word forms shown previously in Tables 7.1 and 7.2). This corpus provides sufficient data to obtain a reasonable sample of common words, that may serve as a rough estimate of children's basic language knowledge.

Across all time periods combined, ReM has around 230,000 tokens belonging to 1647 lemmas (or 5919 word types) that might provide relevant data on velar nasals. Lemmas are considered potentially relevant if any of their normalised wordforms contain a velar nasal sequence, or if any of the attested tokens have spellings that imply a velar nasal even if this is not reflected in the normalised wordform. All tokens belonging to such lemmas are included at first, to capture both velar nasals and anything they might in principle alternate with.

To model the phonemic status of GŊ at some historical stages, ReM was divided into four time periods based on the dates (centuries or half-centuries) assigned to each document. The divisions are formed with the following criteria:

11C texts labelled with dates 11, 11;2, 11;2-12;1.

12C texts dated 12, 12;1, 12;1-12;2, 12;2, 12;2-13;1.

13C texts dated 13, 13;1, 13;1-13;2, 13;2, 13;2-14;1.

14C texts dated 14, 14;1, 14;1-14;2, 14,2.

There are only 52 texts in the 11C division, amounting to 62,000 tokens; due to this relatively small amount of data, the 11C texts are not used for modelling. There are 130 12C texts with a total of 400k tokens; 142 13C texts with 800k tokens; and 58 14C texts with 600k tokens. Therefore, it is possible to assess GŊ learnability in these three time periods, although there is not enough data to further divide each time period into the five dialect regions. Needless to say, the tokens of high-frequency words sampled in each broad time period give an extremely rough estimate of the possible vocabulary composition of some child acquiring German in any one of the included geographic locations. The aim of this work is appropriately limited, only aspiring to model an approximation of some historical reality at least well enough to establish that APL is applicable, interpretable, and informative in this language change scenario.

7.3.2 Methods

While pronunciation is impossible to recover from written text, exact phonetic detail is not required in this case; it is only necessary to detect instances of [ŋ] and determine whether or not the [ŋ] is followed by something like [k]. Fortunately, unlike Early Modern and present-day German, spelling in Middle High German is generally considered phonetic enough to identify from the written form when velar stops are present and un/voiced, as 'most or all surface contrasts were orthographically represented' (e.g. spelling *lanc* -[ŋk] for modern standard spelling *lang*; Gress-Wright 2010: 6; Penzl 1968; Vennemann 1970; Wright 1907). Lemmas are therefore coded by spelling as having any of the following three labels:

(7.12) **k-form** Spellings with *nk*, *nc*: assumed to be pronounced [ŋk], a velar nasal followed by a surface velar stop.

g-form Spellings with *ng*: something that perceptibly alternates with [ŋk], assumed to often be [ŋ] with no stop, but if this assumption is incorrect then [ŋg] (for example) would have the same consequences for acquisition as it also perceptibly alternates with [ŋk].

x-form Forms that have no velar nasal.

It is possible for a lemma to receive all three labels, or any combination of them, depending on the tokens that occur. K-forms and g-forms can alternate with one another within a lexeme, but do not have to. X-forms rarely alternate in paradigms with k-forms or g-forms, and instead usually indicate an absence of velar nasal from the stem; for example, *ewigliche* (modern *ewiglich* 'eternal') has variants with n-insertion before the velar stop (*ewinglichen, ewenclichen*), but in most dialects this word is unrelated to velar nasals (*ewiklichen, ewichlichen, ewigliche, ewiliche*).

Implementations of APL for each of the three time periods 12C, 13C, and 14C, follow the no-assimilation lexeme-based model (−AsL) developed in §7.2. This was determined to have the best cognitive justification and most closely resembles how APL has been applied in previous chapters. Also, of the four model types in §7.2 −AsL provides the most challenging scenario to reject H0, as exceptions accumulate less than they would in word-based models while N is higher than in the +As models. Since H0 is rejected anyway for all three historical time periods, this learning outcome can be considered robust to alternative possible modelling assumptions.

For this model in principle, within the sample from each time period the number of items N subject to H0 (7.10) consists of every ReM-defined lemma where at least one form has a spelling indicative of a surface velar nasal, *ng*, *nc*, or *nk*. This matching is maximally inclusive based on the spellings of both the normalised wordforms and all the attested tokens of them. For the lemmas N that seem to have a velar nasal, a cost e should be assessed for each lemma where at least one wordform has an attested token evidently pronounced with a stop following the nasal and at least one other wordform (e.g. a different inflected form) has a token without that stop. These alternations are identified using the k-form and g-form labels defined in (7.12) to indicate tokens with and without a pronounced stop. The example in §7.2.4 of calculating N and e for alternating k-epenthesis data (Table 7.2) with L-F also shows essentially how this is carried out for the historical data, with the k-form and g-form labels in place of ŋ+ and ŋ_.

To most accurately determine relevance and alternation status, all included high-frequency lemmas have been personally reviewed after the automated coding, using all forms and attested tokens of the lemma for context. Lemmas are coded as alternating only if the k-forms and g-forms appear to alternate across different forms of the lexeme; they are not coded as alternating if the stop variation seems to only exist across different dialects/documents within the same wordform, as in this case there is no justification to suppose any child received input in which they alternated. A few lemmas with occasional $-nc-$ spellings that clearly represent something non-velar are entirely excluded from consideration, e.g. *ganz* 'whole' has attested tokens *gancz, gancen*, among *gantz, ganz, ganzzen, gantcem*; the present-day word is pronounced [gants] and in this context the historical $-nc-$ is a recurring spelling variant for an entirely alveolar cluster.

Finally, for each time period 12C, 13C, and 14C, the sample of lemmas included in the model's calculations is restricted to those that are frequent enough and have sufficient data to show their [lack of] alternation status. This is accomplished by keeping only lemmas that have at least 5 total tokens within the given time period, and where those tokens belong to at least two different wordforms, to eliminate an inflated number of 'non-alternating' lemmas represented by a single form that cannot possibly show alternation with itself.

7.3.3 Results

Table 7.5 shows that the data in all three time periods contains frequent alternations between k-forms and g-forms, requiring acquisition of some productive generalisation that relates the two realisations. Although APL does not predict what this generalisation is, external evidence would suggest that it involves an underlying nasal + stop sequence that can account for the predictably conditioned realisations of both g-forms and k-forms.

Due to the dialect mixture in each time period's data, these sources of model input will not directly implement the real acquisition conditions of some particular children several hundred years ago. However, in all three time periods, results indicate rejection of the simple H0 (7.10) by a very large margin. The only way this mixed-dialect data could be responsible for an incorrect model prediction in this case, i.e. incorrectly rejecting (7.10) when in fact it is tolerable within each dialect individually, would be if each dialect had

Table 7.5: **H0 for ŋ in three time intervals of Middle High German** (tolerable when $e < \theta_N$, $\theta_N = N/ln(N)$)

Interval	g-only	c/k-only	Alternating (e)	Total lexemes (N)	θ_N
12C	59	42	63	164	32
13C	87	69	86	242	44
14C	68	59	79	206	38

about half the alternations in the full dataset and did not have the other half. This is implausible, as the different patterns of alternations that have been described in various dialects are mostly supersets/subsets of one another, not partially-intersecting or disjoint sets as this scenario demands. It remains possible that the prevalence of alternations as modelled in §7.3.2 is a less extreme overestimate or superset of the actual alternations for some of the dialects whose texts form part of the data, but the margin by which H0 (7.10) is rejected in Table 7.5 means that a moderate overestimate of alternations would not have changed the ultimate result. Therefore, while no specific claims can be made about any one of the dialect regions on the basis of the mixed historical corpus data, it appears realistic to take this data as an indication that the prevalence of alternations was high enough to reject H0 during the stated time periods.

To help check for possible problematic effects of data sparsity, present-day descendants of the 100 most frequent lexical items in the historical corpus were also identified in CELEX. Then, the equivalent coding (g-forms, k-forms, forms with neither) and alternation status of these modern SG lexemes were determined from all of their listed forms in CELEX. Finally, after analysing the complete paradigms, their coding and alternation status were re-calculated twice by randomly sampling a subset of the CELEX forms for each lexical item, with the number of forms being proportionate to the frequency of the corresponding word in the historical corpus. This simulates the way that missing word forms in the historical data could lead to an underestimation of the real number of alternations, if the limited sample failed to include both variants for some paradigm. However, in both of the randomised limited samples from CELEX, very few words had their apparent alternation status change either time; the error caused by lacking information on all forms of each lemma would not come close to affecting model outcomes with the large margins in Table 7.5. This gives assurance that concerns of sparsity in the historical sample were adequately addressed with the choice to only include words with at least 5 tokens and 2 forms.

7.4 Discussion: Rule reordering and restructuring

A significant part of previous GŊ analysis has focused on building a case study of rule reordering. In this account of modern standard German, as presented in §7.1.1, the language can maintain an inherited phonemic inventory, same lexical forms, and same set of rules, operating in a different order than they historically did. 'Rule reordering' can now be understood in a very general sense; for example, the relevant grammars might be described in terms of ordered generative transformation rules, but could equally be described in an Optimality Theory framework. In that case, the comparison of a 'reordering' analysis with a restructuring analysis may involve a hypothesised historical re-ranking of constraints (whose effect generates the same output as the transformative rules) without lexicon or phoneme changes, compared with a grammar in which the modern [ŋ] output is faithful to an underlying velar-specified representation /ŋ/. Regardless, APL provides a cognitively motivated alternative grammar showing that the language varieties conventionally described by a reordered grammar with obligatory two-segment representation of [ŋ] can also be described by a grammar that assumes the language has undergone some phonological restructuring. This restructuring analysis implies a process of historical change in which an output of a former allophonic rule became memorised as underlying phoneme in its own right, adding a third nasal phoneme /ŋ/ alongside original /n/ and /m/ in German with corresponding changes to lexical forms where /ŋ/ replaced inherited /ng/ sequences. The clear prediction in §7.2 is that the restructured grammar must be what is learned from the present-day input.

Since the two-segment analysis of modern GŊ has conventionally been used to support the existence of rule reordering as a mechanism of language change, or even demonstrate the operation of extrinsically ordered phonological rules in the first place (Kohrt 1980), the result of APL favouring an analysis in which the language has instead undergone restructuring weakens the status of this evidence. The argument for GŊ language change by rule re-ordering is broadly that a modern language variety must be a possible descendent of the historical variety it derives from; there is a historical stage of German that can be explained by a certain set of phonemes and ordered rules; there is a modern descendent variety that can be be described with the same phoneme inventory and rules by applying the rules in a different order; and therefore a historical process of rule-reordering accounts for the transition from the former to the current language stage. Reordering-based accounts of GŊ are largely missing a compelling idea of what caused the rule-order changes towards modern dialects, or what intermediate stages the language could have passed through that would cause children to learn a new order of rule application and not a new phoneme /ŋ/; it is not clear how the motivations cited in many other proposed examples of rule re-ordering, such as simplifying either the stored structure of the language or the derivational computations of using it, would be relevant in this particular case. There are some attempts to find explanations or causes for GŊ rule reordering, which tend to either rely on unrealistic assumptions about the input environment and abstract away from dialect variation, or else seem to first assume that rule reordering happened (e.g. reject grammars with a descriptively superfluous /ŋ/ phoneme as an undesirable account of the language) and then seek plausible motivation for it given this assumption (Gress-Wright 2010; Kiparsky 1982; Kohrt 1980). Therefore, the rule re-ordering argument mostly stands on the observation that if re-ordering happened then its outcome would look like present-day Standard German. This observation is accurate, but it is only a compelling argument for the actual historical occurrence of re-ordering to the extent that there is no other process whose outcome could also look like present-day Standard German. However, there is an alternate description of present-day German that does not include inherited rules in a novel order, and cognitive modelling based on synchronic learnability gives strong support to this alternate analysis, suggesting that there is a modern descendent language without re-ordered rules but with a restructured phoneme inventory. Therefore, German velar nasal phenomena cannot serve as an adequate argument for the existence of rule reordering as a mechanism of language change.

This is not at all a claim that rule-reordering in general does not or cannot happen: only that evidence for it must be found elsewhere. Although it is not a concern of this dissertation to address the overall possibility of rule reordering in language change, intuitively it seems that as long as children can learn extrinsically ordered rules (or ranked constraints, etc.) in the first place then they can presumably learn a reordered set just the same, since learners have no special access to the mental rule ordering(s) that produced their input data (Kiparsky 1982). The predictions of APL claim only that German [ŋ] specifically does not provide compelling evidence about rule reordering, because in this case the reordered grammar is neither the only nor the most reasonable account of the language varieties it is meant to describe.

Beyond casting doubt on an example of rule reordering, APL makes more concrete and generalisable contributions to understanding restructuring as a mechanism of historical change. As discussed in the introduction to this chapter, restructuring is a 'silent' process at the time it occurs, in that by definition it involves a significant qualitative change in underlying representations without corresponding change in the surface language generated by the grammar. Therefore, it is only possible to see direct evidence that it has happened well after the fact, if there have been subsequent changes made possible by the restructuring; beyond this linguists must generally use holistic judgement of individual cases to try to assess when or whether restructuring happens.

APL yields precise quantitative conclusions about restructuring from a general acquisition model that predicts phonemic/allophonic status based on any input, because it can indicate which extensionally equivalent grammars are or are not learnable without resorting to case-specific subjective judgement. This proposes a solution to the long-standing issue that 'an attempt of quantitative or qualitative weighing of conflicting evidence from different domains of external evidence might easily lead to an impasse, since no canon of argumentation has been established for such a case' (Dressler 1979: 459, writing specifically about the representation of German velar nasals). In the APL framework, restructuring emerges when models applied to successive generations or stages of input have different phoneme-learning outcomes. APL is therefore capable of predicting points of categorical grammar change even where input is only slightly or gradiently different, 'sharpening the synchronic theory so that it will provide the right basis for diachronic explanation' (Kiparsky 1982: 57). As an account of historical change, this explains both when and why children would learn a different grammar than the one(s) that produced their input. It links the intuitions behind the proposed mechanism of restructuring, that 'If there is no synchronic justification for retaining [sound change] rules in the grammar, the changes which they effect are simply incorporated in the base forms and the rules disappear from the language' (Kiparsky 1982: 51), to an empirically-motivated measurable criterion for when input does and does not contain such justification.

Chapter 8

Conclusion

This dissertation has proposed a model of Alternation-sensitive Phoneme Learning, describing how input shapes children's acquisition of abstract representations. This models allophone learning as a function of observed alternations, illuminating the emergence of a qualitative change in children's abstract representations based on a quantitative description of their experience with concrete stimuli. APL yields predictions of a critical factor in the shape of a phonemic inventory: whether surface-distinct segments are underlyingly contrastive or allophonic.

A single model without free parameters is able to account for a variety of learning outcomes, based on quite general principles of child language acquisition. It gives quantitative insight on a scale from individuals' development to change across generations, and is a concretely defined response to one of 'the most central problems in historical linguistics, e.g. when does restructuring take place?' (Kiparsky 1982: 57). Working within this framework requires only initially projecting some set of surface contrasts as underlying contrasts, retaining this grammar (however it is represented) as long as tolerable, or rejecting it if motivated by reaching an input threshold for obviously incompatible surface alternations.

8.1 Summary of contributions

Chapters 1 and 2 justified the utility of developing APL. Allophony, as operationally defined in this dissertation, involves language-specific representations that must be acquired through experience and subsequently play an active role in basic language perception and cognition. The predominant view in current research focuses on two elements of this aspect of acquisition, a well-studied stage of phonetic category acquisition and a much less studied stage of learning how the surface segments are organised into a system of abstract phonemic categories. APL is a proposal about this second abstract phoneme-building stage, and is flexible about possible theories and models of the previous and following parts of first language acquisition.

Chapter 3 sets out the APL model proposal, justifying its components. These include an initial highly concrete phoneme system following the Invariant Transparency Hypothesis, along with a description of how lexical, semantic, and morphological development can lead learners to generalise about systematic correspondences between surface-distinct segments. The Tolerance Principle is used for quantitative evaluation of when such generalisations take place. This chapter also considers how APL as a general model may advance knowledge in linguistics, including which aspects of the acquisition process it does and does not explicitly model, and the range of phonological phenomena whose acquisition it might model.

Chapter 4 uses a speech corpus to investigate how children's productions change throughout the course of allophone acquisition. This longitudinal study provides crucial empirical evidence for the model. Firstly, the children's productions show U-shaped or non-monotonic learning of English alveolar flap allophony. This appears as an early stage of increasingly successful target production, followed by a reduction in medial flap target accuracy and the introduction of novel errors. Intervocalic [ɾ] is shown to develop differently in word-medial and word-final positions, indicating a phonological (not purely phonetic or articulatory) component to the non-linear development, and the large sample of tokens in this study identified flap productions at younger ages than previously reported. U-shaped learning is a characteristic signal of learning rules or generalisations in many domains of language, but has not previously been demonstrated this thoroughly in allophone learning; therefore, this documentation supports APL's assumption that allophony involves making abstract generalisation. Furthermore, beyond the qualitative result of identifying U-shaped allophone learning, this study provides data to be used for precise validation of how APL models the relationship between a learner's vocabulary and their phonological knowledge as both develop over time.

Chapter 5 presents the first of three case studies applying APL to different scenarios. In this chapter, APL accurately traces the time course of U-shaped learning for medial [ɾ]. While Chapter 2 discussed how acquisition of allophonic generalisations influences language processing in ways that can be seen in a wide variety of controlled experiments, it would be impractically expensive and time-consuming to try to run longitudinal repetitions of these studies with small children to try to detect when allophone acquisition occurs; instead, the production data from Chapter 4 contains more easily observable correlates of allophone learning which can provide a useful signal for model evaluation. The model is validated using data from four children, predicting each of their development by initialising with a default H0 that lacks abstract correspondences between surface segments and iteratively evaluating this hypothesis with the Tolerance Principle. As each child's vocabulary accumulates items with multiple surface forms, alternations become prevalent enough to predict rejection of H0 in favour of some generalisation about the alternating segments, and the time of this prediction for each child is shown coincide reliably with the time of their reduction in medial flap production accuracy. In addition, predictions based on aggregated child-directed input are broadly compatible with the general developmental course indicated by previously published group-level studies as well as the new data of Chapter 4; this provides assurance that although predictions derived from the vocabularies of individual children have an especially close correspondence with these children's own observable development, APL can also give useful and informative results even in less ideal data conditions.

Chapter 6 presents this dissertation's second case study, using APL to understand the progression of secondary split in Menominee back vowels. This chapter builds on the previous study's evaluation of the framework, now taking APL as a reliable model of language acquisition and therefore applying it in a scenario with much more uncertainty about the primary language data. APL accounts for the phonemic split between the high and mid back vowels as a regular outcome of child language acquisition, when independent changes affect the appearance of conditioned allophony in the vocabulary. This case study also serves to illustrate the range of possible outcomes of APL and how the model is sensitive to the nature of different input. The Pre-split model variant predicts that Menominee allophonic back vowel raising is learnable when short vowels provide clear triggers for this process, while the *Long-vowel and Split variants show that this process is not likely to be acquired without short vowel cues. A drawback of this case study is the limited availability of relevant language data; therefore, it may be most informative not as a source of any new insight on the Menominee language itself, but rather as an illustration of how to work with and interpret the model in similar cases.

Chapter 7 presents a case study of the representation of velar nasals in German. This case uses APL to offer a learnability-based perspective for evaluation of extensionally equivalent competing analyses, and this is applied to both modern and historical language varieties. Here, APL provides an alternative to the conventional way that 'decisions of a sort that frequently arise in phonology – for example, between adding a segment to the inventory of phonemes and adding a phonological rule to the grammar – are made on the basis of simplicity' (Kiparsky 1982: 85). Since the interpretation of modelling outcomes for the German data is quite unambiguous, this study most strongly brings out theoretical commitments and consequences of APL, including failure to acquire productive allophony for segments that occur in predictable contexts if the segments do not directly alternate.

Finally, this chapter will review possible drawbacks of the current work, as well as how the APL framework could be extended in the future.

8.2 Limitations

Some limitations to the work in this dissertation derive from the particular language data that was used for modelling, or specific decisions that had to be made without complete information on linguistic theory, as necessary to build concrete testable implementations. For example, Chapters 4 and 5 include a corpus study of only four children's English alveolar flap development; while APL is able to account for each child's data, this sample is not large enough to generalise with confidence to North American English-learners overall. Chapter 6 was unable to incorporate any information on Menominee word frequency beyond a binary distinction on whether or not a word was considered common enough to include in a particular dictionary; based on the differences in the frequencies of certain vowels in this dictionary compared to the naturalistic speech sample of Cudworth (2019), it is certainly possible that the dictionary is not a particularly representative sample of vowel segment distributions in speakers' vocabularies. The historical modelling in Chapter 7 was able to use German corpus data, albeit from written rather than spoken sources, but the data does not provide an accurate picture of any particular learners' language input because the model's input was constructed by combining corpus data across different regions as well as 50-year time periods.

In addition, APL in general is designed to predict a specific point in language development and remain as agnostic as possible about the rest of the acquisition process. This has the consequence that in any particular implementation, some details about child language acquisition must be filled in a bit arbitrarily. For example, different methods were used in each case study to decide which words and morphemes learners might understand as corresponding to one another, and therefore which surface alternations a learner is aware of. Chapter 5 used a detailed holistic decision procedure that relies on personal intuition to evaluate a child's productions according to the criteria described in (5.3-9); this is an attempt to construct the most accurate possible model inputs, but is neither objective nor perfectly reproducible if applied to additional data. Chapter 6 likewise relied on personal intuition, combined with information from decisions made by the authors of a dictionary oriented towards intermediate second language learners. However, in this case there is much less language/speaker data available to base each judgement on compared to Chapter 5, and the intuitions are those of a linguist who does not speak the language in question (nor any other language with analogues to some of the relevant Menominee morphophonological processes). Therefore, there is a higher chance for errors or inaccurate model input, although the chapter discussion showed a possible way to incorporate concrete detail from acoustic data to hopefully make APL applications more informative in marginal or ambiguous cases like this. Finally, Chapter 7 used a fully reproducible method of categorising wordforms into related and distinct lexical items, by simply taking this information from the pre-existing CELEX lexicon. However, this method is not available for many data sources, and it cannot be treated as a testable or informative element of a cognitive model unless a language variety happens to have an available lexicon resource that was constructed according to some specific well-described theory about language cognition. The different strategies used in Chapters 5-7 indicate an opportunity for further development of APL, by integrating principled theories or models of young children's lexicons.

There are some other limitations of the APL framework itself, which has been proposed for the first time in this dissertation and may still be considered under development. For example, the scope of APL is not yet fully clear. This framework describes learning with the Tolerance Principle in a similar way to morphological and syntactic acquisition, and therefore, it is possible that this model is especially applicable for phonemes whose allophones are phonetically dissimilar and/or learned fairly late. These are both characteristics of the English alveolar flap allophone system, which was used to validate the model; in addition, the historical case studies in which Menominee and German gained new phonemes also most likely involve rather late allophone acquisition during the critical period of change in each language's history, as discussed in §6.1.3. Considering this, it remains possible that some other mechanism may have an important role during the acquisition of phonemes that are acquired relatively earlier when their allophones have greater phonetic similarity for learners to exploit.

8.2.1 Asymmetry in predictive ability

The most salient limitation of the APL framework is probably an asymmetry in which the model gives a more complete description of phonemic contrast (failure to acquire allophones) than productive allophony. APL is applicable in a strictly defined window of allophone acquisition: up until the point when a learner first realises there is systematic allophony in their language. What happens next is not specified in the model; therefore, APL makes a concrete prediction if the threshold for learning allophony of two segments is not reached, but if the threshold is reached the model cannot fully specify the shape of the resulting phoneme inventory. Even with this limitation, the model has been able to make relevant empirical predictions on a range of phenomena, but the inability to better define the nature of allophonic grammars restricts its scope.

The asymmetry arises because a learner rejecting their early surface-based grammar learns that at least some instances of a certain segment have to be underlyingly something else, but this does not mean that all known instances of the segment must be reanalysed as underlyingly something else. Perhaps a fully allophonic grammar will be learned, with the size of the phonemic inventory reduced, but perhaps some instances of the segment will be interpreted as allophonic outputs while others still receive a simple surface-based interpretation. Given segments x and y in complementary distribution with enough alternations to learn that at least sometimes [y] = /x/, the model cannot necessarily distinguish between the grammar (8.1a) vs. the grammar (8.1b) (in which, for example, [y] in paradigms where it alternates with [x] may tend to be understood as /x/ while [y] in non-alternating paradigms may tend to be interpreted as /y/).

(8.1) a. [x] = /x/ and [y] = /x/

b. [x] = /x/ and some [y] = /y/ and some y = /x/

The uncertainty of the APL model at this point may well reflect underdetermination or individual variation in reality, but even so, there are quite different expectations for the possible historical development when a language contains only /x/ with realisations [x], [y], vs. when a language may contain /n/ and /m/ with surface segments [n], [m] where [m] is an allophone of both /m/ and /n/ (as in English, with phonemic contrast shown by *moon*∼*noon* but allophony shown in place assimilation of [n] in *rain* to [m] in *rainbow*). However, this modelling uncertainty is common to acquisition of other types of linguistic generalisations, which could be useful to consider in parallel to phoneme learning. For example, by the time a child learns enough English regular past-tense verbs to generalise about the dental suffix, they have previously memorised many such forms; the fact that the child could re-analyse their lexicon with the new productive generalisation does not confirm whether they actually do so exhaustively, or whether they potentially retain some memorised forms for any period of time. Further research on lexical access, along the lines of studies described in §3.3, may also indicate when words are reanalysed to be stored with one stem and when (perhaps for the most frequent or earliest-learned words) they may retain multiple forms in the lexicon (Bybee 2006; Hay 2001).

8.3 Future directions

Above all, APL is proposed as a framework for evaluating learnability of linguistic generalisations given learners input. This follows the general representation-learning strategy of starting with a concrete grounding in the input data, and building more complex non-surface structure just when required by failure of the more concrete representation. As developed here, APL builds in specific theories (Invariant Transparency Hypothesis and Tolerance Principle) to specify a formal model in this framework. However, the basic way it operates would permit substitution of a wide variety of other proposals about language learning. The primary goal in proposing APL is for it to be useful, not correct; it may contain incorrect assumptions and therefore give inaccurate results, but as long as the predictions are clear enough to be able to tell when they are wrong and in what way, the model could illuminate something about language or acquisition. Variations or revisions to model components may be expected to retain APL's ability to apply the same model both to track the course of acquisition in individuals, and to track outcomes for a population of individuals which may change over time as input shifts; this characteristic is what makes APL both empirically testable in practise and informative in a wide range of historical scenarios.

In addition, APL can be adapted to explicitly represent language variation in which an individual learner's input is a probabilistic function of a mixture of grammars in their community. In this context young children initially tend to regularise their lexicon according to the form they hear most frequently, so a distribution over learners can be predicted by applying the same model to a population of different probabilistically-generated lexicons (Hudson Kam and Newport 2009; Kodner and Richter 2020). The APL framework would likewise be adaptable to alternative theories of how individuals acquire a lexicon in scenarios of input variation; Chapter 4 justifies the application of the model to individuals once given some lexicon. Considering the general relationship between language variation and change, the ability to model mixed-grammar populations rather than only unrealistic homogeneous input environments is a significant benefit.

Finally, something like the APL framework might be extensible beyond phoneme learning. This is because the essential features of an approach extending APL are (i) an early default hypothesis, presumably with straightforward relation to surface input, and (ii) iterative evaluation of the initial hypothesis, rejecting it if the input requires something more abstract. For example, it could be informative to examine the Case Directionality Hypothesis (CDH; Eythórsson 2009; Eythórsson and Thráinsson 2017; Yip et al. 1987) with such an approach. The CDH is a diachronic observation based primarily in data from several Scandinavian languages, stating that 'quirky' (idiosyncratic) lexical case can be replaced over time by thematic (semantically predictable) lexical case, and lexical case can be replaced by structural case, while historical change does not occur in the other direction of this hierarchy. A fundamental part of the three-part distinction between quirky, thematic, and structural case centres around the different nature of exceptions for each type of case (Eythórsson et al. 2012; Jónsson and Eythórsson 2011). Since the TP is already known to apply to many levels of linguistic structure, the major considerations for implementation of a new model along the lines of APL would be how a 'surface-based default' grammar may be adapted into the domain of case, and identification of relevant subsets of the language in which to (attempt to) generalise about thematic or structural case. To study CDH, default hypotheses may be suggested by the handling of exceptionality at each level involved in the hierarchy, and ideally, the directions of change expected by the CDH might then emerge from the model and be accounted for by very general assumptions about the acquisition of productive generalisation (TP). However, even if it is not possible to account for CDH in a model parallel to APL, the reasons for failure should be informative about the relationship of case acquisition and change, and contribute to an explanation for the scope of APL. For example, it is possible that there are not sensible analogues of ITH at various other levels of language, like case; the acquisition strategy of starting with (and by default keeping) a surface-based representation might be part of phoneme learning precisely because the 'surface-based' categories here come directly from acoustic properties of the speech signal, whereas linguistic categories that are less physically grounded might require a different approach to learning. Additionally, if the framework does support a reasonable account of CDH phenomena, then its benefits of distinguishing between tolerable and learnable grammars may also be relevant beyond phoneme inventory learning.

Bibliography

Albright, A. and Hayes, B. (2003). Rules vs. analogy in english past tenses: A computational/experimental study. *Cognition*, 90(2):119–161.

Archangeli, D. B. and Pulleyblank, D. G. (1994). *Grounded phonology*, volume 25. MIT Press.

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). The celex lexical database (release 2). *Distributed by the Linguistic Data Consortium, University of Pennsylvania.*

Bekemeier, N. (2016). *On the Representation and Processing of Phonological Stem Variants of Complex Words*. Universität Konstanz.

Benua, L. (1997). *Transderivational identity.* PhD thesis, University of Massachusetts Amherst.

Bergelson, E. and Swingley, D. (2015). Early word comprehension in infants: Replication and extension. *Language Learning and Development*, 11(4):369–380.

Berger, A. L., Pietra, V. J. D., and Pietra, S. A. D. (1996). A maximum entropy approach to natural language processing. *Computational linguistics*, 22(1):39–71.

Bergmann, C. and Cristia, A. (2016). Development of infants' segmentation of words from native speech: A meta-analytic approach. *Developmental Science*, 19(6):901–917.

Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., and Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child development*, 89(6):1996–2009.

Bermúdez-Otero, R. (2016). We do not need structuralist morphemes, but we do need constituent structure. *Morphological metatheory*, pages 385–428.

Berwick, R. C. and Weinberg, A. S. (1982). Parsing efficiency, computational complexity, and the evaluation of grammatical theories. *Linguistic Inquiry*, pages 165–191.

Bloomfield, L. (1962). *The Menomini Language*, volume 1. Yale University Press.

Bloomfield, L. (1975). *Menominee lexicon.* Milwaukee Public Museum Press.

Boersma, P., Escudero, P., Hayes, R., et al. (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. In *Proceedings of the 15th international congress of phonetic sciences*, volume 1013, page 1016. Barcelona.

Boomershine, A., Hall, K. C., Hume, E., and Johnson, K. (2008). The impact of allophony versus contrast on speech perception. In *Contrast in phonology*, pages 145–172. De Gruyter Mouton.

Braginsky, M., Yurovsky, D., Marchman, V. A., and Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, 3:52–67.

Braver, A. (2014). Imperceptible incomplete neutralization: Production, non-identifiability, and non-discriminability in American English flapping. *Lingua*, 152:24–44.

Brown, R. (1973). The order of acquisition. In *A first language: The early stages*, pages 254–293. Harvard U. Press.

Buckley, E. (2000). On the naturalness of unnatural rules. In *Proceedings from the Second Workshop on American Indigenous Languages. UCSB working papers in linguistics*, volume 9, pages 1–14.

Burrows, L., Jarmulowicz, L., and Oller, D. K. (2019). Allophony in english language learners: the case of tap in english and spanish. *Language, speech, and hearing services in schools*, 50(1):138–149.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, pages 711–733.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, 15(1-2):39–54.

Calamaro, S. and Jarosz, G. (2015). Learning general phonological rules from distributional information: A computational model. *Cognitive science*, 39(3):647–666.

Capdevila i Batet, M. and Llinàs i Grau, M. (1995). The acquisition of negation in english. *Atlantis*, pages 27–44.

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., and Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, 110(28):11278–11283.

Cazden, C. B. (1968). The acquisition of noun and verb inflections. *Child development*, pages 433–448.

Chambers, K. E., Onishi, K. H., and Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, 87(2):B69–B77.

Chomsky, N. (1957). *Syntactic Structures*. Mouton, The Hague.

Chomsky, N. (1965). *Apects of the theory of syntax*. MIT Press.

Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English*. Harper & Row, New York.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of german inflection. *Behavioral and brain sciences*, 22(6):991–1013.

Clahsen, H., Eisenbeiss, S., Hadler, M., and Sonnenstuhl, I. (2001a). The mental representation of inflected words: An experimental study of adjectives and verbs in German. *Language*, 77(3):510–543.

Clahsen, H., Felser, C., Neubauer, K., Sato, M., and Silva, R. (2010). Morphological structure in native and nonnative language processing. *Language Learning*, 60(1):21–43.

Clahsen, H., Prüfert, P., Eisenbeiss, S., and Cholin, J. (2001b). Strong stems in the German mental lexicon: Evidence from child language acquisition and adult processing. In *More than words. Festschrift for Dieter Wunderlich*, pages 91–112.

Covington, M. A. and McFall, J. D. (2010). Cutting the gordian knot: The moving-average type–token ratio (mattr). *Journal of quantitative linguistics*, 17(2):94–100.

Creel, S. C. (2012). Phonological similarity and mutual exclusivity: on-line recognition of atypical pronunciations in 3–5-year-olds. *Developmental Science*, 15(5):697–713.

Cristia, A. (2018). Can infants learn phonology in the lab? a meta-analytic answer. *Cognition*, 170:312–327.

Cudworth, A. (2019). Menominee vowels: a first acoustic analysis. *Linguistic Society of America Annual Meeting*.

Cui, A. (2020). *The Emergence of Phonological Categories.* PhD thesis, University of Pennsylvania.

Davis, S., Berkson, K., and Strickler, A. (2020). Unlocking the mystery of dialect b: A note on incipient/ai/-raising in fort wayne, indiana. *American Speech*, 95(2):149–172.

Davis, S. and Van Summers, W. (1989). Vowel length and closure duration in word-medial vc sequences. *Journal of Phonetics*, 17(4):339–353.

De Jong, K. (1998). Stress-related variation in the articulation of coda alveolar stops: Flapping revisited. *Journal of Phonetics*, 26(3):283–310.

De Jong, K. (2011). Flapping in American English. *M. van Oostendorp & CJ Ewen & E. Hume & K. Rice (Eds.), Blackwell Companion to Phonology. Oxford*, pages 2711–2729.

de Villiers, J. G. and de Villiers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of psycholinguistic research*, 2(3):267–278.

Demuth, K., Culbertson, J., and Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of english. *Language and Speech*, 49(2):137–173.

Dillon, B., Dunbar, E., and Idsardi, W. (2013). A single-stage approach to learning phonological categores: Insights from Inuktitut. *Cognitive science*, 37(2):344–377.

Dinnsen, D. A. and Gierut, J. A. (2008). Optimality theory: A clinical perspective. *The handbook of clinical linguistics*, page 439.

Dressler, W. U. (1979). External evidence for an abstract analysis of the german velar nasal. *Wiener Linguistische Gazette Wien*, 19:3–28.

Eddington, D. (2007). Flaps and other variants of /t/ in American English: Allophonic distribution without constraints, rules, or abstractions. *Cognitive Linguistics*, 18:234–46.

Eddington, D. and Channer, C. (2010). American english has goʔ a loʔ of glottal stops: Social diffusion and linguistic motivation. *American speech*, 85(3):338–351.

Eddington, D. E. and Brown, E. K. (2020). A production and perception study of/t/-glottalization and oral releases following glottals in the united states. *American Speech: A Quarterly of Linguistic Usage*, 96(1):78–104.

Ervin, S. M. and Miller, W. R. (1963). Language development. *Child psychology: The sixty-second yearbook of the National Society for the Study of Education, Part 1.*

Eythórsson, T., Johannessen, J. B., Laake, S., and Åfarli, T. A. (2012). Dative case in norwegian, icelandic and faroese: Preservation and non-preservation. *Nordic Journal of Linguistics*, 35(3):219–249.

Eythórsson, T. and Thráinsson, H. (2017). Variation in oblique subject constructions in insular scandinavian. *Syntactic Variation in Insular Scandinavian*, 1:53.

Eythórsson, T. (2009). *Frændafundur 6: fyrilestrar frá føroyskari-íslendskari ráðstevnu í Tórshavn 26.-28. juni 2007*, volume 49, chapter Stöðugleiki og breytingar í fallaferfi norrænu eyjamálanna. Faroe University Press.

Feldman, N., Myers, E., White, K., Griffiths, T., and Morgan, J. L. (2011). Learners use word-level statistics in phonetic category acquisition. In *Proceedings of the 35th boston university conference on language development*, pages 197–209.

Fergadiotis, G., Wright, H. H., and Green, S. B. (2015). Psychometric evaluation of lexical diversity indices: Assessing length effects. *Journal of Speech, Language, and Hearing Research*, 58(3):840–852.

Ferguson, C. A. and Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, pages 419–439.

Fertig, D. L. (1996). Phonology, orthography and the umlaut puzzle. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, pages 169–184.

Fikkert, P. and Levelt, C. (2008). How does place fall into place? *Contrast in phonology: Theory, perception, acquisition*, pages 231–268.

Fleischhauer, E. (2013). *Morphological processing in children: an experimental study of German past participles*. Universität Potsdam.

Fort, M., Brusini, P., Carbajal, M. J., Sun, Y., and Peperkamp, S. (2017). A novel form of perceptual attunement: Context-dependent perception of a native contrast in 14-month-old infants. *Developmental cognitive neuroscience*, 26:45–51.

Foulkes, P. and Vihman, M. (2013). First language acquisition and phonological change. In *The Oxford Handbook of Historical Phonology*. Oxford University Press.

Fox, R. and Terbeek, D. (1977). Dental flaps, vowel duration and rule ordering in American English. *Journal of Phonetics*, 5:27–34.

Fukaya, T. and Byrd, D. (2005). An articulatory examination of word-final flapping at phrase edges and interiors. *Journal of the international phonetic association*, 35(01):45–58.

Gerken, L. and McIntosh, B. J. (1993). Interplay of function morphemes and prosody in early language. *Developmental psychology*, 29(3):448.

Goodman, J. C., Dale, P. S., and Li, P. (2008). Does frequency count? parental input and the acquisition of vocabulary. *Journal of child language*, 35(3):515–531.

Gress-Wright, J. (2010). *Opacity and transparency in phonological change.* PhD thesis, University of Pennsylvania.

Guenther, F. H. and Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *The Journal of the Acoustical Society of America*, 100(2):1111–1121.

Hall, T. (2010). Nasal place assimilation in emsland german and its theoretical implications. *Zeitschrift für Dialektologie und Linguistik*, pages 129–144.

Hall, T. A. (1989). German syllabification, the velar nasal, and the representation of schwa. *Linguistics*, 27(5):807–842.

Hanson, H. M. and Shattuck-Hufnagel, S. (2011). An acoustic study of word-medial stop consonants in 2-3 year-old speakers of american english. In *ICPhS*, pages 819–822.

Hart, B. and Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American educator*, 27(1):4–9.

Haugen, E. (1938). Notes on “voiced t” in american english. *Dialect Notes*, 6(16-17):627–634.

Hay, J. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics*, 39(6; ISSU 376):1041–1070.

Hay, J. and Foulkes, P. (2016). The evolution of medial/t/over real and remembered time. *Language*, 92(2):298–330.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies.* University of Chicago Press.

Hayes, B. (2004). Phonological acquisition in optimality theory: the early stages. In *Constraints in phonological acquisition*, pages 158–203. Cambridge University Press.

Hayes, B. and Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic inquiry*, 39(3):379–440.

Herd, W., Jongman, A., and Sereno, J. (2010). An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. *Journal of Phonetics*, 38(4):504–516.

Hoenigswald, H. M. (1960). *Language change and linguistic reconstruction.* University of Chicago Press.

Hohne, E. A. and Jusczyk, P. W. (1994). Two-month-old infants' sensitivity to allophonic differences. *Perception & Psychophysics*, 56(6):613–623.

Hudson Kam, C. L. and Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive psychology*, 59(1):30–66.

Issatschenko, A. (1963). Der phonologische status des velaren nasals im deutschen. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 16:77–84.

Itô, J. and Mester, A. (1999). The phonological lexicon.

Janda, R. D. (1999). Accounts of phonemic split have been greatly exaggerated—but not enough. In *Proceedings of the International Congress of Phonetic Sciences*, volume 14, pages 329–332.

Jarosz, G. (2019). Computational modeling of phonological learning. *Annual Review of Linguistics*, 5:67–90.

Johnson, M. (1984). A discovery procedure for certain phonological rules. In *10th International Conference on Computational Linguistics and 22nd Annual Meeting of the Association for Computational Linguistics*.

Jónsson, J. G. and Eythórsson, T. (2011). Structured exceptions and case selection in insular scandinavian. *Expecting the unexpected: Exceptions in grammar*, 216:213.

Kehoe, M. (2015). Acquisition of velars: A whole-word approach. *Clinical linguistics & phonetics*, 29(12):873–908.

Kimper, W. A. (2011). *Competing triggers: Transparency and opacity in vowel harmony.* PhD thesis, University of Massachusetts Amherst.

Kiparsky, P. (1982). *Explanation in phonology.* Walter de Gruyter.

Kiparsky, P. (2015). Phonologization. In *The Oxford Handbook of Historical Phonology.* Oxford University Press.

Klánová, A. (2016). Sociophonetic study of substitutional glottalization in native english speakers.

Klein, H. B. and Altman, E. K. (2002). The acquisition of medial /t,d/ allophones in bisyllabic contexts. *Clinical linguistics & phonetics*, 16(3):215–232.

Klein, H. B., Altman, E. K., and Tate, J. (1998). Relationship between perceptual and acoustic judgments of children's flap productions. *Contemporary Issues in Communication Sciences and Disorders*, 25:25–31.

Ko, E.-S. (2007). Acquisition of vowel duration in children speaking american english. In *Eighth Annual Conference of the International Speech Communication Association.*

Kodner, J. and Richter, C. (2020). Transparent /aɪ/-raising as a contact phenomenon. *University of Pennsylvania Working Papers in Linguistics*, 25(2):8.

Kohrt, M. (1980). Rule ordering, rule extension, and the history and present state of/ng/-clusters in german. *Linguistics*, 18(9-10):777–802.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044):606–608.

Labov, W. (1994). Principles of language change: Internal factors.

Labov, W. (2011). *Principles of linguistic change, vol. 3: Cognitive and cultural factors*, chapter 1.6 The community orientation of language learning. John Wiley & Sons.

Lahiri, A. and Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38(3):245–294.

Leminen, A., Lehtonen, M., Bozic, M., and Clahsen, H. (2016). Morphologically complex words in the mind/brain. *Frontiers in human neuroscience*, 10:47.

Lewis, M., Cristiano, V., Lake, B. M., Kwan, T., and Frank, M. C. (2020). The role of developmental change and linguistic experience in the mutual exclusivity effect. *Cognition*, 198:104191.

Liberman, M. (1984). Intonational invariance under changes in pitch range and length. In *Language sound structure*. MIT Press.

Liittschwager, J. C. and Markman, E. M. (1994). Sixteen-and 24-month-olds' use of mutual exclusivity as a default assumption in second-label learning. *Developmental Psychology*, 30(6):955.

Linares, R. E., Rodriguez-Fornells, A., and Clahsen, H. (2006). Stem allomorphy in the Spanish mental lexicon: Evidence from behavioral and ERP experiments. *Brain and Language*, 97(1):110–120.

Liu, L., Peter, V., Ong, J. H., and Escudero, P. (2020). Revisiting infant distributional learning using event-related potentials: Does unimodal always inhibit and bimodal always facilitate? *10th International Conference on Speech Prosody.*

Luce, P. A., McLennan, C. T., and Chance-Luce, J. (2003). Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming. *Rethinking implicit memory*, pages 197–214.

Macaulay, M. and Menominee Tribe of Wisconsin (2012). Menominee dictionary ‘intermediate’. *http://www.menominee-nsn.gov/governmentpages/Department.aspx?departmentID=1*.

Macken, M. A. (1995). Phonological acquisition. In *The handbook of phonological theory*, pages 1–23. Blackwell Handbooks in Linguistics, Blackwell Publishers.

MacWhinney, B. (2000). *The CHILDES project: The database*, volume 2. Psychology Press.

Major, D. (1974). *The acquisition of modal auxiliaries in the language of children*. De Gruyter Mouton.

Malécot, A. and Lloyd, P. (1968). The/t:/d/distinction in american alveolar flaps. *Lingua*, 19(3-4):264–272.

Mangold, M. (1990). *Duden" Aussprachewörterbuch": Wörterbuch der deutschen Standardaussprache*. Duden-Verlag.

Markman, E. M. and Wachtel, G. F. (1988). Children’s use of mutual exclusivity to constrain the meanings of words. *Cognitive psychology*, 20(2):121–157.

Markman, E. M., Wasow, J. L., and Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive psychology*, 47(3):241–275.

Marshall, J. (2003). The changing sociolinguistic status of the glottal stop in northeast scottish english. *English world-wide*, 24(1):89–108.

Martin, A., Peperkamp, S., and Dupoux, E. (2013). Learning phonemes with a proto-lexicon. *Cognitive Science*, 37(1):103–124.

Mattys, S. L. and Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, 78(2):91–121.

Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3):B101–B111.

McAllister Byun, T. and Tessier, A.-M. (2016). Motor influences on grammar in an emergentist model of phonology. *Language and Linguistics Compass*, 10(9):431–452.

Menn, L. (1971). Phonotactic rules in beginning speech: a study in the development of english discourse. *Lingua*, 26(3):225–251.

Menn, L. and Vihman, M. (2011). Features in child phonology. *Where do phonological features come from*, pages 261–301.

Milligan, M. (2000). A new look at menominee vowel harmony. *Algonquian Papers-Archive*, 31.

Milligan, M. (2005). *Menominee prosodic structure*. PhD thesis, Wisconsin.

Miner, K. L. (1975). *Interference relations in Menomini phonology*. Indiana University.

Miner, K. L. (1979a). Theoretical implications of the great menominee vowel shift. *Kansas Working Papers in Linguitics*.

Miner, K. L. (1979b). Through the years with a small language: more trouble with data in linguistic theory. *International Journal of American Linguistics*, 45(1):75–78.

Moulton, W. G. (1947). Juncture in modern standard german. *Language*, 23(3):212–226.

Muggeo, V. M. et al. (2008). Segmented: an r package to fit regression models with broken-line relationships. *R news*, 8(1):20–25.

Münte, T. F., Say, T., Clahsen, H., Schiltz, K., and Kutas, M. (1999). Decomposition of morphologically complex words in english: Evidence from event-related brain potentials. *Cognitive Brain Research*, 7(3):241–253.

Nielsen, H. F. (1994). 1.1 on the origin and spread of initial voiced fricatives and the phonemic split of fricatives in english and dutch. In *Speaking in our tongues: Proceedings* *of a colloquium on medieval dialectology and related disciplines*, page 19. Boydell & Brewer.

Oliveira-Guimarães, D. (2013). Beyond early words: word template development in brazilian portuguese. *The emergence of phonology: Whole-word approaches and cross-linguistic evidence*, page 291.

Oxford, W. (2015). Patterns of contrast in phonological change: Evidence from algonquian vowel systems. *Language*, pages 308–358.

Patterson, D. and Connine, C. M. (2001). Variant frequency in flap production. *Phonetica*, 58(4):254–275.

Paul, H. (1916). *Deutsche Grammatik*. Walter de Gruyter.

Pegg, J. E. and Werker, J. F. (1997). Adult and infant perception of two English phones. *The Journal of the Acoustical Society of America*, 102(6):3742–3753.

Penzl, H. (1968). The history of the third nasal phoneme of modern german. *Publications of the Modern Language Association of America*, pages 340–346.

Peperkamp, S. and Dupoux, E. (2002). Coping with phonological variation in early lexical acquisition. *The process of language acquisition*, pages 359–385.

Peperkamp, S., Le Calvez, R., Nadal, J.-P., and Dupoux, E. (2006). The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition*, 101(3):B31–B41.

Peperkamp, S., Pettinato, M., and Dupoux, E. (2003). Allophonic variation and the acquisition of phoneme categories. *Proceedings of BUCLD 27*.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and speech*, 46(2-3):115–154.

Priestly, T. M. (1977). One idiosyncratic strategy in the acquisition of phonology. *Journal of Child Language*, 4(1):45–65.

Prince, A. and Smolensky, P. (2008). *Optimality Theory: Constraint interaction in generative grammar*. John Wiley & Sons.

Prince, A. and Tesar, B. (2004). Learning phonotactic distributions. *Constraints in phonological acquisition*, pages 245–291.

Purse, R. (2019). The articulatory reality of coronal stop “deletion”. In *19th International Congress of Phonetic Sciences*, pages 1595–1599.

Raffelsiefen, R. (2016). Allomorphy and the question of abstractness: evidence from german. *Morphology*, 26(3):235–267.

Rasin, E., Berger, I., Lan, N., and Katzir, R. (2018). Learning phonological optionality and opacity from distributional evidence. In *Proceedings of NELS*, volume 48, pages 269–282.

Rasin, E. and Katzir, R. (2016). On evaluation metrics in optimality theory. *Linguistic Inquiry*, 47(2):235–282.

Reifegerste, J. and Clahsen, H. (2018). Accessing morphosyntactic information is preserved at old age, except for irregulars. *The Mental Lexicon*, 12(3):342–372.

Richter, C. (2017). Learning allophones: What input is necessary? *BUCLD 42*.

Richter, C. (2020). Menominee high back vowel split as a consequence of alternation-sensitive phoneme learning. In *Linguistic Society of America Annual Meeting*. New Orleans.

Riehl, A. (2003). American english flapping: Evidence against paradigm uniformity with phonetic features. In *Proceedings of the 15th International Congress of Phonetic Sciences*, pages 2753–2756.

Ringe, D. and Eska, J. F. (2013). *Historical linguistics: toward a twenty-first century reintegration*. Cambridge University Press.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14(5):465–471.

Rissanen, J. and Ristad, E. S. (1994). Language acquisition in the mdl framework. *Language computations*, pages 149–166.

Samuels, B., Andersson, S., Sayeed, O., and Vaux, B. (2020). Getting ready for primetime: paths to acquiring substance-free phonology. *Canadian Journal of Linguistics*.

Sapir, E. (1921). *Language: An introduction to the study of speech*. Courier Corporation.

Savage, S. L. and Au, T. K.-f. (1996). What word learners do when input contradicts the mutual exclusivity assumption. *Child Development*, pages 3120–3134.

Scarborough, H. S., Rescorla, L., Tager-Flusberg, H., Fowler, A. E., and Sudhalter, V. (1991). The relation of utterance length to grammatical complexity in normal and language-disordered groups. *Applied Psycholinguistics*, 12(1):23–46.

Schuler, K. (2017). *The acquisition of productive rules in child and adult language learners*. Georgetown University.

Schuler, K., Yang, C., and Newport, E. (2016). Testing the tolerance principle: Children form productive rules when it is more computationally efficient to do so. In *The 38th Cognitive Society Annual Meeting*, pages 2321–26. Cognitive Science Society Mahwah, NJ.

Segbers, J. and Schroeder, S. (2017). How many words do children know? a corpus-based estimation of children's total vocabulary size. *Language Testing*, 34(3):297–320.

Seidl, A. and Cristia, A. (2012). Infants' learning of phonological status. *Frontiers in psychology*, 3:448.

Sharf, D. J. (1960). Distinctiveness of 'voiced t' words. *American Speech*, 35(2):105–109.

Smolka, E., Gondan, M., and Rösler, F. (2015). Take a stand on understanding: Electrophysiological evidence for stem access in german complex verbs. *Frontiers in human neuroscience*, 9:62.

Smolka, E., Zwitserlood, P., and Rösler, F. (2007). Stem access in regular and irregular inflection: Evidence from German participles. *Journal of Memory and Language*, 57(3):325–347.

Solomonoff, R. J. (1964). A formal theory of inductive inference. parts i and ii. *Information and control*, 7(1):1–22,224–54.

Son, M. (2008). Pitfalls of spectrogram readings of flaps. *Journal of the Acoustical Society of America*, 123(5):3079.

Song, J. Y., Shattuck-Hufnagel, S., and Demuth, K. (2015). Development of phonetic variants (allophones) in 2-year-olds learning American English: A study of alveolar stop /t, d/ codas. *Journal of Phonetics*, 52:152–169.

Steinberg, J., Jacobsen, T. K., and Jacobsen, T. (2016). Repair or violation detection? pre-attentive processing strategies of phonotactic illegality demonstrated on the constraint of g-deletion in german. *Journal of Speech, Language, and Hearing Research*, 59(3):557–571.

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of child language*, 38(1):1–34.

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1536):3617–3632.

Swingley, D. (2016). Two-year-olds interpret novel phonological neighbors as familiar words. *Developmental psychology*, 52(7):1011.

Swingley, D. and Alarcon, C. (2018). Lexical learning may contribute to phonetic learning in infants: A corpus analysis of maternal spanish. *Cognitive science*, 42(5):1618–1641.

Tesar, B. (2008). Output-driven maps. *Available as ROA-956 from the Rutgers Optimality Archive.*

Tesar, B. and Smolensky, P. (1996). Learnability in optimality theory (long version). *Rutgers Optimality Archive*.

Tesar, B. and Smolensky, P. (2000). *Learnability in optimality theory ch.5*. MIT Press.

Tessier, A.-M. (2013). The nature of regressions in the acquisition of phonological grammars. In *Proceedings of the Annual Meetings on Phonology*, volume 1(1).

Tessier, A.-M. (2019). U-shaped development in error-driven child phonology. *Wiley Interdisciplinary Reviews: Cognitive Science*, 10(6):e1505.

Thornton, R. and Rombough, K. (2015). The syntax-pf interface in children's negative sentences. *Language Acquisition*, 22(2):132–157.

Thornton, R. and Tesan, G. (2013). Sentential negation in early child english. *Journal of Linguistics*, 49(2):367–411.

Tollfree, L. (2001). Variation and change in australian consonants: Reduction of/t. *English in Australia*, 26:45–68.

Tomas, E., Smith-Lock, K., Demuth, K., et al. (2012). Errors of omission and commission in verbal and nominal inflectional morphemes by children with sli: Phonological effects and acoustic analysis. In *Proceedings of the 14th Speech Science and Technology Conference*, pages 101–104. Causal Productions Sydney, Australia.

Touretzky, D. S., Elvgren III, G., and Wheeler, D. W. (1990). Phonological rule induction: An architectural solution. Technical report, Carnegie-Mellon University.

Treiman, R., Cassar, M., and Zukowski, A. (1994). What types of linguistic information do children use in spelling? the case of flaps. *Child development*, 65(5):1318–1337.

Tsuji, S. and Cristia, A. (2014). Perceptual attunement in vowels: A meta-analysis. *Developmental psychobiology*, 56(2):179–191.

Tsuji, S., Cristia, A., Frank, M. C., and Bergmann, C. (2020). Addressing publication bias in meta-analysis. *Zeitschrift für Psychologie*.

Twaddell, W. F. (1938). A note on old high german umlaut. *Monatshefte für deutschen Unterricht*, pages 177–181.

van Lessen Kloeke, W. (1982). *Deutsche Phonologie und Morphologie*, volume 117. Niemeyer.

Vennemann, T. (1970). The german velar nasal. *Phonetica*, 22(2):65–81.

Vihman, M. and Croft, W. (2007). Phonological development: Toward a "radical" templatic phonology.

Vihman, M. M. (2014). *Phonological development: The first two years.* John Wiley & Sons.

Vihman, M. M. (2019). *Phonological templates in development*, volume 4. Oxford University Press, USA.

Vihman, M. M. and Velleman, S. L. (1989). Phonological reorganization: A case study. *Language and Speech*, 32(2):149–170.

Walker, R. (2018). Feature identity and icy targets in menominee vowel harmony. In *Hanabana: a Festschrift for Junko Ito and Armin Mester*. UC Santa Cruz.

Warner, N. (2005). Flapping and reduction of stops: Speech style, phonological environment, and variability. *ASA*.

Warner, N. and Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America*, 130(3):1606–1617.

Werker, J. F. and Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, 7(1):49–63.

Whalen, D. H., Best, C. T., and Irwin, J. R. (1997). Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics*, 25(4):501–528.

White, K. S., Peperkamp, S., Kirk, C., and Morgan, J. L. (2008). Rapid acquisition of phonological alternations by infants. *Cognition*, 107(1):238–265.

Wieczorek, R. (2010). Using mlu to study early language development in english. *Psychology of Language and Communication*, 14(2):59.

Wiese, R. (1986). Schwa and the structure of words in german. *Linguistics*, 24(4):697–724.

Wright, J. (1907). *Historical German Grammar*. Oxford University Press.

Wu, S.-Y., Huang, R.-J., and Tsai, I.-F. (2019). The applicability of d, mtld, and mattr in mandarin–speaking children. *Journal of communication disorders*, 77:71–79.

Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. MIT Press.

Yang, C. (2017). How to wake up irregular (and speechless). *On looking into words (and beyond)*, pages 211–233.

Yeung, H. H. and Nazzi, T. (2013). Ostensive and referential object labeling helps infants learn generalizable sound patterns. *Proceedings of BUCLD 37*.

Yip, K. and Sussman, G. J. (1997). Sparse representations for fast, one-shot learning. *Proceedings of the National Conference on Artificial Intelligence*.

Yip, M., Maling, J., and Jackendoff, R. (1987). Case in tiers. *Language*, pages 217–250.

Yun, S., Choi, J.-Y., and Shattuck-Hufnagel, S. (2020). A landmark-cue-based approach to analyzing the acoustic realizations of american english intervocalic flaps. *The Journal of the Acoustical Society of America*, 147(6):EL471–EL477.

Zue, V. W. and Laferriere, M. (1979). Acoustic study of medial /t, d/ in American English. *The Journal of the Acoustical Society of America*, 66(4):1039–1050.

Zue, V. W. and Seneff, S. (1996). Transcription and alignment of the timit database. In *Recent Research Towards Advanced Man-Machine Interface Through Spoken Language*, pages 515–525. Elsevier.