Draft March 15, 2016

Defying the Stimulus: Acquisition of Complex Onsets in Polish

Gaja Jarosz, University of Massachusetts Amherst

0. Abstract

Behavioural findings indicate that English, Mandarin, and Korean speakers exhibit gradient sonority sequencing preferences among unattested initial clusters (Davidson 2006; Berent *et al.* 2007, 2008; Daland *et al.* 2011; Ren *et al.* 2010). While some have argued these results support an innate principle, recent modelling studies have questioned this conclusion, showing that computational models capable of inducing generalisations using abstract phonological features can detect these preferences from lexical statistics in these languages (Daland *et al.* 2011, Hayes 2011). This paper presents a computational analysis of the development of initial clusters in Polish, which arguably presents a stronger test of these models. We show that 1) the statistics of Polish contradict the Sonority Sequencing Principle (SSP), favouring sonority plateaus, 2) models that succeeded in the other languages do not predict SSP preferences for Polish, and 3) children nonetheless exhibit sensitivity to the SSP, favouring onset clusters with larger sonority rises.

1. Introduction

The Sonority Sequencing Principle (Sievers 1881; Jespersen 1904; Steriade 1982; Selkirk 1984; Clements 1990) is a scale characterising cross-linguistic syllable well-formedness. The SSP favours syllables whose onsets rise in sonority toward the nucleus and fall in sonority from nucleus to coda. The SSP (1) prefers greater sonority rises for complex onsets, favouring large rises [bl] over plateaus [bd], which are in turn preferred to sonority falls [lb].

(1) Sonority Sequencing Principle (SSP)
[jb]ack ≺ [lb]ack ≺ [nb]ack ≺ [bd]ack ≺ [bn]ack ≺ [bl]ack ≺ [bj]ack

Recent studies employing a variety of behavioural tasks across multiple languages have consistently found that the SSP is active in speakers' grammars (Berent 2008; Berent *et al.* 2009, 2007, 2008; Zhao & Berent 2015; Lennertz & Berent 2015; Tamási & Berent 2014; Berent *et al.* 2012, 2011; Daland *et al.* 2011; Ren *et al.* 2010; Davidson 2006). Because the effects of SSP are found for novel clusters that are unobserved in the speakers' language input, these effects have been termed *sonority projection* (see also Hayes 2011). In English, besides initial clusters beginning with [s] (e.g. [st], [sk], [sn], [sm]), only clusters with large sonority rises are permitted (e.g. [pl], [gw, [ʃɹ]). For English, sonority projection effects have been demonstrated with both adults and children in various tasks including production, perception, and non-word acceptability. For example, Berent *et al.* (2007) shows that English speakers are more likely to perceptually repair severe SSP violators like [lb]if than moderately violating forms like [bd]if, which in turn are repaired more often than mildly violating forms like [bn]if. This occurs despite the fact that English speakers have no direct experience with word-initial [lb], [bd], or [bn]. Sonority projection is also observed in Korean and Mandarin, which have a more restricted inventory of initial clusters than English. Both languages lack syllable initial clusters entirely except for obstruent-glide-vowel sequences which are sometimes analysed as combinations of simple onsets and complex nuclei (though see Lee 1994; Duanmu 2002).

That the SSP is active in speakers' grammars is largely uncontroversial. What remains actively debated, however, is the source of this knowledge. How do phonotactic generalisations like the SSP come to be part of speakers' grammars? Must the SSP be provided to the learner as a universal principle or phonetic bias or could speakers induce generalisations about the SSP from their language input? On the one hand there are proponents of a strong universalist interpretation of these results (Berent 2008; Berent *et al.* 2009, 2007, 2008; Zhao & Berent 2015; Lennertz & Berent 2015; Tamási & Berent 2014; Berent *et al.* 2012, 2011; Ren *et al.* 2010). Others, however, question the need to invoke universal principles or phonetic biases to explain these effects, arguing instead that they are predictable from the input, given the right kinds of computational models (Hayes 2011; Daland *et al.* 2011).

There is abundant evidence that speakers' knowledge and learning of phonology and phonotactics is sensitive to the statistical properties of the language input. A substantial body of research indicates that gradient phonotactic acceptability (Coleman & Pierrehumbert 1997; Vitevitch *et al.* 1997; Frisch *et al.* 2000; Bailey & Hahn 2001; Kager & Pater 2012) and productive knowledge of phonological alternations (Ernestus & Baayen 2003; Zuraw 2000; Hayes & Londe 2006) mirror statistical trends in the lexicon. Phonological acquisition is also shaped by the statistical structure of the input. For example, infants below the age of one are sensitive to phonotactic probability in their native language, and this sensitivity affects speech production, speech perception and word segmentation (Coady & Aslin 2004; Jusczyk *et al.* 1994; Mattys & Jusczyk 2001; Zamuner 2009; Saffran *et al.* 1996). The effects of frequency are also evident in acquisition order of phonological patterns, with developmental order varying cross-linguistically depending on input statistics (Edwards & Beckman 2008; Ingram 1988; Jarosz 2010; Levelt *et al.* 2000; Vihman 1993).

Although clusters like [nb], [bd], [bn] are absent word-initially in English, Mandarin, and Korean, speakers of these languages do experience patterns with varying degrees of abstract phonological similarity to these clusters. There is a growing literature demonstrating the impressive abilities that computational models with minimal representational assumptions have to induce phonological generalisations predicting speakers' performance in various behavioural tasks (Coleman & Pierrehumbert 1997; Albright 2009; Bailey & Hahn 2001; Hayes & Wilson 2008; Frisch *et al.* 2004; Daland *et al.* 2011). Recent studies show that these models are more successful when they rely on richer phonological representations like features, syllables, tiers, and metrical grids (Albright 2009; Hayes & Wilson 2008; Frisch *et al.* 2004; Daland *et al.* 2011; Kager & Pater 2012; Hayes 2011). Most pertinent to the current investigations, recent findings reveal that models with certain properties correctly predict sonority projections effects not only for English (Daland *et al.* 2011), but also for languages likes Korean and Mandarin and, in principle, CV languages (Hayes 2011), without the need to posit a built-in SSP principle. If sonority projection can be inferred by the learner with the right representational assumptions, this undermines the strong universalist interpretation of these findings.

At the same time, much research on phonological learning suggests that reference to universal principles or phonetic biases may be necessary. Sonority projection is just one example of a wider range of studies demonstrating *poverty of the stimulus* effects wherein speakers exhibit preferences among novel structures that follow universal scales. Such effects are also observed in second language acquisition and loanword adaptation, for example (Broselow & Finer 1991; Broselow *et al.* 1998), and a number of studies show that these effects are not reducible to basic lexical statistics (see e.g. Berent *et al.* 2007; Davidson 2006). *Surfeit of the stimulus* effects, on the other hand, demonstrate that language learners may fail to learn or learn more poorly generalisations that do not conform to universal principles or phonological naturalness even when the patterns supporting those generalisations are robustly represented in the input (Becker *et al.* 2012, 2011; Hayes *et al.* 2009; Hayes & White 2013). Many of the most striking arguments for the role of universal biases come from developmental findings demonstrating remarkable consistency in developmental trajectories across languages. It has long been observed that typologically rare and marked structures tend to be acquired later by children regardless of their input (Jakobson 1941; Stampe 1969). For example, research on the development of basic syllable structure across languages consistently finds that children acquire more marked syllable shapes (e.g. CVCC) later than less marked syllables shapes e.g. (CV) (Demuth 1995; Fikkert 1994; Jarosz 2010; Levelt *et al.* 2000). Once again, these developmental progressions are often not directly derivable from the statistics of the input (see e.g. Levelt *et al.* 2000). One concrete example is the basic CV syllable in English, which is not the most frequent syllable shape in English child-directed speech (Jarosz 2010), but studies find that children generally acquire open syllables before closed syllables (Demuth 1995; Fee & Ingram 1982). Studies of phonological development also show that systematic preferences observed in children's errors can be attributed to universal principles. For example, when children reduce complex onsets, there is a tendency to retain the lower sonority consonant (Gnanadesikan 1995; Pater & Barlow 2003).

(2) Hierarchy of Increasingly Innatist Hypotheses
- *a.* *Segmental Statistics*
  Lexical Analogy (e.g. Bailey & Hahn 2001)
  Phoneme Co-occurrence (e.g. Vitevitch & Luce 2004)
- *b.* *Structured Generalisation*
  UCLA Phonotactic Learner (Hayes & Wilson 2008)
  Feature-Based Generalisation (Albright 2009)
- *c.* *Substantively Biased Generalisation*
  SG + Universal Grammar (e.g. Prince & Smolensky 1993)
  SG + Phonetically Based Phonology (e.g. Hayes *et al.* 2004; Hayes 1999; Wilson 2006)

Thus, while the role of statistics is indisputable, the necessity, nature, and strength of any universal biases is unclear and actively debated. Complicating the picture is the fact that universal principles are often mirrored by language-particular statistics (Zamuner *et al.* 2005; Jarosz *et al.* to appear; Jarosz 2010; Levelt & van de Vijver 1998), which makes it difficult to disentangle the independent contribution of universal biases and statistics in any particular case. What remains controversial, then, is the extent to which linguistic knowledge can be reduced to experience with the language input. With respect to the SSP, is the language input *sufficient* to support sonority projection? If not, what kind of restrictions or biases must be present to constrain learning? The present paper frames these questions by considering hypotheses falling along an increasingly universalist hierarchy, as schematised in (2). At the bottom end of this hierarchy is *Segmental Statistics*, which posits sensitivity to statistics of unstructured segmental representations. The middle level, *Structured Generalisation*, posits statistical learning coupled with rich phonological representations (such as features, syllables, tiers, grids) and the capacity to state generalisations over these abstract representations. Finally, the highest level, *Substantively Biased Generalisation* (a term inspired by Wilson (2006)), posits, in addition to structured generalisation, learning biases that reflect substantive phonetic or phonological preferences. Note that the category *lexicalist* used elsewhere (e.g. Daland *et al.* 2011) subsumes both Segmental Statistics and Structured Generalisation and that Substantively Biased Generalisation does not necessarily commit to innate substance. Substantively Biased Generalisation is also compatible with the view that the SSP is phonetically grounded, learned from universally shared experience with speech perception and articulation (Hayes *et al.* 2004; Hayes 1999). The present paper abstracts from many important differences within each of these levels, focusing primarily on evidence that can differentiate models falling into the Structured Generalisation level as opposed to the Substantively Biased Generalisation level.

After briefly reviewing the existing evidence that the Segmental Statistics Hypothesis is inadequate, this paper focuses on pushing Structured Generalisation to its limits and ultimately arguing for its inadequacy. The test case is the SSP in Polish, examined through the computational analysis of acquisition. There are several contributions. First, Section 2 shows that Polish lexical statistics contradict the SSP, while the same statistics in English, Mandarin, and Korean mirror the SSP. This establishes Polish as a strong test case for the Structured Generalisation Hypothesis in the domain of sonority sequencing. Second, Section 3 examines the role of the SSP in the development of initial clusters by four children acquiring Polish, showing that SSP is a robust predictor of cluster production accuracy. Next, Section 4 shows that the SSP effect cannot be reduced to lexical statistics of Polish and that the predictions of Structured Generalisation for Polish do not derive the SSP effect. Finally, Section 5 considers several alternative hypotheses, and 6 concludes. Before delving in to the lexical statistics of Polish, the following sections take a closer look at how Structured Generalisation can successfully model sonority projection effects in other cases.

1.1 Existing Sonority Projection Modelling Findings

After demonstrating sonority projection in English experimentally using Likert and head-to-head acceptability ratings tasks, Daland *et al.* evaluate six lexicalist computational models' abilities to capture sonority projection effects. They trained all models on a corpus of both syllabified and unsyllabified phonetically transcribed words to explore the effect of syllable structure. Three of the models fall within the Segmental Statistics class outlined above. The classical bigram model (Jurafsky & Martin 2008) assigns probability to novel forms based on the phoneme bigrams it contains. It has no notion of phonological structure, nor does it rely on features. The Phonotactic Probability Calculator (Vitevitch & Luce 2004) is similar, except that it maintains counts for bigrams and unigrams separately according to the serial position within the word in which they occur. It also has no inherent notion of features or phonological structure. Finally, the Generalised Neighborhood Model (Bailey & Hahn 2001) is an analogical model that assigns scores to novel words based on their similarity to existing lexical items. Similarity in this model is computed in terms of whole-phoneme insertions, deletions, or substitutions. Like the preceding models, this model treats phonemes as atomic, unstructured units and does not represent phonological structure of any kind.

The other three models Daland *et al*. consider fall to at least some degree within the Structured Generalisation class. The syllabic parser (Coleman & Pierrehumbert 1997) relies on syllable structure and stress, but it treats phonemes (and indeed entire onsets and rhymes) as atomic units and has no capacity to generalise using phonological features. The UCLA Phonotactic Learner (Hayes & Wilson 2008) constructs and weights constraints on sequences of natural classes. The Featural Bigram model (Albright 2009) also relies on natural classes, assigning a score to novel strings based on the most probable combinations of natural classes they contain. Both of these models are capable of generalising on the basis of phonological similarity as measured by shared featural representations.

Daland *et al*. show that only the last two models, the UCLA Phonotactic Learner and the Featural Bigram, are able to provide a good fit to the particpants' ratings on unattested clusters, with the Phonotactic Learner performing substantially better. All models perform better when trained on syllabified data although the Phonotactic Learner does quite well with unsyllabified data as well. They conclude that two properties are essential for successful sonority projection: 1) the ability to generalise on the basis of phonological features, in particular features that can represent relative sonority, and 2) the ability to represent and utilise phonological context or syllable structure to differentiate among clusters in different syllabic positions. On the basis of these results and the earlier reported findings (Berent *et al.* 2007; Davidson 2006) that segmental statistics and lexical analogy cannot capture the behavioural results, it is clear that the Segmental Statistics class of hypotheses must be rejected and that at least Structured Generalisation is necessary.

What about Mandarin and Korean? In related work, Hayes (2011) shows that, given appropriate representations, the UCLA Phonotactic Learner can produce sonority projection when trained on toy languages with distributions like those in Mandarin and Korean and even on a toy languages with strict CV syllables. Hayes equips the UCLA Phonotactic Learner with a 'modest UG' (Hayes 2011: 836; Daland *et al.* 2011: 226-227) of 32 constraints referring to binary combinations of single-feature natural classes defined in terms of the features [syllabic], [consonantal], [approximant], and [sonorant], such as *[+sonorant][−approximant]. These constraints effectively militate against sequences of sonority thresholds. The toy language "Bwa" has only CV syllables plus *obstruent+glide+vowel*, while the toy language "Ba" has only CV syllables. Hayes shows that in both cases the learner is able to project SSP preferences on the basis of the input. The basic intuition behind why this is possible is that the learner generalises broadly on the basis of observed sonority sequencing patterns at the beginnings of words, including the rise from a singleton onset into the vowel. The input provides support for word initial sonority rises of varying degrees, with nearly all words providing positive support for sequences like #[−syllabic][−consonantal], some words providing support for larger rises like #[−sonorant][−consonantal], and no words providing support for the reverse sequences like #[−consonantal][−sonorant]. Therefore, without stipulating the direction of sonority sequencing preferences, learners with these representations can detect that Ba and Bwa favour rises at the beginnings of words. Notice, however, that these simulations are set up to focus the learner's attention on word initial sequences up to the vowel, in effect providing syllabic context and a universal bias to treat pre-vocalic sequences separately from post-vocalic sequences, which is part of the substantive content of the SSP. Nonetheless the fact remains that a model with no built-in preference about sonority sequencing in initial position per se, other than to base these preferences on the *onset+vowel* combination, can detect a lexical preference consistent with the SSP.

To summarise, existing modelling results on sonority projection in Mandarin, Korean, and English (Daland *et al.* 2011; Hayes 2011; Berent *et al.* 2007) support rejection of the Segmental Statistics class of hypotheses, showing that sensitivity to syllabic context and the ability to generalise using features is essential, while success of Structured Generalisation shows there is lexical evidence for the SSP in these languages at an abstract level.

1.2 Pushing Structured Generalisation to its Limits

The fact that lexicalist models succeed in projecting SSP preferences in English, Mandarin, and Korean means these languages do not provide a true poverty of the stimulus test for sonority projection. Indeed, Hayes' Ba example reveals there can be no (natural) poverty of the stimulus test case for sonority projection. Cross-linguistically, languages that allow smaller sonority rises also allow larger sonority rises, which forms an essential part of the empirical foundation for the SSP (Sievers 1881; Jespersen 1904; Steriade 1982; Selkirk 1984; Clements 1990). This means that the more impoverished the inventory of consonant clusters is in a language, the more skewed the distribution in that language must be toward favouring sonority rises. The black bars in Figure 1 depict this distributional skew for English, showing the relative proportion of various sonority rise degrees in word initial clusters in the English lexicon. These proportions are based on the type frequencies of English word-initial clusters, estimated from the CMU Pronouncing Dictionary, as reported by Hayes & Wilson (2008).

They utilise the coarse-grained five-level sonority scale, *vowel* > *glide* > *liquid* > *nasal* > *obstruent* (Clements 1990), which is also what Hayes (2011) and Daland *et al*. (2011) used in their simulations[1]. The lexical support for the SSP is evident in this figure: relative frequency dramatically increases with the rise in sonority. At a more abstract level, the lexicon shows substantial support for rises (82.6%), minimal support for plateaus (17.4%) and no support whatsoever for falls (0%). The analogous distribution for Mandarin and Korean is easy to imagine: the entirety of the distribution (100%) would be concentrated on large sonority rises (3), also providing lexical bias in support of SSP.

If even an impoverished input distribution with purely CV syllables supports SSP projection, how can the predictions of Structured Generalisation be disentangled from those of the SSP? The answer is that the models must be tested on cases where the statistical trends in the lexicon actually *contradict* those of the SSP, and for that to be possible, the language input must be rich in SSP violators. The next section shows that Polish provides such a test case. Since Polish allows all sonority combinations, it is possible to examine phonological development of various sonority profiles in Polish to determine whether evidence of a preference for higher rises can be observed. As discussed above, some of the strongest arguments for the necessity of universal biases come from developmental findings that demonstrate children's sensitivity to generalisations that are apparently unsupported by their language input. For lexicalist models to subvert the need for universal biases, they not only have to account for the evidence of universal bias in adults' grammars, but they also have to contend with the evidence of biases in development.

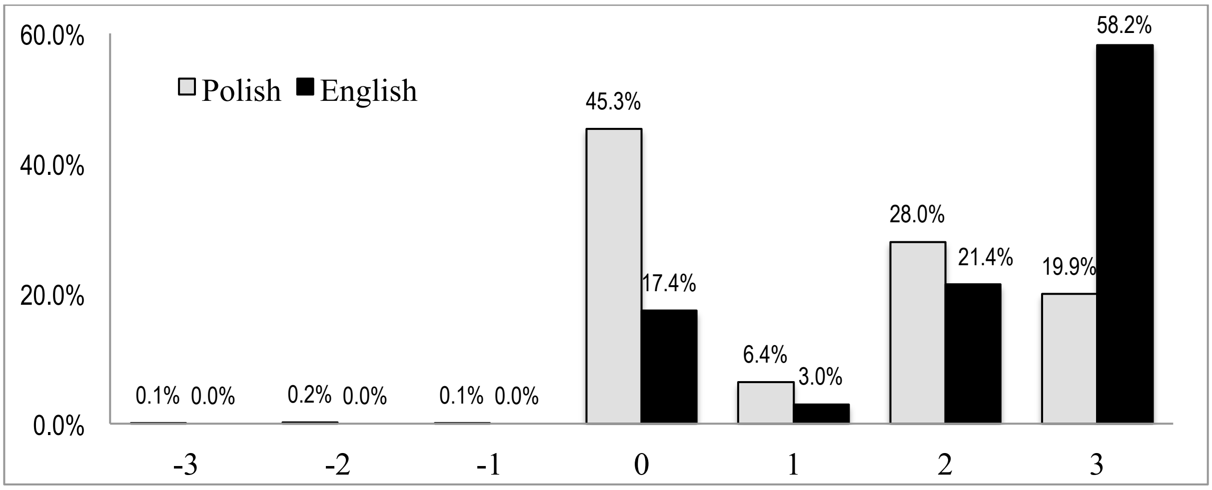

Figure 1 Relative Frequency of Sonority Rises in English and Polish

To summarise, the specific questions addressed in the present work are: a) does Structured Generalisation predict sonority projection effects for Polish, b) do children acquiring Polish exhibit sensitivity to the SSP, and c) do predictions of Structured Generalisation capture children's sonority preferences? The paper argues that the answers (a: no, b: yes, c: no) provide evidence to reject unbiased Structured Generalisation.

2. The Input

Polish is well-known for allowing SSP-violating clusters falling along all points of the scale in (1). Example words are shown in the third column of (3). While it is often acknowledged that sonority falls and other more complex configurations (e.g. trapped sonorants) are rare,

[1] Following Hayes & Wilson (2008) and Daland *et al.* (2011), these counts assume English [ɹ] is categorized as a glide.

often found in only a few lexical items, there is no shortage of obstruent-initial clusters with varying rise degrees in Polish. Nonetheless, formal analyses of syllable structure show that the SSP is active in the phonology of Polish (Jarosz 2006; Gussmann 1992; Bethin 1987; Rubach & Booij 1990, 1990; Bethin 1984, 1992). There are a number of productive phonological and morphological processes that depend on sonority sequencing. For example, voicing assimilation is normally blocked by a sonorant, unless that sonorant is flanked on both sides by obstruents or is word-final preceded by an obstruent, exactly where SSP violations are at stake (Gussmann 1992; Bethin 1984, 1992; Rubach & Booij 1990). There are also two morpho-phonological alternations, comparative and imperative allomorphy, where the choice of the allomorph is driven by sonority sequencing considerations (Rubach & Booij 1990; Bethin 1987; Rubach & Booij 1990).

(3) Sonority Profile and Sonority Rise Type Frequency in CDS

Sonority Profile	*Frequency*		*Example*	*SSP*	*SSP Frequency*	
GO	7	0.1%	[wza] "tear"	-3	7	0.1%
LO	26	0.2%	[lvɨ] "lions"	-2	26	0.2%
NO	10	0.1%	[mʂa] "mass"	-1	14	0.1%
LN	4	0.0%	[lnu] "linen" (gen)			
OO	4919	45.1%	[ptak] "bird"	0	4942	45.3%
NN	23	0.2%	[mɲɛ] "me (inst.)"			
ON	642	5.9%	[ɕɲɛk] "snow"	1	694	6.4%
NL	51	0.5%	[mlɛkɔ] "milk"			
LG	1	0.0%	[ljana] "vine"			
OL	2761	25.3%	[drɔga] "road"	2	3051	28.0%
NG	290	2.7%	[mjut] "honey"			
OG	2176	19.9%	[gwɔva] "head"	3	2176	19.9%

To determine the relative frequency of initial clusters and sonority profiles in Polish, counts were estimated from the largest available frequency dictionary of child-directed speech in Polish (Haman *et al.* 2011). This corpus includes about 800k word tokens and about 44k word types of speech directed at 128 children aged 0;10-6;11. About 34k of these word tokens come from a sample of speech directed at the four children in the Weist-Jarosz corpus whose phonological development is analysed in the next section (Jarosz 2010; Jarosz *et al.* to appear; Weist *et al.* 1984; Weist & Witkowska-Stadnik 1986). The Polish CDS dictionary was manually culled to remove misspellings, abbreviations and acronyms (e.g. “USB”), exclamations (e.g. “mhmmmm”), and a small number of unassimilated foreign words (e.g. “huckleberry”). This resulted in a frequency corpus of about 43k word types.

This sample of child-directed speech was then transcribed phonetically using automatic methods based on standard pronunciation in Polish, whose orthography is highly regular (Demenko *et al.* 2003). The transcription conventions were adapted to match those used in the child speech corpus analysed in the following section. Similar automatic methods have been used to construct many other child-directed speech corpora in a variety of languages (see e.g. Jarosz & Johnson 2013a; Brent & Cartwright 1996; Goldwater *et al.* 2009; Batchelder 2002; Blanchard *et al.* 2010).

This process produced a corpus with nearly 11k word types and 115k word tokens beginning with bi-consonantal clusters. The primary focus of the analyses of the following sections is on type frequency, as it is type frequency that has been most fruitfully used for modelling of phonotactic and phonological generalisation learning in previous work (Albright & Hayes 2002; Becker *et al.* 2011). However, because many prior studies examining spontaneous production in children find token frequency to be predictive of developmental trends (Roark & Demuth 2000; Zamuner *et al.* 2004, 2005; Kirk & Demuth 2005; Stites *et al.* 2004), including a prior analysis of the child speech examined here (Jarosz *et al.* to appear), token frequency is also calculated and considered in Section 5.1. The analysis focuses primarily on the coarse-grained SSP assumed in the previous sonority projection modelling studies. However, the corpus is also used to estimate frequencies for a finer-grained version of the sonority scale, and the consequences of this choice are explored in Section 0. Finally, the corpus is used to estimate the type and token frequencies of the nearly 200 segmental combinations occurring in bi-consonantal initial clusters (see Appendix) and other word-initial onsets that are used for training computational models in Section 4.2.

Table (3) summarises the type frequency estimates for sonority profiles (e.g. "ON"= *obstruent+nasal*) in the second column and sonority rises (e.g. SSP = -2) in the last column. G stands glide, L for liquid, N for nasal, and O for obstruent. The relative frequency of the sonority rises is also depicted graphically in Figure 1 alongside the same statistics previously discussed for English. Clearly, sonority falls are indeed quite rare in Polish, accounting for less than 0.5% of the input combined. On the other hand, the input is rich with obstruents. Nearly all the input (96.4%) consists of obstruent-initial clusters, which is compatible with SSP in principle; however, nearly half (45.1%) of these are *obstruent+obstruent* combinations. Viewed alongside the English distribution, it is not obvious how the Polish distribution could give rise to an SSP preference. If anything, it appears that the input distribution favours most strongly the middle of the SSP scale, that is, the sonority plateaus.

3. Sonority Sequencing Effects in Acquisition

There are several previous studies of syllable structure development in Polish (Jarosz 2010; Jarosz *et al.* to appear; Łukaszewicz 2006; Zydorowicz 2007). Jarosz (2010) examines development of basic CV syllable structure, in particular the relative timing of the development of initial and final clusters, but does not examine the role of sonority sequencing. Łukaszewicz (2006) does examine sonority sequencing effects, but her focus is on extra-syllabic sonorants (sonorants that cannot be syllabified consistently with SSP) and processes like voicing that depend on it. Consistent with the present arguments, she demonstrates a role of SSP in children's development of these particularly marked configurations. However, since extra-syllabic sonorants are both rare and marked, her findings do not specifically speak to the research questions examined presently. Zydorowicz (2007) examines cluster acquisition in all positions in the development of one child; her focus is primarily on whether morphological complexity of clusters affects development. She identifies developmental trends for clusters of varying sonority profiles, but does not identify any general patterns concerning the role of SSP. Finally, Jarosz *et al.* (to appear) examine acquisition of syllable structure and how it relates to input frequency, including showing developmental trends for onset clusters at the sonority level, but they do not systematically explore the role of the SSP. In sum, while there are a number of studies of cluster acquisition in Polish, the role of the SSP in initial cluster development is not well understood.

The data for this analysis comes from the Weist-Jarosz Corpus of spontaneous child speech (Jarosz 2010; Jarosz *et al.* to appear; Weist *et al.* 1984; Weist & Witkowska-Stadnik 1986) available via CHILDES (MacWhinney 2000). The corpus includes 2303 phonetically transcribed spontaneous productions of bi-consonantal initial clusters spoken by four typically-developing children aged 1;7 to 3;2. The corpus provides a word-by-word alignment between children's actual productions and their target pronunciations. All word initial clusters are considered in the analysis with the following exceptions: utterances involving onomatopoeia, wordplay, child-specific forms, incomplete word tokens, wholly or partially unintelligible word tokens, continuations of adult prompts, and repetitions or memorized passages. These cases are systematically marked in the corpus as such and can be removed automatically.

Since the corpus aligns each word with its target pronunciation, it is straightforward to determine whether the initial cluster is produced accurately. Because the focus in this analysis is on acquisition of sonority sequencing patterns word initially, accuracy is coded after translating both the target and actual pronunciations to their respective sonority classes. This means that substitutions along other dimensions like place or voice (e.g. substituting [ɕ] for [ʂ], a common pattern) do not count as errors as long as the child produced the target sonority profile for the initial cluster.

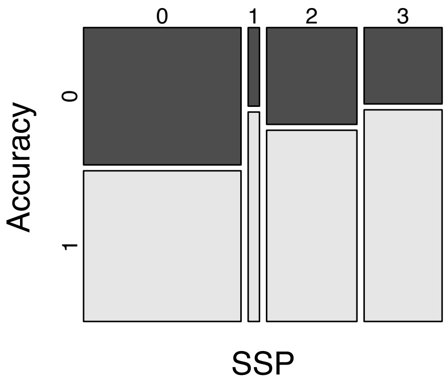

Figure 2 Accuracy By Sonority Rise

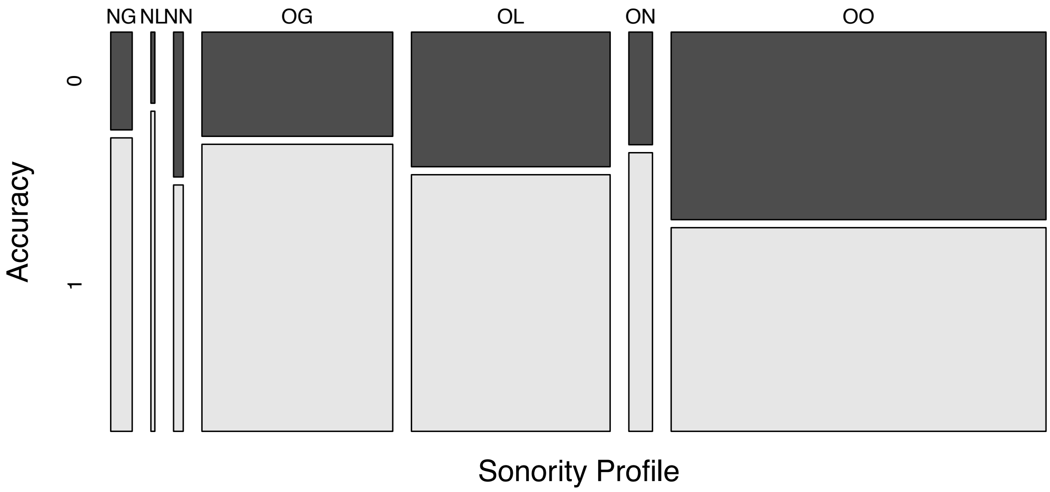

Figure 3 Accuracy By Sonority Profile

The children targeted seven sonority profiles in their productions[2]. Given the rarity of sonority falls in adult speech and in the lexicon, it is not surprising that young children's spontaneous productions include hardly any attempts of these clusters. Nonetheless, the data provides a representative sample of plateaus and rises of various degrees. Overall, children produce clusters with the target sonority profile 61.7% of the time. The vast majority of

[2] The corpus includes 1 token of a *liquid+fricative* cluster and 3 tokens of *stop+nasal* cluster. These were excluded from the analysis due to lack of sufficient data.

errors are deletions of the first consonant (13.4%) or the second consonant (14%) or substitutions in sonority profile for intact clusters (4.0%). Vowel epenthesis occurs in only 2.4% of the cases, and other scattered errors occur in the remaining cases (4.5%)[3]. The mosaic plots in Figure 3 (sonority profile) and Figure 2 (sonority rise degree) show the proportion of accurate responses on the vertical axis, and the cluster type on the horizontal axis. The width of each bar is proportional to the number of tokens in the children's production data corresponding to each cluster type. These figures reveal that accuracy depends on the sonority sequencing of the cluster, with more accurate productions for higher sonority rises.

Since the spontaneous production corpus is not balanced, logistic regression is used to determine whether the effect of SSP is significant after controlling statistically for a number of potentially confounding variables. This approach follows a number of recent studies utilizing regression modelling strategies to analyse spontaneous corpus data (Jarosz & Johnson 2013; Jarosz *et al.* to appear; Roland *et al.* 2006; Bane *et al.* 2010; Jaeger 2010). The logistic regression models are fitted using the lrm() function in the RMS package for R (Harrell 2014), which provides convenient methods for validating the resulting models.

The dependent variable is ACCURACY, coded at the level of sonority as described above, and the predictor of interest is SSP. The data are also coded for a number of control predictors. AGE (a continuous variable measured in months) is included to account for developmental progression over time. To account for individual differences in overall production accuracy, SUBJECT is included as a four-level factor[4]. Because children's phonological development often proceeds at different rates, the interaction of SUBJECT and AGE is also included. To control for the possibility that mere experience with a particular word form affects the accuracy with which it is produced, (log) WORD FREQUENCY is included as a predictor[5]. Also included are four predictors to control for potential effects of the prosodic and morphological contexts in which these clusters occur. This includes WORD LENGTH (counted in number of syllables), STRESS (a binary variable indicating whether the syllable the cluster initiates carries primary stress)[6], FUNCTION WORD (a binary variable indicating whether the cluster occurs in a closed-class word), and PREFIX (a binary variable indicating whether the cluster is morphologically complex, that is, composed of a mono-consonantal fricative prefix, orthographic *z* or *w,* followed by a singleton consonant). Effects of prosodic position and prominence have been repeatedly observed in child production studies (see e.g. Demuth 1995; Fikkert 1994), and morphological factors have also been found to affect production (Zydorowicz 2007) so these predictors are included to ensure the observed effects of SSP are not attributable to imbalances in the distribution of these factors. To summarise, the base model includes seven control variables and one interaction term with the following associated number of parameters for each: Age (1), Subject (3), Age x Subject (3), Word Frequency (1), Word Length (1), Prefix (1), Function Word (1), and Stress (1), for a total of 12 parameters.

The superset model with SSP and the control predictors is shown in (4). The main finding is that SSP is positively associated with production accuracy, indicating that clusters

[3] Examples of how error types vary by sonority are shown in mosaic plots in the Appendix.

[4] Although subject-level predictors are often included in mixed effects regression models as random effects (see e.g. Jaeger 2008) the subject factor in these data has only four levels and hence does not provide sufficient information to estimate group-level variation (Gelman & Hill 2006:247).

[5] Because log(0) is undefined and because a handful of children's targets do not occur in the child-directed speech corpus, we use log(frequency + 1) for all of the frequency predictors.

[6] Primary stress was assigned automatically to penultimate syllables and monosyllables according to the regular stress pattern of Polish lexical stress (Rubach & Booij 1985).

with higher sonority rises are produced more accurately by the children after statistically controlling for all of the above potential confounds. Nested model comparison with the base model shows SSP is highly significant (χ^2 (1) = 52.7; $p < 0.0001$). Although the control predictors are not the focus of this analysis, it is worth noting that their contributions to the model are sensible. Production accuracy increases with age, although the degree to which this occurs varies by child. Clusters that result from the concatenation of mono-consonantal prefixes and singleton onsets are produced significantly less accurately. Since these prefixes express verbal aspect, this is consistent with the well-known tendency for children to omit functional, morpho-syntactic markers in spontaneous speech (and indeed, the only difference in the error patterns for these clusters is a higher rate of deleting the first consonant, i.e. the prefix). Clusters at the beginnings of stressed syllables tend to be more accurate and those in closed-class items less accurate, although these effects only tend toward significance. Finally, word frequency is negatively associated with production accuracy, suggesting that children are more likely to reduce clusters in common words.

(4) Logistic Regression Model Predicting Cluster Production Accuracy

	β	S.E.	Wald Z	Pr(>|Z|)
Intercept	-11.16	1.92	-5.8	**<0.0001**
Subject				
Kub	9.24	2.66	3.47	**0.001**
Mar	9.86	6.02	1.64	0.101
Waw	10.48	1.94	5.4	**<0.0001**
Age	0.47	0.09	5.38	**<0.0001**
Function Word	-0.29	0.17	-1.64	0.100
log(Word Frequency)	-0.11	0.03	-3.51	**0.001**
Stress	0.33	0.20	1.64	0.101
Prefix	-0.65	0.14	-4.47	**<0.0001**
Word Length	0.05	0.12	0.45	0.650
SSP	**0.28**	**0.04**	**7.16**	**<0.0001**
Kub * Age	-0.38	0.11	-3.46	**0.001**
Mar * Age	-0.43	0.30	-1.4	0.163
Waw * Age	-0.42	0.09	-4.79	**<0.0001**

To examine the robustness of the SSP predictor to fluctuations in the data sample, the validate() function in the RMS package was used to obtain bootstrap samples, calculate model optimism, and perform backwards elimination on the bootstrapped models. Out of 200 bootstrap validation samples with backward elimination, SSP was retained in the model 200 times. Furthermore, the model shows a small amount of shrinkage: the original D_{xy} is 45.5, the optimism is 0.011, and the corrected D_{xy} is 44.4. This is substantially higher than the corrected D_{xy} of the base model that excludes SSP (40.9). While these findings nonetheless depend on the particulars of this corpus, the results of the validation procedure provide some reassurance that the role of the SSP is robust in this data sample.

4. Predicting Developmental Effects

This section examines in detail the predictions that Structured Generalisation makes for sonority sequencing in Polish. Three approaches to modelling sonority projection are considered and rejected. Section 4.1 examines the relationship between frequency in the input, the SSP, and accuracy. Section 4.2 generates predictions using the UCLA Phonotactic Learner in two ways: first (4.2.1), by allowing the learner to induce its own constraints based on a corpus of segment-level word-initial onsets, and second (4.2.2), by weighting a pre-specified set of sonority sequencing constraints as in Hayes (2011). For sonority projection to be successful, the models must account for the developmental SSP effect and generate systematic preferences for larger rises over smaller rises, and smaller rises over plateaus. None of the models generate these predictions on the basis of the Polish input distribution.

4.1 Raw Segmental and Class Frequency

The first step in determining whether frequency in the input could account for the developmental results presented in the preceding section is to better understand how input frequency is associated with production accuracy and the SSP. In addition to examining the type frequency distribution over sonority profiles and rises (Table (3)), the section also considers the frequency of initial segmental bigrams (e.g. clusters; shown in the Appendix).

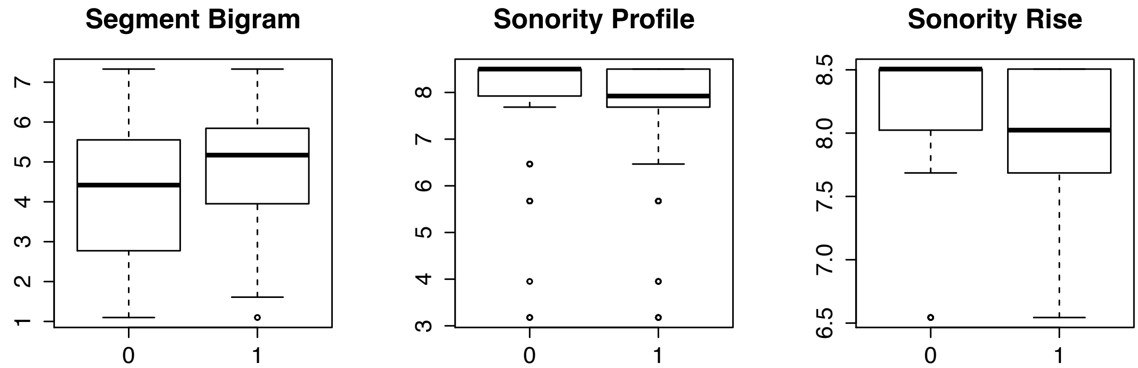

Figure 4 Association Between Accuracy and Log Type Frequency

Figure 4 summarises the association between children's production accuracy and three measures of log type frequency: frequency of initial segmental bigrams, frequency of sonority profiles, and frequency of sonority rises. Each box and whisker plot is a visual summary of the distribution, showing the median (horizontal line), the upper and lower quartiles (the box), and the maximum and minimum (the whiskers), excluding any outliers (the dots). If variation in the production accuracy of initial clusters is driven by one of these measures, higher frequency on that measure should be associated with higher production accuracy. As the box and whisker plots illustrate, a positive association exists only for segmental bigrams. For sonority profiles and rises, lower frequency is associated with higher production accuracy. This is not the association one would expect if frequency of these classes were driving development directly.

Given that higher rises are predictive of higher production accuracy, this suggests that class-level frequency of sonority profiles and sonority rises is unlikely to yield sonority projection. This is confirmed by examining the rank correlation (kendall's τ) between each of these frequency measures and the SSP for all the cluster types targeted by the children. For segmental bigrams there is a weak positive correlation between frequency and SSP (τ = 0.289), while for sonority profiles (τ = −0.657) and sonority rises (τ = −0.737), the correlations are negative. Therefore, while it is in principle possible for sonority projection to occur on the basis of segmental bigram type frequency, the class-level frequencies, at least when used directly, predict anti-SSP preferences. This is not surprising given the earlier figures showing that sonority plateaus are the distribution mode in Polish.

The strongest test of these raw frequency measures is a nested model comparison including all the control predictors discussed above plus each frequency measure compared to a superset model that also includes SSP. If the superset model provides a significantly better fit to the developmental data that warrants the extra degree of freedom (as determined by a χ^2

test), this means that the frequency measure does not account for the effect of SSP on production accuracy. If, on the other hand, a measure has fully accounted for the SSP effect, the SSP predictor in the superset model would be superfluous. Table (5) summarises the results of these nested comparisons for each of the frequency measures. Segmental bigram frequency is a significant predictor of production accuracy after all control predictors are included in the model (χ^2 (1) = 34.8; p < 0.0001). Higher segmental frequency is associated with more accurate production (β = 0.235, z = 5.84, p < 0.0001). However, the superset model with both segmental frequency and SSP is superior (χ^2 (1) = 34.8; p < 0.0001), indicating that segmental bigram frequency does not subsume the effect of SSP on production accuracy. 200 bootstrap validation samples confirm that SSP is always retained in the model.

(5) Summary of Model Comparisons for Type Frequencies

		Segmental Bigram		Sonority Profile		Sonority Rise	
	Base	+Freq	+Freq+SSP	+Freq	+Freq+SSP	+Freq	+Freq+SSP
D_{xy}	0.425	0.445	0.47	0.437	0.459	0.453	0.458
LR	357.4	392.2	427.0	375.77	412.30	388.65	409.83
$\chi^2(1)$		34.8	34.8	18.4	36.5	31.3	21.2
p		< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

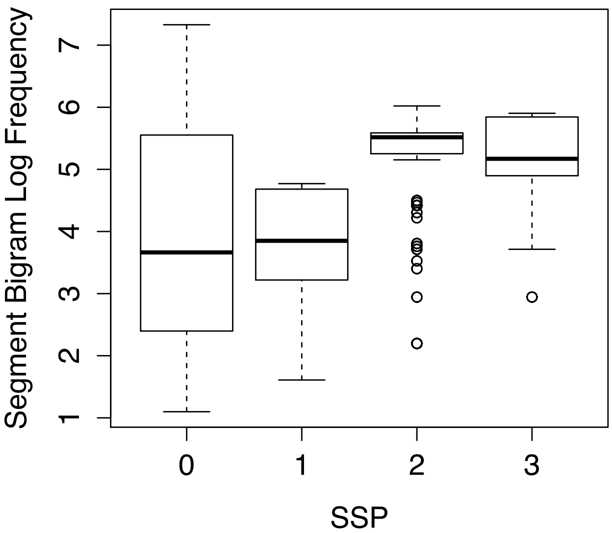

Figure 5 Association Between Segmental Bigram Type Frequency and SSP

Both sonority profiles (β = −0.256, z = −4.05, p < 0.0001) and sonority rises (β = −0.657, z = −5.39, p < 0.0001) are predictive of accuracy, but in the wrong direction, as expected from the above discussion: lower frequency is predictive of higher accuracy. More importantly, in both cases the superset model with SSP is superior, indicating that these measures do not subsume the SSP effect. Finally, the 200 bootstrap validation procedure confirms that SSP is retained on all 200 samples for sonority profiles and 199 samples for sonority rises, while the frequency measures are dropped in each case, indicating they are not reliable predictors of accuracy.

To summarise, none of these measures directly supports sonority projection. Sonority profile and sonority rise class frequency predict anti-SSP preferences. Segmental bigram frequency is predictive of production accuracy, and is weakly positively correlated with SSP. However, this measure reflects properties of clusters that are at least partially orthogonal to the SSP since both predictors remain highly significant in a superset model. The box and whisker plot in Figure 5 shows graphically why segmental bigrams fail to fully account for the SSP effect. While there is a noticeable upward trend as the sonority rise increases (accounting for the positive correlation), the range of frequency values for plateaus (SSP=0) essentially overlaps the entire range of values for each of the other rise degrees, some of which overlap one another almost entirely (SSP=2 and SSP=3). This means it is not possible to reliably predict on the basis of these frequencies that sonority plateaus (SSP=0) should be systematically disfavoured to all the others. Likewise, it is not possible to predict that moderate rises (SSP=2) should be disfavoured compared to large rises (SSP=3).

4.2 UCLA Phonotactic Learner

This section explores whether the computational model that has shown the greatest capacity for sonority projection in previous work on other languages (Daland *et al.* 2011; Hayes 2011), the UCLA Phonotactic Learner (Hayes & Wilson 2008), can capture the observed developmental SSP effect in Polish.

4.2.1 Inducing Constraints from Scratch

This section utilises the modelling strategy Daland *et al.* (2011) used to show sonority projection for English, except that the training data are limited to initial onsets. As discussed earlier, Daland *et al.* found that all models performed better with syllabified data as this enables models to state generalisations separately for onset and coda consonant sequences. Limiting the model's attention to initial onsets simplifies the learning problem. In other respects the set-up follows Daland *et al.*; in particular, all initial onsets together with their segmental type frequencies are included in the training data, including null, singleton, bi-, and longer consonant sequences. The singleton onsets account for the vast majority of the data: while the type frequency of bi-consonantal clusters is about 11k, the type frequency of singleton onsets is over 26k, and the frequency of singleton obstruents alone is over 19k. This presumably gives the model the best chance possible to discover a preference for word-initial rises since *obstruent+vowel* sequences are the largest possible rise and are abundantly represented in the data. Three models are considered with cut-offs on the number of induced constraints at 100, 200, and 300. The trained models are tested on the bi-consonantal clusters targeted by the children and evaluated on their ability to predict accuracy.

Figure 6 visualises the association between the SSP and the scores assigned to bi-consonantal clusters according to the 100 constraint and 200 constraint models. The size of the 'dots' in the scatterplot is proportional to the frequency of those target clusters in the children's productions. The scores are the summed violations of the weighted constraints and should be interpreted as penalties. Therefore, successful sonority projection requires lower scores for higher rises. The regression line plotted in the figure shows that this is indeed the direction of the association predicted by the models. This is because the models pick out particular segmental combinations to penalise, and these tend to be lower SSP clusters. However, the scatterplot also shows that the model assigns a penalty only to a small portion of the clusters targeted by the children: most of the clusters receive no penalties. Comparing the 100 constraint to the 200 constraint model reveals that the 200 constraint model induces additional constraints to penalise specific segmental combinations, and these include some constraints against larger rises as well as additional small rises and plateaus. Not surprisingly, the additional constraints in the 200 constraint model tend to penalise low frequency clusters. As the learner induces more constraints it penalises more and more of the attested clusters, but rather than penalising sonority plateaus as required for successful sonority projection, the model is simply inducing constraints that allow it account for the rarity of low frequency clusters in the segmental type frequency distribution. *Obstruent+obstruent* sequences in Polish have widely varying frequencies (recall the wide distribution in Figure 5), and some are underrepresented. The fact that the correlation between the model scores and the segmental bigram type frequency increases from the 100 constraint model ($\tau = -0.312$) to the 200 constraint model ($\tau = -0.426$) to the 300 constraint model ($\tau = -0.474$) is consistent with this claim. The model is doing what it is designed to do: induce constraints to penalise underrepresented patterns, and while particular sonority plateaus are underrepresented, sonority plateaus *in general* are not.

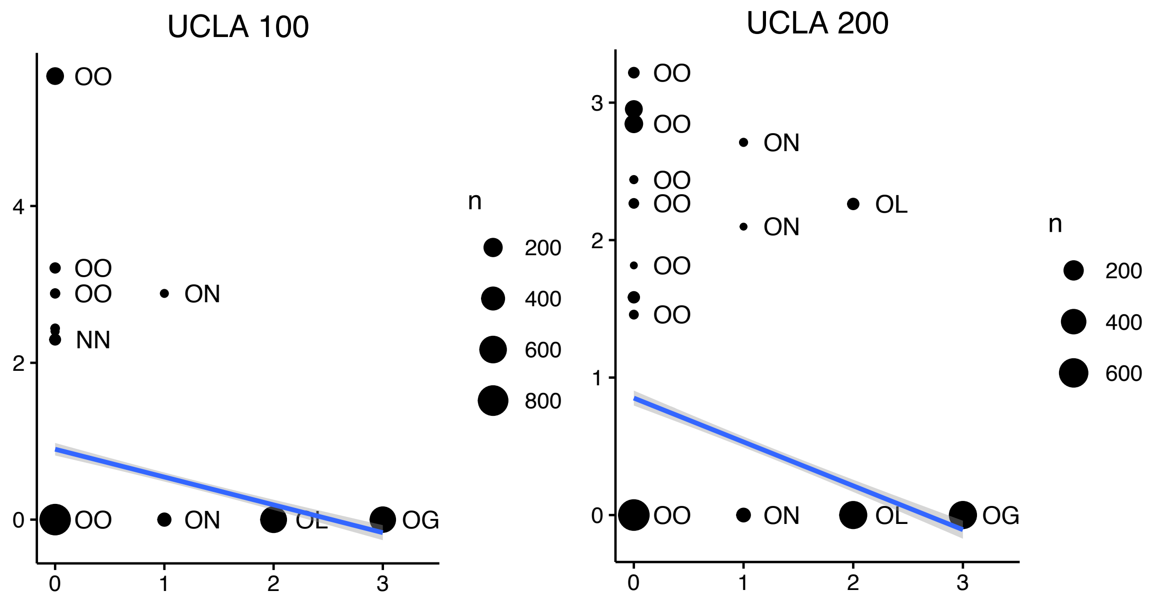

Figure 6 Association Between Model Predictions and SSP

As before, the strongest test of the models is to check whether they can capture the SSP predictor in a nested model comparison. The results of these model comparison evaluations are summarised in the left half of (6). Qualitatively, these models show the same pattern of results as the segmental type frequency of the preceding section. Model scores are significant predictors of children's production accuracy (100:χ^2 (1) = 51.1; 200: χ^2 (1) = 37.1; p < 0.0001), and higher scores (e.g. penalties) are predictive of lower accuracy for both the 100-constraint model ($\beta = -0.276$, z = −6.96, p < 0.0001) and the 200-constraint model ($\beta = -0.352$, z = −6.07, p < 0.0001). However, neither model can capture the SSP effect. SSP is highly significant in a nested model comparison for both the 100-constraint model (χ^2 (1) = 29.4; p < 0.0001) and the 200-constraint model (χ^2 (1) = 33.0; p < 0.0001). For both models, bootstrap validation retained SSP on 200 out of 200 samples. The predictor based on the 100-constraint model was retained on all 200 samples, while the predictor based on the 200-constraint model was somewhat less reliable, retained on 193 samples. The results for the 300-constraint model (not shown) are nearly identical to those for the 200-constraint model.

(6) Summary of Statistical Tests for UCLA Phonotactic Learner Predictions

	Induce 100		Induce 200		Hayes2011 UG32		Hayes2011 UG64	
	+F	+F+SSP	+F	+F+SSP	+F	+F+SSP	+F	+F+SSP
D_{xy}	0.453	0.473	0.447	0.468	0.427	0.46	0.435	0.459
LR	408.48	437.88	394.43	427.41	360.19	412.81	371.97	412.26
$\chi^2(1)$	51.1	29.4	37.1	33.0	2.8	52.6	14.6	40.3
p	< 0.001	< 0.001	< 0.001	< 0.001	0.093	< 0.001	< 0.001	< 0.001

To characterise the models' failure to capture SSP in alternative terms, we can consider just those clusters to which the 200-constraint model assigns no penalties whatsoever. Restricting attention just to these 'perfect' clusters, SSP is a significant predictor (χ^2 (1) = 32.5; $p < 0.0001$). This means there are robust effects of the SSP on production accuracy that these models are failing to capture.

In sum, the UCLA Phonotactic Learner induces constraints that penalise underrepresented clusters. While children are generally less accurate on the penalised clusters, these phonotactic constraints do not reflect the SSP in general, which remains highly significant in nested model comparisons.

4.2.2 Weighting Pre-specified Sonority Threshold Sequencing Constraints

The previous section showed that the UCLA Phonotactic Learner induces constraints for Polish that do predict sonority projection. However, perhaps if the learner is restricted to working with Hayes' 'modest UG' of 32 constraints, which worked for Ba and Bwa, it will be more successful.

Recall that Hayes' constraint set includes constraints on combinations of sonority thresholds as defined by the features [syllabic], [consonantal], [approximant], and [sonorant]. The set includes SSP-abiding constraints like *[+sonorant][−approximant] as well as anti-SSP constraints like the reverse *[−approximant][+sonorant]. It therefore does not directly encode pro-SSP preferences: it allows the opposite preferences to be discoverable in principle. However, because the constraints refer to sonority thresholds, the constraints are in stringency relationships, constraining the learner substantially. The learner cannot reproduce arbitrary frequency distributions over sonority profiles with this constraint set. For example, there is no way for the learner to penalise (only) *obstruent+nasal* combinations, which happen to be underrepresented in Polish. It could weight *[−sonorant][+sonorant] heavily, but this constraint penalises all *obstruent+sonorant* sequences, which are common overall.

The careful reader may have noticed that there are in fact 64, not 32, possible combinations of four binary features in two positions. Hayes' set of 32 constraints includes only those constraints that penalise opposing sonority thresholds such as *[+sonorant][−sonorant] and *[−sonorant][+sonorant] (Daland *et al.* 2011: 226-227). It does not include constraints penalising sonority threshold combinations that refer to the same side of the sonority scale, such as *[+sonorant][+sonorant] or *[−sonorant][−sonorant]. This means the set of 32 constraints is also restricted in that it cannot penalise (only) combinations of high sonority segments or (only) combinations of low sonority segments. Returning to the *obstruent+nasal* example, the set of 32 constraints do not provide a way to penalise this *low+low* sonority combo: constraints such as *[−sonorant][−approximant] are not available.

To examine the consequence of this restriction, the results of simulations with both the 32 constraint set and the full 64 constraint set are presented. As in the previous section and in Hayes' Ba and Bwa simulations, the training set includes all word-initial onsets. The test set is once again the bi-consonantal clusters targeted by the children. Since the constraints distinguish segments on only four manner features, only sonority distinctions are representable by this system. Thus, the segments in the training and test sets can be converted to their respective sonority classes (O, N, L, G, and V) and their frequencies collapsed.

The predicted sonority sequencing preferences of both models are depicted graphically in Figure 7. The model with 32 constraints primarily penalises sonorant-initial clusters. Obstruent-initial clusters receive scores close to zero. Since obstruent-initial clusters account for more than 90% of the data, it is not surprising that the regression line shows hardly any relationship to SSP. If the model has the ability to detect anti-SSP preferences, why did it not do so for Polish? The reason is that the 32 constraint set is so limited that only strong evidence of a statistical preference for sonority *falls* word-initially would be expected to trigger an anti-SSP generalisation. Because initial sonority falls are not common in Polish, accounting for less than 0.5% of the lexicon, the model does not predict an anti-SSP preference. Indeed, an anti-SSP preference would be expected with this constraint set only for languages that predominantly have sonority falls across all onset+vowel combinations.

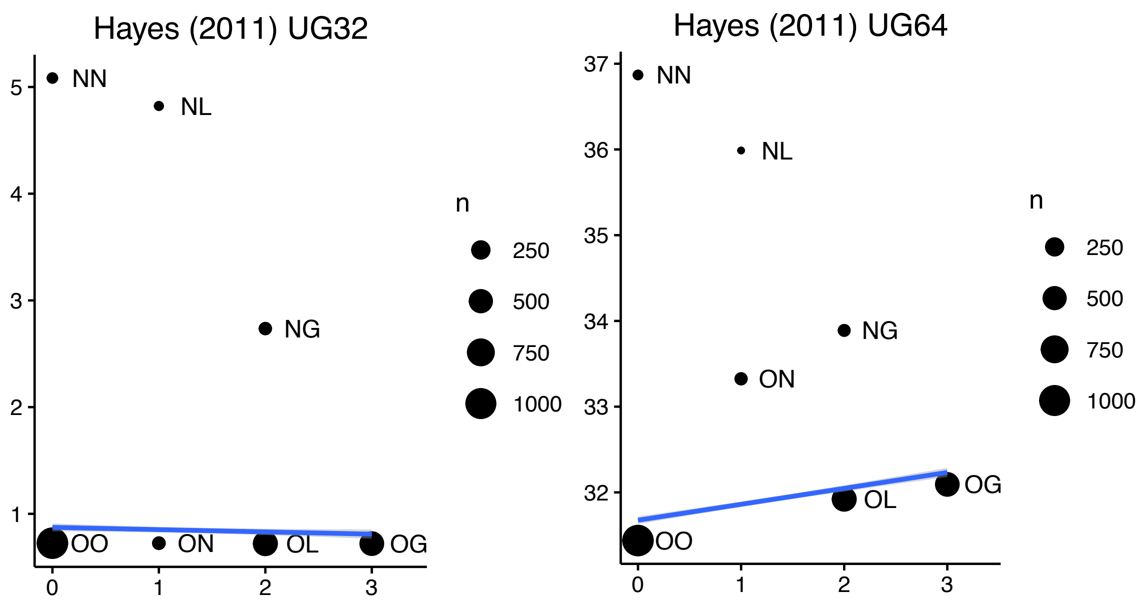

Figure 7 Association Between Model Predictions and SSP

The 64-constraint model has more freedom to match the observed distribution in Polish. It penalises all observed clusters, sonorant-initial clusters most heavily, and it also manages to penalise the underrepresented *obstruent+nasal* clusters. It does this by highly weighting various constraints penalising *high+high* sonority combinations (common to ONV, OLV, OGV, but not OOV) and penalising *low+low* sonority combinations (common to OOV, ONV, and NNV, but not OLV or OGV). This allows ONV to accumulate violations from both sets of constraints while OOV and OLV violate only a subset of these. Importantly, constraints against *high+high* combinations are also supported by the singleton onsets, which are primarily OV. The additional 'uni-directional' constraints in this set are crucial for these predictions, which in this case are unwelcome. Because of the constraints' availability, the learner incorrectly generalises from the singleton onsets that *high+high* combos (e.g. GV) should be disfavoured when they occur in complex onsets (e.g. OGV). As a consequence of the improved fit to the sonority profile frequency distribution in the data, the 64-constraint model makes worse predictions from the perspective of the SSP, predicting higher penalties for higher sonority rises on average. In general, this simulation demonstrates that generalising from the singleton CV transitions does not always produce desirable consequences.

The model comparisons in the right half of Table (6) confirm the above conclusions statistically. The predictor based on the 32-constraint model is only marginally significant (χ^2 (1) = 2.8; p = 0.093). This weak association goes in the wrong direction, and backwards elimination using bootstrap validation drops it. The superset model with SSP included is superior (χ^2 (1) = 52.6; $p < 0.0001$), and SSP is retained on all bootstrap validation samples. Finally, the predictor based on the 64-constraint model patterns similarly, except that it is a significant predictor (in the wrong direction) in the model without SSP ($\beta = 0.238$, $z = 3.6$, $p < 0.001$). On the majority of bootstrap validation backward elimination samples, the predictor is dropped. SSP, on the other hand, is highly significant in the superset model (χ^2 (1) = 40.3; $p < 0.0001$) and is retained in all bootstrap backward elimination samples.

To summarise, two restricted constraint sets penalising combinations of sonority thresholds do not generate sonority projection for Polish like they did for Ba and Bwa. Indeed, the simulations demonstrate that unconstrained generalisation from CV transitions to

CCV sequences can produce the unwelcome consequence of most strongly penalising complex clusters with the largest possible rises (OGV). For the learner to generalise as intended, something would have to prevent it from applying what it learns about CV transitions to the CV portion of CCV sequences. The learner would instead have to be compelled to apply these generalisations (only) to the CC portion of CCV, effectively encoding a built-in SSP bias.

5. Some Alternatives

The previous section showed that the lexicalist approaches that have worked well to generate sonority projection for other languages fail to do so for the Polish input distribution. Before concluding that unbiased generalisation does not project the SSP for Polish, the following sections consider two alternative ways of analyzing the input and the SSP scale, respectively.

5.1 Token Frequency

For completeness, this section demonstrates that relying on token frequency instead of type frequency does not provide a way out for the lexicalist hypothesis. Figure 8 shows the association between accuracy and token frequency measures, analogous to the figures above for the type frequency measures. The results are similar, albeit less promising: token segmental bigram frequency is not positively associated with accuracy.

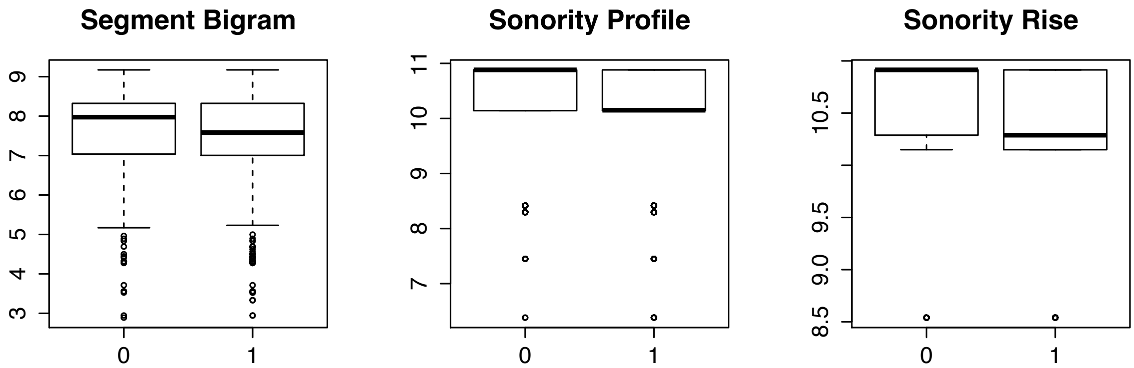

Figure 8 Association Between Accuracy and Token Frequency

(7) Summary of Statistical Tests for Token Frequencies

TOK/SSP4		Segmental Bigram		Sonority Profile		Sonority Rise	
	Base	+Freq	+Freq+SSP	+Freq	+Freq+SSP	+Freq	+Freq+SSP
D_{xy}	0.425	0.425	0.458	0.445	0.459	0.452	0.458
LR	357.4	357.74	410.20	380.07	412.49	385.61	410.03
$\chi^2(1)$		0.366	52.47	22.7	32.41	28.24	24.42
p		0.545	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

A summary of model comparisons analogous to those in Table (5) above is shown for token frequencies in Table (7). The pattern of results is similar, except that token segmental bigram frequency is not predictive of production accuracy. Just like the corresponding type frequencies, sonority profile ($\beta = -0.328$, $z = -4.54$, $p < 0.0001$) and sonority rise ($\beta = -0.569$, $z = -5.07$, $p < 0.0001$) token frequencies are predictive of production accuracy, but in the wrong direction. The superset models with SSP are superior in all cases, and SSP is retained as a predictor in all but one of 200 backwards elimination bootstrap samples. Thus, token frequencies do not provide a way to capture the developmental SSP effect.

5.2 Finer-Grained SSP

The analyses above follow prior modelling studies in assuming the coarse-grained sonority scale that worked well for generating sonority projection in other languages. The granularity of the sonority scale is often debated, however. Could it be that the coarse sonority scale is working against the lexicalist hypothesis by lumping all the obstruents together? This section first demonstrates that the same conclusions about children's sensitivity to the SSP are reached when a finer-grained sonority scale that separates plosives and fricatives is used (Selkirk 1984). Interestingly, the finer-grained sonority scale turns out to be a better predictor of children's production accuracy than the coarse scale. This is unexpected given that formal analyses of Polish phonology explicitly argue for the coarse-grained scale. It is consistent, however, with recent findings suggesting that sonority projection effects differentiate fricatives and stops (Lennertz & Berent 2015; Tamási & Berent 2014). The section then shows that the frequency predictions calculated on a finer-grained scale yield qualitatively similar pattern of results as the coarse-grained scale examined above.

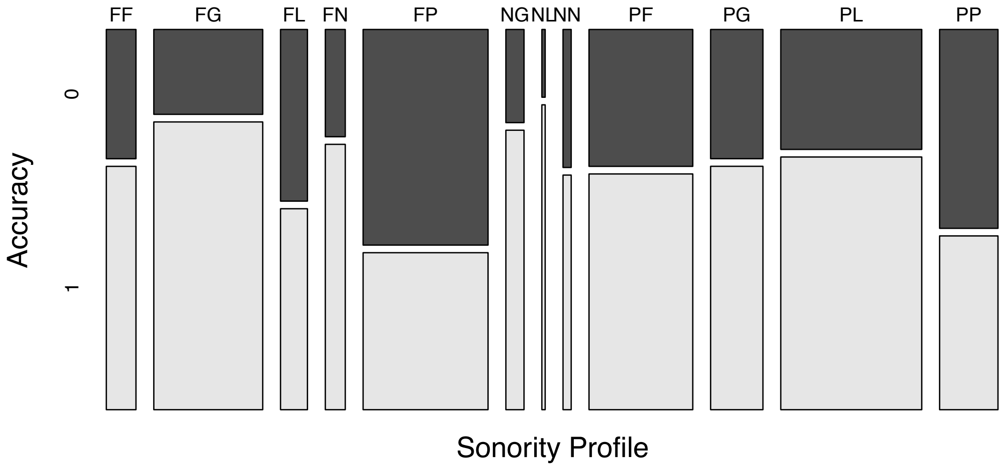

Figure 9 Accuracy By Finer-Grained Sonority Profile

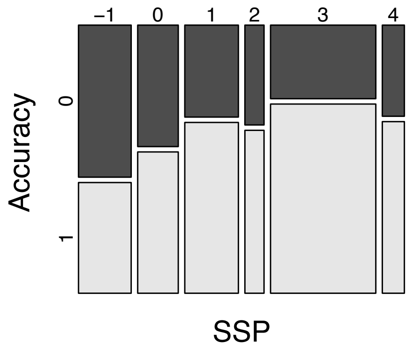

Figure 10 Accuracy By Finer-Grained Sonority Rise

(8) Summary of Statistical Tests for Frequencies Using Finer-Grained Sonority Scale

		Sonority Profile		Sonority Rise	
	Base	+Freq	+Freq+SSP	+Freq	+Freq+SSP
D_{xy}	0.425	0.428	0.471	0.434	0.469
LR	357.4	362.9	429.44	365.91	422.31
$\chi^2(1)$		5.5	66.6	8.54	56.4
p		0.019	< 0.0001	0.003	< 0.0001

Figure 9 and Figure 10 visualise the relationship between accuracy and finer-grained sonority profiles and rises, respectively ('F'=*fricative,* 'P'=*plosive*). The main differences from before are that finer-grained sonority treats FP as a mild sonority fall, PF as a mild rise like FN, PG as a larger rise than FG or PL, and PL as a larger rise than FL. Inspection of Figure 9 reveals that most of these difference align well with accuracy: FP is the least accuracy cluster type, PF is close in accuracy to FN, and children are less accurate on FL than PL. PG does not appear to be favoured by children relative to FG, however. Nonetheless, a nested model comparison shows that a predictor based on the finer-grained SSP (fSSP) is highly significantly predictive of production accuracy after controlling for the various potential confounding variables discussed earlier (χ^2 (1) = 64.7; $p < 0.0001$). As expected, the association is positive: higher fSSP is assossiated with higher accuracy ($\beta = 0.24$, $z = 7.95$, $p < 0.0001$). Out of 200 bootstrap validation samples with backward elimination, fSSP is retained in the model 200 times. Overall, the model shows a small amount of shrinkage: the original D_{xy} is 46.7, the optimism is 0.014, and the corrected D_{xy} is 45.3.

As mentioned earlier, an unexpected finding, is that fSSP is a better predictor of children's accuracy than SSP. This is verified by evaluating a superset model that adds SSP. The superset model is not superior to the model with just fSSP (χ^2 (1) = 0.059; $p > 0.8$), indicating that SSP is superfluous once fSSP is in the model. Accordingly, the opposite nested model comparison reveals that fSSP makes a significant contribution to a model that already has SSP (χ^2 (1) = 12.1; $p < 0.001$). Note that this is not a matter of degrees of freedom since both SSP and fSSP are continuous predictors with one degree of freedom: this just means that the additional distinctions made on the finer-grained sonority scale are reflected in the children's production accuracy. As further verification of this result, the backwards elimination validation procedure retains fSSP and not SSP.

To conclude this discussion, Table (8) shows that calculating frequency along the finer-grained sonority scale does not rescue the lexicalist approach. Just like the corresponding coarse-grained type frequencies, fine-grained sonority profile ($\beta = -0.143$, $z = -2.32$, $p < 0.05$) and sonority rise ($\beta = 0.2725$, $z = 2.93$, $p < 0.01$) type frequencies are predictive of production accuracy. The only difference is that sonority rise frequency is associated in the right direction with accuracy. Just as with coarse-grained frequency, the superset models with SSP are superior in both cases, and SSP is retained as a predictor in all 200 backwards elimination bootstrap samples[7]. Thus, coarse-grained frequencies do not provide a way to capture the developmental SSP effect.

6. General Discussion and Conclusion

To summarise, the major findings presented in this paper are:

(i) The sonority sequencing distribution of initial bi-consonantal clusters in Polish peaks at sonority plateaus, with nearly half of the input involving combinations of obstruents.

(ii) Lexicalist models that have succeeded at predicting sonority projection effects for English, Mandarin, and Korean do not predict systematic SSP-abiding preferences for Polish.

(iii) The spontaneous productions of four typically-developing children acquiring Polish reveal a robust effect of SSP on production accuracy.

(iv) None of the lexicalist models or frequency measures can capture the SSP effect observed in the developmental data.

More generally, returning to the hierarchy of hypotheses laid out in the introduction, the results presented here are consistent with neither the Segmental Statistics nor the Structured Generalisation classes of hypotheses. They therefore suggest that some kind of universal pressure or bias is required to explain the robust effect of SSP on production. While the

[7] These are the numbers using the weaker SSP predictor. Results are similar when fSSP is used.

present findings support the need for universally biased phonological learning, they do not differentiate between various possible sources of this universal bias. The universal bias could be due to innate preferences of the phonological grammar, or it could be induced on the basis of phonetic difficulty. The results are equally compatible with a bias that makes direct reference to the SSP as a grammatical principle as they are with a constellation of lower-level phonetic pressures that give rise to a similar preference scale. The conclusion supported by these findings is that phonological learning makes reference to substance in some way. Adjudicating between alternative notions of substance and universal bias is an important and difficult question that is beyond the scope of the present analyses.

These results significantly expand on and develop the arguments presented in previous work on sonority projection (Berent 2008; Berent *et al.* 2009, 2007, 2008; Zhao & Berent 2015; Lennertz & Berent 2015; Tamási & Berent 2014; Berent *et al.* 2012, 2011; Daland *et al.* 2011; Hayes 2011; Ren *et al.* 2010; Davidson 2006). This paper does not dispute the findings presented by Hayes (2011) and (Daland *et al.* 2011). The findings presented here are consistent with their conclusions that lexicalist models can give rise to sonority projection in English, Mandarin, and Korean. Indeed, the argument emphasised here is that the compatibility between the predictions of lexicalist models and the SSP for these languages is precisely why they are not the right languages to differentiate between these two hypotheses. This is not to say that the sonority projection findings for these languages are not a significant discovery. On the contrary, that sonority projection has been reliably demonstrated in numerous studies provides clear and consistent evidence differentiating between the Segmental Statistics class of hypotheses and the Structured Generalisation class of hypotheses, indicating that at least Structured Generalisation is necessary.

The primary contribution of this work is to expand the scope of this line of research to a language where the predictions of Structured Generalisation fall short. This opens the door to a deeper and more nuanced understanding of how universal bias may interact with Structured Generalisation. The developmental findings presented here provide initial evidence that reference to a universal bias of some sort is needed. Just as importantly, however, this paper highlights sonority sequencing in Polish as an important test case for further investigation. More generally, the results here highlight the importance and utility of integrating computational modelling with behavioural data on human learning. This paper has argued that at least Structured Generalisation is necessary to explain existing results on human learning of phonology and phonotactics and has presented initial evidence that Structured Generalisation should be further enhanced with a universal bias. Given the complexity of these learning models, it is only by understanding the predictions of these and other explicit computational models that concrete hypotheses about learning can be tested and compared. Sonority sequencing in Polish provides an excellent example. It is only through generating predictions of explicit models positing a particular balance between the role of statistics, abstract representations, and substantive biases that the ways in which existing models must be refined can be understood. While the present results have identified a potential universal bias at work in phonological learning in Polish, much remains to be discovered about the nature of this bias, the role it plays in various behavioural tasks, and how its effects unfold over the course of phonological acquisition. Assuming subsequent behavioural investigations confirm the effect of SSP reliably shows up in early productions of children acquiring Polish, it remains to be determined how or if this effect also affects perception, acceptability, and other tasks, and how adult learners ultimately reconcile the preferences of the universal SSP scale with their language experience. Do adult Polish speakers exhibit sonority projection effects for unattested, but licit clusters? Given that there is evidence for an active role of SSP in the phonology of Polish, these results also motivate further examination of how exactly adult speakers' productive knowledge of these processes depends on the combined effects of the SSP and the input statistics. A mutually-informing link between computational modelling and behavioural studies are key to further understanding of these and other essential questions about how phonological learning works.

References

Albright, Adam. 2009. Feature-based generalisation as a source of gradient acceptability. *Phonology* 26 (01): 9-41.

Albright, Adam & Bruce Hayes. 2002. Modeling english past tense intuitions with minimal generalization. In Ed. Michael Maxwell, *Proceedings of the Sixth Meeting of the ACL Special Interest Group in Computational Phonology (SIGPHON)*. Philadelphia: ACL.

Bailey, Todd M & Ulrike Hahn. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44 (4): 568-591.

Bane, Max, Peter Graff & Morgan Sonderegger. 2010. Longitudinal phonetic variation in a closed system. *Chicago Ling. Soc* 46.

Batchelder, Eleanor Olds. 2002. Bootstrapping the lexicon: A computational model of infant speech segmentation. *Cognition* 83 (2): 167 - 206.

Becker, M, A Nevins & N Ketrez. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in turkish laryngeal alternations. *Language* 87 (1): 84-125.

Becker, Michael, Andrew Nevins & Jonathan Levine. 2012. Asymmetries in generalizing alternations to and from initial syllables. *Language* 88 (2): 231-268.

Berent, I, T Lennertz & M Rosselli. 2012. Universal phonological restrictions and language-specific repairs: Evidence from spanish. *The Mental Lexicon* 13:275-305.

Berent, Iris. 2008. Are phonological representations of printed and spoken language isomorphic? Evidence from the restrictions on unattested onsets. *Journal of Experimental Psychology: Human Perception and Performance* 34 (5): 1288.

Berent, Iris, Donca Steriade, Tracy Lennertz & Vered Vaknin. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104 (3): 591-630.

Berent, Iris, Katherine Harder & Tracy Lennertz. 2011. Phonological universals in early childhood: Evidence from sonority restrictions. *Language Acquisition* 18 (4): 281-293.

Berent, Iris, Tracy Lennertz, Jongho Jun, Miguel A Moreno & Paul Smolensky. 2008. Language universals in human brains. *Proceedings of the National Academy of Sciences* 105 (14): 5321-5325.

Berent, Iris, Tracy Lennertz, Paul Smolensky & Vered Vaknin-Nusbaum. 2009. Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology* 26 (01): 75-108.

Bethin, Christina Y. 1984. Voicing assimilation in polish. *International Journal of Slavic Linguistics and Poetics* 29:17-32.

Bethin, Christina Y. 1987. Syllable structure and the polish imperative desinence. *The Slavic and East European Journal* 31 (1): 76-89.

Bethin, Christina Y. 1992. *Polish Syllables: The Role of Prosody in Phonology and Morphology.* Slavica Pub.

Blanchard, Daniel, Jeffrey Heinz & Roberta Golinkoff. 2010. Modeling the contribution of phonotactic cues to the problem of word segmentation. *Journal of Child Language* First View:1-25.

Brent, Michael R & Timothy A Cartwright. 1996. Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition* 61 (1-2): 93 - 125.

Broselow, Ellen & Daniel Finer. 1991. Parameter setting in second language phonology and syntax. *Second Language Research* 7 (1): 35-59.

Broselow, Ellen, Su-I Chen & Chilin Wang. 1998. The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition* 20 (02): 261-280.

Clements, George. 1990. The role of the sonority cycle in core syllabification. In John Kingston & Mary Beckmann (eds), *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*. Cambridge: Cambridge University Press.

Coady, Jeffry A & Richard N Aslin. 2004. Young children's sensitivity to probabilistic phonotactics in the developing lexicon. *Journal of Experimental Child Psychology* 89 (3): 183 - 213.

Coleman, John & Janet Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. *ArXiv Preprint Cmp-lg/9707017*.

Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis & Ingrid Norrmann. 2011. Explaining sonority projection effects. *Phonology* 28 (02): 197-234.

Davidson, Lisa. 2006. Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics* 34 (1): 104 - 137.

Demenko, Grażyna, Mikolaj Wypych & Emilia Baranowska. 2003. Implementation of grapheme-to-phoneme rules and extended SAMPA alphabet in polish text-to-speech synthesis. *Speech and Language Technology* 7:79-97.

Demuth, Katherine. 1995. Markedness and the development of prosodic structure. *Proceedings of the North East Linguistic Society*. GSLA University of Massachusetts, Amherst.

Duanmu, San. 2002. *The Phonology of Standard Chinese.* Oxford University Press, Usa.

Edwards, Jan & Mary E Beckman. 2008. Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language Learning & Development* 4 (2): 122-156.

Ernestus, M & R H Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in dutch. *Language*5-38.

Fee, Jane & David Ingram. 1982. Reduplication as a strategy of phonological development. *Journal of Child Language* 9 (01): 41-54.

Fikkert, Paula. 1994. On the acquisition of prosodic structure. University of Leiden.

Frisch, Stefan A, Janet B Pierrehumbert & Michael B Broe. 2004. Similarity avoidance and the OCP. *Natural Language & Linguistic Theory* 22 (1): 179-228.

Frisch, Stefan A, Nathan R Large & David B Pisoni. 2000. Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language* 42 (4): 481-496.

Gelman, Andrew & Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/hierarchical Models.* Cambridge University Press.

Gnanadesikan, Amelia. 1995. Markedness and faithfulness constraints in child phonology. In René Kager Joe Pater & Wim Zonneveld (eds), *Constraints in Phonological Acquisition*. Cambridge: Cambridge University Press, 2004.

Goldwater, Sharon, Thomas L Griffiths & Mark Johnson. 2009. A bayesian framework for word segmentation: Exploring the effects of context. *Cognition* 112 (1): 21 - 54.

Gussmann, Edmund. 1992. Resyllabification and delinking: The case of polish voicing. *Linguistic Inquiry* 23 (1): 29-56.

Haman, E, B Etenkowski, M Łuniewska, J Szwabe, E Dąbrowska, M Szreder & M Łaziński. 2011. *Polish CDS Corpus.* Alvailable from http://childes.psy.cmu.edu/data/Slavic/Polish/CDS.zip.

Harrell, F E. 2014. RMS: Regression modeling strategies: R package version 4.1-3. available at http://biostat.mc.vanderbilt.edu/wiki/Main/RmS.

Hayes, B, K Zuraw, P Siptár & Z Londe. 2009. Natural and unnatural constraints in hungarian vowel harmony. *Language* 85 (4): 822-863.

Hayes, Bruce. 1999. Phonetically driven phonology: The role of optimality theory and inductive grounding. In Michael Darnell Edith Moravcsik Frederick Newmeyer Michael Noonan & Kathleen Wheatley (eds), *Formalism and Functionalism in Linguistics, Volume 1: General Papers*. Amsterdam: John Benjamins.

Hayes, Bruce. 2011. Interpreting sonority-projection experiments: The role of phonotactic modeling. *Proceedings of the 17th International Congress of Phonetic Sciences.*

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39 (3): 379-440.

Hayes, Bruce & James White. 2013. Phonological naturalness and phonotactic learning. *Linguistic Inquiry* 44:45-75.

Hayes, Bruce & Zsuzsa Cziraky Londe. 2006. Stochastic phonological knowledge: The case of hungarian vowel harmony. *Phonology* 23 (01): 59-104.

Hayes, Bruce, Robert Kirchner & Donca Steriade. 2004. *Phonetically Based Phonology.* Cambridge University Press.

Ingram, David. 1988. The acquisition of word-initial [v]. *Language and Speech* 31 (1): 77-85.

Jaeger, T Florian. 2008. Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59 (4): 434-446.

Jaeger, T Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61 (1): 23-62.

Jakobson, Roman. 1941. *Child Language, Aphasia and Phonological Universals.* The Netherlands: Mouton, The Hague, 1968.

Jarosz, Gaja. 2006. Rich lexicons and restrictive grammars - maximum likelihood learning in optimality theory. Ph.D. Thesis

Jarosz, Gaja. 2010. Implicational markedness and frequency in constraint-based computational models of phonological learning. *Journal of Child Language. Special Issue on Computational Models of Child Language Learning* 37 (3): 565-606.

Jarosz, Gaja & J. Alex Johnson. 2013a. The richness of distributional cues to word boundaries in speech to young children. *Language Learning and Development* 9 (2): 175-210.

Jarosz, Gaja & J Alex Johnson. 2013b. The richness of distributional cues to word boundaries in speech to young children. *Language Learning and Development* 9 (2): 175-210.

Jarosz, Gaja, Shira Calamaro & Jason Zentz. to appear. Input frequency and the acquisition of syllable structure in polish. *Language Acquisition*.

Jespersen, Otto. 1904. *Lehrbuch Der Phonetik.* Leipzig & Berlin: Teubner.

Jurafsky, Daniel & James H Martin. 2008. *Speech & Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* Second edition ed. Prentice Hall.

Jusczyk, Peter W., Paul A. Luce & Jan Charles-Luce. 1994. Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language* 33 (5): 630 - 645.

Kager, René & Joe Pater. 2012. Phonotactics as phonology: Knowledge of a complex restriction in dutch. *Phonology* 29 (01): 81-111.

Kirk, Cecilia & Katherine Demuth. 2005. Asymmetries in the acquisition of word-initial and word-final consonant clusters. *Journal of Child Language* 32 (04): 709-734.

Lee, Yongsung. 1994. Onset analysis of korean on-glides. *Theoretical Issues in Korean Linguistics*133.

Lennertz, Tracy & Iris Berent. 2015. On the sonority levels of fricatives and stops. *The Mental Lexicon* 10 (1): 88-132.

Levelt, Claartje & Ruben van de Vijver. 1998. Syllable types in cross-linguistic and developmental grammars. *Constraints in Phonological Acquisition*. Cambridge: Cambridge University Press, 2004.

Levelt, Claartje, Niels Schiller & Willem Levelt. 2000. The acquisition of syllable types. *Language Acquisition* 8:237-264.

Łukaszewicz, Beata. 2006. Extrasyllabicity, transparency and prosodic constituency in the acquisition of polish. *Lingua* 116 (1): 1 - 30.

MacWhinney, B. 2000. *The CHILDES Project: Tools for Analyzing Talk.* Lawrence Erlbaum Associates.

Mattys, Sven L & Peter W Jusczyk. 2001. Phonotactic cues for segmentation of fluent speech by infants. *Cognition* 78 (2): 91 - 121.

Pater, Joe & Jessica Barlow. 2003. Constraint conflict in cluster reduction. *Journal of Child Language* 30 (03): 487-526.

Prince, Alan & Paul Smolensky. 1993. Optimality theory : Constraint interaction in generative grammar . Malden, MA : Blackwell Pub, 2004.

Ren, Jie, Liqun Gao & James L Morgan. 2010. Mandarin speaker's knowledge of the sonority sequencing principle. *20th Colloquium on Generative Grammar, University of Pompeu Fabra, Barcelona.*

Roark, Brain & Katherine Demuth. 2000. Prosodic constraints and the learner's environment: A corpus study. In S. C. Howell Sarah S. A. Fish & Thea Keith-Lucas (eds), *Proceedings of the 24th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Roland, Douglas, Jeffrey L Elman & Victor S Ferreira. 2006. Why is that? Structural prediction and ambiguity resolution in a very large corpus of english sentences. *Cognition* 98 (3): 245-272.

Rubach, Jerzy & Geert Booij. 1990. Syllable structure assignment in polish. *Phonology* 7 (1): 121-158.

Rubach, Jerzy & Geert E Booij. 1985. A grid theory of stress in polish. *Lingua* 66 (4): 281 - 320.

Rubach, Jerzy & Geert E Booij. 1990. Edge of constituent effects in polish. *Natural Language & Linguistic Theory* 8 (3): 427-463.

Saffran, Jenny R, Richard N Aslin & Elissa L Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274 (5294): 1926-1928.

Selkirk, Elisabeth. 1984. On the major class features and syllable theory. In Mark Aronoff & Richard R. T. Oehrle (eds), *Language Sound Structure*. Cambridge MA: MIT Press.

Sievers, Eduard. 1881. *Grundzüge Der Phonetik Zur Einführung in Das Studium Der Lautlehre Der Indogermanischen Sprachen.* Leipzig: Breitkopf & Hartel.

Stampe, David. 1969. The acquisition of phonetic representations. *Papers From the Fifth Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society.

Steriade, Donca. 1982. Greek prosodies and the nature of syllabification. PhD Dissertation, MIT.

Stites, J, K Demuth & C Kirk. 2004. Markedness vs. Frequency effects in coda acquisition. *Proceedings of the Boston University Conference on Language Development.*

Tamási, Katalin & Iris Berent. 2014. Sensitivity to phonological universals: The case of stops and fricatives. *Journal of Psycholinguistic Research* 44 (4): 359-381.

Vihman, Marilyn May. 1993. Variable paths to early word production. *Journal of Phonetics* 21:61-82.

Vitevitch, Michael S & Paul A Luce. 2004. A web-based interface to calculate phonotactic probability for words and nonwords in english. *Behavior Research Methods, Instruments, & Computers* 36 (3): 481-487.

Vitevitch, Michael S, Paul A Luce, Jan Charles-Luce & David Kemmerer. 1997. Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech* 40 (1): 47-62.

Weist, Richard M & Katarzyna Witkowska-Stadnik. 1986. Basic relations in child language and the word order myth. *International Journal of Psychology* 21 (1-4): 363-381.

Weist, R M, H Wysocka, K Witkowska-Stadnik, E Buczowska & E Konieczna. 1984. The defective tense hypothesis: On the emergence of tense and aspect in child polish. *Journal of Child Language* 11 (02): 347-374.

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30 (5): 945-982.

Zamuner, Tania S. 2009. Phonotactic probabilities at the onset of language development: Speech production and word position. *Journal of Speech, Language & Hearing Research* 52 (1): 49-60.

Zamuner, Tania S, LouAnn Gerken & Michael Hammond. 2005. The acquisition of phonology based on input: A closer look at the relation of cross-linguistic and child language data. *Lingua* 115 (10): 1403 - 1426.

Zamuner, T S, L A Gerken & M Hammond. 2004. Phonotactic probabilities in young children's speech production. *Journal of Child Language* 31 (3): 515-536.

Zhao, Xu & Iris Berent. 2015. Universal restrictions on syllable structure: Evidence from mandarin chinese. *Journal of Psycholinguistic Research*1-17.

Zuraw, Kie. 2000. Patterned exceptions in phonology. Ph.D. Thesis

Zydorowicz, Paulina. 2007. Polish morphonotactics in first language acquisition. Ed. Florian Menz and Marcus Rheindorf. *Wiener Linguistische Gazette*, no. 74:24-44.

7. Appendix

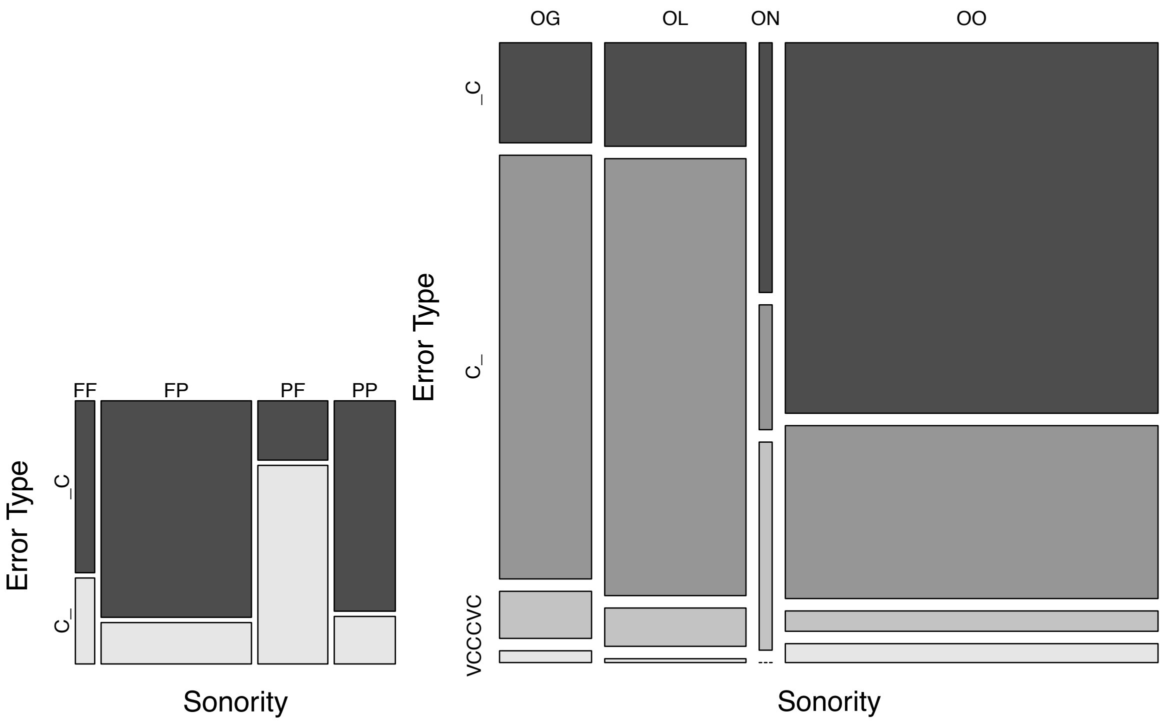

Figure 11 Error Types by Finer-Grained (left) and Coarser-Grained (right) Sonority

(9) Initial Segment Bigram Type Frequency in CDS

pʂ	1522	zm	117	xw	49	zl	29	xm	17	ʐw	9	wz	5	tn	2
st	427	ʂt͡ʂ	116	ʂp	48	ml	29	zʐ	17	d͡ʑv	9	t͡sn	5	ft͡s	2
kr	411	zn	107	zɲ	47	d͡zv	28	ft͡ʂ	17	tl	9	ʐm	4	vd͡ʑ	2
sp	370	sm	106	xr	46	kt	28	dm	16	sn	9	st͡ʂ	4	t͡ʂx	2
pj	365	kw	102	tw	46	pɕ	27	t͡ɕf	15	sl	9	db	4	bʑ	2
vj	344	tʂ	99	kf	46	pt	27	ʑr	15	dɲ	8	tɲ	4	vm	2
pr	340	ɕt͡ɕ	94	ɕɲ	46	fʂ	27	ʂf	15	vz	8	pɲ	4	mx	2
tr	266	ɕl	89	ɕr	44	fɕ	27	vb	15	rv	8	fk	4	ln	2
sk	257	zb	88	tf	42	vʑ	25	mɲ	15	dl	8	st͡ɕ	4	kt͡ɕ	2
gr	249	zg	86	mw	42	ft͡ɕ	25	zd͡ʑ	14	ʂw	8	xj	3	lɲ	2
mj	248	vr	86	xl	40	ʂn	24	px	13	d͡zb	8	ʐv	3	vɲ	1
sw	227	zr	82	dw	40	ʂm	23	ɕp	13	mn	8	t͡ɕm	3	ʐb	1
br	206	zj	78	vʐ	39	dʐ	23	t͡ʂw	12	t͡sf	8	rd͡z	3	wb	1
kl	196	zv	78	sf	38	ʂt	23	st͡s	12	t͡sm	8	ʐɲ	3	wg	1
dr	190	kʲj	75	fl	38	mr	22	zgʲ	12	lv	8	rj	3	zz	1
pw	175	kɕ	74	gɲ	37	t͡ʂf	22	gʲj	12	gn	7	lʐ	3	ʐn	1
pl	172	bl	73	ʂk	36	dj	20	ft	12	xt͡s	7	t͡ʂk	3	pt͡ɕ	1
vw	146	fr	67	fs	35	sr	20	xʂ	11	tj	7	t͡ʂt͡ɕ	3	km	1
zw	140	sx	60	ɕf	34	fx	20	tx	11	tk	7	ʂr	3	mʐ	1
bj	133	fp	59	xf	33	vn	20	ɕm	11	mʂ	7	rʐ	2	d͡ʐd͡ʐ	1
gw	121	dv	58	vl	33	gd	19	vd	11	kn	6	gm	2	vv	1
kʂ	121	ps	54	xt͡ɕ	31	ʂl	18	ss	10	bz	6	rt	2	vg	1
zd	119	bw	52	gv	29	t͡ʂt	18	gd͡ʑ	10	sʂ	6	ʑl	2	lj	1
gʐ	118	bʐ	51	gl	29	fj	18	sɕ	10	gz	5	dn	2		