Noname manuscript No.
(will be inserted by the editor)

Prosodic end-weight reflects phrasal stress

(draft of September 2017; under review)

the date of receipt and acceptance should be inserted later

Abstract Prosodic end-weight refers to the well-documented tendency of prosodically heavier constituents to be preferred at the ends of domains when other factors (e.g. semantics, accessibility, and syntactic complexity) are controlled. Various explanations for prosodic end-weight have been put forth, including complexity deferral, final lengthening, rhythm, phonotactics, and nuclear stress. This article adduces several new arguments for phrasal stress as a unified explanation for prosodic end-weight and proposes a constraint-based theory of the stress-weight interface in sentential prosody.

Keywords weight · stress · prosody · complexity · end-weight · word order

1 Introduction

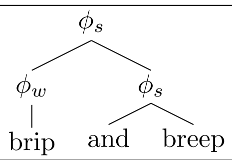
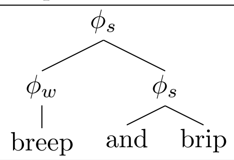
End-weight (after Quirk et al. 1972) refers to the tendency of heavier constituents to be localized later in sentences, all else being equal, as documented extensively for English and several other languages (though some languages exhibit the opposite, beginning-weight tendency; §5.1). End-weight has been recognized since antiquity (e.g. Pāṇini 2.2.32ff, Quintilian 9.4.22ff), and was famously formulated as the Gesetz der wachsenden Glieder ('law of increasing members') by Behaghel (1909). More than a hundred scholarly works have investigated end-weight in the context of several constructions, including coordination, echo reduplication, dvandvas, heavy NP shift, extraposition, adjective-noun order, stacked adjuncts, particle verbs, and the so-called dative, genitive, and locative alternations. For reasons of space, this article focuses on coordination, the best-studied end-weight phenomenon phonologically. However, the proposed theory of prosodic end-weight extends to all relevant constructions, as discussed in §4 and §7.

Address(es) of author(s) should be given

Prosodic end-weight (PEW) refers to the specifically phonological aspect of end-weight, as emerges when when other factors, including semantics, accessibility, rhythm, and syntactic complexity, are controlled. Building on previous work, seven principles of PEW are put forth in §3. Specifically, the following properties pattern as heavier: longer vowels, lower vowels, longer onsets, less sonorous onsets, longer codas, more sonorous codas, and more syllables. All of these properties accord with the general phonology of weight across languages and systems (§4). For example, while increasing sonority contributes to weight in the rime, it detracts from it in the onset, an asymmetry found both in end-weight and in stress and meter.

PEW is argued here to be motivated by phrasal stress, in that heavier constituents are attracted to loci of greater sentential prominence. Extensive previous work has treated the stress-weight interface as it affects word stress. This article proposes that the same interface is also operative in sentential prosody, as revealed by PEW. A stringent constraint family is put forth to implement this interface, generalizing from similar constraints for word phonology (§4). For example, consider a nonsense binomial such as *brip and breep*, which is preferred to *breep and brip*, though both orders are grammatical. In either order, greater stress falls on the second conjunct. A constraint $*VV/\phi_w$ penalizes a heavy element, in this case, a long vowel (VV), in a weak prosodic setting (prosodic word, phrase, or higher, notated ϕ_w). Thus, *brip and breep* and similar binomials are preferred to their transpositions, the strength of the preference being a function of the constraint's weight.[1] But no language can favor a heavy element in a weak prosodic setting, all else being equal, since there is no constraint like $*\breve{V}/\phi_w$ or $*VV/\phi_s$. It further follows that prosodically head-initial contexts cannot exhibit PEW.

(1)

			$*VV/\phi_w$
(favored)	a.		
(disfavored)	b.		$*!$

Other possible explanations for PEW, including final lengthening, complexity deferral, rhythm, and phonotactics, are argued to be either orthogonal to PEW or too parochial to explain its core properties (§6). For example, consider

[1] In Optimality Theory, candidate (b) would be harmonically bounded in this illustration, as only one constraint is shown. If one employs maxent Harmonic Grammar to implement variation, however, the weight of a single constraint can modulate the strength of a preference even in the absence of competing constraints (§4).

the proposal that PEW involves ordering constituents in terms of increasing complexity. Such a proposal cannot motivate sonority effects (including the onset-rime reversal), gradient effects of duration, or beginning-weight (§6.4).

PEW is not a linguistic universal. In verb-final languages such as Japanese and Turkish, it is at least significantly weaker, if not reversed, relative to languages like English and French. Lohmann and Takada (2014), for instance, find almost no PEW tendency in Japanese (vs. a strong tendency in English by the same metric), and I suggest in §5.1 that Turkish favors prosodic beginning-weight rather than end-weight, judging by effects of both vowel quality and syllable count. Thus, any theory of PEW must also explain its confinement to certain languages. Phrasal stress is a promising diagnostic: PEW is evidently found only in prosodically head-final contexts (though some constructions call for further investigation; §5.3).

Eleven predictions of the stress-weight interface analysis of PEW are enumerated in (2). As subsequent sections elaborate, arguably all of these predictions are borne out.

(2) 1–4. Syllable weight in PEW should be treated identically to (gradient) syllable weight in stress and meter, such that the following contribute to weight:
   1. Length or complexity in any part of the syllable
   2. A lower or more peripheral vowel
   3. A more sonorous coda
   4. A less sonorous (especially voiceless) onset

5. Additionally, in line with weight elsewhere, weight should be more sensitive to the nucleus than to the margins (especially onset).
6. Adding syllables should add to weight, with no upper bound.
7. PEW should be more sensitive to stressed than unstressed syllables.
8. PEW should be found only in languages with phrase-level stress.
9. Prosodically head-initial contexts should not exhibit PEW.
10. PEW should be sensitive to constituent-medial structure, not just structure at or near the edges of constituents.
11. Prosodic boundary structure should be irrelevant for PEW.

These predictions are for the most part not made by non-stress-based analyses of PEW, though there is some overlap, as shown in (3). A checkmark indicates that the analysis makes the prediction. Parentheses indicate that the prediction is made only in certain contexts, with incomplete coverage. For example, final lengthening could potentially explain some of the syllable weight effects, but only for p-phrase-final syllables. These alternative explanations are considered in more detail in §6.

	Stress-Weight	Complexity	Final Lengthening	Phonotactics	Rhythm
1.	✓	✓	(✓)	(✓)	
2.	✓		(✓)		
3.	✓		(✓)	(✓)	
4.	✓			(✓)	
5.	✓				
6.	✓	✓			(✓)
7.	✓				
8.	✓				✓
9.	✓				
10.	✓	✓			✓
11.	✓	✓			

Fig. 1: Coverage of five explanations for PEW of the generalizations in (2).

2 End-weight

2.1 End-weight in coordination

A coordinate pair such as *X and Y* is usually synonymous with its transposition *Y and X*, and is usually grammatical in both orders. Nevertheless, conjuncts tend to be organized from lightest to heaviest, all else being equal (Jespersen 1905, 1961, Abraham 1950, Malkiel 1959, Jakobson 1960, Bolinger 1962, Gustafsson 1974, Cooper and Ross 1975, Gustafsson 1975, Pinker and Birdsong 1979, Oden and Lopes 1981, Ross 1982, Oakeshott-Taylor 1984, Kelly 1986, Allan 1987, Fenk-Oczlon 1989, McDonald et al. 1993, Wright and Hay 2002, Wright et al. 2005, Benor and Levy 2006, Wolf 2008, Mollin 2012, 2013, Lohmann and Takada 2014, Morgan and Levy 2016). Some English illustrations of end-weight orders are given in (3) (for hundreds of additional examples, see the appendices of Fenk-Oczlon 1989 and Benor and Levy 2006). Coordinate (or similar; cf. (h)) pairs are here termed BINOMIALS, regardless of fixity; end-weight also applies to multinomials.

(3)
a. kit and caboodle
b. trials and tribulations
c. friends, Romans, countrymen
d. lock, stock, and barrel
e. Joan and Margery
f. trick or treat
g. slip and slide
h. tit for tat

The question of how exactly 'heaviest' is defined is treated in §4; for now, note that (3) (a–e) increase in syllable count and (f–h) increase in vowel duration. As the work just cited makes clear, end-weight applies as a significant tendency in both idiomatically frozen binomials (e.g. *kit and caboodle*) and relatively free or novel ones (e.g. *Joan and Margery*), is found in both written corpora and spontaneous speech, is widespread crosslinguistically (though perhaps not universal; see §5.1), and is robust under various experimental paradigms, including nonce word ordering tasks (§3).

2.2 Non-phonological factors in end-weight

Before turning to the contribution of phonology to end-weight in §3, this subsection surveys non-phonological factors affecting ordering in end-weight constructions such as binomials. First, items tend to decrease in frequency, presumably reflecting the priority of more accessible items (Bock 1982, Kelly 1986, Kelly et al. 1986, Allan 1987, Fenk-Oczlon 1989, McDonald et al. 1993, Griffin and Bock 1998, Golenbock 2000, Wright et al. 2005, Benor and Levy 2006, Mollin 2012, Shih and Zuraw 2016). For example, in *kit and caboodle*, while *kit* is the lighter conjunct, it is also the more frequent. In general, frequency and weight are negatively correlated (Zipf 1936), meaning that much of the observed tendency for end-weight could in principle reflect frequency rather than weight. Indeed, many studies deconfounding the two factors support an independent contribution of frequency (*op. cit.*). That said, when frequency is controlled, either through wug-testing or regression, weight remains a clear effect. In corpus studies with large sets of predictors, frequency is usually either non-significant or weak compared to structural predictors (Benor and Levy 2006, Mollin 2012, Lohmann and Takada 2014, Shih et al. 2015).

Second, semantics and pragmatics influence ordering, as studied extensively for binomials. One particularly influential early treatment of semantic factors is Cooper and Ross's (1975) article 'World order' (sic), which attempts to subsume several semantic predictors under a principle termed Me First, according to which initial position favors properties associated with the prototypical speaker. See also Allan (1987), Benor and Levy (2006), and Lohmann and Takada (2014) for more recent surveys of semantic and information-theoretic predictors. Some notable factors include animacy (more animate first; Byrne and Davidson 1985, McDonald et al. 1993, Shih et al. 2015), proximity (nearer in time or space first, or 'own before other'; Jespersen 1961, Cooper and Ross 1975), iconicity (reflecting temporal or other scales; Malkiel 1959, Benor and Levy 2006, Lohmann and Takada 2014), gender (male before female; Malkiel 1959, Cooper and Ross 1975, Wright and Hay 2002, Wright et al. 2005), concreteness (Bock and Warren 1985), specificity (Karimi 2003, Faghiri and Samvelian 2014), prototypicality (Benor and Levy 2006), and subjectivity (Scontras et al. 2017). Priority also tends to favor more active, agentive, positive, powerful, or culturally important elements (Malkiel 1959, Cooper and Ross 1975, Allan 1987). A further pragmatic principle is 'old before new,' that is, given information tends to precede new information (Bock 1977, Fenk-Oczlon 1989, Wasow 2002, Ferreira and Yoshita 2003, Wasow and Arnold 2003, Benor and Levy 2006, Lohmann and Takada 2014). Focus can also affect word order (e.g. Quirk et al. 1972, Zubizarreta 1998, Büring and Gutiérrez-Bravo 2001, Szendroi 2001, Arregi 2002, Samek-Lodovici 2005, Vogel 2006, Selkirk 2011, Büring 2013).

Finally, end-weight is often analyzed in terms of syntactic complexity. Indeed, for higher-level phenomena such as extraposition and heavy NP shift, most studies reckon weight in terms of word count or syntactic complexity, without considering phonological form (though of course syntactic and phonological complexity are highly correlated). Syntactic weight is usually operationalized in terms of word, node, or phrase count; see Wasow (2002:§2), Szmrecsányi (2004), and Shih et al. (2015) for comparisons of metrics. Accounts of weight from the processing literature typically rely on syntactic complexity as it affects locality or the cost of integration (e.g. Hawkins 1990, 1994, 2004, Gibson 1998, 2000, Temperley 2007). Studies directly comparing syntactic to phonological criteria for higher-level end-weight suggest that the former are dominant at that scope of complexity (e.g. Shih and Grafmiller 2013).

3 Seven core properties of PEW

To isolate the contribution of phonology to end-weight, one can control for the non-phonological factors in §2.2 through wug-testing or regression. A WUG, being a nonce word (Berko 1958), does not (necessarily) have properties such as frequency, meaning, or morphosyntactic complexity, rendering the factors in §2.2 moot. For example, one might test a speaker's preference for *glip and badooza* vs. *badooza and glip*. Several wug tests have been conducted for binomial ordering (Bolinger 1962, Pinker and Birdsong 1979, Oden and Lopes 1981, Oakeshott-Taylor 1984, Parker 2002). Moreover, several corpus studies of binomials have sought to isolate phonology's contribution using logistic regression (Wright and Hay 2002, Wright et al. 2005, Benor and Levy 2006, Grafmiller and Shih 2011, Lohmann and Takada 2014, Shih et al. 2015).

Seven phonological properties are established as contributing to prosodic weight, at least in the context of binomials. These are vowel length, vowel lowness, onset complexity, onset obstruency, coda complexity, coda sonority, and syllable count. These properties (except coda complexity) were posited as an ensemble first by Cooper and Ross (1975:71), and have continued to receive support from subsequent corpus-based and experimental research. (Rhythm and phonotactics can also affect word order, though they are largely orthogonal to end-weight, as addressed in §6.) This section comprises four subsections, addressing the nucleus, onset, coda, and syllable count, respectively. At this point, the emphasis remains empirical; analysis and explanation follow in subsequent sections.

3.1 The nucleus

First, long(er) vowels pattern as heavier than short(er) vowels, where 'heavier' in the context of end-weight refers to favoring final position. This tendency is first noted by the ancient Sanskrit grammarian Kātyāyana (re Pāṇini 2.2.34; Vasu 1898). It is one of the core generalizations of Cooper and Ross (1975) and

Ross (1982) in their studies of (mainly) English binomials: Long/tense vowels tend to favor second position over short/lax vowels, as in *trick or treat* and *slip and slide*. Müller (1997) demonstrates the same for German. Pinker and Birdsong (1979) wug-test English and French nonce binomials, analyzing vowel length/tenseness separately from quality. They find both length and quality to be significant predictors of end-weight in both languages. Minkova (2002) concludes the same for English. Oakeshott-Taylor (1984) conducts a detailed study of the correlation between several phonetic factors and end-weight propensity in largely nonsensical binomials (e.g. *peat and poot*) in South African English, Dutch, and Afrikaans, finding phonetic duration to be a significant predictor of end-weight in all three. Consider, for example, the English front vowels in Figure 2, where end-weight propensity refers to the percentage of the time that the vowel was chosen to be in the second item in a balanced binomial ordering task. As the Figure illustrates, end-weight propensities reflect gradient duration, not just binary length.

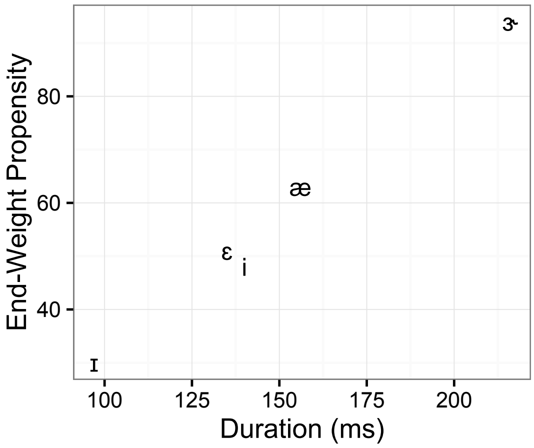


Fig. 2: End-weight propensity vs. duration for five English front vowels, based on wug-test data from Oakeshott-Taylor (1984:228). End-weight propensity is the percentage of the time that the vowel was chosen for final position in a balanced ordering task.

Benor and Levy (2006) and Mollin (2012), in their respective corpus studies of English binomials, test for effects of vowel length, with largely, though not entirely, null results.[2] Nevertheless, these studies are not designed to probe specifically phonological factors, as those in the previous paragraph were. Benor and Levy (2006), for instance, evaluate 411 binomials against 19 predictors, meaning that phonological tendencies might easily be swamped by semantic factors or otherwise be poorly instantiated by the selection of data. A null result, after all, is not the same as a negative result; it can arise from

[2] Mollin (2012:93) does find a significant contribution of length to end-weight, but only in a subset of the data in which semantics, rhythm, and syllable count are held constant.

a test that is not sufficiently powerful, or a paucity of relevant forms. Finally, Lohmann and Takada (2014) find a significant effect of mora count in Japanese, which is related to length, though their model leaves it unclear whether there is a nucleus length effect per se.

Vowel quality is also widely documented to correlate with end-weight, in ways that likely also ultimately reflect duration. English exhibits a well-known tendency for a high(er), front(er) vowel to precede a low(er), back(er) vowel, as in *tit for tat*, *brain and brawn*, and *wend and wander* (Wheatley 1866, Biese 1939, Thun 1963, Gustafsson 1975, Campbell and Anderson 1976, et seq.). Thun (1963), Wescott (1970), and Minkova (2002) provide frequency tables, supporting a general tendency for pairs to increase in F1 (lowness). Jespersen (1961), Marchand (1969), and Shih (2016) also single out vowel height. Cooper and Ross (1975) initially suggest that backness is decisive, but Ross (1982) revises this to height, suggesting that backness is decisive only when height is held constant. Pinker and Birdsong (1979) and Oden and Lopes (1981) both demonstrate that the height effect is productive in binomial wug-tests. Beyond English, a high-before-low tendency has been described for German (Müller 1997), Hungarian (Pordany 1986), and Jingpho (Mortensen 2006). Jespersen (1961:176) mentions also Greek, Lithuanian, and Bantu (sic).[3] Increasing F1 jibes with the more general short-before-long tendency, given that F1 positively correlates with duration crosslinguistically (Lehiste 1970). The tendency for lower vowels to be longer may arise from the greater jaw displacement that they require (Westbury and Keating 1980).

3.2 The onset

Two generalizations concerning onset effects in end-weight approach consensus in the literature. First, greater complexity patterns as heavier. Second, lower sonority patterns as heavier. These two generalizations are considered here in turn.

Cooper and Ross (1975) are perhaps first to articulate the onset complexity effect, based on a number of English examples such as *meet and greet* and *fair and square*. This factor has generally since been corroborated. It is significant in a wug-test by Oden and Lopes (1981). Ross (1982) notes that the contrast between a null and simple onset is greater than that between a simple and complex onset, though both hold. The corpus studies of Benor and Levy (2006) and Mollin (2012) also support the onset size effect, though it is not highly significant in either (but see the caveat in §3.1). Wright and Hay (2002), however, find no significant effect of onset size in their study of binomials of personal names.

In order to further test onset complexity, a short experiment is newly conducted here via Amazon's Mechanical Turk (Daland et al. 2011, Gibson et al. 2011, Sprouse 2011, Yu and Lee 2014). Participants were presented with 15

[3] Languages with beginning-weight show the reverse, low-first tendency; see §5.1.

forced-choice prompts in orthography, of which 10 were fillers and 5 critical items. Fillers consisted of real binomials in which one order is clearly natural (e.g. *near and far*, *black and white*). Critical items are given in Figure 3; these were randomly interspersed with fillers (but never first or last). For critical items and fillers alike, both orders of the binomial were presented in random order as radio-button options. Participants were paid $0.34 for this approximately two-minute task and analyzed only if they were located in the U.S., identified by the service as 'masters,' had a 97% or higher rating based on at least 100 previous tasks, and erred on at most one filler. The fillers served in part as a check that the participants were paying attention. Critical items were monosyllabic wugs, identical in pronunciation except for onset complexity. Spelling was slightly altered so that both conjuncts had the same number of letters, ruling out any possible interference from visual size. In the aggregate, of 27 usable participants, the (phonologically) longer onset was preferred in second position 2.6 times as frequently as in first position, a significant departure from the 50% chance baseline (goodness-of-fit $\chi^2(1) = 25.8$, $p < .0001$). This test therefore corroborates the results in the previous paragraph.

Binomial (in preferred order)		N Agree	N Disagree	% Agree
1.	beck and brek	23	4	85%
2.	keph and klef	20	7	74%
3.	phum and frum	16	11	59%
4.	spimm and sprim	19	8	70%
5.	temm and trem	19	8	70%

Fig. 3: Results for a simple Amazon Turk experiment on onset complexity.

Among other languages, Pordany (1986) supports the onset complexity effect for Hungarian, and Müller (1997) supports it for German. Shih and Zuraw (2016) find it in Tagalog, in that empty onsets are preferred in first position, though they speculate that this might be due to resyllabification avoidance. The situation is the same in Sanskrit: Pāṇini 2.2.33 states that a vowel-initial item goes first; however, it is not clear whether this tendency is due to onset weight or sandhi avoidance (Wackernagel 1905). However, the resyllabification/sandhi explanation does not carry over to languages like English and German, where the same tendency is found. For example, in Figure 3, phonotactics favors the complex-first order (Wright et al. 2005:541), but the observed preference is the opposite.

Beyond segment count, obstruent onsets pattern as heavier than sonorant onsets, as is perhaps clearest when other factors are held constant, as in *wear and tear*, *wheel and deal*, and *huff and puff*. Cooper and Ross (1975), Ross (1982), and independently Campbell and Anderson (1976) state the sonority generalization clearly as such, though indications of it can also be found in earlier work such as Wheatley (1866), Biese (1939), Abraham (1950), and Marchand (1969). Frequency tables in Campbell and Anderson (1976) underscore the strength of the effect. Müller (1997) demonstrates it for German. The onset sonority effect has been further upheld by several experiments. Pinker and Birdsong (1979) find it to be significant in English, though not in French. Moreover, onset sonority is one of the stronger effects in Oden and Lopes (1981), ahead of onset size. Parker (2002) confirms the effect in doublets of the type *roly-poly*, generally supporting falling sonority (with isolated exceptions such as *k* before *g*). Finally, in their study of name pairs, Wright et al. (2005) find a nonsignificant trend for final onset obstruency.

Potentially related to sonority, *h* shows a strong tendency to occupy the first position in English (Wheatley 1866, Biese 1939, Abraham 1950, Marchand 1969, Campbell and Anderson 1976, Parker 2002). In other words, *h* patterns with sonorant onsets and null onsets, as if light. If *h* is regarded as highly sonorous, its behavior can be subsumed by the sonority principle. Laryngeals such as *h* and *ʔ*, after all, are well known to be ambivalent in sonority (e.g. Parker 2002:224). Alternatively, *h* may pattern as light due to its proximity to nullity (e.g. its lack of supralaryngeal place). Onset *h* can comprise a natural class with the null onset, as for example in Ancient Greek resyllabification (West 1987). Some other featural tendencies have been reported more sporadically, such as a preference for labials in second position, though these are less well-established and therefore put aside here.

3.3 The coda

In the coda, unlike in the onset, greater sonority correlates with greater weight, as in *thick and thin*, *beck and call*, and *push and pull*. This generalization is recognized by Cooper and Ross (1975) and Ross (1982). It is also supported by experimental data, with strong results in Bolinger (1962) and Wright et al. (2005) (Pinker and Birdsong 1979, for their part, do not test it). Additionally, Mollin (2012) finds a highly significant effect of coda sonority in English corpus data, though Benor and Levy (2006) obtain a null result (though once again, see §3.1). Moreover, although it has not been investigated systematically, post-hoc analysis of wug-test data from Bolinger (1962) suggests that this sonority effect is gradient. As the sonority difference between the codas increases, the more sonorant coda is increasingly favored in second position, as illustrated in Figure 4 for three sonority classes (obstruent, nasal, and liquid). Paired wugs were identical except for their codas (e.g. *skrit* and *skrill*). The y-axis is the percentage of the time that Bolinger's subjects (in Test 4) preferred the more sonorant-final order. The fact that nasals fall approximately halfway between obstruents and liquids is not surprising in terms of sonority hierarchies (Parker 2002), but suggests that the effect is not driven by [±sonorant] alone.

A few studies (Bolinger 1962, Cooper and Ross 1975, Mollin 2012) further claim that open syllables pattern as heavier than closed syllables, as in *hem and haw* and *lock and key*. This generalization likely folds in with the coda sonority generalization, as the open syllables in question, usually monosyllables, typically contain long vowels or diphthongs. Put differently, the second

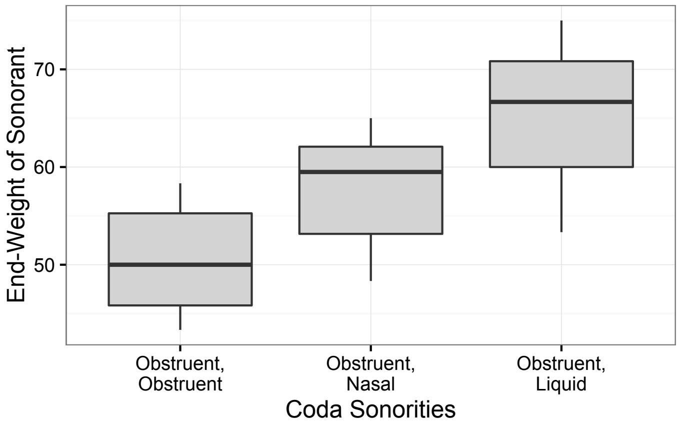


Fig. 4: The greater the sonority difference between codas, the more the higher sonority coda is preferred in final position, based on wug-test data from Bolinger (1962). Once again, the y-axis is the percentage of time that the more sonorous coda was chosen for final position in a balanced ordering task.

timing slot of C_0VV is more sonorous than that of $C_0\breve{V}C$, and greater sonority, wherever it occurs in the rime, contributes greater weight.

Sonority aside, greater coda complexity correlates with greater weight. Ross (1982) reports this tendency for English binomials, and also mentions Arabic as complying (also Pinker and Birdsong 1979:506).[4] Pinker and Birdsong (1979) find a significant finality-favoring effect of coda complexity in their wug-test of French, though in their English test, it is only a nonsignificant trend. The English corpus studies of Benor and Levy (2006) and Mollin (2012) obtain null results for coda complexity. Thus, on the whole, coda complexity appears to be a relatively weak predictor, possibly subordinate even to onset complexity. One possible explanation for its weakness is that longer codas tend to cooccur with shorter vowels, both phonemically (e.g. *betwixt and between*) and durationally (e.g. *beet* has a shorter vowel than *bee*). There is a partial compensatory timing relationship between the durations of the nucleus and coda that does not obtain to the same extent between the nucleus and onset (Abercrombie 1967, Maddieson 1985, Browman and Goldstein 1988, Katz 2010, Ryan 2014). Therefore, increasing coda size may not contribute to the overall duration of the syllable to the same extent that increasing the nucleus size or onset size does.

[4] Ross (1982) corrects Cooper and Ross (1975) on this point. The latter had tentatively suggested that greater coda complexity is favored in initial position (as in *betwixt and between*), though they also cited counterexamples. Most of their positive examples involved vowel length confounds.

3.4 Syllable count

This subsection surveys the evidence for what is here termed the SYLLABLE-COUNT EFFECT (SCE), that is, the tendency for the syllabically longer item to be preferred finally. SCE is argued to be robust as such (at least in the context of binomials) and not merely an artifact of rhythm or frequency, as is sometimes suggested.

SCE is the best known phonological principle of ordering, being cited by the vast majority of studies that consider the prosodic aspect of end-weight (e.g. Pāṇini 2.2.34, Jespersen 1905 et seq., Behaghel 1909, Abraham 1950, Malkiel 1959, Jakobson 1960, Bolinger 1962, Hetzron 1972, Gustafsson 1975, Cooper and Ross 1975, Pinker and Birdsong 1979, Ross 1982, Kelly 1986, Pordany 1986, Fenk-Oczlon 1989, McDonald et al. 1993, Müller 1997, Wasow 2002, Wright and Hay 2002, Wright et al. 2005, Benor and Levy 2006, Mollin 2012, 2013, Wolf 2008, Ingason and MacKenzie 2011, MacKenzie 2012, Thuilier 2012, Lohmann and Takada 2014, Shih and Zuraw 2016). Focus here is on the subset of these studies that address SCE while controlling for rhythm and frequency. To illustrate these possible confounds, consider the binomial *salt and pepper*. At least three motivations for this ordering are conceivable. First, it might reflect SCE. Second, it might reflect frequency, as *salt* is more frequent (§2.2). Third, it might reflect rhythm, as *pepper and salt* violates LAPSE and NONFINALITY, while the preferred order violates neither. Thus, to demonstrate that SCE exists as a weight effect, such possible confounds need to be controlled. Note that many binomial types permit SCE and rhythm to be disentangled. For example, whether a monosyllable is favored before or after an iamb depends on whether SCE or rhythm is dominant.

Benor and Levy (2006) analyze a corpus of binomials using logistic regression, finding that SCE is not only highly significant but also the strongest non-semantic predictor. Importantly, their models also include predictors for frequency and rhythm, and while these predictors are also significant, the fact that SCE remains powerful with the other predictors in the model supports SCE as an independent effect. Kelly (1986) and Mollin (2012) also find SCE to be highly significant in their corpus studies, separately from rhythm and frequency, though they do not use regression. Lohmann and Takada (2014), in their study of English binomials, find SCE to be a powerful predictor ($z = 7.0$), while frequency, also in the model, is not significant ($z = 1.0$); nevertheless, they do not include rhythmic predictors. Wright and Hay (2002) and Wright et al. (2005), in their corpus studies of personal name binomials, consistently support SCE, though they also fail to deconfound it from rhythm.

Evidence from experiments and other languages likewise broadly supports SCE in binomials, with one exception to be discussed below. Pinker and Birdsong (1979), for one, wug-test SCE in English and French, finding it to be a strong effect in both, indeed, the strongest of five phonological predictors tested. Because the words are wugs, frequency is moot. Moreover, rhythm is not a confound. This is trivially the case for French: French stress is (if anything) word-final (modulo focus and schwa; Walker 1975). Thus, rhythm favors (if anything) the long-first order, which ameliorates Lapse. Even in English, Pinker and Birdsong (1979) argue that their SCE result is 'not simply an artifact of stress patterns'; for example, it obtains for pairs in which rhythm is neutral or favors the long-first order (e.g. *plúp over geplúp*). Further evidence for SCE in French comes from adjective-noun order (Forsgren 1978, Abeillé and Godard 1999, Thuilier 2012). For example, (short) adjectives that are usually postnominal become increasingly felicitous prenominally for longer nouns (Thuilier 2012:110; cf. also Shih and Zuraw 2016 on Tagalog). SCE is also reported for Sanskrit (Pāṇini), Latin (Behaghel 1909), Ancient Greek (*ibid.*), Modern Greek (Kiparsky 2009), Finnish (Kiparsky 1968), German (Malkiel 1959, Fenk-Oczlon 1989, Müller 1997), Hebrew (Hetzron 1972), Hungarian (Pordany 1986), and other languages (e.g. Malkiel 1959:151).

Although the English forced-choice experiments of Bolinger (1962) probe only monosyllable plus trochee combinations, which are ostensibly not diagnostic between SCE and rhythm, some evidence from the study suggests that SCE tends to be decisive when the two conflict. Bolinger (1962) tests adjectival binomials such as *frank and candid* before initially stressed nouns (e.g. *statement*) vs. initially unstressed nouns (e.g. *appraisal*). He finds that the SCE-compliant order is strongly preferred in both cases (91% and 84%, respectively), despite its not being rhythmically improving in the latter: *Fránk and cándid appráisal* and *cándid and fránk appráisal* both have one lapse.[5] Bolinger's results also cannot be explained by frequency (e.g. *candid* is more frequent than *frank*). The forced-choice experiments of Wright et al. (2005), for their part, also support SCE, but do not control for rhythm.

The binomial ordering experiments of McDonald et al. (1993) have been mentioned as defeating SCE (Shih et al. 2015), though in fact their results are mixed. Seven experiments are presented. The first five involve a recall-and-production task. None finds a significant effect of SCE, but with three caveats raised by the authors. First, every pair tested in this series differs both in length and animacy (e.g. *child* and *music*). Animacy is always a strong effect, and may eclipse phonology. Second, the tests rely on free-form writing, sometimes reconstructing entire sentences (e.g. *The music soothed the child* vs. *The child was soothed by the music*), not necessarily contexts conducive to SCE (in this case, nobody has suggested that passivization is affected by SCE). Finally, the recall methodology might have competed with euphonic principles, as explained presently.

The final two experiments of McDonald et al. (1993) are more relevant here. The sixth probes inanimate noun binomials in two rhythmic conditions, namely, monosyllable-trochee and monosyllable-iamb. SCE was null in the first and negative in the second: Participants slightly but significantly preferred iambs before monosyllables (albeit at only a 4% deviation from the

[5] One might reply that perhaps adjective order is optimized before the noun is available. But this position is arguably untenable. First, adjectival binomials are indeed sensitive to the stress pattern of the noun (Bolinger 1962, Kelly and Bock 1988). Second, altering the phrasing slightly so that the first adjective is syntactically external, as in [*frank*, [*candid statement*]], does not eliminate the ordering preference.

chance baseline of 50%). One explanation for this finding is that rhythm dominates SCE when the two conflict (NB. this does not imply that SCE is inert), though there was no evidence for rhythm or SCE in the first condition. Another possible explanation is that recall may have interfered with euphony, in that longer words may have been more accessible. While shorter words are more accessible in general (Kelly 1986) and lists of uniformly short words are more easily recalled than lists of uniformly long words (Baddeley et al. 1975), in the context mixed-length recall tasks, the opposite sometimes holds: Longer words can stand out, especially if they are in the minority (Levelt and Maassen 1981, McDonald et al. 1993:223, Cowan et al. 2003). For example, in a free recall task using lists of words varying in length, Katkov et al. (2014) find that length in syllables significantly positively (sic) correlates with recall. Indeed, consistent with a possible damping effect of recall on end-weight, the seventh experiment, which features ratings rather than recall, offers some positive results for SCE. The binomials from the six previous experiments were recycled for a ratings task. SCE is significant in two of the sets, near-significant in a third, and null (but not negative) in the remaining three (all of which have the animacy confound). Moreover, in sets designed to probe rhythm, metrical structure is never significant while length sometimes is, indirectly lending some support to SCE as independent from rhythm. In sum, with the exception of one condition in one of seven experiments, the findings of McDonald et al. (1993) are a mix of null (not to be confused with negative) and positive for SCE.

In general, then, insofar as SCE can be disentangled from frequency, rhythm, and other factors based on existing corpus-based, typological, and experimental studies, it is broadly supported. This section now concludes with a short follow-up experiment, new here, on monosyllable-iamb binomials as an additional check on a key case mentioned above. The study was conducted via Amazon's Mechanical Turk, with the same selection criteria and design as the Turk study in §3.2. Prompts included 7 critical items and 14 fillers. Fillers were real binomials, as in §3.2. Critical items were pairs of female names, presented in both orders (as before, orders were randomized both within and across prompts). Each critical item combined one monosyllable from {*Bree*, *Deb*, *Fay*, *Jade*, *Kai*, *Pam*, *Trish*} with one iamb from {*Annette*, *Denise*, *Diane*, *Elaine*, *Louise*, *Michelle*, *Nicole*}. Words were sampled randomly from each set without replacement, yielding 49 combinations. 35 participants are analyzed. Among critical items, iamb-final orders were chosen almost twice as frequently as iamb-initial orders (157 vs. 88; goodness-of-fit $\chi^2(1) = 19.4$, $p < .0001$). Frequency is not a confound: The eight monosyllables were chosen to be uniformly less frequent than the eight iambs. Therefore, both frequency and rhythm favor the monosyllable-final order. But the iamb-final order predominates. These results therefore corroborate the others in this section in support of SCE as distinct from frequency and rhythm.

4 The stress-weight interface in sentential prosody

Stress above the prosodic word (p-word) is normally right-oriented in English.[6] Following Chomsky and Halle (1968:15–24), Liberman (1975), Liberman and Prince (1977), and many others, this generalization is often termed the NUCLEAR STRESS RULE, as in Selkirk (1995:562): ‘The most prominent syllable of the rightmost constituent in a phrase P is the most prominent syllable of P.’ Nothing here hinges on the question of whether this ‘rule’ is in fact language-specific or rather the reflex of a universal (cf. Cinque 1993, Zubizarreta 1998, Arregi 2002). Only the fact that default phrasal stress is right-oriented in end-weight constructions is relevant here. Consider the example in Figure 5 (cf. Anttila et al. 2010:955). Phrasal stress might be represented, for example, as a bracketed grid, as in (a), where column height indicates relative sentential stress level (only relative, not absolute, column heights matter here), or (b) as a tree, depicting constituency along with the S(trong) vs. W(eak) status of each constituent (overviews and references in e.g. Hayes 1995 and Samek-Lodovici 2005). The lowest level shown in Figure 5 is the p-word. As both representations make explicit, prosodic constituents are right-headed above the p-word.

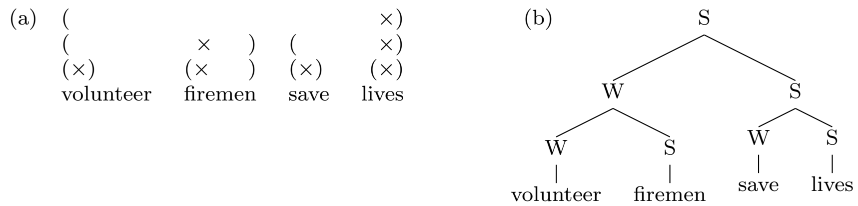


Fig. 5: Two illustrations of relative phrasal prominence.

The phrasal stress explanation of end-weight maintains that greater weight is preferred in final position because that is normally also the locus of greater stress. In short, PEW reflects end-stress. As such, PEW folds in with the more general interface between weight and stress, as documented extensively for the syllable (e.g. Hayes 1995, Gordon 2002, 2006, Ryan 2016). For example, *trick or treat* in Figure 6 is taken as an example of vowel length correlating with weight in binomials. In (a), the attested order, the long vowel is associated with greater stress than the short vowel. In (b), the opposite holds. Thus, (a) better aligns weight with stress than (b).

This explanation of PEW, though not new here, remains overlooked by the vast majority of PEW studies. Surveys by Wright et al. (2005) and Wolf (2008), for instance, mention final lengthening, rhythm, and phonotactics as possible causes of end-weight, but say nothing about rising phrasal stress. Some earlier studies suggest that sentential prosody is relevant, but do not clearly

[6] Certain conditions under which this is not the case are addressed in §5.3–5.4.

```
(a) (              ×)        (b) (              ×)
    (×)   (        ×)            (×)   (        ×)
    trick  or  treat             treat  or  trick
```

Fig. 6

distinguish among independent aspects of prosody such as nuclear stress, final lengthening, and nonfinality (Bolinger 1962, Oakeshott-Taylor 1984). Additionally, several studies have linked focus to nuclear stress, some arguing for movement (or ‘p-movement’) of a focused constituent to final position so that focus and stress coincide (cf. Zubizarreta 1998, Zubizarreta and Vergnaud 2000, Büring and Gutiérrez-Bravo 2001, Samek-Lodovici 2005, Vogel 2006). Nevertheless, the stress-focus interface is orthogonal to PEW. For example, it cannot motivate the weight of long vowels in all-new, broad-focus binomials, as in the *trick or treat* example just given, nor any of the generalizations in §3 under similar conditions.[7]

Two previous studies explicitly propose a link between nuclear stress and PEW.[8] First, Benor and Levy (2006) argue for nuclear stress over final lengthening as an explanation for end-weight on the grounds that binomials exhibit end-weight regardless of their prosodic boundary context. This was one of several arguments raised against final lengthening above in §6.1. Benor and Levy (2006) mention it as a functional principle, though it plays no formal role in their analysis. Second, Anttila et al. (2010) formalize the role of nuclear stress in end-weight via the constraint Stress-to-Stress: ‘Each lexical stress occurs within the prosodic phrase that receives sentence stress.’ They employ this constraint to drive a p-word count effect in the English dative alternation, whereby a coargument of the verb is increasingly likely to be realized finally as its size in p-words increases. In Figure 7, for instance, phrasing (a) receives three violations of Stress-to-Stress, since three lexical stresses (bottom layer) occur in the nonfinal prosodic phrase. Meanwhile, (b) incurs only two violations, and is therefore predicted to be favored. In a weighted-constraints framework, Stress-to-Stress predicts a stronger end-weight tendency as the prosodic word count discrepancy between the direct and indirect objects increases.

Anttila et al. (2010) highlight three novel predictions of Stress-to-Stress. First, function words, lacking lexical stress, are predicted to be ignored for end-weight, an asymmetry that they support for the dative alternation (but cf. §3.4 on binomials). Second, if nuclear stress is lured away, non-end-weight orders should be ameliorated, as in *never send someone them in the mail éither* (with nuclear stress on *either*) and other real tokens collected by Bresnan

[7] Zubizarreta (1998:148), for instance, admits that English has weight-driven p-movement that does not serve to align focus with nuclear stress (e.g. heavy NP shift), but does not formalize an analysis.

[8] This statement puts aside studies such as MacDonald (2015) that cite one of these articles for the nuclear stress point without adding to the argument. Müller (1997) invokes nuclear stress, but uses it only in the context of rhythmic constraints, not as a correlate of weight.

(a)

(				×)
(		×)	(	×)
(×	×	×)	(	×)
gave	critical	backing	to	Bush

(b)

(			×)
(	×)	(	×)
(×	×)	(×	×)
gave	Bush	critical	backing

Fig. 7: Illustration of the dative alternation (grids after Anttila et al. 2010:955).

and Nikitina (2003:20); cf. **never send someone thém*. Finally, if a language has left- instead of right-oriented sentential stress, it is predicted to exhibit beginning-weight. As Anttila et al. (2010) admit, it is unclear whether this last prediction is borne out, though I support it in §5. While Stress-to-Stress capitalizes on the insight that end-weight reflects nuclear stress, it does not address weight per se, and therefore fails to motivate the seven principles in §3. That said, however, the same rationale can be extended to weight, as in the remainder of this section.

The constraint family proposed here penalizes heavy elements under weak phrasal stress. Weak phrasal stress is operationalized as ϕ_w, where ϕ denotes a prosodic node at or above the level of the p-word and w denotes that it is weak (i.e. not a head). A partial stressability hierarchy is defined in (4) (cf. Zec 1988, 1995, 2003). VV notates any long vowel or diphthong, V any vowel, N_μ any moraic sonorant, X_μ any moraic segment, and X any segment. Categories are formulated stringently (Prince 1999, de Lacy 2002, 2004), such that each is a subset of the category listed below it in (4), as diagrammed by Figure 8.

(4) a. $*VV/\phi_w$
b. $*V/\phi_w$
c. $*N_\mu/\phi_w$
d. $*X_\mu/\phi_w$
e. $*X/\phi_w$
Definition: For every element η_i, assign a violation for every node ϕ_{wj} such that ϕ_{wj} dominates η_i.

Each constraint in (4) penalizes a heavy element of the specified type in a weak prosodic setting ϕ_w. These constraints are thus akin to Weight-to-Stress, which penalizes a heavy syllable lacking stress (Prince 1983, Prince and Smolensky 1993/2004, Smith 2002), except that stress is now interpreted phrasally and weight is atomized. Anttila et al.'s (2010) constraint Stress-to-Stress could be folded into this scheme as $*\acute{V}/\phi_w$, though it is omitted from the tableaux below, which exemplify only single-word binomials. The hierarchy in (4) is incomplete; additional sonority effects are addressed later in this section.

This use of generic ϕ_w permits end-weight to emerge in contexts not involving the stress maximum of the sentence. After all, end-weight is still observed in binomials in which neither conjunct contains the matrix nuclear stress. Be-

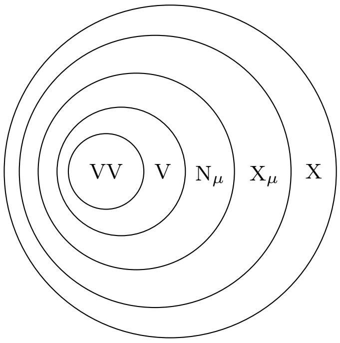
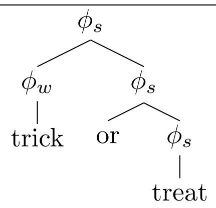
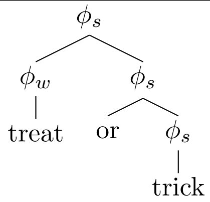


Fig. 8: A stringent weight hierarchy, in which the most embedded set is the heaviest.

cause ϕ_ws can be embedded, a single segment can incur multiple violations from a single constraint. No harm comes from this eventuality, since these extra violations are constant across reorderings of the binomial. Finally, because the constraints in (4) are stringent, the factorial typology excludes grammars that negate weight universals. For example, a consonant is putatively never heavier than a vowel (Gordon 2006, Ryan 2016). If the constraints had been stated as *V/ϕ_w and *C/ϕ_w, the latter could be ranked above the former to implement the counter-universal. But with *V/ϕ_w and *X/ϕ_w, a consonant always receives a subset of the violations of a vowel, and therefore can never be more penalized in ϕ_w.

As an illustration, consider once again *trick or treat*, presented alongside its transposition in the tableau in (5). For additional cases of constraint-based grammar adjudicating between word orders and further discussion of the architectural issues that such an approach raises, see recently Elfner (2012, 2015), Shih (2014), Bennett et al. (2016), Clemens and Coon (2016), Shih and Zuraw (2016), among others (cf. also the TRY-AND-FILTER approach; e.g. Büring 2013:872). Space precludes a detailed review of this literature here. In (5), *VV/ϕ_w is decisive, which (b) alone violates due to the tense vowel in the weak branch of the binomial. The constraints are not crucially ranked at this point, hence the dashed dividers. The issues of ranking and variation are discussed below.

		*VV/ϕ_w	*V/ϕ_w	*N$_\mu$/ϕ_w	*X$_\mu$/ϕ_w	*X/ϕ_w
☞	a. 		*	*	**	****
	b. 	*!	*	*	**	****

(5)

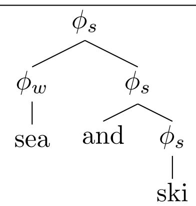
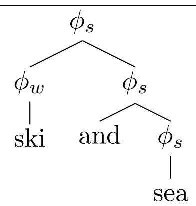
In general, at least for the simple binomials considered here, prosodic structures can be assumed to align with their syntactic structures, consistent with highly-ranked MATCH constraints (Selkirk 2011). Thus, the prosodic structure on the left in Figure 9, as in (5), matches its syntactic structure on the right (cf. Munn 1993, Wagner 2005). The conjunction, lacking stress, is tentatively taken to adjoin to a recursive p-word, but this analysis of clitics is not critical. The only critical assumption about prosodic phonology here is that the right branch of a binomial is prosodically strong, which is a matter of consensus.

Fig. 9

Couched in Optimality Theory (Prince and Smolensky 1993/2004), this analysis predicts that candidate (a) in (5) wins categorically. In the case of *trick or treat*, this is acceptable. In general, however, the length effect is only a tendency in binomials (§3.1). To encode a tendency, the constraints could be numerically weighted in a probabilistic grammar (Goldwater and Johnson 2003, Hayes and Wilson 2008, Pater 2009, Boersma and Pater forthcoming). Because violation vectors are potentially subsetted, as in (5), maximum entropy Harmonic Grammar (Goldwater and Johnson 2003, Wilson 2006, Hayes and Wilson 2008) is perhaps the most appropriate framework, since it lacks harmonic bounding (Jesney 2007). That said, no attempt is made here to determine precise weights of these constraints in English. With any nonzero weights, the generalizations in §3 emerge, which is the main concern here. The question of how exactly the constraints are weighted when they conflict with each other (e.g. vowel length vs. syllable count in *city or town* or *beer and butter*) is put aside here.[9] Another issue left to future work is that of lexicalization. While some binomials are relatively fixed as idioms (e.g. *trick or treat*), others are less so (e.g. *clip or pleat*). Presumably, a faithfulness constraint overrides the productive grammar for idioms, though the implementation of idioms is outside of the scope of the present article (see Zuraw 2010 and Morgan and Levy 2016 for discussion, the latter treating specifically binomials).

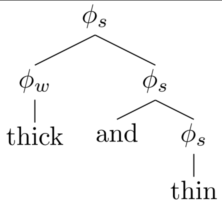
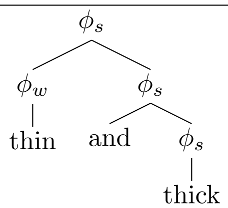
Constraint family (4) captures the effects of margin complexity and coda sonority, as in (6) and (7), respectively. Though there is no dedicated constraint for the syllable-count effect (SCE), SCE emerges from the fact that additional syllables incur additional violations of the other constraints (e.g. *baba* incurs double the violations of *ba*). The vowel height effect, whereby lower vowels pattern as heavier, is not handled by the constraints so far, but the analysis easily extends to it. $*V_{[+low]}/\phi_w$ and similar constraints can be added, with as much featural detail as is justified empirically.

(6)

<table>
<tr><th></th><th></th><th>$*VV/\phi_w$</th><th>$*V/\phi_w$</th><th>$*N_\mu/\phi_w$</th><th>$*X_\mu/\phi_w$</th><th>$*X/\phi_w$</th></tr>
<tr><td>☞</td><td>a. 
</td><td>*</td><td>*</td><td>*</td><td>*</td><td>**</td></tr>
<tr><td></td><td>b. 
</td><td>*</td><td>*</td><td>*</td><td>*</td><td>***!</td></tr>
</table>

[9] Previous work addressing ranking or weighting in some form includes Cooper and Ross (1975), Pinker and Birdsong (1979), Oden and Lopes (1981), Benor and Levy (2006), and Mollin (2012), among others.

(7)

	$*VV/\phi_w$	$*V/\phi_w$	$*N_\mu/\phi_w$	$*X_\mu/\phi_w$	$*X/\phi_w$
☞ a. 		*	*	**	***
b. 		*	**!	**	***

The onset sonority effect, whereby obstruent onsets are heavier (as in *wear and tear*), can likewise be implemented by $*T_{ons}/\phi_w$, where T is any obstruent, or by multiple such constraints if further detail is justified. This constraint, as just defined, stands outside of the stringency hierarchy proposed in this section. For instance, a language could assign weight to $*T_{ons}/\phi_w$ and no weight to the remaining constraints, in which case only onset sonority would matter, or vice versa, in which case onset sonority would be ignored. This may well be a case of overgeneration. Nevertheless, given the paucity of cases of onset sonority in the weight typology, it is not empirically clear how onset sonority should be integrated into the stringency hierarchy. One might begin with the hypothesis that onset sonority is subordinate to all aspects of rime structure. This is suggested by Pirahã (Everett and Everett 1984, Everett 1988), in which onset voicing affects weight, but vowel length takes precedence (e.g. TV outweighs DV, but DVV outweighs TV). Indeed, this hypothesis is supported by numerous languages in which weight is based on the the rime alone, ignoring the onset. On the other hand, the onset appears to be superordinate to the rime in the stress systems of Karo (Gabas 1999) and Tümpisa Shoshone (Dayley 1989), in which onset voicing affects weight, but the presence or absence of a coda does not. I tentatively leave T_{ons} outside of the stringency hierarchy, pending a better understanding of implicational relationships in the weight typology. This question concerns weight in general, not just in PEW.

All of the other constraints in this section, however, are formulated stringently, since their implicational relationships are clearer. And all of the constraints (including onset sonority) are consistent with the phonology of weight more generally across languages and systems, hence the slogan: PROSODIC END-WEIGHT IS WEIGHT. First, nucleus and coda size are well known as canonical determinants of weight (Hayes 1995, Gordon 2006). Second, greater coda sonority and vowel lowness are also widely reported to correlate with weight, as they increase the duration or energy of the rime (e.g. Zec 1988, 1995, 2003, Kenstowicz 1994, Anttila 1997, Prince 1999, Gordon 2002, 2006, de Lacy 2004, Crowhurst and Michael 2005, Nevins and Plaster 2008, Carpenter 2010, Garcia 2017b). For example, Garcia (2017a) finds that coda sonority contributes to weight for English word stress, and Carpenter (2010) finds that English and French speakers are sensitive to vowel height when rating word stress patterns, such that low vowels are better hosts for stress. Third, the treatment of onsets in end-weight, including both complexity and obstruency, is convergent with the onset weight typology across systems (Gordon 2005, Topintzi 2010, Ryan 2011, 2014, 2016). The stress system of Pirahã, mentioned above, famously exhibits both effects, but they are also found elsewhere in stress and meter (Ryan 2011, 2014). On the phonetic rationale for the reversal in the treatment of sonority between the onset and rime, see Gordon (2005) and Ryan (2014). In conclusion, the phrasal stress account of PEW, unlike the other explanations in §6, correctly predicts that the PEW typology should exhibit the same features as the weight typology in other systems.

5 Prosodically head-initial languages and constructions

5.1 Beginning-weight

While phrasal stress is usually right-oriented in English (§4), this is not always the case in other languages. In many verb-final languages, stress is left-oriented in phonological phrases (p-phrases), such that the leftmost p-word heads its p-phrase, as in Bengali (Hayes and Lahiri 1991), Persian (Kahnemuyipour 2003), and Turkish (Kabak and Vogel 2001). Left-headedness does not necessarily continue above the p-phrase. For example, Hayes and Lahiri (1991:55) claim that while p-phrases are prosodically head-initial in Bengali, intonation groups are head-final under neutral focus (similarly for Persian; Kahnemuyipour 2003:337).

Following the phrasal stress account of PEW, insofar as such languages exhibit left-oriented phrasal stress and weight polarity, they are predicted to exhibit BEGINNING-WEIGHT rather than end-weight. A few caveats are in order. First, given the possible inconsistency of prosodic headedness just mentioned, prosodic beginning- vs. end-weight is not necessarily a language-wide parameter. Second, syntactic weight and prosodic weight might not behave identically, as discussed below for Japanese. Finally, the phrasal stress account may not extend to languages in which headedness is not realized as stress (again, this nuance may be relevant for Japanese).

The evidence for prosodic (as opposed to syntactic) weight at the phrasal level in prosodically head-initial contexts is limited at this point, but what evidence exists tentatively favors the phrasal stress account of weight. In the context of binomials, Tungus (Swadesh 1962, Wescott 1970) and Turkish (Marchand 1952, 1969, Pinker and Birdsong 1979) show the reverse vocalism as English, that is, low-before-high, ostensibly a case of heavy-before-light. A few Turkish reduplicatives and binomials illustrating this trend are given in (8) (Marchand 1952). The initial members of these compounds are their prosodic heads, as with p-phrases and compounds more generally in Turkish (Kabak and Vogel 2001). Kabak and Vogel (2001) analyze binary compounds as comprising two prosodic words, even though little to no stress is perceived on the second member. Thus, a compound such as $(\textit{takur})\phi_s(\textit{tukur})\phi_w$ does not violate $^*\mathrm{V}_{[+low]}/\phi_w$, while its transposition does. See the next section concerning syllable-count based beginning-weight in Turkish.

(8)
a. takur tukur 'harsh'
b. cak cuk 'noisy'
c. yamuk yumuk 'swollen'
d. çalı çırpı 'wood chip'
e. para pul 'money'

Other cases of low-before-high vocalism, though perhaps not previously discussed as such, come from South Asian languages such as Tamil, which, like Bengali, is verb-final and prosodically left-headed in the word and phrase (Keane 2003, 2006). This vocalism is clear as a stereotype in at least two contexts in Tamil, including deictic oppositions and echo reduplication. First, in deictic pairs, the distal is favored before the proximal, as in (9). This is the opposite of the order found in English, where the proximal usually goes first (e.g. *this and that*, *here and there*, *now and then*; Cooper and Ross 1975). In Tamil, the distal form differs from its proximal counterpart in substituting [ɑ] for [i] (Asher and Annamalai 2002:18). The proximity reversal vis-à-vis English is therefore potentially motivated by stress-weight alignment, that is, once again, $^*\mathrm{V}_{[+low]}/\phi_w$.[10]

(9)
a. atu itu 'that [and] this'
b. avaṉ ivaṉ 'he (distal) [and] he (proximal)'
c. aňkē iňkē 'there [and] here'
d. appōtu ippōtu 'then [and] now'
e. appaṭi ippaṭi 'in that way [and] in this way'

A low-first vocalism is also encountered in echo reduplication in Tamil. This construction involves copying a word, but replacing the initial CV of the second copy with *ki* [gi] or *kī* [giː], matching the phonemic vowel length of the base, as in (10) (Keane 2001). Because [i] is the shortest stressed vowel in Tamil (and [iː] the shortest stressed long vowel), this vocalism guarantees that the vowel in initial position, the locus of greatest stress, is phonetically longer than (or equal to) the vowel in second position. Turkish and Tamil are thus like mirror images of English (cf. *bric-a-brac*, *tit for tat*) when it comes to binomial/echo vocalism, a fact that must be stipulated if their mirror-image stress patterns are ignored.

[10] Beyond Tamil, Cooper and Ross (1975:101) offer a few examples of Hindi binomials suggesting that 'Hindi contradicts the English ordering fairly systematically.' Judging by their data, it is not clear whether the discrepancy arises from phonology or semantics.

(10)	a.	taṇṇīr-kiṇṇīr	'water and such'
	b.	pāmpu-kīmpu	'snakes and such'
	c.	puli-kili	'tigers and such'
	d.	pai-kī	'bag and such'

Beyond vocalism, the case for syntactic beginning-weight has been made for at least four languages, always in head-final contexts: Japanese (Dryer 1980, Hawkins 1994, 2004, Yamashita and Chang 2001, 2006, Chang 2009, Jaeger and Norcliffe 2009), Korean (Choi 2007), Cantonese (Matthews and Yeung 2000), and Persian (Faghiri and Samvelian 2014) (cf. also Hawkins 2004:131 on Hungarian). Japanese and Korean, like Bengali and Tamil, are rigidly head-final. Insofar as ordering alternatives are available in these languages (e.g. for coarguments of the verb), longer elements are favored earlier, the mirror image of English. Although Persian is not rigidly head-final, the situation is evidently similar for its preverbal field (Faghiri and Samvelian 2014). Finally, while Cantonese has a basic word order of subject-verb-object, Matthews and Yeung (2000) show that prenominal relative clauses, a head-final context, exhibit beginning-weight.

Thus, beginning-weight appears to be associated with head-finality. Hawkins (1994, 2004) capitalizes on this correlation to explain beginning-weight in terms of minimizing dependency distances (§6.4; cf. Yamashita and Chang 2001 for a different approach). For example, in a VO (verb-object) language like English, placing the shorter of two post-verbal arguments first entails that the left edges of both arguments are closer to their head than they would be in the reverse order. In an OV language, by contrast, locating the shorter argument second maximally aligns the arguments' (now right) edges with their head. Existing studies of beginning-weight in East Asian languages and Persian (*op. cit.*) all conceive of weight in terms of syntactic complexity (or word count),[11] and their functional explanations follow suit by invoking processing or conceptual factors that relate to syntactic or lexical complexity. It is not presently clear whether prosodic beginning-weight also exists in these languages, if syntactic complexity were controlled.

One study that addresses an aspect of this question for Japanese is Lohmann and Takada (2014). This article analyzes Japanese binomial ordering using logistic models that include predictors for syllable count, mora count, frequency, and pragmatic factors. Syllable count is nonsignificant in their data, while mora count weakly favors end-weight, not beginning-weight. However, caution is warranted, as mora count is fairly borderline ($p > .01$) and its effect size is small compared to that of syllable count in English, which they also test using the same model. Thus, while the evidence for syntactic beginning-weight in Japanese is clear, there is no support from this study for prosodic beginning-weight; if anything, PEW is supported.

Nevertheless, this discrepancy between syntactic beginning-weight and a null or weakly reverse effect in the prosody may not be surprising, given the

[11] Choi (2007) measures phrasal length in syllables, but interprets it as a proxy for word count, and therefore does not disentangle syntactic and phonological weight.

phrasal stress account of PEW. After all, the explanations for beginning-weight put forth by Hawkins (1994, 2004) and Yamashita and Chang (2001) apply to syntactic or lexical complexity, and are thus moot for single-word binomials. Meanwhile, Japanese is a pitch accent language, lacking in nuclear stress qua stress, although prominence can be signaled through pitch phenomena. As Venditti et al. (2008) emphasize, "standard Japanese has no analog to the notion 'accent' when it is used as a synonym for 'nuclear stress' in these Germanic languages." In this sense, the near-null result for prosodic weight in Japanese in Lohmann and Takada (2014) becomes a point in favor of the phrasal stress account rather than a liability for it. A head-final language without stress is predicted to exhibit syntactic beginning-weight, but not prosodic polarization. At any rate, stress languages with left-oriented p-phrases, such as Persian and Turkish, are better test cases for prosodic beginning-weight. Further research is required in order to draw any firm conclusions on prosodic beginning-weight.

5.2 A pilot study of prosodic beginning-weight in Turkish

Turkish was mentioned in the previous section as being a candidate for prosodic beginning-weight, given that its p-phrases are prosodically head-initial. Some support was adduced there from vocalism. Here, I provide further support from syllable count. Binary conjunctions of simplex color terms were investigated in the 3.4-billion-word 'TenTen' corpus at `www.sketchengine.co.uk` (accessed April 15, 2017). The eleven color terms in (11) — the most frequent in Turkish — were checked in all 55 (11 choose 2) nonrepeating combinations in both orders with *ve* 'and' or *veya* 'or' as a conjunction (*gri ve beyaz*, *beyaz ve gri*, etc.) to probe whether longer items are favored initially in binomials, controlling for frequency.[12]

[12] *Kahverengi* 'coffee-colored, i.e., brown' is excluded because due to its transparent complexity. Because the corpus lacks a treebank, initial results included matches that were not self-contained constituents. For example, sometimes the pair is part of a longer list, or the first member is modified. Results were therefore hand-checked to exclude nonconstituents. For queries with more than 20 hits, the first 20 were checked, and the mean constituency rate (always around 75%) was extrapolated to the remaining hits. In practice, there is no reason to suspect that nonconstituency, even if left uncontrolled, would be a pernicious confound as opposed to harmless noise. Nonconstituency rates are similar across pairs, meaning that hit counts were reduced by a near-constant factor, leaving the relative frequencies of pairs close to what they would have been had constituency not been addressed.

(11)

	Word	IPA	Gloss	Syllables	Log Frequency
a.	*gri*	[ˈgriː]	‘gray’	1	4.66
b.	*mor*	[ˈmoɾ]	‘purple’	1	4.52
c.	*beyaz*	[beˈjɑz]	‘white’	2	5.59
d.	*siyah*	[siˈjɑh]	‘black’	2	5.39
e.	*mavi*	[ˈmɑːvi]	‘blue’	2	5.12
f.	*yeşil*	[jeˈʃil]	‘green’	2	5.44
g.	*sarı*	[sɑˈɾɯ]	‘yellow’	2	5.24
h.	*pembe*	[pemˈbe]	‘pink’	2	4.78
i.	*kırmızı*	[kɯɾmɯˈzɯ]	‘red’	3	5.49
j.	*lacivert*	[ˈlɑd͡ʒiveɾt]	‘navy’	3	4.11
k.	*turuncu*	[tuˈɾund͡ʒu]	‘orange’	3	4.36

As Figure 10 shows, in the aggregate, the longer the color term is in syllables, the more likely it is to be first in its binomial. For one vs. two syllables, frequency is a possible confound, since the disyllables are more frequent than the monosyllables. Therefore, the long-first tendency in that subset could in principle reflect a frequent-first effect. However, frequency makes the opposite prediction for two vs. three syllables. Trisyllables are aggregately less frequent than disyllables, and yet trisyllables are favored initially.

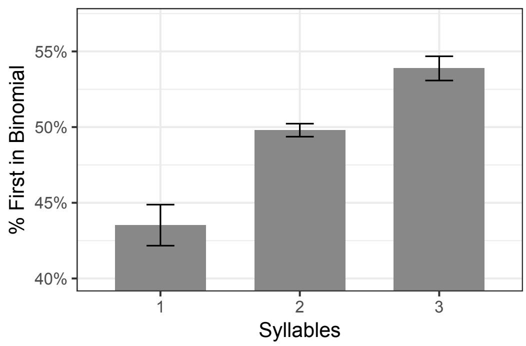


Fig. 10: As the number of syllables in the color term increases, it is more likely to be initial in its binomial.

Regressing on syllable count and frequency, as in (12), reveals both to be significant. Each binomial token is a taken to be a datum in an intercept-free logistic regression (as in Lohmann and Takada 2014). ‘Syllables’ is coded as the number of syllables in the first member minus that of the second member, and similarly for ‘Frequency.’ The two factors are scaled so that their magnitudes can be compared. Both are significant, with syllable count favoring initial position and frequency (perhaps surprisingly) favoring final position, albeit more weakly. The interaction of the two is also included, but not significant. Both main effects are reversed vis-à-vis languages like English and

French, suggesting that prosodic beginning-weight exists in Turkish not just for vocalism, but more generally. This study is tentative, however, in that it addresses only color terms in a corpus.

(12)

	Factor	β	SE	z	p
a.	scale(Syllables)	0.1583	0.0112	14.188	$< .00001$
b.	scale(Frequency)	−0.0831	0.0112	−7.440	$< .00001$
c.	interaction of a:b	0.0023	0.0086	0.268	$= .79$

5.3 Focus and stress manipulations

In §4, it was observed that phrasal stress normally falls on the final member of a binomial in English. This section considers some exceptions to this generalization and their implications for the present proposal. First, binomials with extender phrases such as *and such* and *or something* receive greater stress on the preceding member. These are immaterial here because their order is independently fixed (*trees and such*; **such and trees*).

A second situation in which greater stress can appear on the first member of a binomial in English concerns the Rhythm Rule, which optionally demotes a phrase-medial stress to ameliorate clash, as in *Míssissìppi múd* (cf. *Mìssissíppi* in isolation; Liberman and Prince 1977, Hayes 1984, *inter al.*). The same rule can affect binomials, as in *róugh-and-tùmble pláy* (cf. *ròugh and túmble*).[13] The phrasal stress theory of PEW predicts, under certain architectural assumptions, that beginning-weight should be favored when the Rhythm Rule applies, all else being equal.[14] As before, however, other factors must be held constant. *Rough and tumble*, for instance, may resist reordering because it is lexicalized (see Morgan and Levy 2016 on freezing in binomials). Similarly, *Bill and Melinda* in (Rhythm Rule-eligible) *Bill and Melinda Gates Foundation* is both lexicalized and confounded by semantics (e.g. gender ordering; Wright et al. 2005).

A better test case for the Rhythm Rule involves wug binomials, as in (13) (from Bolinger 1962). As Bolinger (1962) confirms, end-weight is favored prenominally for such binomials (2:1 for (13)). The question is whether (or how often) the Rhythm Rule applies to them. My own intuition is that it is unlikely, though dialects may vary. One can at least say that demoting the second member of a novel pair such as *flope and flome* is less likely than it is for a lexicalized pair such as *rough-and-tumble*, perhaps because the latter more readily accepts a compound treatment adnominally (as implied by English orthography, which hyphenates *rough-and-tumble* as an adjective but not as a noun). *Flópe and flòme grín* to my ear implies that the wug pair is being analyzed as a compound rather than as two independent predicates, as

[13] This issue was brought to my attention by Adam Albright and an anonymous referee.

[14] Under some architectures, the Rhythm Rule is predicted to have no effect. For example, if one adopts a derivational approach whereby the order of the binomial is established before the Rhythm Rule applies in the context of the larger phrase, then the Rhythm Rule is predicted to have no effect on end-weight preferences.

intended. Compare *black-and-white issues*, which favors a compound reading and stress pattern, to *make gray paint by combining black and white paint*, which resists a compound reading/prosody. At any rate, this question calls for a production study with a control condition, which is beyond the reach of the present article.

(13)
a. Get rid of that flope and flome grin on your face.
b. Get rid of that flome and flope grin on your face.

Beyond the Rhythm Rule, focus manipulations of binomials are potentially relevant. Given that coordinated elements are normally of the same type, narrow focus on only one member, as in (14-a), is unlikely out of the blue. But it is possible under contrastive focus, as in (b) and (c). The focused element can be initial (b) or final (c), though it must be parallel between binomials. The present account does not distinguish between (b) and (c), since greater stress is in both cases associated with greater length, but related manipulations might yield a diagnostic case.

(14)
a. ?I want APPLES and pears.
b. I said I want APRICOTS and pears, not APPLES and pears.
c. I said I want pears and APRICOTS, not pears and APPLES.

An anonymous referee points to another potentially relevant focus manipulation: The binomial as a whole can be focused, as in (15-a). Assuming that focusing the whole binomial disproportionately affects its main stress, APPLES AND PEARS has a greater stress differential than *apples and pears*. If this premise holds, the stress account predicts that end-weight should be at least as strong in type (a) as it is in type (b) or (c), without contrastive focus. This test calls for a more nuanced experiment than the simple binomial ratings tasks above.

(15)
a. I said I want APPLES AND PEARS, not APRICOTS AND PLUMS.
b. I said I want apples and pears.
c. I said I want apples and pears, not apricots and plums.

Given the present proposal, one might also expect end-weight to be attenuated under post-stress deaccentuation, as in (16-a) relative to the control condition in (b).

(16)
a. I *ate* the apples and pears; I didn't merely admire them from a distance.
b. I ate the apples and pears.

As mentioned in §4, Anttila et al. (2010) provide some support for the attenuation of end-weight under deaccentuation conditions similar to (16-a). Based on corpus data cited in Bresnan and Nikitina (2003), they observe that otherwise illicit light-final orders of two objects of a ditransitive verb become more acceptable when nuclear stress is lured away from the objects, as in (17).

This is another area in which a systematic study is a desideratum, but the suggested generalization favors the present approach.

(17)
a. never send someone them in the mail either
b. cf. *never send someone them
c. not to give children it to avoid possible allergies
d. cf. *not to give children it

5.4 Compounds

Compounds are another context in which stress above the p-word is not necessarily right-oriented. The phrasal stress account of end-weight predicts that insofar as compounds comprise multiple p-words, they should exhibit weight polarity mirroring their prosodic headedness. Binary compounds in English are usually prosodically left-headed (e.g. *kétchup factory*, *bláckboard*),[15] though in more complex compounds, prosodic headedness can be affected by syntax (Chomsky and Halle 1968, Liberman and Prince 1977, Arregi 2002). Nevertheless, the order of most compounds is syntactically determined, rendering end-weight moot. For example, because *ketchup factory* means something different from *factory ketchup*, prosodic weight has no opportunity to assert itself. That said, there is one major type of compound in English for which end-weight is relevant, namely, echo reduplication. Syntax has nothing to do with, say, *hoity-toity* being preferred to *toity-hoity*. Following Dienhart (1999), echo compounds include ONSET REDUPLICATIVES, in which onsets vary (e.g. *hoity-toity*), and ABLAUT REDUPLICATIVES, in which vowels vary (e.g. *dilly-dally*). Because their stress patterns differ, these types are treated separately here.

For onset reduplicatives, stress is normally on the initial element if the elements are monosyllables (e.g. *hóbnob*), and otherwise on the final element (e.g. *artsy-fártsy*). Consistent with the phrasal stress theory of PEW, end-weight is observed in onset reduplicatives when they are disyllable pairs, but not when they are monosyllable pairs. Specifically, judging by Dienhart's (1999) data, when monosyllable pairs differ in onset complexity (e.g. *crúmbum*), the longer onset is second 55% of the time, essentially chance (6 of 11; goodness-of-fit $\chi^2(1) = .1$, $p = .76$). For disyllable pairs, the longer onset is second 89% of the time, significantly greater than chance (17 of 19; $\chi^2(1) = 11.8$, $p < .001$). When monosyllable pairs comprise members beginning with simplex onsets differing in sonority, the second element begins with the less sonorous onset 45% of the time, again, essentially chance (14 of 31; $\chi^2(1) = .29$, $p = .59$). For disyllable pairs, the less sonorous onset is second 90% of the time (28 of 31; $\chi^2(1) = 20.2$, $p < .0001$). In sum, monosyllable pairs, which have left-oriented stress, lack end-weight, while disyllable pairs, which have right-oriented stress,

[15] Accents here indicate the stress maxima of compounds, not word stress in general.

observe end-weight. This asymmetry, visualized in Figure 11, is precisely what the phrasal stress theory of PEW predicts.[16]

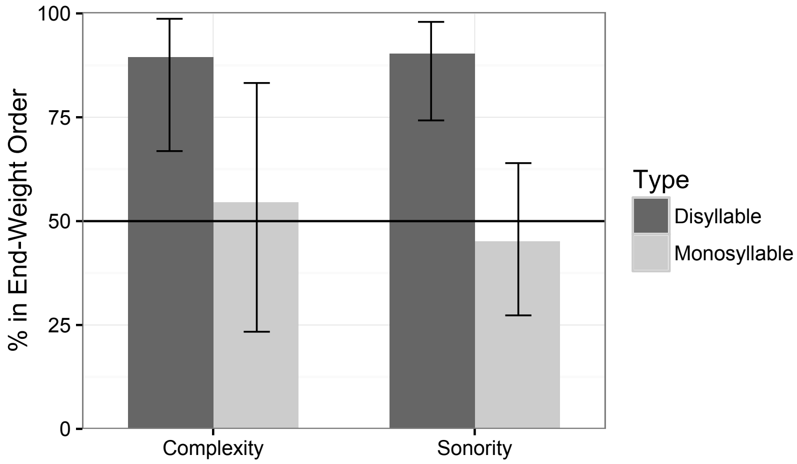


Fig. 11: End-weight among onset reduplicatives consisting of disyllables vs. monosyllables, based on data from Dienhart (1999). Disyllable pairs exhibit end-weight in terms of both complexity and sonority, while monosyllable pairs exhibit neither.

For ablaut reduplicatives, stress is usually initial for both monosyllable and disyllable pairs (e.g. *chítchat*, *dílly-dally*) (Dienhart 1999), though it is final for polysyllables and more complex disyllables (e.g. *twinkum-twánkum*, *Jiminy Jáminy*, *clickety-cláck*) (Minkova 2002). These compounds are at first glance problematic for the phrasal stress analysis, in that they exhibit a high-before-low tendency even when peak stress is initial (*ibid.*). However, initially stressed ablaut reduplicatives do not comprise multiple p-words, rendering phrasal stress moot. For example, ϕ_w constraints do not affect the internal organization of $((dilly)f_s(dally)f_w)\phi$. This foot parse is supported by the fact that stress shifts to the final member in forms in which the members are too large to be their own feet, as in $((twinkum)\phi_w(twankum)\phi_s)\phi$. These longer cases are prosodified as separate p-words, and end-weight applies as expected.

Nevertheless, the question remains: If short ablaut reduplicatives such as *dilly-dally* are not subject to ϕ_w constraints, why do they exhibit a high-before-low vocalism consistent with end-weight? The proposal here is that this vocalism does not synchronically reflect weight, but is rather a case of

[16] Monosyllable pairs exhibit neither beginning- nor end-weight. Phrasal stress predicts that they should exhibit beginning-weight if they are parsed into multiple p-words. However, monosyllable pairs are plausibly parsed into feet (f), for example, $((hob)f_s(nob)f_w)\phi$, in which case phrasal stress correctly predicts that weight is ignored. Meanwhile, a longer form such as *artsy-fartsy* is prosodified as separate p-words — it has the same prosody as *artsy and fartsy* — and ϕ_w constraints apply, favoring end-weight.

FIXED SEGMENTISM (Alderete et al. 1999). One way to distinguish between end-weight and fixed segmentism is to assess the range of vowels involved. If the range is highly limited (at the extreme, one vowel per position), it is more symptomatic of fixed segmentism than of the free operation of end-weight. Coordinate binomials, for their part, must exhibit end-weight as opposed to fixed segmentism, since the vowels entering into the binomial are unrestricted. Similarly, onset reduplicatives must exhibit end-weight, since the onsets involved vary widely. For ablaut reduplicatives, by contrast, the situation is more akin to fixed segmentism. Judging by the table in Minkova (2002:150), 92.2% of ablaut reduplicatives have a first member with /ɪ/. /i/ is the next most frequent at 5%, but as Minkova (2002) points out, it is often driven by length agreement. Meanwhile, in the second element of the reduplicative, the vowel is almost always (95.4%) /æ/ or /ɑ∼ɔ/.

In short, ablaut reduplicatives normally exhibit the limited vocalism /ɪ – æ∼ɑ∼ɔ/, a situation resembling fixed segmentism. If it were end-weight alone, one would expect vowels to combine more freely (albeit in orders favored by end-weight), as with hypothetical compounds such as *booty-boaty* (/u-o/), *babble-bobble* (/æ-ɑ/), and *matey-Matty* (/e-æ/), all of which have the longer vowel in second position, though they flout the fixed vocalism. To claim that /ɪ – æ∼ɑ∼ɔ/ is synchronically fixed is not to deny that the specific vowels in that formula might ultimately be influenced by end-weight. After all, /ɪ/ is the shortest full vowel in English, and /æ∼ɑ∼ɔ/ are located opposite from it in the vowel space; these facts hardly seem arbitrary. However, the vocalism might have been more transparently related to stress in earlier stages of its entrenchment. For example, it was always reinforced by binomials (e.g. *tit-for-tát*, *riff for ráff*), which remain stressed on their final elements. Moreover, some pairs were borrowed from French, where they were finally stressed, before being reanalyzed as initially stressed at some point in English (e.g. *ríffraff* < *rif et ráf*; *bríc-a-brac* < *bric à brác*). Even among Germanic compounds, Dienhart (1999) and Minkova (2002) mention that stress can vary for some ablaut reduplicatives, such as *tip-top*, depending on part of speech and other factors. Thus, changing stress patterns may have partially opacified the conditions under which the fixed segmentism originated. At any rate, as long as there are not multiple p-words involved, these cases are outside of the purview of the phrasal stress theory proposed in §4.

6 Other explanations for PEW

This section offers a critical appraisal of four alternative explanations for PEW, namely, final lengthening, rhythm, phonotactics, and complexity.

6.1 Final lengthening

First, phrase-final lengthening (FL) is sometimes suggested as an explanation for PEW (Pinker and Birdsong 1979, Ross 1982, Oakeshott-Taylor 1984, Minkova 2002, Wright et al. 2005, Wolf 2008). FL refers to the phenomenon whereby prosodic constituents of a certain level are prolonged at their right edges (Delattre 1966, Lindblom 1968, Wightman et al. 1992, Turk and Shattuck-Hufnagel 2000, 2007, Lunden 2006). FL applies clearly at the level of the phonological phrase and higher; it is less clear whether it applies to prosodic words that are not phrase-final (Turk and Shattuck-Hufnagel 2000). The domain of lengthening is to a first approximation the final syllable, though this description oversimplifies in two respects. First, different parts of the final syllable are stretched to different degrees. For example, the final onset, though perhaps not wholly unaffected, is not affected to nearly the same extent that the rime is, and within the rime, the coda is lengthened more than the vowel (*op. cit.*). Second, a small amount of lengthening affects material preceding the final syllable, though it is comparatively minor — perhaps an order of magnitude less in English — and may be confined to stressed syllables (Turk and Shattuck-Hufnagel 2007).

Concerning end-weight, the insight is generally that placing the longer item second better aligns inherent length with the locus of FL, though this rationale has not been formalized, and faces several problems. As is argued here, FL plays no role in end-weight. First, end-weight applies just as strongly to binomials in which the second item is not final in a prosodic phrase (e.g. *check and discipline himself*), a point raised and supported by corpus data by Benor and Levy (2006); it is also demonstrated experimentally by Bolinger (1962) (see §3.4). Bolinger finds that end-weight strongly applies to adjectival binomials before a noun, a context in which no prosodic phrase boundary normally intervenes (e.g. *frank and candid appraisal*). In fact, end-weight is significantly stronger when the adjectival binomial is prenominal than sentence-final (*ibid.*:40), the opposite of what FL predicts. Moreover, trinomials also exhibit end-weight, despite pause after every conjunct. One might counter that the final boundary is the strongest, but just as with binomials, this is not necessarily true. Pause often follows every conjunct except the final one, apparently without affecting end-weight, much less reversing it (e.g. *the lock, stock, and barrel of a gun*). Thus, end-weight is not determined by the boundary structures of binomials and multinomials.

A second point against the FL explanation is that FL is almost entirely confined to the phrase-final syllable (Lindblom 1968, et seq.), but end-weight remains highly (if not more) sensitive to nonfinal syllables. For instance, in binomial experiments, end-weight is not attenuated when the critical modulation is two or more syllables from the ends of the words, as with the trisyllabic pairs of the type *neeminy-nominy* in Oden and Lopes (1981). Additionally, I ran a small Amazon Turk experiment, set up as in §3.2, to check whether binomial ordering is more sensitive to a medial, stressed vowel or to a final, unstressed vowel when the two make crossed predictions for end-weight. For example, consider the wug pair *climmo* ∼ *clamma*. If end-weight is more sensitive to the final vowel, as predicted by FL, *climmo* should be preferred finally, since [oʊ] is heavier than [ə]. If, however, it is more sensitive to the stressed vowel, *clamma* should be preferred finally, since [æ] is longer than [ɪ]. Five such pairs were tested on 37 qualifying participants, screened as in §3.2. The results in Figure 12 show each binomial in its most frequently selected order, which is always the stress-sensitive order. In the aggregate, speakers comply with stress-sensitive end-weight 1.45 times more frequently than ultima-sensitve end-weight ($\chi^2(1) = 6.6$, $p = .01$). This suggests that speakers are generally more sensitive to stressed syllables than to final syllables in assessing end-weight, contradicting the FL account. Indeed, this same conclusion is implicit in Benor and Levy (2006) and Mollin (2012), who code the vowel features of their binomials for stressed as opposed to final vowels, as well as in experiments such as Pinker and Birdsong (1979) and Oden and Lopes (1981), which modulate stressed as opposed to final vowels. To be sure, the claim here is not that speakers ignore properties of unstressed final syllables in assessing end-weight; it is just that they attribute greater importance to stressed syllables.

Binomial (in preferred order)		N Agree	N Disagree	% Agree
1.	bitnaw and batnee	19	18	51%
2.	brimminaw and brimminee	22	15	59%
3.	climmo and clamma	21	16	57%
4.	minto and monta	28	9	76%
5.	pihvo and pahva	20	17	54%

Fig. 12: Binomial orders are generally more sensitive to stressed syllables than to final syllables when the two make crossed predictions for end-weight.

Third, FL cannot explain the syllable-count effect, one of the best-established principles of PEW (§3.4). Given that FL is almost entirely confined to the ultima, it is not clear why adding more syllables to the word would assist FL. In fact, FL predicts the opposite: Another well-established timing principle is polysyllabic shortening (Lehiste 1972, et seq.), whereby final syllables are progressively shorter in words with progressively more syllables (Turk and Shattuck-Hufnagel 2000:403). Because the locus of FL is more compressed in longer words, FL predicts (if anything) a reverse syllable-count effect in end-weight. This argument is not undermined by the fact that FL can also slightly affect nonfinal syllables in English. As mentioned, Turk and Shattuck-Hufnagel (2007) find a nonfinal effect of FL only for the stressed syllable in English, and even then the magnitude of FL is roughly a tenth of what it is for the ultima. But as shown in §3.4, adding unstressed nonfinal syllables significantly contributes to end-weight, as seen, for instance, with monosyllable-iamb pairs in English, or with any uneven binomial in French. Not only can FL not motivate this effect; if anything, it predicts a long-first tendency.

A fourth point against FL-driven end-weight is that FL and PEW diverge with respect to the treatment of subsyllabic factors (§3). In FL, the ultimate coda is lengthened to a greater extent than the ultimate nucleus, on the order of 1.5 to four times as much (based on Dutch, English, and Hebrew; Turk and Shattuck-Hufnagel 2007:462). Meanwhile, the ultimate onset is unaffected or at most slightly affected by FL. Turk and Shattuck-Hufnagel (2007), for one, find no consistent effect of onset lengthening in English FL. Thus, if end-weight reflected FL, one would expect coda structure to have the strongest effect on end-weight, followed (substantially) by the nucleus, followed by little to no effect of the onset. This is almost the mirror image of the hierarchy actually observed. In end-weight, the nucleus is the most decisive factor (as made explicit by Pinker and Birdsong 1979:502, Oden and Lopes 1981:677, and others). Meanwhile, among margins, both the onset and coda contribute significantly (if anything, the onset is more decisive than the coda, though see §3.3 on why coda weakness might be apparent only). These facts are consistent with weight: In gradient weight systems, nuclei take priority over onsets and codas, but both of the latter also contribute significantly (e.g. Ryan 2014, 2016:726–8).

Fifth, FL misses the mark for onsets in another respect. While onsets are not (substantially) lengthened in the ultima, they are substantially lengthened prosodic phrase-initially (Turk and Shattuck-Hufnagel 2000:402 and references therein). Thus, insofar as the left edges of binomials coincide with breaks, one would expect longer onsets to be preferred initially, given a timing-based approach to end-weight. This is is the opposite of what is found. Longer onsets are consistently favored finally, including when the binomial follows a break (§3.2). Once again, weight makes the correct prediction here: Longer onsets are associated with (if anything) greater weight (Ryan 2014), and are therefore predicted to be (if anything) favored in final position, the locus of greater stress.

Sixth, the FL account predicts that only PEW, not beginning-weight, should be attested typologically, given that there is no phrase-initial analog of FL in any language.[17] Preliminary indications, however, support the existence of languages exhibiting prosodic beginning-weight (see §5.1). If these cases hold up, it is an additional point against FL-driven weight effects. If they do not, the previous points are unaffected.

In sum, six empirical problems are raised here for the FL explanation of end-weight. First, end-weight is not affected by boundary structure, while FL is. Second, end-weight is more sensitive to stressed syllables than to final syllables. Third, FL cannot motivate SCE. Fourth, FL predicts the primacy of the coda and the irrelevance of the onset, both erroneously. Fifth, if phrase-initial consonant lengthening is also considered, longer onsets are predicted to be favored initially by a timing-based account, also erroneously. Finally, insofar as prosodic beginning-weight exists, the FL-based explanation does not go through, while the weight-based explanation remains viable. Beyond these empirical points, it is worth noting that the logic of the FL explanation has never been explicated in a formal model, which may conceal further issues of implementation and typology. For example, if FL serves a communicative function (White 2014), it could conceivably better serve that function by aligning the

[17] To be sure, domain-initial lengthening occurs, but it is not the mirror image of FL, as it only affects the domain-initial onset (Fougeron and Keating 1997, Fougeron 1998, Byrd 2000, et seq.).

locus of FL with more expandable elements as opposed to intrinsically longer elements, as the two are distinct (Turk and Shattuck-Hufnagel 2007).

6.2 Rhythm

Aside from final lengthening, rhythm has been suggested as an explanation for end-weight, particularly in the context of the syllable-count effect (SCE). Rhythm subsumes at least three principles of metrical optimization, namely, the avoidance of clash (adjacent stressed syllables) and lapse (adjacent unstressed syllables), and possibly also of final stress (on the latter in binomials, see Bolinger 1962, Müller 1997, Benor and Levy 2006, Mollin 2012). For example, a monosyllable-trochee order such as *salt and pepper* might be preferred over its transposition because the latter violates lapse and nonfinality, while the former violates neither (cf. Jespersen 1905). Thus, in some cases, ostensible end-weight might actually reflect eurythmy. Ehret et al. (2014) and Shih (2014) find the English genitive alternation to be sensitive to rhythm, though the former identify it as 'a minor player' compared to weight, 'a very crucial factor' (298).

More generally, PEW is clearly a distinct phenomenon from eurythmy. As established in §3.4, the maximum possible explanatory scope of rhythm is limited compared to the full range of SCE. In English, for example, favored long-final orders are not always metrically optimizing; indeed, they are often worse rhythmically, as with monosyllable-iamb combinations, which evidently still abide by end-weight. In nearly all studies permitting SCE and rhythm to be disentangled, SCE remains significant (§3.4). Moreover, rhythm cannot explain any aspect of SCE in certain other languages, such as French, where it is equally robust. Thus, rhythm cannot motivate the SCE in general. What's more, beyond SCE, rhythm cannot explain any of the six remaining core principles of PEW enumerated in §3, which concern subsyllabic effects. But all of these principles, including SCE, are amenable to a unified explanation in terms of weight. This is not to deny that rhythm plays a role in word order. But it is not at the core of PEW.

6.3 Phonotactics

The situation is similar for phonotactics, which can affect word order, though mostly in ways that are orthogonal to end-weight. For example, the English genitive alternation favors orders that avoid (near-)adjacent sibilants, all else being equal (e.g. *seats of the bus* ≻ *the bus's seats*; cf. Zwicky 1987, Hinrichs and Szmrecsányi 2007, Ehret et al. 2014, Shih et al. 2015). Shih and Zuraw (2016) find a significant avoidance of (near-)adjacent nasals in Tagalog noun-adjective ordering (e.g. *na itím* 'LINK black' ≻ *itím na*). Furthermore, vowel-vowel (HIATUS) sequences are sometimes avoided (Gunkel and Ryan 2011, Shih 2014, 2016).

While instances of longer onsets patterning as heavier in certain languages could in principle be due to the avoidance of hiatus or resyllabification (§3.2), phonotactics is far from a general solution to end-weight. For one thing, it does not cover the same effect in languages like English and German. In fact, phonotactics predicts that increasingly complex onsets should be increasingly avoided after a consonant-final conjunction, the opposite of the observed effect. For example, *bef or bref* is if anything worse phonotactically than *bref or bef*, given the more complex medial cluster in the former. But onset complexity is favored in final position (§3.2).

As a second case attributed to phonotactics, the onset sonority effect, whereby less sonorous onsets are favored finally (§3.2), is claimed by Parker (2002) to be driven by a SYLLABLE CONTACT LAW that favors deep sonority troughs between nuclei. Parker (2002) bases this conclusion on reduplicative compounds such as *roshy-toshy* $\succ$ *toshy-roshy*. But the same effect is found in coordination; for example, type *roshy and toshy* is preferred to type *toshy and roshy* (§3.2). Parker's analysis is refuted by cases in which the relevant onsets are medial in conjuncts, as with *a roshy and a toshy* or *maróshy and matóshy*. Syllable Contact incorrectly predicts the onset sonority effect to vanish in such cases, since its violations are identical in both orders. To corroborate the intuition that the sonority effect nevertheless persists in such cases, five diagnostic binomials were tested on Amazon's Mechanical Turk, with the same design as in §3.2. Wugs were constructed to be most naturally stressed on their peninitial syllables (e.g. *ayárma*), such that the critical onsets tend to occupy stressed syllables. 18 participants qualified. As summarized in Figure 13, they aggregately favor obstruent-final orders 1.8 times as frequently as sonorant-final orders ($\chi^2(1) = 7.5$, $p = .006$). This result cannot be explained by Syllable Contact, but is expected under the weight-based account, given that greater stress falls on the second conjunct and onset obstruency generally correlates with stress (§4). In conclusion, phonotactics cannot explain away end-weight with respect to the onset sonority effect, one of the only aspects of PEW for which a phonotactic explanation has been suggested.

Binomial (in preferred order)		N Agree	N Disagree	% Agree
1.	ayarma and akarma	10	8	56%
2.	aloompt and atoompt	12	6	67%
3.	lemonte and leponte	12	6	67%
4.	mameert and mapeert	14	4	78%
5.	siroof and sicoof	10	8	56%

Fig. 13: The onset sonority effect persists when the critical onset is medial in the conjunct.

6.4 Complexity

As a final logically possible non-weight-based explanation for PEW, consider the hypothesis that the more phonologically complex item (in segments, features, or prosodic nodes) is favored finally (or, equivalently, that the simpler item is favored initially). One could imagine various processing motivations for such a tendency, borrowing from the literature on the role of syntactic complexity in (non-prosodic) end-weight (e.g. Hawkins 1994, Gibson 1998, Wasow 2002, Temperley 2007). First, speakers might tend to defer complexity, postponing elements anticipated to have a high processing cost (cf. Wasow 2002:56 on syntactic weight). Second, speakers might seek to minimize dependency distances between a head and its arguments (cf. Hawkins 1994, 2004, Gibson 1998, and Chang 2009 on syntactic weight). For example, for a ditransitive verb such as *give*, placing the shorter argument first means that the left edge of the second argument is closer to the verb than it would be in the alternative order.[18] Finally, placing the more complex item adjacent to pause (e.g. sentence-finally) might facilitate comprehension, favoring a more uniform processing load over time, given that medial position is more taxing for processing (Pinker and Birdsong 1979:507).

While complexity deferral, whatever its motivations, might play some role in PEW, it is not a viable explanation for most of the core principles of PEW identified in §3, which have weight-like properties that cannot be derived from complexity. First, sonority correlates with weight in the coda, whereas it is not generally the case that more sonorous segments are more complex than less sonorous ones. Second, this generalization is reversed in the onset, which again eludes complexity, but is explicable in terms of weight (§4). Third, weight correlates gradiently with the durations of vowels (§3.1), which is not a function of their complexity. Fourth, vowel effects are generally stronger than consonant effects (§6.1), a well-established principle of weight, but orthogonal to complexity. Fifth, a rime comprising a long or tense vowel patterns as heavier than one comprising a short vowel followed by a consonant (i.e. Vː > V̆C; see §3.3). In weight systems more generally, Vː is virtually always heavier than V̆C if the two are distinguished (Ryan 2016). In this case, the greater number of segments corresponds with less, not more, weight.

Additional problems for the phonological complexity approach apply more specifically to the individual processing explanations just enumerated. For instance, dependency distance minimization predicts that if a binomial exhibits end-weight, it should do so only as a right-side complement, for example, as the object of a verb or preposition. Meanwhile, a binomial in preverbal position is expected to exhibit beginning-weight on this account, and binomials standing alone are predicted to lack polarity. These predictions are erroneous; binomials exhibit end-weight across the board in end-weight languages. Pause-facilitated processing, for its part, does not account for the fact that end-weight occurs even in non-pause-adjacent positions, as discussed in §6.1. Moreover, like dependency minimization, it incorrectly predicts beginning-weight sentence-initially and ambivalence for stand-alone binomials. An account in terms of postponing less accessible constituents would need to address PERSISTENT EU-

[18] To be clear, neither of these processing accounts has been proposed to apply to prosodic complexity. This subsection merely considers some problems that they would face if they were to be applied to it.

PHONY, that is, the fact that binomials are heavy-final not only in spontaneous production, but also when one bears both orders in mind for a ratings task (§3). It also fails to motivate beginning-weight languages, in which heavier items are preferred initially (§5.1).

In conclusion, complexity is not a viable motivation for PEW, mainly because it cannot explain its weight-like properties, in addition to issues of implementation. It therefore joins final lengthening, rhythm, and phonotactics as being a possible mechanism by which phonology can influence word order, but one that is largely orthogonal to PEW per se. The phrasal stress analysis developed in the next section, by contrast, provides a unified account of the weight-like properties of end-weight without any of the shortcomings of the four approaches outlined in this section.

7 Conclusion and further issues

Prosodically heavier constituents are favored finally in variable-order constructions in end-weight languages, even when one controls for other factors such as semantics, accessibility, rhythm, and syntactic complexity. Seven properties are associated with PEW, namely, longer vowels, lower vowels, longer codas, more sonorant codas, longer onsets, less sonorant onsets, and more syllables. All of these properties, including the reversal in the treatment of sonority between onsets and codas, are consistent with phonological weight more generally across weight-sensitive systems and languages.

Building on previous work, this article proposes that prosodic weight is attracted to final position in end-weight constructions because that is also the locus of greatest stress. PEW therefore constitutes a phrase-level instantiation of the same stress-weight interface that has been documented previously for syllable weight in word stress and meter. Other possible explanations for PEW, including final lengthening, complexity deferral, and rhythmic or phonotactic optimization, are shown not to motivate most of its core properties. The proposed account is further argued to accord with the treatment of prosodic weight in compounds and in verb-final languages, though these areas require further empirical investigation, as do focus and Rhythm Rule manipulations of end-weight constructions (§5.3).[19]

This article has consistently distinguished between prosodic and syntactic end-weight. On the one hand, it is clear that PEW exists independently of syntactic end-weight, as can be seen when morphosyntactic structure is held constant and phonology is manipulated. But this holds in the opposite direction as well: If one controls for phonology, syntactic end-weight is evident in languages like English. Thus, one might wonder whether there is a common cause of both forms of end-weight. As argued in §6.4, a complexity-based account is not viable for PEW. Rather, the common cause appears

[19] See also Blumenfeld (2016), who addresses short-final constructions in verse in languages that otherwise favor end-weight.

to be related to basic word order. Syntactic end-weight tends to be associated with head-initial constructions (such as VO verb phrases) and syntactic beginning-weight with head-final (e.g. OV) constructions (Hawkins 1994, et seq.). At the same time, VO languages tend to have right-oriented nuclear stress, while OV languages often have head-initial phrasal stress, at least in p-phrases. Thus, it is no coincidence that a language like English has both syntactic and prosodic end-weight, as both can be derived from its basic SVO word order and concomitant basic stress pattern. Japanese is a good test case for the present proposal, since it exhibits syntactic beginning-weight, but lacks phrasal stress, and is therefore expected to exhibit neither prosodic beginning- nor end-weight. Indeed, based on Lohmann and Takada (2014), it appears that Japanese comes close to being polarity-neutral. They find no evidence for a syllable-count effect in either direction, and only a borderline effect of mora count, but in the end-weight direction.

Finally, this article has focused on binomials. As mentioned at the outset, numerous other constructions exhibit end-weight effects. For higher-level end-weight phenomena such as heavy NP shift and the dative alternation, the contribution of phonology is evidently more modest. For example, Anttila et al. (2010), in their corpus study of the English dative alternation, find a strong effect of p-word count as distinct from syntactic word count, but little to no effect of syllable count. Similarly, Shih and Grafmiller (2013) compare syntactic vs. phonological operationalizations of weight for the English dative and genitive alternations, concluding that syntactic factors are dominant in ‘higher-level constituent ordering.’ Insofar as prosodic weight affects such constructions, they find, it is mostly in terms of p-word count; syllable count is marginal (see also Ingason and MacKenzie 2011, 2012).

Thus, a kind of depth-of-field effect appears to be in evidence, such that the available grain of detail scales with the complexity of the constituents involved. For simple constituents, such as monosyllables, segmental properties are readily available (§3). For more complex constituents, such features are more likely to be ignored. Some form of depth of field is predicted by the constraint-based theory advocated in §4 if it is instantiated in a weighted-constraints framework such as maxent HG. In such a setting, individual segments are never wholly ignored, but as their number increases the relative impact of each dwindles to a point where it might be difficult to discern a measurable effect. Depth of field might also arise from processing. For increasingly complex phrases, it becomes increasingly infeasible for the speaker to plan every phonological detail in advance. Indeed, at a certain level of complexity, the speaker might not even completely plan both phrases at the point at which their order is determined (Wasow 2002:45–6). A more detailed description and implementation of depth of field must be left for future work, in part because its empirical properties are not well established for prosody. For one, processing complexity predicts that depth of field should be complexity-dependent, not construction-dependent. For example, segmental detail may become available for dative alternants consisting of monosyllables, and it may become unavailable for binomials with complex conjuncts. The task may also be relevant, in that, for instance, ratings might diverge from production.

References

Abeillé, Anne, and Danielle Godard. 1999. La position de l'adjectif épithète en français: le poids des mots. *Recherches Linguistiques de Vincennes* 28: 9–32.

Abercrombie, David. 1967. *Elements of general phonetics.* Edinburgh: Edinburgh University Press.

Abraham, Richard D. 1950. Fixed order of coordinates: A study in comparative lexicography. *The Modern Language Journal* 34 (4): 276–287.

Alderete, John, Jill Beckman, Laura Benua, Amalia Gnanadesikan, John McCarthy, and Suzanne Urbanczyk. 1999. Reduplication with fixed segmentism. *Linguistic Inquiry* 30 (3): 327–364.

Allan, Keith. 1987. Hierarchies and the choice of left conjuncts (with particular attention to English). *Journal of Linguistics* 23: 51–77.

Anttila, Arto. 1997. Deriving variation from grammar. In *Variation, change and phonological theory*, eds. Frans Hinskens, Roeland van Hout, and Leo Wetzels, 35–68. Amsterdam: John Benjamins.

Anttila, Arto, Matthew Adams, and Michael Speriosu. 2010. The role of prosody in the English dative alternation. *Language and Cognitive Processes* 25 (7–9): 946–981.

Arregi, Karlos. 2002. Focus on Basque movements. PhD diss, MIT.

Asher, Ronald E., and Elayaperumal Annamalai. 2002. *Colloquial Tamil: The complete course for beginners.* London: Routledge.

Baddeley, Alan D., Neil Thomson, and Mary Buchanan. 1975. Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior* 14 (6): 575–589.

Behaghel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25: 110–42.

Bennett, Ryan, Emily Elfner, and James McCloskey. 2016. Lightest to the right: An anomolous displacement in Irish. *Linguistic Inquiry* 47: 169–234.

Benor, Sarah, and Roger Levy. 2006. The chicken or the egg? A probabilistic analysis of English binomials. *Language* 82 (2): 233–278.

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14: 150–177.

Biese, Yrjo Mooses. 1939. Neuenglisch tick-tack und Verwandtes. *Neuphilologische Mitteilungen* 40: 146–205.

Blumenfeld, Lev. 2016. End-weight effects in verse and language. *Studia Metrica et Poetica* 3 (1): 7–32.

Bock, J. Kathryn. 1977. The effect of a pragmatic presupposition on syntactic structure in question answering. *Journal of Verbal Learning and Verbal Behavior* 16: 723–734.

Bock, J. Kathryn. 1982. Toward a cognitive psychology of syntax: Information processing contributions to sentence formation. *Psychological Review* 89: 1–47.

Bock, Kathryn, and Richard Warren. 1985. Conceptual accessibility and syntactic structure in sentence formulation. *Cognition* 21: 47–67.

Boersma, Paul, and Joe Pater. To appear. Convergence properties of a gradual learning algorithm for Harmonic Grammar. In *Harmonic grammar and harmonic serialism*, eds. John McCarthy and Joe Pater. London: Equinox Press.

Bolinger, Dwight L. 1962. Binomials and pitch accent. *Lingua* 11: 34–44.

Bresnan, Joan, and Tatiana Nikitina. 2003. On the gradience of the dative alternation. MS., Stanford University.

Browman, Catherine P., and Louis Goldstein. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica* 45: 140–155.

Büring, Daniel. 2013. Syntax, information structure, and prosody. In *The Cambridge handbook of generative syntax*, ed. Marcel den Dikken, 860–896. Cambridge: Cambridge University Press.

Büring, Daniel, and Rodrigo Gutiérrez-Bravo. 2001. Focus-related constituent order variation without the NSR: A prosody-based crosslinguistic analysis. *Syntax at Santa Cruz* 3: 41–58.

Byrd, Dani. 2000. Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica* 57: 3–16.

Byrne, Brian, and Elizabeth Davidson. 1985. On putting the horse before the cart: Exploring conceptual bases of word order via acquisition of a miniature artificial language. *Journal of Memory and Language* 24 (4): 377–389.

Campbell, Mary Ann, and Lloyd Anderson. 1976. Hocus pocus nursery rhymes. In *Papers from the 12th Regional Meeting of the Chicago Linguistic Society*, eds. Salikoko S. Mufwene, Carol A. Walker, and Sanford B. Steever, 72–95. Chicago: Chicago Linguistic Society.

Carpenter, Angela C. 2010. A naturalness bias in learning stress. *Phonology* 27: 345–392.

Chang, Franklin. 2009. Learning to order words: A connectionist model of heavy NP shift and accessibility effects in Japanese and English. *Journal of Memory and Language* 61 (3): 374–397.

Choi, Hye-Won. 2007. Length and order: A corpus study of Korean dative-accusative construction. *Discourse and Cognition* 14: 207–227.

Chomsky, Noam, and Morris Halle. 1968. *The sound pattern of English.* Cambridge, Massachusetts: Massachusetts Institute of Technology Press.

Cinque, Guglielmo. 1993. A null theory of phrase and compound stress. *Linguistic Inquiry* 24: 239–297.

Clemens, Lauren, and Jessica Coon. 2016. Deriving verb-initial word order in Mayan. MS., University at Albany, SUNY and McGill University, submitted.

Cooper, William E., and John R. Ross. 1975. World order. In *Papers from the parasession on functionalism*, eds. R. Grossman, L. J. San, and T. Vance, 63–111. Chicago: Chicago Linguistic Society.

Cowan, Nelson, Alan D. Baddeley, Emily M. Elliott, and Jennifer Norris. 2003. List composition and the word length effect in immediate recall: A comparison of localist and globalist assumptions. *Psychonomic Bulletin & Review* 10 (1): 74–79.

Crowhurst, Megan J., and Lev D. Michael. 2005. Iterative footing and prominence-driven stress in Nanti (Kampa). *Language* 81 (1): 47–95.

Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis, and Ingrid Norrmann. 2011. Explaining sonority projection effects. *Phonology* 28: 197–234.

Dayley, Jon Philip. 1989. *Tümpisa (Panamint) Shoshone grammar*. Vol. 115 of *University of california publications in linguistics*. Berkeley, CA: University of California Press.

de Lacy, Paul. 2002. The formal expression of markedness. PhD diss, University of Massachusetts-Amherst.

de Lacy, Paul. 2004. Markedness conflation in Optimality Theory. *Phonology* 21 (2): 145–199.

Delattre, P. 1966. A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics* 4: 183–198.

Dienhart, John M. 1999. Stress in reduplicative compounds: Mish-mash or hocus-pocus? *American Speech* 74: 13–38.

Dryer, Matthew S. 1980. The positional tendencies of sentential noun phrases in universal grammar. *Canadian Journal of Linguistics* 25: 123–195.

Ehret, Katharina, Christoph Wolk, and Benedikt Szmrecsányi. 2014. Quirky quadratures: On rhythm and weight as constraints on genitive variation in an unconventional data set. *English Language and Linguistics* 18 (2): 263–303.

Elfner, Emily. 2012. Syntax-prosody interactions in Irish. PhD diss, University of Massachusetts, Amherst.

Elfner, Emily. 2015. Recursion in prosodic phrasing: Evidence from Connemara Irish. *Natural Language and Linguistic Theory* 33: 1169–1208.

Everett, Daniel. 1988. On metrical constituent structure in Pirahã. *Natural Language and Linguistic Theory* 6: 207–246.

Everett, Daniel, and Keren Everett. 1984. On the relevance of syllable onsets to stress placement. *Linguistic Inquiry* 15: 705–711.

Faghiri, Pegah, and Pollet Samvelian. 2014. Constituent ordering in Persian and the weight factor. In *Empirical issues in syntax and semantics*, ed. Christopher Piñón, Vol. 10, 215–232. Paris: CNRS.

Fenk-Oczlon, Gertraud. 1989. Word frequency and word order in freezes. *Linguistics* 27 (3): 517–556.

Ferreira, V. S., and H. Yoshita. 2003. Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research* 32 (6): 669–692.

Forsgren, Mats. 1978. *La place de l'adjectif épithète en français contemporain, étude quantitative et sémantique*. Stockholm: Almqvsit & Wilksell.

Fougeron, Cécile. 1998. *Variations articulaires en début de constituants prosodiques de différents niveaux en Français*. Paris: Université Paris III-

Sorbonne Nouvelle.

Fougeron, Cécile, and Patricia Keating. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101 (6): 3728–3740.

Gabas Jr., Nilson. 1999. A grammar of Karo (Tupi). PhD diss, University of California, Santa Barbara.

Garcia, Guilherme D. 2017a. Repairing the lexicon: A probabilistic approach to weight and stress. MS., McGill University, submitted.

Garcia, Guilherme D. 2017b. Weight effects on stress: Lexicon and grammar. PhD diss, McGill University.

Gibson, Edward. 1998. Linguistic complexity: Locality and syntactic dependencies. *Cognition* 68: 1–76.

Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In *Image, language, brain*, eds. Y. Miyashita, A. Marantz, and W. O'Neil, 95–126. Cambridge, Massachusetts: Massachusetts Institute of Technology Press.

Gibson, Edward, Steve Piantadosi, and Kristina Fedorenko. 2011. Using Mechanical Turk to obtain and analyze English acceptability judgments. *Language and Linguistics Compass* 5 (8): 509–524.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a Maximum Entropy model. In *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, eds. Jennifer Spenader, Anders Eriksson, and Osten Dahl, 111–120.

Golenbock, Janice. 2000. Binomial expressions — does frequency matter? MS., Carnegie Mellon University.

Gordon, Matthew. 2002. A phonetically-driven account of syllable weight. *Language* 78: 51–80.

Gordon, Matthew. 2005. A perceptually-driven account of onset-sensitive stress. *Natural Language and Linguistic Theory* 23: 595–653.

Gordon, Matthew. 2006. *Syllable weight: phonetics, phonology, typology.* New York, NY: Routledge Press.

Grafmiller, Jason, and Stephanie Shih. 2011. New approaches to end weight. Paper presented at Variation and Typology: New Trends in Syntactic Research, 25–7 August 2011, Helsinki.

Griffin, Zenzi M., and J. Kathryn Bock. 1998. Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language* 38 (3): 313–338.

Gunkel, Dieter, and Kevin Ryan. 2011. Hiatus avoidance and metrification in the Rigveda. In *Proceedings of the 22nd Annual UCLA Indo-European Conference*, eds. Stephanie W. Jamison, H. Craig Melchert, and Brent Vine, 53–68. Bremen: Hempen.

Gustafsson, Marita. 1974. The phonetic length of the members in Present-Day English binomials. *Neuphilologische Mitteilungen* 75: 663–677.

Gustafsson, Marita. 1975. *Binomial expressions in present-day English.* Turku: Annales Universitatis Turkunsis.

Hawkins, John A. 1990. A parsing theory of word order universals. *Linguistic*

Inquiry 21: 223–261.
Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge, U.K.: Cambridge University Press.
Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
Hayes, Bruce. 1984. The phonology of rhythm in English. *Linguistic Inquiry* 15 (1): 33–74.
Hayes, Bruce. 1995. *Metrical stress theory: Principles and case studies*. Chicago, IL: University of Chicago Press.
Hayes, Bruce, and Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language and Linguistic Theory* 9 (1): 47–96.
Hayes, Bruce, and Colin Wilson. 2008. A Maximum Entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39: 379–440.
Hetzron, Robert. 1972. Phonology in syntax. *Journal of Linguistics* 8: 251–262.
Hinrichs, Lars, and Benedikt M. Szmrecsányi. 2007. Recent changes in the function and frequency of Standard English genitive constructions: a multivariate analysis of tagged corpora. *English Language and Linguistics* 11 (3): 437–474.
Ingason, Anton Karl, and Laurel MacKenzie. 2011. 'Heaviness' as evidence for a derive-and-compare grammar. Poster presented ate the 19th Manchester Phonology Meeting.
Jaeger, T. Florian, and Elisabeth J. Norcliffe. 2009. The cross-linguistic study of sentence production. *Language and Linguistics Compass* 3: 866–887.
Jakobson, Roman. 1960. *Linguistics and poetics*. Cambridge, Massachusetts: Massachusetts Institute of Technology Press.
Jesney, Karen. 2007. The locus of variation in weighted constraint grammars. Poster presented at the Workshop on Variation, Gradience, and Frequency in Phonology, Stanford, CA.
Jespersen, Otto. 1905. *Growth and structure of the English language*. Stuttgart: Teubner Verlag.
Jespersen, Otto. 1961. *A Modern English grammar on historical principles. Part VI, Morphology*. London, U.K.: George Allen and Unwin Ltd..
Kabak, Baris, and Irene Vogel. 2001. The phonological word and stress assignment in Turkish. *Phonology* 18 (3): 315–360.
Kahnemuyipour, Arsalan. 2003. Syntactic categories and Persian stress. *Natural Language and Linguistic Theory* 21: 333–379.
Karimi, Simin. 2003. On object positions, specificity, and scrambling in Persian. In *Word order and scrambling*, ed. Simin Karimi, 91–124. Oxford: Blackwell Publishing.
Katkov, Mikhail, Sandro Romani, and Misha Tsodyks. 2014. Word length effect in free recall of randomly assembled word lists. *Frontiers in Computational Neuroscience* 8: 129.
Katz, Jonah. 2010. Compression effects, perceptual asymmetries, and the grammar of timing. PhD diss, Massachusetts Institute of Technology.
Keane, Elinor. 2001. Echo words in Tamil. PhD diss, University of Oxford.

Keane, Elinor. 2003. Word-level prominence in Tamil. In *Proceedings of the 15th International Congress of the Phonetic Sciences*, Vol. 2, 1257–1260.

Keane, Elinor. 2006. Prominence in Tamil. *Journal of the International Phonetic Association* 36: 1–20.

Kelly, Michael H. 1986. On the selection of linguistic options. PhD diss, Cornell University.

Kelly, Michael H., and J. Kathryn Bock. 1988. Stress in time. *Journal of Experimental Psychology: Human Perception and Performance* 14 (3): 389–403.

Kelly, Michael H., J. Kathryn Bock, and Frank C. Keil. 1986. Prototypicality in a linguistic context: effects on sentence structure. *Journal of Memory and Language* 25 (1): 59–74.

Kenstowicz, Michael. 1994. Sonority-driven stress. Rutgers Optimality Archive ROA–33.

Kiparsky, Paul. 1968. Metrics and morphophonemics in the Kalevala. In *Studies presented to Roman Jakobson by his students*, ed. Charles Gribble. Cambridge, MA: Slavica.

Kiparsky, Paul. 2009. Verbal co-compounds and subcompounds in Greek. *MIT Working Papers in Linguistics* 57.

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.

Lehiste, Ilse. 1972. The timing of utterances and linguistic boundaries. *The Journal of the Acoustical Society of America* 51 (6): 2018–2024.

Levelt, Willem, and Ben Maassen. 1981. Lexical search and order of mention in sentence production. In *Crossing the boundaries in linguistics*, eds. W. Klein and W. Levelt, 221–252. Dordrecht: Reidel.

Liberman, Mark. 1975. The intonational system of English. PhD diss, Massachusetts Institute of Technology.

Liberman, Mark, and Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8: 249–336.

Lindblom, Björn. 1968. Temporal organization of syllable production. In *Speech transmission laboratory quarterly progress*, Vol. 2–3, 1–6. Stockholm, Sweden: Royal Institute of Technology.

Lohmann, Arne, and Tayo Takada. 2014. Order in NP conjuncts in spoken English and Japanese. *Lingua* 152: 48–64.

Lunden, Anya. 2006. Weight, final lengthening and stress: A phonetic and phonological case study of Norwegian. PhD diss, University of California, Santa Cruz.

MacDonald, Jonathan E. 2015. A movement analysis of some double object constructions. In *Proceedings of the 32nd West Coast Conference on Formal Linguistics*, eds. Ulrike Steindl, Thomas Borer, Huilin Fang, Alfredo García Pardo, Peter Guekguezian, Brian Hsu, Charlie O'Hara, and Iris Chuoying Ouyang, 276–285. Somerville, MA: Cascadilla Proceedings Project.

MacKenzie, Laurel. 2012. Location variation above the phonology. PhD diss, University of Pennsylvania.

Maddieson, Ian. 1985. Phonetic cues to syllabification. In *Phonetic linguistic essay in honor of Peter Ladefoged*, ed. Victoria Fromkin, 203–221. New

York: Academic Press.

Malkiel, Yakov. 1959. Studies in irreversible binomials. *Lingua* 8: 113–160.

Marchand, Hans. 1952. Alliteration, Ablaut und Reim in den Türkischen Zwillingsformen. *Oriens* 5: 60–69.

Marchand, Hans. 1969. *The categories and types of Present-Day English word-formation*. Wiesbaden, Germany: Harrossowitz.

Matthews, Stephen J., and L. Y. Y. Yeung. 2000. Processing motivations for topicalization in Cantonese. In *Cognitive-functional linguistics in an East Asian context*, eds. Kaoru Horie and Shigeru Sato, 81–102. Tokyo: Kurosio Publishers.

McDonald, Janet L., Kathryn Bock, and Michael H. Kelly. 1993. Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology* 25: 188–230.

Minkova, Donka. 2002. Ablaut reduplication in English: The criss-crossing of prosody and verbal art. *English Language and Linguistics* 6: 133–169.

Mollin, Sandra. 2012. Revisiting binomial order in English: Ordering constraints and reversibility. *English Language and Linguistics* 16 (1): 81–103.

Mollin, Sandra. 2013. Pathways of change in the diachronic development of binomial reversibility in Late Modern American English. *Journal of English Linguistics* 41 (2): 168–203.

Morgan, Emily, and Roger Levy. 2016. Abstract knowledge versus direct experience in processing of binomial expressions. *Cognition* 157: 384–402.

Mortensen, David. 2006. Logical and substantive scales in phonology. PhD diss, University of California, Berkeley.

Müller, Gereon. 1997. Beschrankungen für Binomialbildungen im Deutschen. *Zeitschrift für Sprachwissenschaft* 16 (1/2): 5–51.

Munn, Alan. 1993. Topics in the syntax and semantics of coordinate structures. PhD diss, University of Maryland.

Nevins, Andrew, and Keith Plaster. 2008. Review of Paul de Lacy, Markedness: reduction and preservation in phonology. *Journal of Linguistics* 44: 770–81.

Oakeshott-Taylor, John. 1984. Phonetic factors in word order. *Phonetica* 41: 226–237.

Oden, Gregg C., and Lola L. Lopes. 1981. Preference for order in freezes. *Linguistic Inquiry* 12: 673–679.

Parker, Steve. 2002. Quantifying the sonority hierarchy. PhD diss, University of Massachusetts, Amherst.

Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33: 999–1035.

Pinker, Steven, and David Birdsong. 1979. Speakers' sensitivity to rules of frozen word order. *Journal of Verbal Learning and Verbal Behavior* 18 (4): 497–508.

Pordany, Laszlo. 1986. A comparison of some English and Hungarian freezes. *Papers and Studies in Contrastive Linguistics* 21: 117–125.

Prince, Alan. 1983. Relating to the grid. *Linguistic Inquiry* 14: 19–100.

Prince, Alan. 1999. Paninian relations. Handout, University of Marburg.

Prince, Alan, and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in Generative Grammar*. Malden, MA: Blackwell. Technical Report, Rutgers University and University of Colorado at Boulder, 1993. Revised version Blackwell, 2004..

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1972. *A grammar of Contemporary English*. London: Longman.

Ross, John Robert. 1982. The sound of meaning. In *Linguistics in the morning calm*, ed. The Linguistic Society of Korea, 275–290. Seoul: Hanshin Publishing Co..

Ryan, Kevin M. 2011. Gradient weight in phonology. PhD diss, University of California, Los Angeles.

Ryan, Kevin M. 2014. Onsets contribute to syllable weight: Statistical evidence from stress and meter. *Language* 90 (2): 309–341.

Ryan, Kevin M. 2016. Phonological weight. *Language and Linguistics Compass* 10: 720–733.

Samek-Lodovici, Vieri. 2005. Prosody-syntax interaction in the expression of focus. *Natural Language and Linguistic Theory* 23: 687–755.

Scontras, Greg, Judith Degen, and Noah D. Goodman. 2017. Subjectivity predicts adjective ordering preferences. *Open Mind: Discoveries in Cognitive Science* 1 (1): 53–65.

Selkirk, Elisabeth O. 1995. Sentence prosody: Intonation, stress, and phrasing. In *The handbook of phonological theory*, ed. John A. Goldsmith, 550–569. Oxford: Blackwell Publishing.

Selkirk, Elisabeth O. 2011. The syntax-phonology interface. *The Handbook of Phonological Theory*.

Shih, Stephanie. 2014. Towards optimal rhythm. PhD diss, Stanford University.

Shih, Stephanie, and Jason Grafmiller. 2013. Weighing in on end weight. Paper presented at the LSA 85th Annual Meeting, 6–9 January 2011, Pittsburgh, PA.

Shih, Stephanie, and Kie Zuraw. 2016. Phonological conditions on variable adjective-noun word order in Tagalog. MS., submitted, UC Merced and UCLA, available at `ling.auf.net/lingbuzz/002796`.

Shih, Stephanie S. 2016. Phonological influences in syntactic alternations. In *The morphosyntax-phonology connection: Locality and directionality at the interfaces*, eds. Vera Gribanova and Stephanie S. Shih, 223–254. Oxford: Oxford University Press.

Shih, Stephanie, Jason Grafmiller, Richard Futrell, and Joan Bresnan. 2015. Rhythm's role in genitive construction choice in spoken English. In *Rhythm in cognition and grammar*, eds. Ralf Vogel and Ruben van de Vijver, 207–34. Berlin: De Gruyter Mouton.

Smith, Jennifer L. 2002. Phonological augmentation in prominent positions. PhD diss, University of Massachusetts, Amherst.

Sprouse, Jon. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* 43 (1): 155–167.

Swadesh, Morris. 1962. Archaic doublets in Altaic. In *American studies in Altaic linguistics*, ed. Nicholas Poppe. Vol. 13 of *Uralic and Altaic series*, 293–330. Bloomington, Indiana: Indiana University Publications.

Szendroi, Kriszta. 2001. Focus and the syntax-phonology interface. PhD diss, University College London.

Szmrecsányi, Benedikt M. 2004. On operationalizing syntactic complexity. *Journées internationales d'Analyse statistique des Données Textuelles (JADT-04)* 2: 1032–1039.

Temperley, David. 2007. Minimization of dependency length in written English. *Cognition* 105: 300–333.

Thuilier, Juliette. 2012. Contraintes préférentielles et ordre des mots en français. PhD diss, Université Paris Diderot.

Thun, Nils. 1963. *Reduplicative words in English.* Lund, Sweden: Carl Bloms Boktryckeri.

Topintzi, Nina. 2010. *Onsets: suprasegmental and prosodic behaviour.* Cambridge, U.K.: Cambridge University Press.

Turk, Alice E., and Stefanie Shattuck-Hufnagel. 2000. Word-boundary-related duration patterns in English. *Journal of Phonetics* 28: 397–440.

Turk, Alice E., and Stefanie Shattuck-Hufnagel. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics* 35 (4): 445–472.

Vasu, Srisa Chandra. 1898. *The Ashtádhyáyí of Páṇini, translated into English by Śríśa Chandra Vasu*, Vol. 8. Allahabad: Indian Press.

Venditti, Jennifer J., Kikuo Maekawa, and Mary E. Beckman. 2008. Prominence marking in the Japanese intonation system. In *Handbook of Japanese linguistics*, eds. Shigeru Miyagawa and Mamoru Saito, 456–512. Oxford: Oxford University Press.

Vogel, Ralf. 2006. Weak function word shift. *Linguistics* 44: 1059–1093.

Wackernagel, Jakob. 1905. *Altindische Grammatik*, Vol. II, 1. Göttingen: Vandenhoeck & Ruprecht.

Wagner, Michael. 2005. Prosody and recursion. PhD diss, Massachusetts Institute of Technology.

Walker, Douglas C. 1975. Word stress in French. *Language* 51 (4): 887–900.

Wasow, Thomas. 2002. *Postverbal behavior.* Stanford, California: CSLI Publications.

Wasow, Thomas, and Jennifer Arnold. 2003. Post-verbal constituent ordering in English. In *Determinants of grammatical variation in English*, eds. G. Rohdenburg and B. Mondorf, 119–154. Mouton.

Wescott, Roger W. 1970. Types of vowel alternations in English. *Word* 26: 309–343.

West, M. L. 1987. *Introduction to Greek metre.* Oxford University Press.

Westbury, John, and Patricia Keating. 1980. Central representation of vowel duration. *Journal of the Acoustical Society of America* 67 (Suppl. 1): 37.

Wheatley, Henry B. 1866. *A dictionary of reduplicated words in the English language*, Vol. 17. Transactions of the Philological Society.

White, Laurence. 2014. Communicative function and prosodic form in speech timing. *Speech Communication* 63–64: 38–54.

Wightman, C. W., Stefanie Shattuck-Hufnagel, M. Ostendorf, and P. J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 92: 1707–1717.

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

Wolf, Matthew. 2008. Optimal interleaving: Serial phonology-morphology interaction in a constraint-based model. PhD diss, University of Massachusetts-Amherst.

Wright, Saundra, and Jennifer Hay. 2002. Fred and Wilma: A phonological conspiracy. In *Gendered practices in language*, eds. Sarah Benor, Mary Rose, Devyani Sharma, Julie Sweetland, and Qing Zhang, 175–191. Stanford, California: CSLI Publications.

Wright, Saundra, Jennifer Hay, and Tessa Bent. 2005. Ladies first? Phonology, frequency, and the naming conspiracy. *Linguistics* 44: 531–561.

Yamashita, Hiroko, and Franklin Chang. 2001. "Long before short" preference in the production of a head-final language. *Cognition* 81 (2): 45–55.

Yamashita, Hiroko, and Franklin Chang. 2006. Sentence production in Japanese. In *Handbook of East Asian psycholinguistics*, eds. Mineharu Nakayama, Reiko Mazuka, and Yasuhiro Shirai, Vol. 2, 291–297. Cambridge: Cambridge University Press.

Yu, Alan C. L. , and Hyunjung Lee. 2014. The stability of perceptual compensation for coarticulation within and across individuals: A cross-validation study. *The Journal of the Acoustical Society of America* 136 (1): 382–388.

Zec, Draga. 1988. Sonority constraints on prosodic structure. PhD diss, Stanford University.

Zec, Draga. 1995. Sonority constraints on syllable structure. *Phonology* 12 (1): 85–129.

Zec, Draga. 2003. Prosodic weight. In *The syllable in Optimality Theory*, eds. Caroline Féry and R. van de Vijer. Cambridge, U.K.: Cambridge University Press.

Zipf, George Kingsley. 1936. *The psycho-biology of language: An introduction to dynamic philology.* Cambridge, MA: MIT Press.

Zubizarreta, Maria-Luisa. 1998. *Prosody, focus and word order.* Cambridge, MA: MIT Press.

Zubizarreta, Maria-Luisa, and Jean-Roger Vergnaud. 2000. Phrasal stress and syntax. MS., University of Southern California.

Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28 (2): 417–472.

Zwicky, Arnold M. 1987. Suppressing the zs. *Journal of Linguistics* 23 (1): 133–148.