

ARTICLE

Trochaic bias overrides stress typicality in English lexical development

Klaus HOFMANN* and Andreas BAUMANN

University of Vienna, Austria

*Corresponding author: Department of English and American Studies, Spitalgasse 2-4, Hof 8.3, University of Vienna, Austria Vienna, 1090 Austria. E-mail: klaus.hofmann@univie.ac.at

(Received 16 October 2018; revised 14 December 2019; accepted 11 June 2020;
first published online 7 October 2020)

Abstract

This paper investigates whether typical stress patterns in English nouns and verbs are available as a prosodic cue for categorisation and accelerated word learning during first language acquisition. The stress typicality hypothesis states that left-stressed nouns and right-stressed verbs should be acquired earlier than the reverse configurations if stress effectively signals lexical class membership. In this view, class-typical stress patterns are expected to facilitate learning of novel items. A series of generalized additive models (GAMs) based on a comprehensive set of lexical data (CELEX) as well as a large set of age-of-acquisition (AoA) and concreteness ratings reveals that stress typicality plays a minor role in early acquisition, as it is generally superseded by a preference for left-hand (or ‘trochaic’) patterns in both nouns and verbs. This may be explained by general cognitive constraints (perceptual salience and recency) or exposure to the dominant pattern in the ambient language.

Keywords: phonological typicality; age of acquisition; stress; trochaic bias

1 Introduction

English word stress placement has received a fair amount of attention in the linguistic literature, both synchronically (Chomsky & Halle, 1968; Hayes, 1982; Burzio, 1994) and diachronically (Halle & Keyser, 1971; Lass, 1992; Minkova, 1997). Yet, one of the most striking features of English stress has remained markedly underexplored. This is the differential behaviour of disyllabic and polysyllabic nouns and verbs regarding the position of primary stress. Nouns display a strong tendency to be stressed on syllables at, or closer to, the beginnings of words (‘left-hand stress’), while verbs tend to be stressed word-finally (‘right-hand stress’). This paper investigates the potential of the stress contrast as an indicator of word class (McCully, 2002). Specifically, it pursues the question whether class-specific stress can assist language learners in their acquisition of nouns and verbs, respectively, by acting as a phonological cue for their categorisation and storage in the mental lexicon (henceforth ‘stress typicality hypothesis’). Stress typicality represents a language-internal mechanism as it covers

© The Author(s), 2020. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

the (language-specific) interaction between two linguistic domains (phonology and morpho-syntax). The stress typicality hypothesis is contrasted with two alternative hypotheses based on reported learning constraints: namely, the nominal and trochaic biases. Our study draws on an extensive database of Age-of-Acquisition (AoA) ratings (Kuperman, Stadthagen-Gonzalez & Brysbaert, 2012) enriched with lexical and phonological information supplied by the lexical database CELEX (Baayen, Piepenbrock & Gulikers, 2001). As we will show, the analysis of these data speaks against a stress typicality bias and instead attributes a prominent role to a trochaic bias. We propose that a trochaic bias in language acquisition represents the combined effect of two more general cognitive pressures: left-stressed (i.e., 'trochaic') items are preferred because their stressed syllables benefit from perceptual salience while their unstressed syllables benefit from recency.

The remainder of this paper is structured as follows. Section 1 provides an overview of the noun-verb stress contrast in English and formulates the hypotheses to be tested on acquisition data. Section 2 introduces the dataset and lays out the modelling procedure. In Section 3, the empirical results are presented, and their implications for the relationship between prosodic learning biases and lexical acquisition are discussed in Section 4. Section 5 sums up the main findings.

1.1 The noun-verb stress contrast in English

The point of departure is the observation that English exhibits a strong correlation between patterns of word prosody and morpho-syntactic class. Nouns are characterised by left-hand stress, while verbs tend towards right-hand stress. The stress difference is independent of phonological properties, notably syllable weight, as evidenced in examples where nouns and verbs display parallel syllable structures but asymmetrical stress patterns (e.g., *incest* [N] vs. *infést* [V]). This becomes most conspicuous in noun-verb homographs that are identical apart from their stress patterns and ensuing phonological reductions (e.g., *cóntest* vs. *contést*).

The most common explanation for the stress difference in the generative literature is that nouns and verbs follow different extrametricality rules during metrical foot construction. Depending on word class, different parts are 'invisible' to a stress placement rule (final syllables in nouns, final consonants in verbs; Chomsky & Halle, 1968; Hayes, 1982; Burzio, 1994). Thereby, the problem essentially becomes a matter of lexical indexation and thus uninteresting to generative accounts. In contrast, the point of view of this paper is that the ubiquity of the stress contrast justifies further investigation: for example, from a functional perspective.

Dictionary analyses confirmed that differential stress patterns are indeed a pervasive feature of the English vocabulary (cf. Table 1). Kelly and Bock (1988) found that disyllabic nouns exhibit left-hand stress at a rate of up to 94 per cent, while disyllabic verbs are left-stressed at only 31 per cent. In verbs introduced to English since 1700, left-hand stress is even more uncommon at as little as 5 per cent (McCully, 2002).¹ More important for our purposes is the converse finding that up to 95 per cent of all left-stressed items in Kelly and Bock's (1988) sample turned out to be nouns, while up to 85 per cent of right-stressed items were verbs. In other words, if one were to guess the word class from a word's stress pattern, the success rate of this criterion would be very high.

¹These numbers exclude conversion and other types of class-shifting derivation.

Table 1. Dictionary analyses

	Sample	Nouns (left-stressed)	Verbs (left-stressed)
Kelly and Bock (1988)	Disyllabic nouns and verbs (excl. conversion)	89% (94%)	46% (31%)
McCully (2002)	Polysyllabic verbs 1700 (excl. noun derivations)	---	37% (5%)
Present study (cf. 2.1)	Disyllabic nouns and verbs	92%	29%

The stress contrast is not only pervasive but also productive. Studies investigating pronunciation guidelines in historical dictionaries report that the stress-distinction has been increasing in homographs (Sherman, 1975; cf. Sonderegger & Niyogi, 2010). Experimental evidence suggests that disyllabic pseudowords (Kelly & Bock, 1988) and categorically ambiguous words (Serenio & Jongman, 1995) receive left-hand stress more often when interpreted as nouns, while the opposite is true if words are interpreted as verbs. Guion, Clark, Harada and Wayland (2003) found that for disyllabic pseudowords lexical class is by far the strongest predictor of stress placement in a multifactorial logistic regression model.

Finally, the noun-verb stress contrast also seems to be cognitively functional as it facilitates online processing, particularly in individuals with limited vocabularies (Kawamoto, Farrar & Overbeek, 1993, cited in Serenio & Jongman, 1995; Arciuli & Cupples, 2003, 2006). These findings lend empirical support to McCully's (2002: 343) assessment that "[s]tress marking is being used signally [...]".

1.2 Word class and stress in first language acquisition

Given the extent of the noun-verb stress contrast in English, and considering its productivity and effects on processing, one remaining question is under what circumstances the purported stress marking of morpho-syntactic class becomes functionally relevant. One possible answer is that stress typicality may facilitate the parsing and acquisition of unknown words. This hypothesis can be seen in the context of a growing body of research showing that phonology plays an active role in lexical development. For instance, one of the benefits of prosodic regularities in the vocabulary lies in their capacity to signal word boundaries and thereby bootstrap lexical acquisition during infancy (Cutler & Norris, 1988; Jusczyk, Houston & Newsome, 1999; Thiessen & Saffran, 2003, 2007). Crucially, phonological characteristics of words assist pre-school children in forming hypotheses about the syntactic class of new lexical items. Various phonological features have been proposed as cues for distinguishing word classes in this way, including word length, syllable number, vowel type and stress pattern (Kelly, 1992; Monaghan, Chater & Christiansen, 2005). It has also been demonstrated that class-typical phonological features can influence the learnability of unfamiliar words (Cassidy & Kelly, 2001; Fitneva, Christiansen & Monaghan, 2009).

The noun-verb stress contrast could serve a similar end: namely, to facilitate categorisation and memorisation. Two empirical expectations regarding acquisition

data follow from this. First, given its spread in the English vocabulary, the noun-verb stress contrast is expected to figure prominently in data representing children's growing vocabularies. That is, word class and stress pattern should correlate in the lexicon of first language learners as in the lexicon generally – otherwise the stress contrast would hardly function as a cue to begin with. Second, if the stress contrast is utilised as a phonological cue fast-tracking acquisition, nouns and verbs are expected to be acquired faster when they conform to their typical stress patterns compared to class-atypical patterns. Left-stressed nouns and right-stressed verbs should, on average, be learned before right-stressed nouns and left-stressed verbs, even when confounding factors, such as lexical token frequencies, are considered.

In addition to, and contrasting with, the stress typicality hypothesis just outlined, there are other logical possibilities of how word class and word prosody could relate to lexical development. The order of acquisition might be simply random regarding these parameters. In that case, class and stress would either not correlate with age and order of word acquisition at all, or any such relationship would be sufficiently explained as a corollary of other factors.

However, other scenarios involving word class and stress pattern are also conceivable. First, a well-documented observation in word learning is that in early vocabularies nouns generally dominate (Gentner, 1978; Bornstein & Cote, 2005). This nominal bias has been attributed to prototypical noun referents (i.e., objects) being easier to conceptualise than verb referents (i.e., actions) (Gentner, 1982; McDonough, Song, Hirsh-Pasek, Golinkoff & Lannon, 2011).

A second learning bias reported in the literature is the preference for one prosodic pattern over the other. In children's speech, a left-hand stress pattern ('trochee') appears to function as a default template when the complexity of an utterance surpasses a child's planning capacity (McGregor & Johnson, 1997). Studies have shown that children indeed perform markedly better in the production of left-hand patterns compared to right-hand patterns (James, Van Doorn & McLeod, 2008; Ballard, Djaja, Arciuli, James & van Doorn, 2012). These results are in accordance with the trochaic bias hypothesis, which states that strong-weak syllable sequences are favoured by a universal rhythmic constraint (Allen & Hawkins, 1980). Yet, the literature is divided on the existence of such a constraint (Snow, 1998; Vihman, DePaolis & Davis, 1998; Rose & Champdoizeau, 2008; Adam & Bat-El, 2009). One methodological problem is that studies are often based on languages in which trochees dominate in absolute numbers (such as English; Cutler & Carter, 1987; Vihman et al., 1998; Bijeljac-Babic, Hohle & Nazzi, 2016). Thus, an acquisition preference could simply be due to greater exposure rather than a cognitive predisposition.

Nonetheless, the existence of a cognitive bias is plausible on perceptual grounds. Word-prosodic tendencies in a language may be interpreted as the outcome of interacting domain-general cognitive constraints. In this view, a sequence of a strong and a weak syllable represents the optimal trade-off between two well-reported perception and processing biases: namely, salience and recency. Salience generally describes the phenomenon by which entities which stand out from their background are perceived more easily and retained more stably in working memory (Pedale & Santangelo, 2015). Applied to phonological prominence, this principle suggests that the sonority events represented by stressed syllables stand out more prominently from the rest of the speech stream by virtue of phonetic properties such as loudness, length and pitch (Lehiste, 1970), and can be segmented and recognised more successfully. This prioritises them for a function as carriers of semantic content.

Various researchers have suggested that signalling the location where the lexical access process should be initiated is one of the main functions of stress for language acquisition (Cutler & Norris, 1988; Echols & Newport, 1992; Jusczyk et al., 1999; Thiessen & Saffran, 2003, 2007). The second principle, i.e., recency, refers to a processing advantage due to temporal immediacy rather than prominence (Slobin, 1973). Other things being equal, more recent events are better represented in memory than earlier ones. Recency is a defining characteristic of the 'phonological loop': a specialised, fast-decaying type of working memory, whose main function has been suggested to be word learning (Baddeley, Gathercole & Papagno, 1998). Additionally, structural finality is often assisted by other prosodic cues such as lengthening, especially in child-directed speech, which once again renders these sonority events more prominent (Shady & Gerken, 1999).

In left-stressed disyllabic words both principles apply. The first syllable is stressed, thus serving as a salient prominence event. The second syllable is unstressed, but as it constitutes the more recent sonority peak, often exaggerated phrase-finally, it is held longer in the phonological loop. Thereby, both syllables may attain a higher likelihood of successful transmission to longer-term memorisation and processing compared to unstressed earlier sonority events. In iambic words, on the other hand, prominent and recent sonority events coincide in the same syllable, thus impeding successful perception of the whole sequence and thwarting its transmission to more permanent storage.

The existence of such a cognitive bias would go some way towards explaining the uneven distribution of prosodic systems in the world's languages, where trochaic patterns form the overwhelming majority (Goedemans & van der Hulst, 2013). At the same time, it raises the question why iambic systems even exist if they are at a cognitive disadvantage. First, it needs to be acknowledged that prosodic prominence can be conveyed by a range of different phonetic cues, including amplitude, pitch and length (Lehiste, 1970). In English, these cues converge in the phenomenon commonly referred to as 'stress'. Second, there seems to be a typological correlation between word prosody patterns and the cue used to signal them. Hayes (1995) claims that trochaic systems primarily rely on amplitude as a marker of word prominence (as in English), while iambic systems mark prominence by phonological length. Importantly, trochaic systems have turned into iambic or mixed systems through processes that are historically contingent. Latin, for example, changed from a trochaic left-stress system to a quantity-sensitive system where stress became calculated from the right edge (Bybee, Chakraborti, Jung & Scheibman, 1998). In the case of Latin, a second cognitive bias conflicted with and eventually outranked the trochaic preference: namely, the perceptual constraint to perceive structurally heavy syllables as more prominent than light ones.

1.3 Empirical predictions

If the stipulations of either the nominal learning bias or the trochaic bias, or both, are accurate, they should materialise in empirical acquisition data in a way that conflicts with the stress typicality hypothesis. Instead of words with class-typical stress patterns being acquired first, the nominal bias suggests that nouns are acquired before verbs, while the trochaic bias predicts that left-stressed words are acquired before right-stressed words, irrespective of stress pattern.

Table 2. Logical possibilities of the interplay between word class and stress pattern as predictors of acquisition order. Braces indicate no expected difference between two factors (ini=initial stress, fin=final stress, see Section 2.1., fn. 4).

Hypothesis	Acquisition order
a Stress typicality	{ N ini, V fin } < { N fin, V ini } or N ini < { V fin, N fin } < V ini
b Nominal bias	{ N ini, N fin } < { V ini, V fin }
c Trochaic bias	{ N ini, V ini } < { N fin, V fin }
d Nominal and trochaic bias	N ini < { V ini, N fin } < V fin
e No effect	{ N ini, V ini, N fin, V fin }

The foregoing discussion generates a number of logical possibilities regarding the order in which left-stressed and right-stressed nouns and verbs are acquired by first language learners (Table 2): (a) in line with the stress typicality hypothesis, left-stressed nouns and right-stressed verbs are generally acquired before the reverse combinations; (b) in line with the nominal bias hypothesis, nouns of any stress pattern are generally acquired before verbs; (c) in line with the trochaic bias hypothesis, left-stressed words of any class are generally acquired before right-stressed words; (d) the nominal and the trochaic biases both apply, so that nouns are generally acquired before verbs and left-stressed words are acquired before right-stressed patterns; (e) grammatical class and stress pattern do not influence word acquisition, i.e., words are acquired randomly with regard to these parameters.

1.4 Additional factors

Word class, stress pattern and age of acquisition all correlate with several other variables, which could potentially distort the outcome of an empirical study. In order to isolate the effects of class/stress pairings on lexical acquisition data, some of these factors need to be taken into account.

First on this list is the frequency of occurrence of the individual lexical items ('token frequency'). *Ceteris paribus*, the more frequent a form, the earlier it is learned (Ambridge, Kidd, Rowland & Theakston, 2015). Since word class or stress pattern may be associated with different segments of the frequency spectrum, token frequency needs to be controlled.

Second, the learnability of new lexical items crucially depends on how accessible the referents are to perception. In short, it is easier to store new words in memory when the objects, actions or concepts they denote can also be seen, heard, touched or otherwise perceived with one's senses, distinguishing such items from abstract concepts (Brybaert, Warriner & Kuperman, 2014). Since it has been suggested that the nominal bias in lexical acquisition is largely a result of easier imageability of noun referents (i.e., objects) compared to prototypical verb referents (i.e., actions) (McDonough et al., 2011), perception-based measures of ease of conceptualisation are also worth including in the present study. There are several measures that capture perception effects, among which imageability and concreteness are the most prominent. Although there are some important differences between the two (Richardson, 1976), it is generally agreed that they are closely related, not least

because in either the visual mode of perception is often foregrounded (Connell & Lynott, 2012). To remedy this dominance, Brysbaert et al. (2014) provide a comprehensive list of concreteness measures in which the impact of different senses on the elicited ratings is more balanced.

Third, morphological complexity has been identified as a factor influencing the learnability of words. Morphologically simple words are assumed to be acquired before morphologically complex ones because they involve fewer meaning elements and are thus easier to decode (Anglin, 1993). Possible correlations between morphological complexity and word class or stress pattern may affect the interpretation of the data and therefore need to be considered as well.

Fourth, word length has been named as one of the phonological cues for distinguishing nouns from verbs (Cassidy & Kelly, 2001; Monaghan et al., 2005). Longer words are also more difficult to memorise (Hulme et al., 2006). Moreover, it is worth noting that word length correlates positively with morphological complexity (Anglin, 1993) and negatively with word frequency (Zipf, 1935).

Finally, syllable structure holds a prominent place within the study of prosody. Generative accounts derive the position of stress from the segmental weight of syllables in some form or other (Chomsky & Halle, 1968; Hayes, 1982; Burzio, 1994). The importance of this factor has been confirmed by experimental and corpus-based methods (Guion et al., 2003; Domahs, Plag & Carrol, 2014). Since different types of syllables may predominate during different stages of lexical acquisition (e.g., complex syllables being acquired later, Monaghan et al., 2005), syllable structure is also worth controlling for in a multifactorial model.

2 Data and methodology

2.1 Data

Our data come from three different sources. Age-of-acquisition (AoA) data were taken from Kuperman et al. (2012), who compiled a list of age-of-acquisition ratings for about 30,000 English words in a crowd-sourced data collection effort, where participants were asked to estimate the age at which a given word was learned. The data represent estimates of receptive rather than productive lexical knowledge. However, this does not pose a problem to our research since we are interested in strategies to exploit phonological cues for forming hypotheses about the functions of new words.

AoA ratings are frequently employed to account for latencies in lexical access tasks, alongside factors such as token frequency, word length or neighbourhood effects. Although intrinsically subjective and introspective in nature, AoA ratings do correlate highly with objectively collected data (Brysbaert & Biemiller, 2017), and the two data types are also comparable regarding variance size in lexical access tasks (Ellis & Morrison, 1998; Morrison & Ellis, 2000).² In terms of size and availability, however, AoA ratings far surpass other types of acquisition data. Also, the crowd-sourced data from Kuperman et al. (2012) show high correlations with ratings obtained under laboratory conditions and can thus be regarded as reliable. The ratings underestimate vocabulary size in the early acquisition period (before about 5 or 6 years), which will

²We also checked Kuperman et al.'s (2012) against American English acquisition norms provided as part of the Wordbank project (Frank et al., 2017) and found a robust correlation for nouns and verbs (Spearman's $\rho = 0.57$) despite the fact that the latter only recorded acquisition data up to the age of 30 months.

be acknowledged in our modelling procedure (Section 3.3). However, since we are primarily interested in the relative order in which class/stress configurations are acquired (rather than absolute acquisition ages), the fact that the ages in this period are slightly off the mark does not pose a problem to our approach.

The second source of data is the CELEX database³ (Baayen et al., 2001), from which a list of English lemmas was extracted, together with corresponding frequency measures for spoken discourse ('CobS'),⁴ consonant-vowel transcriptions ('PhonCV'), stress patterns ('StrsPat'), syllable count ('SylCnt'), word class ('Class'), and morphological information ('MorphStatus' and 'StrucLab').

Finally, concreteness measures were adopted from Brysbaert et al. (2014), providing a list of 40,000 common English words. This list was preferred over others (such as imageability ratings) for two principal reasons: on the one hand, the ratings were based on the explicit effort to counterbalance the dominance of the visual sense in previous concreteness (and imageability) measures; on the other hand, in terms of size, item selection and collection procedure (via an online survey), this list is maximally similar to the AoA ratings by Kuperman et al. (2012) described above.

All data sets were combined and subsequently filtered in the following way: (i) in order to prevent overly rare words from skewing our analysis, we limited our lexical base set down to those items occurring at least once in the spoken reference corpus that informed the frequency measures provided by CELEX ('CobS'); (ii) from this subset, all disyllabic nouns and verbs were extracted; (iii) cases of conversion (N-to-V or V-to-N) are potentially harmful to the study design because stress patterns are often carried over from the original form into the new word class (e.g., *collapse*, V-to-N; *to distance* N-to-V; Kiparsky, 1982; Kelly & Bock, 1988). In order to eliminate this potential confound from the analysis, derived forms were discarded based on the morphological analysis provided by CELEX ('StrucLab'). As a consequence, 453 items were omitted, i.e., roughly 18 per cent of the data at that stage; (iv) Finally, the remaining items were collated with AoA ratings from Kuperman et al. (2012). In total, 2430 target words were retrieved and annotated in this way.

The distribution of word class and word stress among these items is shown in Table 3. Lexical development with respect to these categories is displayed in Figure 1. As noted above, AoA raters tended to underestimate the size of their earliest vocabularies, which is why acquisition data is only provided from Age 4 onwards in Figure 1b. However, the relative order of acquisition is more informative than the actual age estimates. We observe, as expected, that the disyllabic nouns in our dataset are overwhelmingly marked by left-hand ('initial')⁵ stress (92 per cent) and that this pattern dominates in nouns throughout the acquisition process (Figure 1b). More

³We used the freely available online version WebCelex (<http://celex.mpi.nl>).

⁴To ensure that the frequency measures we used are reasonably representative of the ambient language a child would be exposed to, we correlated the occurrence frequencies for nouns and verbs from CELEX (which are also available for word forms in addition to lemma frequencies) to frequencies derived from child-directed speech corpora (available on CHILDES <https://childes.talkbank.org/derived/parentfreq.cdc>, cf. MacWhinney, 2000). We found a robust correlation between the two (Pearson's $r=0.51$). The fact that the CHILDES frequency list was only available for word forms rather than lemmas whereas our AoA dataset is based on lemmas prevented us from using the CHILDES frequencies themselves.

⁵For the disyllabic items in our set the descriptors 'left-hand stress' and 'right-hand stress' are synonymous with 'initial stress' and 'final stress'. The latter labels were used in the statistical models simply for descriptive convenience.

Table 3. Distribution of word class and stress pattern in the dataset.

Word class	Left ('initial') stress	Right ('final') stress
Nouns (% of class % of stress type)	1662 (90% 92%)	178 (10% 29%)
Verbs (% of class % of stress type)	144 (24% 8%)	446 (76% 71%)

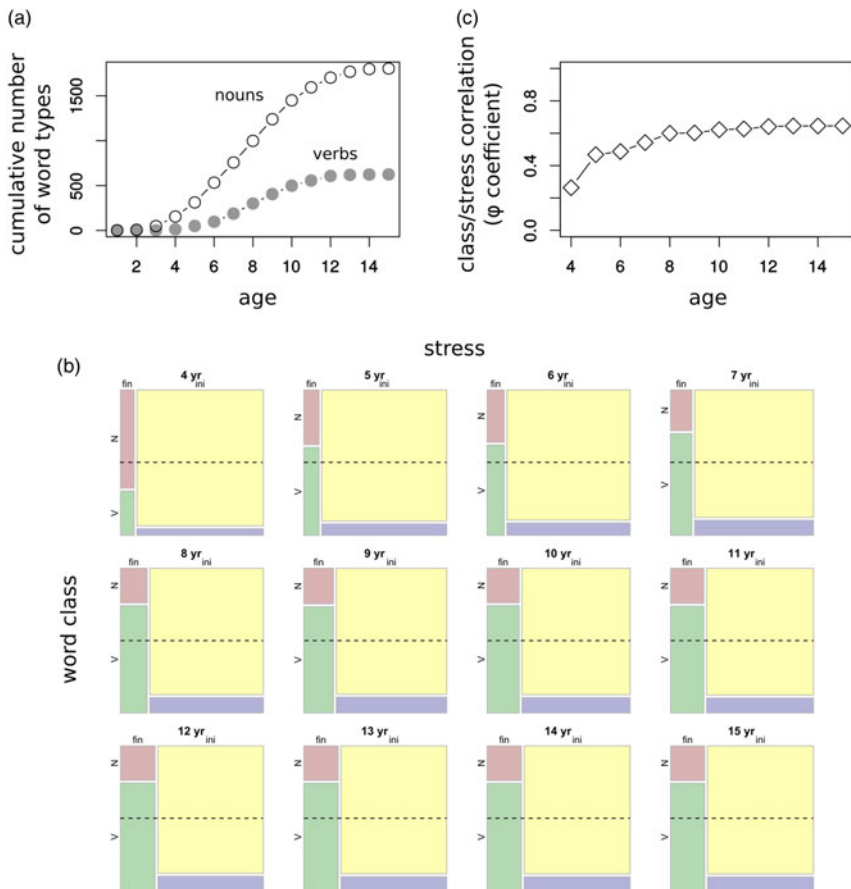


Figure 1. Descriptive account of the data: (a) Growth of acquired number of nouns (white) and verbs (grey). (b) Development of distribution of word class and stress among all acquired items per year of age shown as series of mosaic plots (from age of 4 yr onwards). Dashed line denotes equal 1:1 distribution of nouns/verbs across stress patterns. Color code: red/green: proportion of finally stressed nouns/verbs; yellow/blue: proportion of initially stressed nouns/verbs, respectively. Initial stress clearly signals nouns while final stress signals verbs from 5 yr onwards. (c) Trajectory of strength of correlation (φ -coefficient) of the respective distributions shown in (b). Over time, the correlation among word class and stress gets stronger, ultimately approaching $\varphi = 0.65$.

importantly, an initially stressed word in our dataset has a 90 per cent likelihood of being a noun. The expected relationship between verbs and right-hand (i.e., 'final') stress is also clearly attested. Overall, 71 per cent of disyllabic verbs exhibit final

stress and about three quarters of finally stressed words are verbs (76 per cent). However, the latter correlation only seems to develop over time. As the vocabulary grows, final stress becomes more clearly associated with verbs. This is reflected in an increasing phi coefficient, which measures the strength of the correlation between word class and stress pattern (Figure 1c).

These findings by themselves already represent an important addition to our understanding of the relationship between stress and language acquisition. Thus, it is intriguing to see that, while class-typical stress patterns distinguish the lexical classes well in later years, this correlation is not yet clearly established during the critical early years of acquisition. This conflicts with our expectation that the stress contrast should be especially salient in early vocabulary for it to function as a marker of word class and thereby bootstrap acquisition. However, the descriptive statistical overview does not yet take into account various other relevant factors, nor their interactions with the primary predictors. For that, more sophisticated statistical modelling is required.

2.2 Variables

Since our aim was to investigate the acquisition order of left-stressed and right-stressed disyllabic nouns and verbs, we first assigned labels to 2430 disyllabic items in our dataset representing the four configural types of our main predictor variable *class/stress pattern* (N ini, N fin, V ini, V fin). As described in the previous section, we adopted the age-of-acquisition ratings provided by Kuperman et al. (2012) for AoA measures for each entry of each type. We observed that AoA scores were distributed in a symmetric and approximately normal fashion around the overall mean age of acquisition ($\mu = 8.77$, $\sigma^2 = 2.51$). For reasons laid out in Section 1.4, we considered concreteness word frequency (*frequency*; Box-Cox transformed spoken token frequency estimates taken from CELEX; cf. Box & Cox, 1964), word length in terms of number of phonemes (*length*; based on CV transcriptions from CELEX), as well as morphological complexity (*morphology*) as control variables. The last of these was defined as a three-valued variable based on the morphological information from CELEX, counting monomorphemic ('M') words as *simple* (e.g., *attic*), polymorphemic ('C') words as *complex* (e.g., *cupboard*), and all other morphological labels as *opaque* (e.g., *ladder*). Finally, we categorised the syllable structure of each lexeme based on the CV transcription in CELEX as heavy-heavy (HH; e.g., *bedside*), heavy-light (HL; e.g., *carbon*), light-heavy (LH; e.g., *detect*), or light-light (LL; e.g., *atom*). In keeping with the generative literature, we disregarded word-final consonants ('final consonant extrametricality', Hayes, 1982), so that the first syllable was counted as light if of the form $(C_1 \dots C_n)V$, where $n \geq 1$, and heavy otherwise, while the second syllable was counted as light if of the form $(C_1 \dots C_n)V(C)$, where $n \geq 1$, and heavy otherwise.

2.3 Modelling procedure

Our aim was to assess in what way the combined effects of word class and stress pattern influence a word's (relative) age of acquisition, other things being equal. Our statistical modelling approach unfolds in three steps.⁶ First, we analyzed in what way AoA is predicted by *class/stress pattern*, initially considered on its own, and then

⁶All computations were done in R (R Development Core Team, 2018).

alongside concreteness, frequency, length, morphology and syllable structure as additional variables. For this, we used generalized additive models (GAM, Wood, 2006).⁷ We opted for this model family instead of generalized linear models (GLM) to detect any nonlinear effects of the three metric variables (concreteness, frequency, length), which were implemented as smooth terms (Wood, 2006). One of the strengths of GAMs is that they are not restricted to linear dependencies among variables. When plotting non-linear terms of a GAM, these terms become visible as curves rather than straight lines.

To determine the explanatorily most valuable model of AoA with our set of predictors, AICc-driven model selection was employed (Burnham & Anderson, 2002; Burnham et al., 2011). We computed a GAM for every possible subset of the predictors *class/stress pattern*, *concreteness*, *frequency*, *length*, *morphology* and *syllable structure*, additively entered into the model. At this stage, no interactions among the variables were considered. For each of the resulting 63 candidate models (disregarding the null model), AICc and Akaike weight w were determined. The Akaike information criterion (AIC) is a measure of a model's goodness of fit that also considers complexity. It can be interpreted as approximated loss of information in a given model relative to reality. AICc is a version of AIC adjusted for small sample sizes (Johnson & Omland, 2004). A model's Akaike weight can be interpreted as the probability (or strength of evidence) of a model for the data given the set of candidate models. The best model is the one with the lowest AICc score and, consequently, the highest Akaike weight.

In a second step, interactions between *class/stress pattern* and the remaining factors were added to the model to address the question whether the impact of stress and word class on AoA depends in any way on *concreteness*, *frequency*, *length*, *morphology* or *syllable structure*. For example, in the upper part of the frequency spectrum the acquisition order of *N ini*, *N fin*, *V ini*, and *V fin* may differ from that found for rarer lexemes. Once again, the AICc score was used to identify the most informative model, i.e., the one in which the addition of interactions makes a relevant contribution to the model's informativity. For continuous variables such as *frequency*, a significant interaction effect results in four separate curves, one for each *class/stress* configuration.

To facilitate comparisons of estimated coefficients for factor variables, 84 per cent confidence intervals were computed for each term, instead of the usual 95 per cent confidence intervals. While the latter are used to assess significant deviation from trivial behaviour (i.e., a zero effect) at a 95 per cent level of confidence, the former are used for assessing whether two terms differ significantly from each other at a 95 per cent confidence level by checking if the corresponding 84 per cent confidence intervals do not overlap (Payton, Greenstone & Schenker, 2003; Knol, Pestman & Grobbee, 2011; MacGregor-Fors, Payton & Quince, 2013; this threshold holds under the assumption of roughly equal standard errors, which is met in most of the models presented in this paper). We deliberately decided against procedures that would further adjust confidence levels for multiple pairwise comparisons (e.g., Bonferroni correction) due to the considerable amount of criticism these techniques have received (e.g., Perneger, 1998; Nakagawa, 2004).

⁷For the computation of GAMs the *mgcv* package in R was used (Wood, 2006).

Finally, we examined whether the acquisition order of *N ini*, *N fin*, *V ini*, and *V fin* in any way depends on the acquisition period itself. More precisely, we tested if the ordering of combinations of word class and stress in very early acquisition differs from later acquisition. As Kuperman et al. (2012: 987) point out, their AoA ratings overestimate actual AoA before the age of five or six, so that vocabulary size is underestimated below that threshold, which is problematic for identifying an early period for word learning in the data. We thus decided to set the boundary between early and late acquisition at six years as this age approximates the point at which vocabulary size increases substantially (cf. Figure 5 in Kuperman et al., 2012). In practical terms, the decision was suggested by the small number of disyllables before the age of four to five years, which would have rendered any reliable statistical analysis impossible. Two separate models of AoA depending on *class/stress pattern* were fit to the data, one for the early period and one for the later period. To facilitate cross-model comparisons, normalized regression coefficients (β coefficients) were computed as the two samples differ substantially in size. Confidence intervals were computed as described in the previous paragraph.

3 Results

3.1 Main effects

Before any control variables are added to the model, *class/stress pattern* by itself affects AoA in the following way (Figure 2; cf. Table A1). Stress-initial items are acquired before stress-final items, with stress-initial verbs slightly ahead of stress-initial nouns, but only by a marginally significant difference. Both of the former configurations are acquired significantly earlier than stress-final nouns and stress-final verbs. Schematically, the acquisition order can be represented as *V ini* < *N ini* <^{*} *N fin* < *V fin* (the asterisk denoting a significant inequality at the 95 per cent confidence level). This sequence fits the trochaic bias hypothesis in language acquisition, since left-stressed forms of both classes receive significantly lower AoA ratings than right-stressed ones. The evidence for a nominal bias, on the other hand, is at best ambiguous as stress-initial verbs actually receive the lowest AoA ratings overall, while the head start of stress-final nouns over stress-final verbs is insignificant. Most importantly, the stress typicality hypothesis does not find any support in this initial model. Class-typical left-stressed nouns are not acquired earlier than atypical left-stressed verbs, and class-typical right-stressed verbs are acquired significantly later than atypical left-stressed verbs. In fact, right-stressed verbs are marked by higher AoA averages than any other category. The question at this point is whether the ordering of the four *class/stress* configurations is modified when other potential factors also enter the model.

When considering the effect of *class/stress pattern* on AoA alongside the variables *concreteness*, *frequency*, *length*, *morphology* and *syllable structure*, the optimal model turns out to be the one that includes all six predictor variables ($w = 0.999$). Overall, the optimal model explains only about one third of the variation in AoA (adjusted $R^2 = 0.337$), which – unsurprisingly – suggests that additional factors not considered in this study determine the acquisition of words to a large degree, notably their communicative utility for the child.

The model reveals multiple significant effects of the predictor variables on AoA (Figure 2, cf. Table A4 and Table A5). As expected, AoA values decrease with rising *frequency*, and increase with *length*. Morphological complexity also correlates

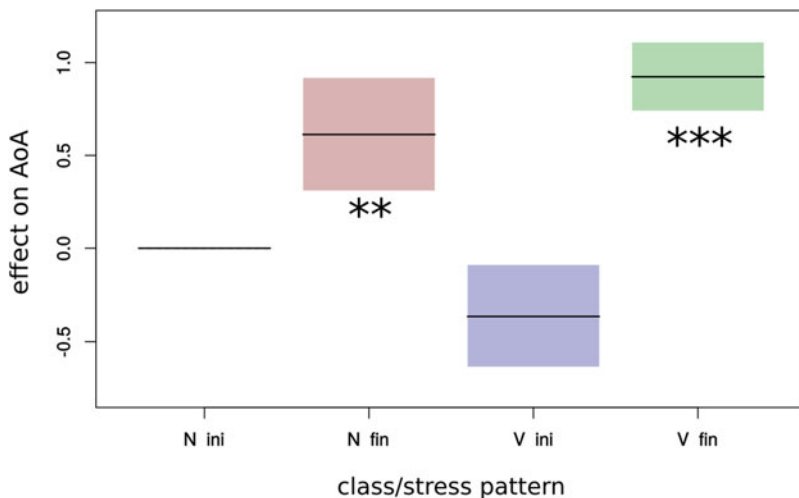


Figure 2. Linear model of AoA depending solely on the primary predictor `class/stress pattern`. Coloured areas denote 84% confidence intervals corresponding to predictor coefficients.

with AoA, but not in the straightforward way that was expected. Items labelled as complex actually receive significantly lower AoA ratings than both simple and opaque ones, which goes against the prediction that AoA should increase with complexity. Looking more closely at the items themselves, it emerges that disyllables tagged as morphologically complex are often compounds which are transparently compositional, including *birthday*, *bathroom* or *pancake*. Morphologically simple words, on the other hand, also include erudite vocabulary such as *object*, *respond* or *bonus*: that is, loan words whose compositionality (if there is any) may only be unravelled with some knowledge of Romance or Latin. With respect to syllable structure, it seems that items featuring a second syllable labelled as light (i.e., with a short vowel and at most one coda consonant, Section 3.2) are acquired on the whole earlier than items with a second syllable labelled as heavy (i.e., with long vowels or at least two coda consonants). Concreteness has the expected effect that the more concrete (i.e., accessible to the senses) a word-referent relationship, the lower its AoA ratings.

It is interesting to see how the addition of these variables affects the ranking within our primary predictor: in effect, they account for a good part of the differences observed in Figure 2 above, such that nouns of either stress pattern now receive similar AoA averages. This suggests that any effects that stress might have on the acquisition of nouns may sufficiently be explained by the combined effect of other phonological (length, syllable structure), distributional (frequency), morphological, and conceptual (concreteness) characteristics. Somewhat surprisingly, stress still plays an independent role for verbs. However, it is again the class-atypical stress-initial verbs that are acquired earlier than the remaining class/stress configurations, regardless of frequency, morphology, length, syllable structure and concreteness. This effect is highly significant and suggests a robust effect of stress pattern on acquisition age.

Since up to this point only main effects have been taken into account without considering possible interactions, the explanatory value of this model is still limited.

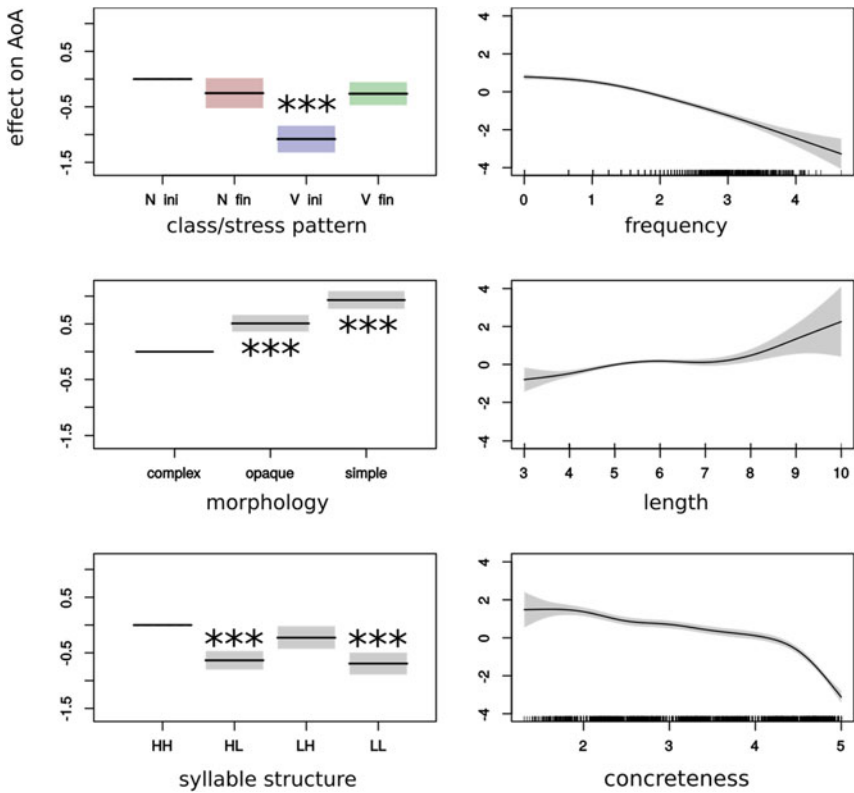


Figure 3. GAM modelling AoA depending on concreteness, frequency, length, class/stress pattern, morphology, and syllable structure. Grey/colored areas denote 84% confidence intervals for factor variables (left) and 95% confidence areas for continuous variables (right).

Nevertheless, what can be seen from the present model is that class/stress pattern substantially affects AoA, the difference between the averages of the earliest and the latest acquired patterns being c. 1.1 years. This is comparable to the effects of morphology and syllable structure, while concreteness and frequency account for the largest ranges between AoA ratings.

3.2 Interactions between variables

Next, we consider the effect of concreteness, frequency, length, morphology and syllable structure on the acquisition order of class/stress configurations: that is, we look for possible interactions between the primary and additional predictor variables. The best model among all possible combinations of predictors and interactions according to the AICc once again includes main effects for all six variables as well as interactions between all three metric variables (concreteness, frequency, length) and the primary predictor class/stress pattern (Akaike weight $w = 0.098$). Interactions with morphology and syllable structure were dropped based on the model selection procedure.

Every interaction in Figure 4 is represented by four curves, one for each class/stress configuration, illustrating the effect of the metric variables on AoA for each class/stress

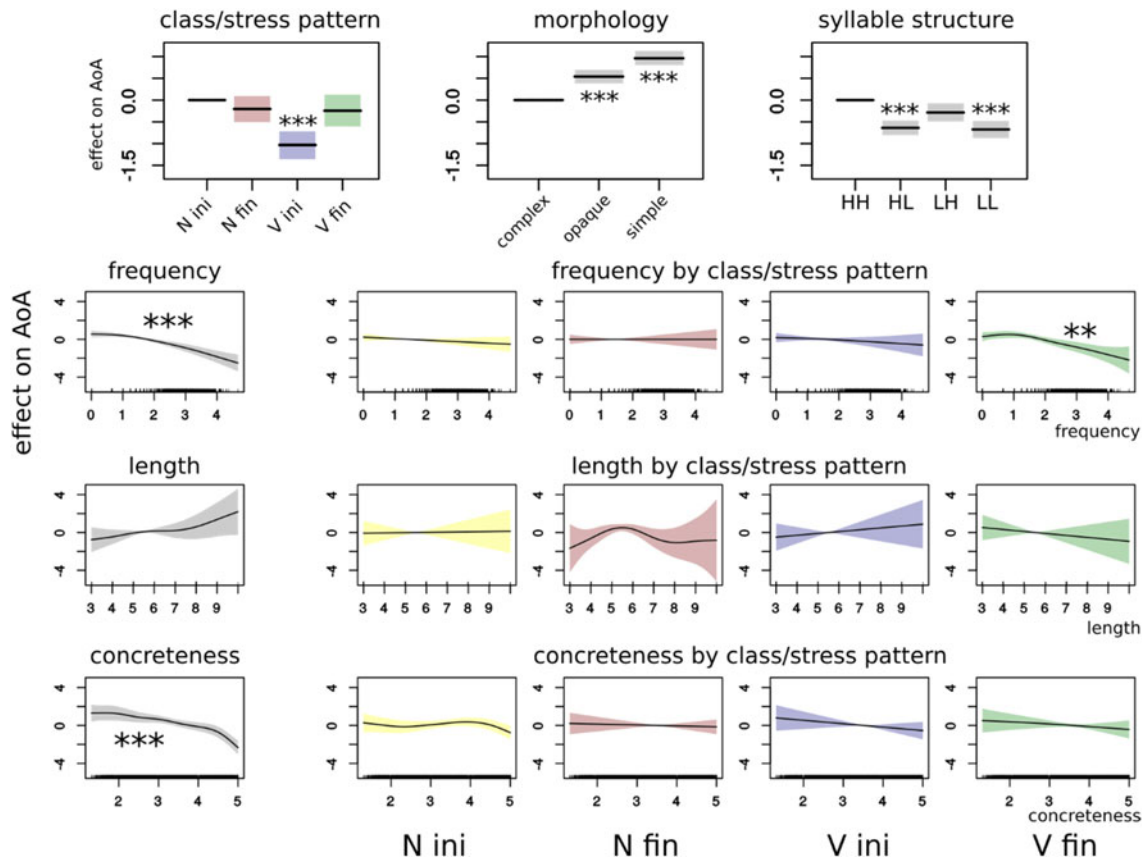


Figure 4. Effect of class/stress pattern on AoA (predictor coefficients and terms shown). Shaded areas denote 84% confidence intervals (factor variables, top) and 95% regions (smooth predictors, bottom), respectively. Color code: yellow 'N ini'; red: 'N fin'; green: 'V fin'; blue: 'N fin'. Significant differences with respect to baselines ('N ini'; 'complex'; 'HH') or trivial predictor behavior (i.e., no effect on AoA; length and frequency) indicated by: **: $p < 0.05$; ***: $p < 0.01$; ****: $p < 0.001$.

configuration separately. The most important result emerging from this detailed analysis is that the main effects observed in the previous section are largely robust even in the presence of interactions. Specifically, neither concreteness nor length affects the various class/stress configurations in substantially different ways. The general rule that higher concreteness averages imply earlier AoA ratings applies to nouns and verbs, both stress-initial and stress-final alike. In the full model, length even loses all predictive power. Only stress-final verbs behave somewhat exceptionally with regard to frequency. While the basic effect of frequency is the same for all configurations, lowering AoA values, for stress-final verbs this dependence seems to be more strongly pronounced than for other patterns.

Importantly, however, the main effect of class/stress pattern is stable despite the increased complexity of the model. Initially stressed verbs remain the only pattern that displays significantly earlier AoA ratings in the presence of the other predictors.

3.3 Comparing early and later lexical development

Finally, it may be argued that the bootstrapping effect of stress typicality, if it exists, should be strongest in early acquisition when morphological and syntactic cues for categorisation may still be less firmly established. In this view, allowing words into the analysis that are acquired as late as AoA 15 may conceal these effects. Although the descriptive analysis in Section 2 actually suggested a weaker correlation between word class and stress pattern at the very beginning of vocabulary acquisition, the findings that extremely frequent items seem to behave slightly differently compared to the rest of the lexicon, and also that concreteness and frequency are inversely correlated with AoA, suggest that it is worth considering the relationship between class/stress pattern and AoA for the early and later-acquired vocabulary separately.

The cut-off point here was defined as AoA 6, which roughly corresponds to the age of school entry. Whereas late acquisition fits the general picture (V ini <* N ini < N fin <* V fin; Figure 5, lower panel; Table A2), the configuration in the earlier period is rather surprising (Figure 5, upper panel; Table A3) as the ordering N fin <* N ini < V ini < V fin seems to apply. This result does not readily integrate with any of our initial hypotheses. While stress-final verbs do display significantly higher AoA ratings than nouns of any stress pattern, as is consistent with a trochaic stress bias, stress-final nouns emerge as the configuration with the lowest AoA averages, contradicting a preference for trochees in early acquisition.

Relatively robust results that can be derived from the above analyses are (i) that, globally speaking, initially stressed words are acquired earlier on average than finally stressed words; (ii) that additional variables influencing AoA, including frequency and concreteness, can account for the learning advantage in stress-initial nouns, but not in stress-initial verbs; (iii) that early and late acquisition differ considerably with regard to which patterns show lower AoA ratings. In sum, there is little evidence in favour of a decisive role of stress typicality in lexical acquisition. A trochaic bias, on the other hand, is generally supported, albeit not necessarily for early acquisition. A nominal bias for disyllables is also not independently supported when distributional and conceptual factors such as frequency and concreteness are taken into account.

4 Discussion

Two empirical aims were formulated at the outset of this paper, both relating to the relationship between class-typical stress patterns and the acquisition ages of lexical

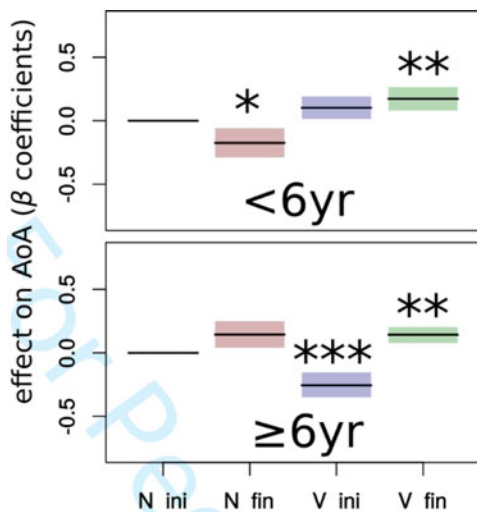


Figure 5. Models of AoA depending on class/stress pattern without interaction for early (younger than 6 years; upper graph) and late learning period (6 years or older; lower graph). Normalised coefficients (β) shown. Shaded areas denote 84% confidence intervals.

items. The first aim concerned the question whether left-stressed nouns and right-stressed verbs were the dominant types at the beginning of and throughout the process of word learning, as they are in the vocabulary at large. The descriptive analysis presented in Section 2.1 indicates that stress typicality characterises the vocabulary of young language learners only to a certain extent. The correlation between class and stress is shown to grow stronger with age and an expanding lexicon. This is mostly due to the accelerating inflow of right-stressed verbs to the average learner's vocabulary, whereas in nouns class-typical left-hand stress is the predominant pattern throughout.

The descriptive analysis did not exclude the possibility that class-typical stress patterns may exert a decisive effect on AoA when other structural, conceptual and distributional factors are controlled for and interactions are considered. For example, the assumed functionality of class-typical stress for fast-tracking acquisition may only operate within a specific subset of the lexicon, such as words within a certain segment of the frequency spectrum or words of a specific structure. The statistical models presented in Section 3 have approached this question from different angles. Overall, these models concur that configurations of class and stress do have a decisive effect on the acquisition age of disyllabic words. However, the effect is not the one predicted by the stress typicality hypothesis. The convergent outcome of the analyses is that left-stressed words have a head start over right-stressed ones. However, it is only in verbs that this preference comes out clearly, regardless of the level of concreteness, the frequency or the linguistic structure of a word.

Among the literature-backed acquisition constraints introduced earlier, the trochaic bias is thus the only one that finds consistent support in the acquisition data. Support for a nominal bias, on the other hand, is at best ambiguous. This is very likely due to the fact that nominal bias, as initially formulated (Gentner, 1978, 1982), relates to early word acquisition in general. Since monosyllabic words, which dominate early vocabularies, were excluded in the present study, the reported results do not invalidate earlier findings regarding this question. It can only be noted that a

preference for nouns, to the extent that it exists (Bornstein & Cote, 2005), seems to be less clear in disyllabic words than in monosyllabic ones.

Despite the fact that trochaic patterns in general seem to be acquired earlier compared to iambic patterns, it is intriguing to find that the ratings for early acquisition did not display these tendencies in a clear way. To the contrary, the models in Section 3.3 suggest that at least stress-final nouns are relatively well attested in early vocabularies to the extent that they received the lowest AoA estimate of all class/stress configurations. On closer inspection, this group is found to contain items such as *cartoon*, *balloon*, *shampoo* or *giraffe*, which are indeed words one would expect to be part of the early acquired vocabulary. At first blush, this result does not readily integrate with a cognitive bias that disfavors iambic patterns on perceptual grounds as discussed in the introduction. However, it is worth keeping in mind that one aspect that the current study could not control for is that the form that a language learner hears and the one that is subsequently stored in memory need not be identical. That is, even if trochaic patterns are more likely to be stored *intact* due to a confluence of salience and recency effects, this still does not exclude the possibility that originally iambic patterns can be successfully learned, but in a form that is *unfaithful* to the input. This interpretation agrees well with the observation that early iambic words, such as *giraffe*, are frequently clipped in children's first productions, with only the second syllable remaining (i.e., **raff*; McGregor & Johnson, 1997; also consider the common clipping *'toon* for *cartoon*).

Thus, though it is true that trochaic patterns dominate in English overall and that in general they are also learned at an earlier age, the present study falls short of supplying clear evidence that this imbalance is caused by a cognitive advantage that such patterns may create for young language learners. The most plausible interpretation of the data is simply to assume a distributional bias. While early acquisition seems to be relatively insensitive to prosody, allowing for the acquisition of nouns of either pattern, later acquisition more clearly favours trochaic patterns over iambic ones. Rather than relying on an innate bias for trochaic patterns, a trochaic template may thus simply emerge through exposure to the overwhelming pattern in the input language. The fact that the trochaic bias only emerges with some delay in the AoA averages may point in this direction. As a template based on distributional input data rather than a cognitive constraint, the trochaic preference in this case would have the status of a generalisation which needs time to develop.

Whatever its ultimate motivation, if there is a general preference for trochaic patterns as suggested by the data, what does this mean for non-trochaic patterns? If they do not serve a discernible purpose in signalling class membership for first language learners, this raises the question why they even exist. Why have they not given way to trochaic ones or simply been reduced to monosyllabic stems (as children's early productions would suggest)? Why have non-word-initial stress patterns survived over generations of first-language learners, despite their failing to align prominence peaks with word onsets, thereby benefiting speech segmentation? This particularly concerns verbal iambic patterns, which, contrary to the predictions of the stress typicality hypothesis, turn out to be the class/stress pattern configuration that is acquired latest on average.

One possible answer to this is that, regardless of acquisition, stress typicality may still offer a slight processing advantage in communication, as it allows interlocutors to parse speech efficiently. This may suffice to ensure the diachronic stability of iambic patterns. At least some effect was found if the phonological signal was presented to subjects as

incomplete (Arciuli & Cupples, 2004). Based on these results, stress typicality could be seen as a feature of functional redundancy, stepping in when segmental information fails to disambiguate.

Another explanation holds that the rhythm of English speech encourages the existence of different stress patterns in the lexicon to ensure rhythmicity in syntagmatic combinations (Baumann & Ritt, 2017). According to this view, diversity of stress patterns helps reduce the likelihood of stress clashes in the speech stream, which are dispreferred due to articulatory or perceptual constraints (for discussion cf. Schlüter, 2005; Patel, 2010). As Kelly and Bock (1988) and Kelly (1989) have argued, differential stress in nouns and verbs may have partly evolved to absorb the impact of such rhythmic shocks.

In addition to these accounts, one seemingly minor outcome of the present study may also contribute to the understanding of why particularly verbs have remained iambic through time. Although the interactions studied in Section 3.2 generally turned out non-predictive, there was one interaction that displayed a highly significant effect: namely, that between frequency and stress-final verbs. Thus, iambic verbs at the high end of the frequency spectrum are learned at a substantially earlier age than less frequent ones, even when the main effect of frequency was also taken into account. In other words, while high occurrence frequency positively promotes the acquisition of all words regardless of class or stress pattern, it especially does so in iambic verbs. This may be consequential, considering that generally right-stressed verbs were found to be acquired later on average than the other configurations. This would suggest that, in high-frequency items, iambic verbs are at least relatively well-represented among earlier learned vocabulary, including such lemmas as *begin*, *behave*, *allow* or *forget*.

Additionally, considering that left-stressed patterns overwhelmingly map onto nouns during all stages of acquisition, a 60 per cent chance of right-stressed items being verbs (i.e., roughly the probability at AoA 5 to 6; Section 2.1) may suffice as a heuristic for word class membership that is arguably more accurate than assuming no prosodic cue for class at all. Ignoring prosody altogether (and disregarding for the sake of the argument other morphological and syntactic cues) would result in a baseline of successful categorisation of verbs at a much lower rate of c. 26 per cent (i.e., the proportion of verbs in the dataset regardless of stress pattern). These factors may have guaranteed the survival of iambic verbs in defiance of the trochaic pressure, despite the fact that they fail to materialise in significantly lower AoA ratings overall.

On the other hand, iambic patterns in verbs may also have been preserved as a result of their constituting a particular group of items. Their sharp increase during later acquisition relative to other disyllabic nouns and verbs suggests that this might be the case. One characteristic that many of the iambic verbs in the sample share is that they are of Latinate origin (e.g., *connect*, *delete*). In contrast, many (but not all) early acquired iambic verbs are of native Germanic origin (e.g., *begin*, *forget*). Latinate words tend to form part of more educated registers, where social pressures could account for keeping the original Latinate right-oriented patterns (similar to the way that *cacti* is retained as an irregular plural). After all, many iambic verbs can easily be replaced by low-register monosyllabic synonyms of non-Latinate origin if the situation permits it (e.g., *receive* vs. *get*, *depart* vs. *leave*), thus reserving the 'exotically' stressed iambic counterparts for more elevated usage. Even so, the fact that relatively late-acquired iambic verbs like *connect* can build upon a tentative iambic template established by a small number of frequent, early acquired verbs such as *begin* cannot

be ruled out as a factor securing their diachronic stability (cf. Sherman, 1975; Kastovsky, 1992: 361).

5 Conclusion

The present study investigated the relationship between stress pattern and lexical class in lexical development. It was found that stress typicality in English nouns and verbs holds only limited potential for functioning as a prosodic cue during word acquisition. Instead, a general preference for trochaic patterns was found, which is stable at least in verbs when a range of confounding variables including concreteness, frequency as well as other morphological and phonological factors are added to the model. Early acquisition did not actually follow this tendency, however, which points to the trochaic bias being an emergent rather than an innate characteristic in English lexical development. The stability of non-trochaic patterns in verbs is speculatively attributed to the impact of a small number of frequent and early-acquired words, which may suffice to establish a prosodic template for later-acquired Latinate verb lemmas to build on.

These findings go somewhat against the grain of a number of studies which have identified various bootstrapping effects of phonological features in language acquisition and processing. The results also call into question a strong interpretation of the signalling function of class-specific stress patterns. More experimental and comparative work is needed to further explore the functional roles that phonological features may take and to investigate the cognitive or environmental nature of the trochaic bias in first language acquisition.

Acknowledgements. We would like to thank two anonymous reviewers and the editors for their valuable comments on a first draft of this paper.

Supplementary Material. For supplementary material accompanying this paper, visit <https://doi.org/10.1017/S030500092000046X>

References

- Adam, G., & Bat-El, O. (2009). When do universal preferences emerge in language development? The acquisition of Hebrew stress. *Brill's Annual of Afroasiatic Languages and Linguistics* 1. 255–282.
- Allen, G. D., & Hawkins, S. (1980). Phonological rhythm: Definition and development. In Yeni-Komshian, G. H., J. F. Kavanagh & C. A. Ferguson (eds.). *Child phonology. Vol 1: Production*, 227–256. New York: Academic Press.
- Ambridge, B., Kidd, E., Rowland, C.F., & Theakston, A.L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language* 42. 239–273.
- Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development* 58. 1–166.
- Arciuli, J., & Cupples, L. (2003). Effects of stress typicality during speeded grammatical classification. *Language and Speech* 46. 353–374.
- Arciuli, J., & Cupples, L. (2004). Effects of stress typicality during spoken word recognition by native and nonnative speakers of English: Evidence from onset gating. *Memory & Cognition*. 21–30. 10.3758/BF03195817.
- Arciuli, J., & Cupples, L. (2006). The processing of lexical stress during visual word recognition: Typicality effects and orthographic correlates. *Quarterly Journal of Experimental Psychology* 59. 920–948.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (2001). *WebCelex*. <http://celex.mpi.nl>.
- Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review* 105. 158–173.

- Ballard, K. J., Djaja, D., Arciuli, J., James, D. G. H., & van Doorn, J. (2012). Developmental trajectory for production of prosody: Lexical stress contrastivity in children ages 3 to 7 years and in adults. *Journal of Speech Language and Hearing Research* 55, 1822–1835.
- Baumann, A., & Ritt, N. (2017). On the replicator dynamics of lexical stress: accounting for stress-pattern diversity in terms of evolutionary game theory. *Phonology* 34, 439–471.
- Bijeljic-Babic, R., Hohle, B., & Nazzi, T. (2016). Early prosodic acquisition in bilingual infants: The case of the perceptual trochaic bias. *Frontiers in Psychology*, 10.3389/fpsyg.2016.00210.
- Bornstein, M. H., & Cote, L. R. (2005). Expressive vocabulary in language learners from two ecological settings in three language communities. *Infancy* 7, 299–316.
- Box, G., & Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B* 26(2), 211–252.
- Brysaert, M., & Biemiller, A. (2017). Test-based age-of-acquisition norms for 44 thousand English word meanings. *Behavior Research Methods* 49, 1520–1523.
- Brysaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods* 46, 904–911.
- Bybee, J., Chakraborti, P., Jung, D., & Scheibman, J. (1998). Prosody and segmental effect: Some paths of evolution for word stress. *Studies in Language* 22, 267–314.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference a practical information-theoretic approach*. New York, NY: Springer.
- Burnham, K. P., Anderson, D. R., & Huyvaert, K.P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology* 65, 23–35.
- Burzio, L. (1994). *Principles of English stress*. Cambridge: Cambridge University Press.
- Cassidy, K. W., & Kelly, M. H. (2001). Children's use of phonology to infer grammatical class in vocabulary learning. *Psychonomic Bulletin & Review* 8, 519–523.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.
- Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition* 125(3), 452–465.
- Cutler, A., & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language* 2, 133–142.
- Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology-Human Perception and Performance* 14, 113–121.
- Domahs, U., Plag, I., & Carrol, R. (2014). Word stress assignment in German, English and Dutch: Quantity-sensitivity and extrametricality revisited. *Journal of Comparative Germanic Linguistics* 17, 59–96.
- Echols, C. H., & Newport, E. L. (1992). The role of stress and position in determining first words. *Language Acquisition* 2, 189–220.
- Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology-Learning Memory and Cognition*, 24, 515–523.
- Fitneva, S. A., Christiansen, M. H., & Monaghan, P. (2009). From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of Child Language*, 36, 967–997.
- Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language* 44(3), 677–694.
- Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child Development* 49, 988–998.
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In Kuczaj, Stan (ed.). *Language development. Vol. 2: Language, thought and culture*, 301–334. Hillsdale, NJ: Erlbaum.
- Goedemans, R., & van der Hulst, H. (2013). Rhythm types. In Dryer, Matthew S., & Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/17>, Accessed on 2019-12-14.)
- Guion, S. G., Clark, J. J., Harada, T., & Wayland, R. P. (2003). Factors affecting stress placement for English nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech* 46, 403–427.

- Halle, M., & Keyser, S. J. (1971). *English stress: its form, its growth, and its role in verse*. New York: Harper & Row.
- Hayes, B. (1982). Extrametricality and English stress. *Linguistic Inquiry* 13, 227–276.
- Hayes, B. (1995). *A metrical theory of stress: Principles and case studies*. Chicago: University of Chicago Press.
- Hulme, C., Neath, I., Stuart, G., Shostak, L., Surprenant, A. M., & Brown, G. D. A. (2006). The distinctiveness of the word-length effect. *Journal of Experimental Psychology* 32, 586–594.
- James, D. G. H., Van Doorn, J., & McLeod, S. (2008). The contribution of polysyllabic words in clinical decision making about children's speech. *Clinical Linguistics & Phonetics*, 22, 345–353.
- Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, 19, 101–108.
- Jusczyk, P. W., Houston, D. W., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology* 39, 159–207.
- Kastovsky, D. (1992). Semantics and vocabulary. In Hogg, Richard M. (ed.). *The Cambridge History of the English Language. Vol. 1: The beginnings to 1066*, 290–408. Cambridge: Cambridge University Press.
- Kawamoto, A., Farrar, W.T., & Overbeek, M. (1993). *Effects of grammatical category and stress pattern on on-line processing of bisyllabic words*. Unpublished manuscript.
- Kelly, M. H. (1989). Rhythm and language change in English. *Journal of Memory and Language* 28, 690–710.
- Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review* 99, 349–364.
- Kelly, M. H., & Bock, J. K. (1988). Stress in time. *Journal of Experimental Psychology* 14, 389–403.
- Kiparsky, P. (1982). From cyclic phonology to lexical phonology. In Hulst, Harry van der & Norval Smith (eds.). *The Structure of Phonological Representations*, 131–175. Dordrecht: Foris.
- Knol, M., Pestman, W., & Grobbee, D. (2011). The (mis)use of overlap of confidence intervals to assess effect modification. *European Journal of Epidemiology* 26, 253–254.
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods* 44, 978–990.
- Lass, R. (1992). Phonology and morphology. In Blake, Norman (ed.). *The Cambridge History of the English Language. Vol. 2: 1066 to 1476*, 23–155. Cambridge: Cambridge University Press.
- Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Mass: MIT Press.
- MacGregor-Fors, I., Payton, M. E., & Quince, C. (2013). Contrasting diversity values: Statistical inferences based on overlapping confidence intervals. *PLoS ONE*, 10.1371/journal.pone.0056794.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.
- McCully, C. (2002). Exaptation and English stress. *Language Sciences* 24, 323–344.
- McDonough, C., Song, L., Hirsh-Pasek, K., Golinkoff, R. M., & Lannon, R. (2011). An image is worth a thousand words: why nouns tend to dominate verbs in early word learning. *Developmental Science* 14, 181–189.
- McGregor, K. K., & Johnson, A. C. (1997). Trochaic template use in early words and phrases. *Journal of Speech Language and Hearing Research*, 40, 1220–1231.
- Minkova, D. (1997). Constraint ranking in Middle English stress-shifting. *English Language and Linguistics*, 1, 135–176.
- Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, 96, 143–182.
- Morrison, C. M., & Ellis, A. W. (2000). Real age of acquisition effects in word naming and lexical decision. *British Journal of Psychology*, 91, 167–180.
- Nakagawa, S. (2004). A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behavioral Ecology* 15, 1044–1045.
- Patel, A. D. (2010). *Music, language, and the brain*. Oxford: Oxford Univ. Press.
- Payton, M. E., Greenstone, M. H., & Schenker, N. (2003). Overlapping confidence intervals or standard error intervals: what do they mean in terms of statistical significance? *Journal of Insect Science* 3.
- Pedale, T., & Santangelo, V. (2015). Perceptual salience affects the contents of working memory during free-recollection of objects from natural scenes. *Frontiers in Human Neuroscience* 9.

- Perneger, T. V.** (1998). What's wrong with Bonferroni adjustments. *BMJ: British Medical Journal* 316. 1236–1238.
- R Development Core Team.** (2018). *R: A language and environment for statistical computing*. Vienna, Austria. R Foundation for Statistical Computing. <https://www.r-project.org/>.
- Richardson, J. T. E.** (1976). Imageability and concreteness. *Bulletin of the Psychonomic Society* 7(5). 429–431.
- Rose, Y., & Champdoizeau, C.** (2008). There is no innate trochaic bias: Acoustic evidence in favour of the neutral start hypothesis. In Gavarró, A. & M. J. Frieas (eds.). *Language acquisition and development: Proceedings of GALA 2007*, 359–369. Newcastle: Cambridge Scholars Publishing.
- Schlüter, J.** (2005). *Rhythmic grammar: The influence of rhythm on grammatical variation and change in English*. Berlin; New York: Mouton de Gruyter.
- Serenio, J. A., & Jongman, A.** (1995). Acoustic correlates of grammatical class. *Language and Speech* 38. 57–76.
- Shady, M., & Gerken, L.** (1999). Grammatical and caregiver cues in early sentence comprehension. *Journal of Child Language*, 26. 163–175.
- Sherman, D.** (1975). Noun-verb stress alternation: An example of the lexical diffusion of sound change in English. *Linguistics* 13(159). 43–71.
- Slobin, D. I.** (1973). Cognitive prerequisites for the development of grammar. In Ferguson, C. A., & Dan I. Slobin (eds.). *Studies of child language development*, 175–208. New York: Holt, Rinehart & Winston.
- Snow, D.** (1998). A prominence account of syllable reduction in early speech development: The child's prosodic phonology of tiger and giraffe. *Journal of Speech Language and Hearing Research* 41. 1171–1184.
- Sonderegger, M., & Niyogi, P.** (2010). Combining data and mathematical models of language change. *48th Annual Meeting of the Association for Computational Linguistics. Uppsala, Sweden, 11–16 July 2010*. 1019–1029.
- Thiessen, E. D., & Saffran, J. R.** (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology* 39. 706–716.
- Thiessen, E. D., & Saffran, J. R.** (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development* 3. 73–100.
- Vihman, M. M., DePaolis, R. A., & Davis, B. L.** (1998). Is there a 'trochaic bias' in early word learning? Evidence from infant production in English and French. *Child Development* 69. 935–949.
- Wood, S. N.** (2006). *Generalized additive models: An introduction with R*. Boca Raton, FL: Chapman & Hall/CRC.
- Zipf, G.** (1935). *The psychobiology of language: An introduction to dynamic philology*. Cambridge, Mass.: M.I.T. Press.

Appendix

The following significance codes apply: $p < 0.1$: '.'; $p < 0.05$: '*'; $p < 0.01$: '**'; $p < 0.001$: '***'. For the predictor class/stress pattern the shorthand pattern is used. Baseline categories in all models: N_ini, complex, HH.

Table A1. Reference model (linear model) of AoA depending on class/stress pattern. *R-squared (adj.)* = 0.0236, $n = 2430$. Beta coefficients shown.

term	estimate	SE	t	p	significance
(Intercept)	8.59475	0.06092	141.084	< 0.001	***
pattern_N_fin	0.61289	0.21574	2.841	0.00454	**
pattern_V_ini	-0.36390	0.19586	-1.858	0.06330	.
pattern_V_fin	0.92366	0.13244	6.974	< 0.001	***

Table A2. Linear model of AoA depending on pattern for early acquired items (up to 6 yr). *R-squared* (*adj.*) = 0.0312, *n* = 362. Beta coefficients shown.

term	estimate	SE	t	p	significance
(Intercept)	-1.54188	0.01792	-86.034	< 0.001	***
pattern_N_fin	-0.17428	0.08174	-2.132	0.03367	*
pattern_V_ini	0.10198	0.06317	1.614	0.10731	
pattern_V_fin	0.17294	0.06554	2.639	0.00869	**

Table A3. Linear model of AoA depending on pattern for early acquired items (6 yr and older). *R-squared* (*adj.*) = 0.0131, *n* = 2064. Beta coefficients shown.

term	estimate	SE	t	p	significance
(Intercept)	0.24868	0.02211	11.249	<0.001	***
pattern_N_fin	0.14350	0.07523	1.907	0.05660	.
pattern_V_ini	-0.25543	0.06984	-3.657	0.000261	***
pattern_V_fin	0.14223	0.04549	3.126	0.001794	**

Table A4. Parametric terms of optimal GAM of AoA depending on pattern + s(frequency, *k* = 10) + s(length, *k* = 6) + morphology + syllable_structure + s(concreteness, *k* = 10). *R-squared* (*adj.*) = 0.337, *n* = 2430.

term	estimate	SE	t	p	significance
(Intercept)	8.8702	0.1102	80.463	< 0.001	***
pattern_N_fin	-0.2538	0.1925	-1.318	0.1876	
pattern_V_ini	-1.0808	0.1711	-6.316	< 0.001	***
pattern_V_fin	-0.2638	0.1474	-1.790	0.0735	.
morphology_opaque	0.5089	0.1081	4.707	< 0.001	***
morphology_simple	0.9305	0.1150	8.090	< 0.001	***
syllable_structure_HL	-0.6350	0.1194	-5.320	< 0.001	***
syllable_structure_LH	-0.2269	0.1456	-1.558	0.1193	
syllable_structure_LL	-0.6933	0.1406	-4.929	< 0.001	***

Table A5. Smooth terms of optimal GAM of AoA depending on pattern + s(frequency, *k* = 10) + s(length, *k* = 6) + morphology + syllable_structure + s(concreteness, *k* = 10). *R-squared* (*adj.*) = 0.337, *n* = 2430.

smooth term	edf	ref.df	F	p	significance
s(frequency)	2.891	3.559	105.161	< 0.001	***
s(length)	3.706	4.256	9.106	< 0.001	***
s(concreteness)	6.357	7.501	90.156	< 0.001	***

Table A6. Parametric terms of optimal controlled GAM of AoA depending on class/stress pattern + s (frequency, $k = 10$) + s(frequency, $k = 10$, by = pattern) + s(length, $k = 6$, by = pattern) + s(length, $k = 6$) + morphology + syllable_structure + s(concreteness, $k = 10$, by = pattern) + s(concreteness, $k = 10$). R -squared (adj.) = 0.346, $n = 2430$.

term	estimate	SE	t	p	significance
(Intercept)	8.8347	0.1109	79.678	< 0.001	***
pattern_N_fin	-0.2060	0.2025	-1.017	0.3092	
pattern_V_ini	-10.339	0.2261	-4.573	< 0.001	***
pattern_V_fin	-0.2441	0.2618	-0.932	0.3513	
morphology_opaque	0.5422	0.1094	4.957	< 0.001	***
morphology_simple	0.9601	0.1161	8.273	< 0.001	***
syllable_structure_HL	-0.6405	0.1189	-5.389	< 0.001	***
syllable_structure_LH	-0.2829	0.1477	-1.915	0.0556	.
syllable_structure_LL	-0.6746	0.1400	-4.820	< 0.001	***

Table A7. Smooth terms of optimal controlled GAM of AoA depending on class/stress pattern + s (frequency, $k = 10$) + s(frequency, $k = 10$, by = pattern) + s(length, $k = 6$, by = pattern) + s(length, $k = 6$) + morphology + syllable_structure + s(concreteness, $k = 10$, by = pattern) + s(concreteness, $k = 10$). R -squared (adj.) = 0.346, $n = 2430$.

smooth term	edf	ref.df	F	p	significance
s(frequency)	2.5467	3.182	11.382	< 0.001	***
s(frequency):pattern_N_ini	0.8000	0.800	1.969	0.20961	
s(frequency):pattern_N_fin	0.8000	0.800	0.000	0.99316	
s(frequency):pattern_V_ini	0.9890	1.149	0.951	0.39689	
s(frequency):pattern_V_fin	2.5252	3.157	4.904	0.00163	**
s(length):pattern_N_ini	0.8000	0.800	0.016	0.90986	
s(length):pattern_N_fin	2.5454	3.063	2.298	0.06735	.
s(length):pattern_V_ini	0.8000	0.800	0.577	0.49686	
s(length):pattern_V_fin	0.8000	0.800	0.768	0.43310	
s(length)	3.2310	3.811	2.474	0.11165	
s(concreteness):pattern_N_ini	3.3456	4.190	1.987	0.07697	.
s(concreteness):pattern_N_fin	0.8000	0.800	0.171	0.71115	
s(concreteness):pattern_V_ini	0.8000	0.800	1.705	0.24297	
s(concreteness):pattern_V_fin	0.9143	1.007	0.769	0.37519	
s(concreteness)	5.1917	6.280	9.074	< 0.001	***

Cite this article: Hofmann K, Baumann A (2021). Trochaic bias overrides stress typicality in English lexical development. *Journal of Child Language* 48, 645–669. <https://doi.org/10.1017/S030500092000046X>