

## RESEARCH

## How allomorphic is English article allomorphy?

Marjorie Pak

Emory University, 532 Kilgo Circle Ste. 202C, Atlanta, GA 30322, US  
mgpak@emory.edu

The question of how to distinguish allomorphy from phonology plays a central role in morphological theory. While English *a/an* has been cited as a hallmark case of “phrasal” allomorphy, the parallel alternation between /ði/ and /ðə/ in the English definite article (/ðə/ *book*, /ði/ *apple*) has not received as much attention. This paper provides a formal analysis of English *a/an* and /ði/~ðə/. I argue that: (i) /ði/~ðə/ is *not* allomorphic, but is derived by the same phonological rules (tensing and vowel-reduction) as other V~ə alternations in English; and (ii) *a/an* involves a two-tiered derivation: first allomorphy establishes a basic split between V and Vn; then the same phonological rules involved in *the* (tensing and vowel-reduction) derive the four surface variants /ej/, /æɪn/, /ə/, and /ən/. In other words, in addition to its basic n~∅ alternation, *a/an* also participates in the same V~ə alternation as other strong-weak function-word pairs in English (/ði/~ðə/ in *the*, /kæn/~kən/ in *can*, etc.). This two-tiered serialist approach is incompatible with analyses of *a/an* as surface phonological optimization, or TETU (Mascaró 1996b). Accordingly, I provide evidence from emphatic glottal stops, /h/-deletion and pause-fillers showing that despite initial appearances, neither *a/an* nor /ði/~ðə/ is driven by surface syllable well-formedness constraints.

**Keywords:** allomorphy; phonological optimization; serialist vs. OT architectures

## 1 *A/an, the, allomorphy and phonology*

The English *a/an* alternation can be described as follows: *an* is used if the immediately following word starts with a vowel; otherwise *a* is used.

- (1) a. **an** apple, **an** interesting book
- b. **a** book, **a** very red apple

Despite the apparent simplicity of this generalization, the *a/an* alternation has been notoriously difficult for linguists to understand, and has often been featured as a problematic case study in discussions of the phonology-morphology interface (see e.g. Rotenberg 1978; Kaisse 1985; Zwicky 1986; Hayes 1990; Spencer 1991; Mascaró 1996b; Joseph 1997; Asudeh & Klein 2002; Yang 2004; Lee 2009; Nevins 2011). *A/an* presents a paradox: it is restricted to a single morpheme, which suggests that it is a morphological phenomenon (viz. allomorphy), yet it depends crucially on information about the *following* word, and thus cannot be characterized as a strictly word-internal process (Spencer 1991: 127–129). While it is possible to derive *a/an* phonologically, by either /n/-insertion or /n/-elision (Hurford 1972; 1974; Perlmutter 1970; Venneman 1974), the obvious weakness of such an approach is that it requires postulating a special phonological rule that applies to only one morpheme.<sup>1</sup> More recent analyses have therefore treated *a/an* allomorphically, and

<sup>1</sup> The *n~∅* alternations in possessive articles *my/mine* and *thy/thyn* (e.g. *mine eyes~my child*) are now obsolete (Crisma 2009: 137–141; Gramley 2012: ch4). The *n~∅* alternation in the Greek-derived prefix *a(n)-*, as in *a-typical/an-aerobic*, can be assumed not to be productive until adulthood, if then.

have taken care of the non-word-boundedness paradox either by having *a(n)* cliticize to the following word (as proposed here; see §3), or by admitting some kind of “phrasal” or “external” allomorphy into the grammar (e.g. Hayes 1990; Mascaró 1996b; Asudeh & Klein 2002).

Here I look at *a/an* alongside a strikingly similar but far less studied phenomenon from English: the alternation between /ði(j)/ and /ðə/ in the definite article (henceforth *the*). The distribution of the alternants is almost identical to that found with *a/an* (albeit somewhat less regular):<sup>2</sup> use /ði(j)/ if the immediately following word starts with a vowel; otherwise use /ðə/ (Ladefoged 1975: 91–92).<sup>3</sup>

- (2) a. /ði/ apple, /ði/ interesting book  
b. /ðə/ book, /ðə/ very red apple

Like *a/an*, *the* cannot be derived by an across-the-board phonological rule; word-final prevocalic /i/ alternates with preconsonantal /ə/ only in *the*, not in *lovely*, *happy*, *agenda*, etc.:

- (3) a. /ði/ apple ~ /ðə/ book  
b. lovel/i/ apple ~ \*lovel/ə/ book  
part/i/ animal ~ \*part/ə/ time  
\*agend/i/ item ~ agend/ə/ change

Should *the* then be analyzed allomorphically on par with *a/an*? There has not to my knowledge been a formal analysis that addresses this question head-on. Given their structural and distributional similarities, a unified treatment of *a/an* and *the* seems desirable; correspondingly, some previous studies simply adopt the terms *allomorphy* and *allomorphs* to refer to *the* (e.g. Jurafsky et al. 1998: 3; Newton & Wells 1999: 74; Britain & Fox 2009: 180ff). A unified allomorphic treatment might be especially appealing in the context of a theory of *allomorphy as the emergence of the unmarked* (TETU) (Mascaró 1996a; b; 2007). Both *a/an* and *the* appear to be hiatus-avoiding alternations (4), and this kind of phonological optimization can be directly explained by theories where allomorph choice is determined by surface constraints like ONSET and NO-CODA (see §4). Analyzing *the* in this way would imply, of course, that *the* was also an allomorphic alternation.

- (4) a. a egg / (\*VV) vs. an egg (V.CV)  
b. th/ə/ egg / (\*VV) vs. th/ij/ egg (V.CV)

Nevertheless, this paper argues that *the* is phonological, not allomorphic, in nature – specifically, the /ðə/ variant of the definite article is derived by phonological *vowel reduction*. Vowel reduction is also responsible for weak/strong alternations in other English function words, e.g. *for*, *have*, *can*; to explain why this otherwise word-internal rule appears to apply across a word boundary in e.g. *the dog*, I propose that English articles *cliticize* to the following word. I furthermore argue that article cliticization is what enables allomorphy in *a/an* (as suggested in passing in Spencer 1991: 128). In the serialist architecture adopted here, *a/an* is derived in three steps:

<sup>2</sup> In my speech, for example, prevocalic /ðə/ (*th/ə/ orange*) is acceptable while prevocalic *a* (*a orange*) is not (see also Todaka 1992; Keating et al. 1994: 136). However, many varieties of English have variable *a/an* as well (see Gabrielatos et al. 2010 and references cited there). The analysis that I present in §3 assumes categorical *a/an* and *the*, for simplicity of exposition, but in §5 I show how the analysis is compatible with a competing-grammars account of inter- and intraspeaker variation.

<sup>3</sup> I use /ði(j)/ to represent any instance of *the* with a high front vowel, and /ðə/ to represent instances of *the* with a central vowel ([ə], [ɨ] or [ʌ]) (see Todaka 1992; Keating et al. 1994: 136; Fox Tree & Clark: 152).

- (i) article cliticization;
- (ii) allomorphy (insertion of either /e/ or /æŋ/);
- (iii) then phonological vowel-reduction – notably, the same vowel-reduction that derives /ðə/ in the definite article – which derives the weak variants /ə/ and /ən/ from the strong variants /e/ and /æŋ/ (respectively).

The weak/strong distinction in the indefinite article is not widely recognized in previous work, and presents a complication for proposals that treat *a/an* as uniformly allomorphic. I show that the distribution of the four alternants (/e/, /æŋ/, /ə/, /ən/), and their very regular parallels with other weak/strong function-word alternations, are best explained in a hybrid, multi-step model where *a/an* involves both allomorphy and phonology and *the* involves only phonology.

The analysis has at least two important theoretical consequences. First, it underscores the need for at least some “special” phonology (in this case, word-internal phonology) – meaning that the fact that an alternation does not apply across-the-board cannot in itself be taken as evidence that the alternation is allomorphic. Second, if it is true that allomorphy strictly precedes phonology as in the model adopted here, then allomorphy cannot have direct access to the surface phonological structure (*contra* TETU-based approaches). In §4 I show that *a/an* is *not* always phonologically optimizing on the surface; *an* is sometimes selected even when its /n/ syllabifies as a coda (e.g. *an* /ʔ/ápple), and *a* is sometimes selected even when it is immediately followed by a vowel (e.g. *I want a um...*). Implications of these findings for theories of the phonology-morphology interface are discussed in §4.4 and §6.

The paper is laid out as follows. In §2 I review some criteria for distinguishing allomorphy from phonology and show that there are non-trivial differences between *a/an* and *the*. In §3 I lay out my analysis of *a/an* and *the* (briefly sketched above). In §4 I present evidence that neither *a/an* nor *the* is uniformly phonologically optimizing. An account of inter- and intraspeaker variability in *a/an* and *the* is given in §5, and §6 concludes the paper.

## 2 Criteria for distinguishing allomorphy from phonology

In this section I review some well-known criteria for distinguishing allomorphy from phonology and hold up *a/an* and *the* individually to these criteria. We will see that, despite initial appearances, there are non-trivial differences between *a/an* and *the* that would present problems for a uniformly allomorphic treatment of either alternation.

Throughout this paper I use the term *allomorphy* to refer to a situation where a single morpheme has two or more distinct phonological forms, each of which is memorized and stored (i.e. where neither form is derived from the other). Since allomorphy involves storing distinct forms, it is most clearly at work when (i) the alternants in question have very different pronunciations, *and* (ii) the alternation is restricted to a single morpheme. English *go/wen(t)*, for example, is an obvious candidate for an allomorphic treatment by both measures: /go/ and /wɛn/ are so dissimilar that it is certainly easier to store them separately than to learn the series of rules that would be needed to derive one from the other phonologically – especially since these rules would have to be restricted to a single morpheme, the light verb  $v_{go}$ .

The alternation between /ætəm/ and /ærəm/ in the (American) English word *atom*, on the other hand, is uncontroversially a case of phonological “tweaking” rather than allomorphy: unlike with *go/wen(t)*, (i) the alternants are nearly identical phonetically, and (ii) the distinguishing segments (/t/ and /r/) alternate not just in the word *atom* but

in virtually any context that meets the conditions for Flapping (Nespor & Vogel 1986: ch8, among many others), even across XP boundaries (e.g. *We gave the fruitba[r] a shower*; Kaisse 1985: 26). It is obviously far less burdensome to learn a phonological rule of Flapping than to memorize a variant pronunciation for every word with a potential t~r alternation (along with its conditioning contexts).

The two criteria we have been considering are summarized below:

(5) Allomorphy or phonology?

Criterion	<i>Allomorphy</i>	<i>Phonology</i>
A. Degree of phonetic resemblance	Very little resemblance (e.g. <i>go/wen(t)</i> )	Very close resemblance (e.g. t/d/r)
B. Degree of lexical/structural restrictedness	Restricted to one morpheme (e.g. <i>go/wen(t)</i> )	Potentially across the board (e.g. <i>atom, at 'em,...</i> )

I have intentionally begun with very obvious examples. What makes them so obvious is that they are oriented at the opposite endpoints of both the Criterion A scale and the Criterion B scale. But it is important to keep in mind that these criteria *do* involve scales, not binary choices, and that many (perhaps most) cases fall somewhere in the middle of one or both scales. Different kinds of French liaison, for example, are situated at different points along the Criteria A/B scales and have been analyzed both allomorphically and phonologically with varying degrees of success (see Tranel 1990 for a review; see Hayes 1990; Pak 2008; Siddiqi 2013 and §4 for additional cases and discussion). *A/an* and *the* are also both in-between cases, and since they are so similar to each other in basic respects (same language, same morphosyntactic category, same phonological conditions), any differences in terms of *where* on each scale the alternations are situated will be particularly informative.

Let's consider how *a/an* measures up with respect to Criteria A and B.

With respect to Criterion A (degree of phonetic resemblance), *a* and *an* are more similar than /go/ and /wen/, and it is possible to derive one from the other phonologically via a single rule.<sup>4</sup> However, as pointed out by Rotenberg (1978: 27ff), this rule would need to be a rule of /n/—insertion (*a* → *an*) rather than the more phonologically natural /n/—elision (*an* → *a*) (*pace* Perlmutter 1970; Hurford 1972). This is because *a*, rather than *an*, is the form that appears in “elsewhere” contexts like (6), where the indefinite article is structurally separated from its nP complement.

- (6) a. I had a (*silence*) oh, what do you call it, an epiphany.
- b. I'd like a, um, a large coffee and a croissant.
- c. [This is] a, although I hate to admit it, very silly idea. (Rotenberg 1978: 39)

So while a phonological analysis of *a/an* is possible, it would require an unnatural rule arbitrarily inserting /n/ before a vowel (∅ → n / \_\_#V).<sup>5</sup>

<sup>4</sup> Assuming that *a/an* is a two-way alternation; see §3.2 for problems with this assumption.

<sup>5</sup> Unlike in contemporary English, /n/-elision was likely responsible for the *a/an* alternation during its initial development in the 13<sup>th</sup> century, when Old English *ān* began to be systematically pronounced as *a* before consonants and *an* before vowels. During this period, utterances like \**an book*, which are no longer attested, were common (Crisma 2009: 132–133). The development of the *a/an* alternation occurred at roughly the same time as a change in the semantics of *a(n)*, which in Old English denoted the numeral ‘one’ or was used as a presentative marker (Hopper & Martin 1985).

With respect to Criterion B (degree of lexical/structural restrictedness), *a/an* is an isolated case, applying to only one morpheme. In this respect it is similar to *go/wen(t)*.

Now let's see how *the* measures up.

With respect to Criterion A, /ði/ and /ðə/ are identical except that one has a full vowel /i/ where the other has /ə/. It is possible to derive /ðə/ from /ði/ phonologically via *unstressed-vowel reduction*, which (unlike /n/-insertion) is a cross-linguistically well-established phenomenon (Crosswhite 2004).

With respect to Criterion B, as well, *the* is closer to the phonological end of the spectrum than *a/an*. While it is true that word-final /i/ and /ə/ do not alternate in *happy*, *party*, *agenda*, etc. (as noted in §1), V~ə alternations are found in many *word-internal* contexts in English. For example, affixation and other word-formation processes yield well-known alternations between full vowels and /ə/ ((7)a) (Chomsky & Halle 1968; Marvin 2002). Furthermore, there is both inter- and intra-speaker variation in the pronunciation of unstressed vowels in *behave*, *eleven*, and other words beginning with orthographic *re-*, *de-*, *e-*, *be-*, *pre-* ((7)b) (Wells 2008; Nádasdy 2013). Finally, many monosyllabic function words – including *the* – have a “strong” (stressed) full-vowel variant as well as a “weak” (stressless) variant with /ə/ or a syllabic consonant ((7)c) (Selkirk 1995; Jurafsky et al. 1998 among others).

- (7) More V~ə alternations in English:
- a. *Word-formation*: *beaut/i/~beaut/ə/ful*, */ə/xpl/e/n~/ε/xpl/ə/nation*
  - b. *Stylistic variation*: *believe*, *behave*, *relax*, *emergency*, *eraser*, *eleven*
  - c. *Monosyllabic function words*:
    - i. You *c/æ/n* finish early, but you won't. ~ You *c/ə/n* dó it.
    - ii. I voted *f/ɔ/r* it, not against it. ~ I voted *f/ə/r* Jóhn.
    - iii. John wrote *th/í/* paper on Lincoln. ~ John wrote *th/ə/* páper.

So while both *a/an* and *the* are “in-between” cases with respect to Criteria A and B, *the* is closer to the phonological end of both scales than *a/an*. In the remainder of this paper I pursue the hypothesis that these differences, while not extreme, are nevertheless significant, pointing to a fundamental difference in the grammatical status of *a/an* and *the*. As shown in the next section, *the* can be wholly subsumed under a phonological analysis involving unstressed-vowel reduction, while *a/an* calls for allomorphy in addition to vowel reduction.

There are at least two additional preliminary indications that *the* should be analyzed phonologically. First, the phonological conditions for *the* are very similar to those for V~ə alternations in word-formation: just as the vowel in *the* (usually) does not reduce before vowels (/ðə/ *book*, /ði/ *apple*), the vowel at the end of a root or “stem” does not reduce before a vowel-initial suffix (*beaut/ə/-ful*, *beaut/i/-ous*) (Chomsky & Halle 1968: 111). This parallel suggests that *the* might be produced by the same (word-internal) phonological rule(s) as the alternations in (7)a. If we treated *the* allomorphically, the fact that the /ðə/ “allomorph” shows up in just those contexts where /ə/ is generally allowed in English (unstressed, before consonants) would be a mere coincidence.

Interestingly, there is also some evidence that children acquire *the* earlier than *a/an*. Table 1 shows results from a study of 36 North American 3- to 7-year-olds and their adult caregivers in the CHILDES corpus (MacWhinney 2000). In contexts where the definite article was prevocalic (e.g. *the apple*), children in this study used the expected form /ði/ 49% of the time, but in contexts where the indefinite article was prevocalic (e.g. *a(n) apple*), children used the expected form *an* only 38% of the time (see §5 for more details; see Newton & Wells 1999 for similar results from an experimental study). This

apparent lag in the acquisition of *a/an* is particularly striking given that the adult caregivers showed the opposite pattern, using the expected prevocalic form *less* frequently with *the* than with *a/an* (see Healy et al. 1998; Raymond et al. 2002 for further evidence that adults have more variability with *the* than *a/an*).<sup>6</sup>

	% <i>an</i>	<i>an</i> /( <i>a+an</i> )	% <i>ði</i>	<i>ði</i> /( <i>ðə+ði</i> )
children	38%	133/347	49%	457/935
adults	96%	1952/2033	90%	773/863

**Table 1:** Frequency of prevocalic *an* and /ði/ in 36 North American 3- to 7-year-olds and their adult caregivers in CHILDES.<sup>7</sup>

One possible explanation for this contrast is directly related to the proposal advanced here: that *the* is one reflex of a more general vowel-reduction rule in English. The idea would be that children are noticing at least some of the other  $V \sim \text{ə}$  alternations in (7) and making connections among these alternations, *the*, and phonological vowel reduction, whereas with *a/an* they have no such precedent and must acquire the alternation as an isolated case (see §§5–6 for further discussion).<sup>8</sup> This line of explanation rests on the assumption that *the* is a *phonological* alternation between  $V$  and /ə/ rather than an allomorphic alternation between /ði/ and /ðə/. If /ði/ and /ðə/ were allomorphs, they would be stored independently and inserted as atoms, just as e.g. /ði/ and /ma/ would be, and any parallels with other  $V \sim \text{ə}$  alternations in the grammar would have to be seen as coincidental rather than potentially informative for acquisition.

### 3 Analysis

For clarity of exposition, I will first lay out an analysis of *a/an* and *the* that assumes that both alternations are categorical (contrary to fact). Then, in §5, I will show how this analysis can be incorporated into a competing-grammars framework to account for various kinds of attested inter- and intraspeaker variation.

#### 3.1 The *as a phonological alternation*

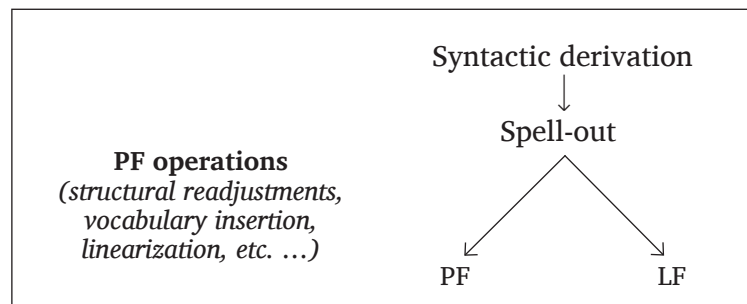
I assume an architecture in which surface phonetic forms are derived by a strictly ordered series of operations in a post-syntactic PF component (Halle & Marantz 1993, among others). PF operations include *linearization*, *vocabulary insertion* (i.e. insertion of the phonological content of functional heads, including allomorphically alternating heads), limited structural *readjustments* (e.g. certain kinds of “cliticization,” or local dislocation), and *phonological rules* of various kinds (see Embick & Noyer 2001 et seq.) (Figure 1).

Consistent with much recent work in Distributed Morphology, I assume that morpho-syntactic structures – including internally complex words – are spelled out in chunks (or cycles) instead of all at once (Marvin 2002; Embick 2010 see also note 15). Furthermore, I assume that phonological rules apply as these chunks of increasing size are spelled out and linearized, and thus have access to different kinds of information (Marvin 2002; Pak 2008).

<sup>6</sup> To explain these unexpected prevocalic forms in adult speech (e.g. /ðə/ *apple*), I propose in §5 that in addition to the grammar that produces the *ði/ðə* alternation, many speakers also have access to a grammar with  $D[+ \text{def}]$  realized as invariant /ðə/. For the adults in Table 1, for example, this “invariant /ðə/” grammar is chosen about 10% of the time.

<sup>7</sup> Table 1 includes data from the Braunwald, MacWhinney, Nelson, Providence, Sawyer, and Snow corpora; see Appendix and MacWhinney (2000) for more information.

<sup>8</sup> ‘Schwa-strengthening’ errors in acquisition (e.g. *el/o/phat* (Sawyer corpus 2-28-92); /i//*lgeric to eggs, pay /e//tention* (observed by author)) might also support the idea that children have hypothesized a rule producing  $V \sim \text{ə}$  alternations; see Levelt (2008) for similar data from Dutch.



**Figure 1:** Distributed Morphology architecture.

In §2 we reviewed a number of facts suggesting that *the* is just one reflex of a more general vowel-reduction process in English. The analysis laid out here therefore advances the following hypothesis:

- (8) *the* is derived by the same (vowel-reduction-based) phonological analysis as other V~ə alternations in English (e.g. (7)a, (7)c).

An initial challenge to such a unified analysis, as noted in §1, is that *the* appears to be an exception to the strictly *word-internal* nature of vowel-reduction: /i/ does not usually reduce word-finally (Chomsky & Halle 1968: 111).

- (9) \*lovel/ə/ book, \*part/ə/ time, \*carr/ə/ babies, \*craz/ə/ kids  
th/ə/ book, th/ə/ time, th/ə/ babies, th/ə/ kids, beaut/ə/ful

Informally speaking, *the* “acts like part of the following word” for the purposes of vowel reduction. Accordingly, I propose that English D[±def] is part of the following word, by virtue of *Local Dislocation* (Embick & Noyer 2001; Embick 2010) – a post-syntactic (PF) operation that takes two linearly adjacent words<sup>9</sup> and turns them into a single word by adjoining (or “cliticizing”<sup>10</sup>) one to the other. Local Dislocation has also been argued to apply to the definite article in French (Embick 2007: 328ff; 2010: 87ff), where its effects are manifested in the phonology as irregular vowel deletion.

- (10) a. l’arbre ‘the tree’ (\*le arbre), l’école ‘the school’ (\*la école)  
b. le chien ‘the.MASC dog’, la fille ‘the.FEM girl’

A Local Dislocation rule for English D[±def] is given in (11). This rule takes as its input two linearly adjacent words spelled out in the same cycle, where the first is D[±def] and the second is any word X, and yields a single word (in square brackets) with D[±def] adjoined to X.

- (11) English Article Local Dislocation: D[±def]  $\widehat{\text{X}}$  → [D[±def] [X]]<sup>11</sup>

Unlike syntactic head-movement, Local Dislocation does not require that one word be the head of the complement of the other; this is why English D[±def] can adjoin to an adverb

<sup>9</sup> I use the term *word* to refer to a (potentially complex) X<sup>0</sup> that is not dominated by any further X<sup>0</sup> (a maximal word, or *M-word* in Embick & Noyer 2001). See Embick (2007; 2010) for more examples of Local Dislocation.

<sup>10</sup> Although I use the term *cliticize* here, I do not intend for this analysis to be applicable to every phenomenon that has fallen under the rubric of *cliticization* in the previous literature. While French *l’* and English contracted auxiliaries have been analyzed as instances of Local Dislocation (Embick 2010; Mackenzie 2012, resp.), other ‘clitics’ have been attributed to lowering (e.g. Bulgarian D[def]; Embick & Noyer 2001), syntactic head-raising, or other processes.

<sup>11</sup> A Concatenation statement X  $\widehat{\text{Y}}$  is read ‘X is left-adjacent to Y.’ Concatenation is assumed to be an operation that establishes linear order between each pair of (M-)words within a cycle (Embick & Noyer 2001). The linear order of morphemes within a complex (M-)word is established by a similar operation, which is also assumed to apply to the output of Local Dislocation rules (Embick 2007: 321–322).

in e.g. (12). On the other hand, there must be *some* word X spelled out in the same cycle as D[±def] in order for Local Dislocation to apply; if this condition is not met, as in (13) (see also (6)), Local Dislocation does not occur and D[±def] remains an independent word.

- (12) a. {ði/an} unusually large baby  
b. {ðə/a} surprisingly large baby

(13) I'd like {the/a}...I don't know...

By effectively making D[±def] word-internal, Article Local Dislocation allows /ði/~/ðə/ to be potentially subsumed under the same phonological analysis as *beauty*~*beautiful* and other word-internal V~ə alternations, consistent with (8). To illustrate how such a unified approach might work, I use a slightly modified version of Chomsky & Halle's (1968: 111ff) analysis of word-internal vowel reduction. To the extent that SPE-style approaches to vowel reduction have been revised in subsequent work (e.g. Rhodes 1996; Marvin 2002), the current proposal could also be revised without introducing any problems that I am aware of. The important point here, again, is to show that *the* can be analyzed as one reflex of a more general vowel-reduction rule.

The basic form of the definite article is assumed to be /ðɪ/, inserted by the Vocabulary Insertion rule in (14):

- (14) Vocabulary Insertion: D[-def] ↔ ðɪ

The /ði/~/ðə/ alternation is then produced by two word-internal phonological rules, Tensing and Vowel Reduction:<sup>12</sup>

- (15) a. Tensing (cyclic): V[-low -stress] → [+tense] / \_{V,#}  
b. Vowel Reduction (non-cyclic): V[-stress -tense] → ə

The rules work together roughly as follows: the final vowel in *the*, *crazy*, *happy*, *beauty*, etc. is underlyingly [-tense] and [-stress]; it becomes [+tense] by rule (15)a if it is prevocalic or final, as in *beaute-ous* and *beauty*, which in turn makes it immune to Vowel Reduction (15)b. If the vowel precedes a consonant (as in *beauti-ful*) then Tensing has no effect and rule (15)b subsequently reduces the vowel to /ə/. The need to specify Tensing as cyclic and Vowel Reduction as non-cyclic will be explained shortly (see (18)-(21) and surrounding discussion, including footnote 17).

Let us see how the proposal works with some sample derivations. Consider first the DP *the crazy kid*, in a grammar where there is a reduced vowel in *the* but not in *crazy*.<sup>13</sup>

- (16) [DP [D the] [nP crazy kid]] /ðə krezi kɪd/

Within the nP, the words *crazy* and *kid* are individually spelled out, and the rules of Tensing and Vowel Reduction apply within each word. The vowel at the end of *crazy* becomes [+tense] by rule (15)a; this [+tense] feature then prevents the vowel from undergoing reduction (rule (15)b).

On the DP cycle, the definite article (D[+def]) is introduced. First D[+def] cliticizes to *crazy* by Article Local Dislocation (11), then D[+def] is spelled out as /ðɪ/ by the Vocabulary Insertion rule in (14), and then the phonological rules in (15) apply. The context for Tensing

<sup>12</sup> In Chomsky & Halle (1968), these rules are intended to explain various gaps in the distribution of lax vowels in English, e.g. (i) lax vowels are never found prevocally and (ii) the only unstressed vowels allowed word-finally are the tense non-low /i/, /u/, /e/, and /o/, as well as /ə/.

<sup>13</sup> In Traditional RP and some other dialects, *happy*, *city*, etc. have a lax final vowel (Wells 1982), as Chomsky & Halle also recognize (1968: 74, note 22). Although a detailed analysis of these non-'happy-tensing' varieties is beyond the scope of this paper, one possibility is that these varieties have Tensing of /i/ only in the context \_V, rather than in the context \_{V,#}.



is *not* met by the vowel in *the* here, because this vowel is preconsonantal. Therefore, the vowel remains [-tense] and subsequently undergoes Vowel Reduction.

- (17) a. *Article Local Dislocation* (11):  
 $D[+def] \widehat{[}_a [\sqrt{\text{CRAZY}}] \emptyset] \rightarrow [D[+def] [}_a [\sqrt{\text{CRAZY}}] \emptyset]]$   
 b. *Vocabulary Insertion* (14):  $D[+def] \leftrightarrow \delta i$   
 c. *Tensing* (15)a: NA (because of following /k/)  
 d. *Vowel Reduction* (15)b:  $\delta i \rightarrow \delta \text{ə}$

In *the ugly kid*, the DP-cycle derivation proceeds exactly as in (17) except that since *ugly* is vowel-initial, the context for Tensing is met. The vowel in *the* becomes [+tense] and thus immune to Vowel Reduction, so the final pronunciation is /ði/ *ugly kid*.

Now consider the individual words *beauty*, *happy*, *beautiful* and *happiness*. For many speakers,<sup>14</sup> *beauty* and *happy* both end in /i/, but while this vowel reduces to /ə/ in *beautiful*, it is a tense /i/ in *happiness*. How can we account for this contrast?

In the spirit of Marvin (2002) I take this contrast as a sign that *beautiful* and *happiness* have different internal structures. *Beautiful* is spelled out in a single cycle (-ful attaches directly to the root (18)), while *happiness* is spelled out in two (word-internal) cycles: first the root  $\sqrt{\text{HAPPY}}$  combines with a null category-defining *a(dj)* head; then this derived adjective combines with [<sub>n</sub> -ness] (19):<sup>15</sup>

- (18) *Cycle 1*: [<sub>a</sub> [ $\sqrt{\text{BEAUTY}}$ ] -ful]  
 (19) *Cycle 1*: [<sub>a</sub> [ $\sqrt{\text{HAPPY}}$ ]  $\emptyset$ ]  
*Cycle 2*: [<sub>n</sub> [<sub>a</sub> [ $\sqrt{\text{HAPPY}}$ ]  $\emptyset$ ] -ness]

Since *beautiful* is spelled out on a single cycle,<sup>16</sup> Tensing “sees” that the /i/ in  $\sqrt{\text{BEAUTY}}$  is followed by a consonant, and accordingly does not apply. Once all word-internal content has been spelled out (in this case, after Cycle 1), the non-cyclic rule of Vowel Reduction applies.

- (20) *Cycle 1*: bjutɪ ful  
*Tensing*: NA (because of following /f/)  
*Word-level*:  
*Vowel Reduction*:  $i \rightarrow \text{ə}$  (and  $u \rightarrow \text{ə}$ )

<sup>14</sup> Inter- and intraspeaker variation in word-internal  $i \sim \text{ə}$  alternations could be accounted for in a number of ways: differences in underlying morphological structure (e.g. grammars with tense /i/ in *beautiful*, etc. could have these words spelled out in two cycles instead of one); underlying vowel quality (e.g. some grammars could have underlying tense /i/ in *eleven*, *remember*, and other words with ‘presume-tensing’; Nádasdy 2013), or differences in rule variability (e.g. some speakers could have optional rather than categorical Tensing, variably allowing lax vowels in *happy*, *happiness*, *the only*, *beauteous*, etc.). See also the previous note. Each solution would of course make distinct predictions that remain to be tested.

<sup>15</sup> The notion of the word-internal cycle is featured not only in Chomsky & Halle (1968) and Marvin (2002), but also in Lexical Phonology and Morphology (where it plays a key role) and Stratal OT (e.g. Bermúdez-Otero 2004 et seq.). The question of whether these theories are in some cases too “aggressively decompositional” (Haugen & Siddiqi 2013), in proposing morpheme breakdowns that no longer occur in contemporary speakers’ mental grammars, is for the most part orthogonal to the current proposal. If it turns out that *beautiful* and *beauteous* are monomorphemic, for example, then the main consequence for this paper is that the examples in (7)a would be only apparent parallels to *the*, and could not be viewed as evidence supporting a phonological treatment of *the*. The other arguments for treating *the* phonologically – e.g. its parallels to the strong/weak function-word pairs in (7)c – would still hold.

<sup>16</sup> While -ful often attaches to (apparent) nouns, there are exceptions, e.g. *forgetful*, *fretful*, *grateful*, *baleful*. It can also yield non-transparent meanings typical of root-attached affixes (e.g. *merciful* means ‘full of mercy’ but *awful* and *dreadful* do not mean ‘full of awe/dread’; the roots in *artful*, *fruitful* have only their archaic meanings). The suffix in *beautiful* is not to be confused with the suffix in *handful*, *mouthful*, etc., which has very different structural properties: -ful in *handful* attaches only to nouns (not to category-neutral roots) and produces a new noun (not an adjective). Notice the corresponding contrast in vowel-reduction between *beautiful* /ə/ and *bellyful* /i/.

Since *happiness* is spelled out on two cycles, on the other hand, Tensing never “sees” the /ɪ/ in *happy* and the /n/ in *-ness* at the same time. Tensing is a cyclic rule, applying once on each word-internal cycle. When Tensing applies on the first cycle, the only material available is the adjective *happy*; since nothing follows the /ɪ/ in *happy* at this point, Tensing assigns [+tense] to it, thus making it immune to (non-cyclic) Vowel Reduction.<sup>17</sup>

- (21) *Cycle 1:* hæpɪ  
 Tensing: ɪ → i (because nothing follows *happy* at this stage)  
*Cycle 2:* < hæpi > nəs  
 Tensing: NA  
*Word-level:*  
 Vowel Reduction: ε → ə (but /i/ does not reduce because it is [+tense])

### 3.2 Vowel reduction also applies in a/an

One advantage of viewing *the* as a phonological alternation is that it allows us to understand certain aspects of *a/an* as well. While *a/an* is often implicitly assumed to be a two-way alternation, many adult speakers actually have four surface variants: /ə/, /ən/, e(j) and either /æn/ or /ɛn/ (see also Bloomfield 1935: 186; Jurafsky et al. 1998; Asudeh & Klein 2002; Clark & Fox Tree 2002: 102). The full-vowel variants /ej/ and /æn-ɛn/ are used in careful speech, as citation forms, or when they bear nuclear sentence stress (e.g. (22)) – all of which, plausibly, are contexts where D[-def] bears at least some stress. Notice that in these contexts, only the /æn-ɛn/ variant is used prevocally.

- (22) a. Not *a* /ej/ house, but *the* house.  
 b. Not *an* /ɛn/ uncle, but *her* uncle. (Bloomfield 1935: 186)

Putting these observations together, we can conclude that *a* and *an* each have a “strong” form with a full vowel and a “weak” form with /ə/, distributed in the same way as the other monosyllabic function-words pairs we saw in §2 – including, of course, *the*.<sup>18</sup>

- (23) a. You c/æ/n finish early, but you won’t. ~ You c/ə/n dó it.  
 (*repeated from (7)c*)  
 b. I voted f/ɔ́/r it, not against it. ~ I voted f/ə/r Jóhn.  
 c. John wrote th/í/ paper on parentheticals. ~ John wrote th/ə/ páper.

(24) Strong/weak function-word pairs in English

	<i>can</i>	<i>for</i>	<i>to</i>	D[+ def]	D[-def]	
					_V	<i>else</i>
<i>strong</i>	kæn	fɔr	tu	ði	æn-ɛn	e
<i>weak</i>	kən	fər	tə	ðə	ən	ə

In the previous subsection I used two word-internal rules, Tensing and Vowel Reduction, to derive /ði/ and /ðə/ in the definite article. As expected under the unified-analysis hypothesis in (8), Tensing and Vowel Reduction can be used to derive the

<sup>17</sup> An alternative analysis, with cyclic vowel-reduction and no reference to [+tense] (V[-stress] → ə / \_C), turns out to be problematic. As pointed out by Chomsky & Halle (1968: 113), Vowel Reduction cannot *itself* be cyclic because then it would apply to inner cycles in *solid*, *brutal*, *president*, etc. and leave no way to recover the full vowels when stress-shifting affixes are added on later cycles (*solid-ify*, *brutal-ity*, *president-ial*).

<sup>18</sup> See §5 for an account of grammars that allow the ‘weak’ forms of D[± def] to bear pitch-accent, e.g. *I read /ɔ́/ book, not th/ə́/ book*.

other alternations in (24) as well.<sup>19</sup> In contexts where the function word bears at least some stress, it will automatically be immune to Vowel Reduction ((15)b). The fact that vowel-final function words (*the*, *a*, *to*) surface with tense vowels when stressed can be attributed to an additional tensing rule (Tensing 2), which assigns [+tense] to a stressed morpheme-final vowel;<sup>20</sup> Tensing 2 also applies word-internally and cyclically, and is responsible for the absence of English words ending in stressed lax vowels (\*pe, \*stæ, \*ki, etc.).

(25) Tensing 2:<sup>21</sup> V[+stress] → [+tense] / \_\_ ]x<sup>0</sup>

In contexts where a function word is stressless, the Tensing and Vowel Reduction rules from (15) derive the weak forms, as shown with the following derivations.

On the DP cycle of *a book* (with unstressed *a*), Article Local Dislocation, Tensing and Vowel Reduction all apply exactly as in *the crazy kid* (17). The main difference between *a/an* and *the* involves step (b), Vocabulary Insertion: for *a/an*, there is an allomorphy rule that inserts /æɪn/ (or /ɛn/) before vowels and /ɛ/ elsewhere.

- (26) Derivation of *a book* (DP cycle):
- a. *Article Local Dislocation* (11):  
D[-def] [̂ [n [√BOOK] Ø] → [D[-def] [n [√BOOK] Ø]]
  - b. *Vocabulary Insertion*: D[-def] ↔ æɪn / \_\_ V  
↔ ɛ elsewhere  
(/ɛ/ inserted here because the following segment is the consonant /b/)
  - c. *Tensing* (15)a: NA (because of following /b/)
  - d. *Vowel Reduction* (15)b: ɛ → ə

I also assume that a Diphthongization rule applies sometime after step (c), inserting a front glide /j/ after a tense front vowel (ði → ðij, e → ej).

In the DP *an apple* (with unstressed *an*), the derivation proceeds exactly as in (26) except that since *apple* is vowel-initial, the /æɪn/ allomorph is selected at step (b). Once /æɪn/ is inserted, it behaves exactly like *can*, *for*, and other monosyllabic function words with lax vowels in closed syllables – i.e., it escapes Tensing and undergoes Vowel Reduction.

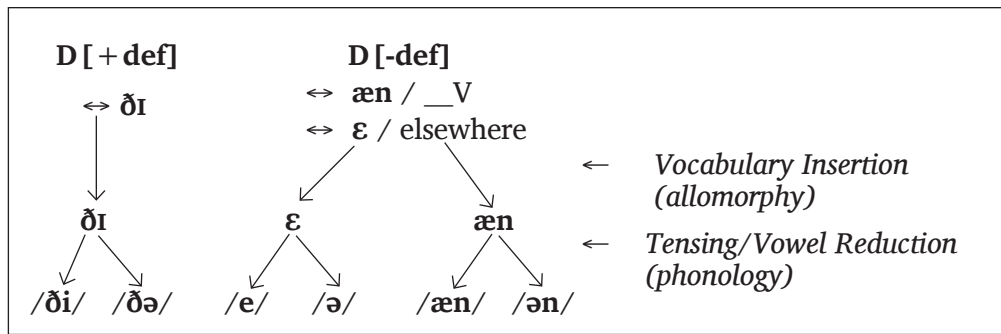
<sup>19</sup> As a reviewer points out, Tensing is predicted to apply to prevocalic *to* (e.g. *to add*) – assuming also that *to* cliticizes onto the following word. This prediction appears to be borne out in my speech and in at least some other varieties (see Britain & Fox 2009), but I have not yet examined the *to* alternation as part of my CHILDES North American English corpus study.

<sup>20</sup> Chomsky & Halle (1968: 74) have a single rule with multiple disjunctions that includes all of the conditions for my Tensing and Tensing 2. One question that arises is whether it might be simpler to replace the Tensing rule(s) with a single, relatively simple Laxing rule (V[-stress] → [-tense] / \_\_ C); under this modified approach, the final vowel in *crazy*, *happy*, etc. would be underlyingly tense. Either type of analysis can be used for the derivations presented in this section. The Laxing-based analysis does not explain the distributional gaps described in note 12, but makes a different prediction: that a stressless tense vowel cannot immediately precede a consonant introduced in the same cycle. This prediction seems to be borne out for the most part, assuming that the tense vowels in *loquacious*, *jujitsu*, *vacation*, *phonological*, *relocate*, etc. bear at least some stress. However, the Laxing-based analysis does not provide an obvious way to account for ‘presume-tensing’ in *behave*, *eleven*, *remember*, etc. ((7)b; Nádasy 2013). See §5 for an account of *presume-tensing* under the Tensing-based analysis.

<sup>21</sup> A reviewer points out that this rule makes reference to a morphosyntactic category (X<sup>0</sup>). There is an extensive literature debating whether the phonology applies directly to morphosyntactic constituents (direct reference), or has access only to a hierarchy of derived prosodic constituents including Prosodic Word, Phonological Phrase, etc. (indirect reference) (see Elordieta 2008 for a review). In other work (Pak 2008) I argue for a direct-reference model where there are no prosodic constituents, and I follow this principle in (25). However, the direct-reference assumption is not crucial for the current paper; my analysis of *a/an* and *the* will work whether the domains for Tensing and Vowel Reduction are defined morphosyntactically (X<sup>0</sup>) or prosodically (ω).

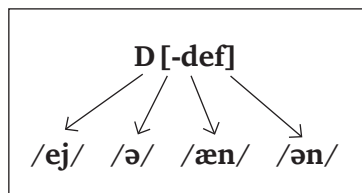
- (27) Derivation of *an apple* (DP cycle):
- a. *Article Local Dislocation* (11):  
 $D[-def] \widehat{[}_n [\sqrt{APPLE}] \emptyset] \rightarrow [D[-def] \widehat{[}_n [\sqrt{APPLE}] \emptyset]]$
  - b. *Vocabulary Insertion*:  $D[-def] \leftrightarrow \text{æn} / \_V$   
 $\leftrightarrow \text{ɛ} \text{ elsewhere}$   
 (/æn/ inserted here because the following segment is a vowel)
  - c. *Tensing* (15)a: NA (vowel in /æn/ is followed by a consonant)
  - d. *Vowel Reduction* (15)b:  $\text{æn} \rightarrow \text{ən}$

This analysis captures the observation that /e(j)/~/ə/ are similar to each other in the same way as /æn/~/ən/ and /ði(j)/~/ðə/ – specifically, each /ə/ form can be derived from a full-vowel form by Vowel Reduction. At the same time, this analysis captures an important difference between *a/an* and *the*, apparent in (24) and also schematized in Figure 2: While *the* is a two-way alternation that can be attributed to phonology alone, *a/an* is a four-way alternation that involves both phonology and allomorphy.<sup>22</sup>



**Figure 2:** Two-tiered analysis of *a/an* and *the*.

If we wanted to pursue instead a single-tiered, uniformly allomorphic treatment of *a/an* (see e.g. Asudeh & Klein 2002), we would have to adopt something like Figure 3: four-way allomorphy for *a/an*, with spellout rules that insert full-vowel forms when [+ stress] and /ə/ variants elsewhere – *but leave this correspondence unexplained*.



**Figure 3:** Uniform “flat” allomorphy for the English indefinite article (rejected).

<sup>22</sup> The initial split between /ɛ/ and /æn/ in D[-def] is treated as allomorphic rather than phonological in accordance with both Criterion A (little phonological resemblance between /ɛ/ and /æn/) and Criterion B (restricted to a single morpheme). A purely phonological analysis of /ɛ/~/æn/ would be compatible with the current proposal, but it would require *two* idiosyncratic rules that were restricted to the morpheme D[-def]: /n/ insertion ( $\emptyset \rightarrow n / \_V$ ; see §2) and either vowel lowering ( $\text{ɛ} \rightarrow \text{æ}$ ) or vowel raising ( $\text{æ} \rightarrow \text{ɛ}$ ) (although grammars with /ən/ as the strong form of *an* would require only /n/-insertion). For current purposes, the important point is that these morpheme-specific rules would still need to *precede* the more general rules of Tensing and Vowel Reduction, since Tensing ‘sees’ and is bled by the /n/ in *an* (see (27)). In other words, the derivation of the four variants of D[-def] requires two tiers, as shown in the diagram above, whether the initial split is allomorphic or phonological.

- (28) D[-def] ↔ æn / \_\_V if D[-def] is [+stress]  
 e if D[-def] is [+stress]  
 ən / \_\_V  
 ə

In the two-tiered approach I have proposed, which makes crucial use of phonological vowel reduction, the systematic correspondences between [ $\pm$ stress] and  $V \sim \text{ə}$  alternations is explained. Furthermore, this treatment takes care of *the* for free. Thus, we have yet another piece of evidence that *the* is a phonological rather than allomorphic alternation.

Before moving on, notice that the fact that *a/an* tends to yield unmarked syllables (*an* before vowels and *a* before consonants, rather than the other way around), is not “explained in the grammar” under my analysis – i.e., the analysis accounts for the pattern but does not incorporate a principled reason for it. This is not necessarily problematic, since the pattern likely has a historical explanation: *a/an* started out as phonological /n/-elision and was reanalyzed over time as an allomorphic alternation with *a* as the default (see note 5). As we will see in the next section, *a/an* does not always yield optimal syllables in any case.

#### 4 Rule-ordering effects

In the previous section I used a two-tiered model – allomorphy, then phonology – to explain the distribution of the various surface realizations of the English definite and indefinite articles. This approach is clearly at odds with a TETU-based analysis of *a/an* (e.g. Mascaró 1996b), where allomorphy can “see” the surface phonological structure and be influenced by surface well-formedness constraints. Mascaró’s (1996b) analysis of *a/an* works as follows: *a* and *an* are listed as allomorphs, both of which are considered as potential candidates for insertion wherever the indefinite article is used. Since *a* and *an* are equally faithful candidates, the choice between them is determined by the low-ranked constraints ONSET and NO-CODA.

- (29) TETU analysis of *a/an* (Mascaró 1996b)

{a,an} book	ONSET	NO-CODA	{a,an} egg	ONSET	NO-CODA
☞ a.book	*	*	a.egg	**!	*
an.book	*	**!	☞ a.n egg	*	*

The idea is that even though English generally allows codas and onsetless syllables, a preference for unmarked CV.CV structures “emerges” in just those contexts where there are multiple, equally faithful underlying forms for a single vocabulary item.

An advantage Mascaró claims for his approach is that it explains why so many cases of (apparent) external allomorphy appear to involve hiatus avoidance or some other kind of phonological optimization (see (30) for a sample of cases cited in the literature).<sup>23</sup> As noted in §1, since tense vowels are diphthongized in English, *the* could be easily incorporated into Mascaró’s framework, and might even be viewed as an additional source of support for a theory of allomorphy as TETU (31).

- (30) Other proposed cases of allomorphy as TETU (Mascaró 1996a; b; 2007; Lee 2009)
- French **bo** \_C vs. **bɛl** \_V (*beau mari*, *bel enfant* ‘good-looking husband/child’) (also *nouveau/nouvel* ‘new’, *ce/cet* ‘this’, *ma/mon* ‘my’, etc.)

<sup>23</sup> Cases of apparent non-optimizing and anti-optimizing allomorphy (e.g. Haitian Creole definite suffix, Korean conjunctive suffix) are discussed in Embick (2010) and Bonet, Lloret & Mascaró (2007), among others.

- b. Catalan personal definite: **ən** \_C vs. **l** \_V (*en* Wittgenstein, *l'Einstein*)
- c. Northwest Catalan **lo** \_C vs. **l** \_V (*lo* pá, *l* ámo, 'the owner/bread')
- d. Ribagorçan Catalan **ésto/ íʃo** \_C vs. **ést/ íʃ** \_V (*ésto* λιβρε, *ést* ɔme 'this book/man')
- e. Moroccan Arabic C\_ **-u** vs. V\_ **-h** (*ktab-u*, *x<sup>ts</sup>a-h* 'his book/error' (also *i/ja* in 1SG))
- f. Korean C\_ **-i** vs. V\_ **-ka** (*sok-i* 'inside-NOM', *so-ka* 'cow-NOM') (Lee 2009)
- g. Basque N\_ **du**, else **tu** (*ilun-du* 'darken', *argi-tu* 'clear up') (also *dar/tar*, *ko/go*, *tik/dik*)

(31) TETU analysis of *the* (to be rejected)

{ðə, ðij} <b>book</b>	ONSET	NO-CODA	{ðə, ðij} <b>egg</b>	ONSET	NO-CODA
☞ ðə.book		*	ðə.egg	*!	*
ðij.book		**!	☞ ði.j egg		*

In Mascaró's (1996b) analysis, *an* is selected iff its /n/ is syllabified as an onset on the surface. This idea is unformulable in the architecture I adopt in §3. Since Vocabulary Insertion (allomorphy) strictly precedes phonological rule application in this model (see Figure 2), there is no way for an allomorphy rule to “see” the final phrase-level syllable structure – or, more specifically, for the /n/ in *an* to “know” that it will ultimately be syllabified as an onset. On the other hand, the model I adopt allows for a different kind of scenario: one where, *after* allomorph insertion, the phonology renders additional changes to D[-def] and surrounding material, possibly enough to disrupt the expected optimal syllable structures. The following subsections provide evidence for exactly this kind of post-allomorphic phonological meddling. First I show that the /n/ in *an* is not always syllabified as an onset on the surface (§4.1). Then I show that *an* sometimes fails to be selected even when it is followed by a vowel on the surface (§4.2–§4.3).

#### 4.1 Emphatic glottal stops

It is well-known that vowel-initial words in English are frequently pronounced with an initial glottal stop. Whether or not /ʔ/ appears depends on a number of factors; it is more likely to be inserted before a stressed vowel at the beginning of an utterance, but it is also possible in connected speech if the vowel-initial syllable has special “prominence” or “emphasis” (Borrorff 2007: 166; Garellek 2012; 2013: ch5).

- (32) a. He'll fall asleep /ʔ/ánywhere.  
 b. She's from /ʔ/Óregon, not Washington.  
 c. I haven't seen John in for/ʔ/éver.

I will refer to the glottal stop in examples like (32) as *Emphatic Glottal Stop*, in order to distinguish it from the (optional) glottal stop in e.g. *Tuscaloosa* /ʔ/*Alabama*, which seems to serve as a hiatus-breaker, but this distinction is not crucial to my analysis.

Notably, emphatic glottal stop can occur between the *an* variant of D[-def] and its complement:

- (33) a. That's an /ʔ/éxcellent idea.  
 b. What an /ʔ/ídiot.

Examples like (33) are by no means odd or unnatural; in our CHILDES corpus study, for example, 25% of adults' connected-speech utterances of *an* + V had an emphatic glottal

stop before the vowel.<sup>24</sup> But under Mascaró's (1996b) TETU-based analysis of *a/an*, these utterances present a problem. The /n/ in *an* here must be a coda, since English does not allow [nʔ] onsets. But if allomorph choice is truly determined by surface syllable well-formedness, /an.ʔidiot/ should always be beaten by either /a.ʔidiot/ or /a.n idiot/, which have fewer NO-CODA violations:<sup>25</sup>

(34)

	ONSET	NO-CODA
an.ʔidiot	**	**
a.n idiot	**	*
a.ʔidiot	**	*

A similar problem would arise under a TETU-based analysis of *the*, since /ði(j)/, like *an*, can be followed by /ʔ/. In our CHILDES corpus study (see §5), for example, 21% of adults' prevocalic /ði(j)/ (161/773) were followed by /ʔ/, and Keating et al. (1994: 137) and Todaka (1992: 46) report /ʔ/ after 30% of prevocalic /ði(j)/ in the TIMIT corpus. Before other consonants, however, unstressed /ði(j)/ is much less frequent (Todaka 1992: 41).

- (35) a. That was /ði(j) ʔ/óther guy.  
b. Turn on the /ði(j) ʔ/áir conditioner.

- (36) ?\* He's walking /ði(j)/ dog.

One might try to save the TETU account of *a/an* by proposing that the glottal stop does not really count as a consonant, or that it is somehow “outside of the grammar” altogether. However, while it is certainly debatable whether glottal stop is a phoneme, segment, feature or gesture (see Borroff 2007 for discussion), the problem here has to do with the *distribution* of emphatic /ʔ/, which is highly systematic and clearly grammar-internal. Consider the following contrasts:

- (37) That's /ənʔo/.  
a. ✓ That's an 'O.'  
b. \* That's a 'no.'

- (38) a. an ʔapple, Joan ʔAllen, %unʔethical, %inʔoperable  
b. \* banʔana, \*anʔalysis, \*connʔection, \*menʔorah, \*internʔational

The relevant generalization seems to be that emphatic /ʔ/ *must be syllable-initial*, and therefore can occur between a consonant C and a vowel V only if C is morpheme-final and thus potentially syllable-final. In other words, emphatic /ʔ/ needs to “see” that the /n/ in *an apple* can be a coda, unlike the /n/ in *analysis*, and it is not clear how this information would be accessible if the emphatic /ʔ/ were not part of the same system as the regular phonology and morphology. Under a TETU account, the problem is that the /n/ in *an apple* is crucially *not* supposed to be a coda.

<sup>24</sup> 238 of 961 utterances in the Braunwald, Ervin-Tripp, MacWhinney, Nelson, Providence, Sawyer and Snow corpora. See Appendix.

<sup>25</sup> A similar problem for Korean *i/ka* ((30)f) is described by Lee (2009) the *-i* allomorph is chosen after roots ending with /ŋ/ (waŋ-i ‘king-NOM’), but /ŋ/ is not a possible onset in Korean, so allomorph choice cannot be driven by ONSET and No-CODA alone (♣/wa.ŋi/). To solve this problem, Lee proposes a DEFAULT constraint, ranked above No-CODA, which identifies the phonologically simpler form (in this case *-i*) as the preferred form. This solution will not work for English *a/an*, however, because the unexpectedly attested form here is *an* rather than the phonologically simpler *a*, and because there are independent reasons to treat *a*, not *an*, as the default (see (6)).

Under my proposal, the fact that *an* and /ði/ show up before emphatic /ʔ/ can be straightforwardly explained as a rule-ordering effect: emphatic /ʔ/ is inserted relatively late, after Vocabulary Insertion, Tensing and Vowel Reduction have applied, and therefore does not count as a consonant for the purpose of *a/an* or *the*.

- (39)
- |    |  |              |              |
|----|--|--------------|--------------|
| a. | Vocabulary Insertion                       | /æ̃n/ idiot  | /ð̃i/ idiot  |
| b. | Tensing / Vowel Reduction                  | /ə̃n/ idiot  | /ð̃i/ idiot  |
| c. | Emphatic /ʔ/ Insertion ( <i>optional</i> ) | /ə̃n ʔ/idiot | /ð̃i ʔ/idiot |

At the stage when Emphatic /ʔ/ is added, the /n/ in *an* is still a coda (by virtue of being morpheme-final) and the first syllable of *idiot* is onsetless. (This VC.V syllable structure is what enables Emphatic /ʔ/ Insertion in *an*. /ʔ/a.lley but not in \*a.n/ʔ/a.ly.sis.) Resyllabification does not apply until after Emphatic /ʔ/ Insertion, and in (39) the resyllabification of /n/ in *an* is blocked by the epenthesized /ʔ/.

Resyllabification in turn precedes other PF processes. Flapping, which has been independently treated as a late-stage phenomenon in various serialist models,<sup>26</sup> applies post-resyllabification and, as expected, turns out to behave very differently from *a/an* with respect to whether it “sees” an epenthesized emphatic /ʔ/. Flapping applies only if /t/ or /d/ is immediately followed by a vowel *on the surface*, with no intervening segment or silence – which in turn suggests that the flap must surface as an onset (Kaisse 1985; Bermúdez-Otero 2007). Notably, an emphatic /ʔ/ that intervenes between /t/ and a vowel blocks Flapping:

- (40)
- |    |  |
|----|--|
| a. | That's Fa[r] Albert.                         |
| b. | That's Fa[t̚, *r] ʔÁlbert, not Flat Stanley. |

- (41)
- |    |                  |
|----|------------------|
| a. | wha[r]ever       |
| b. | wha[t̚, *r]ʔever |

Flapping “sees” and is blocked by emphatic /ʔ/ while *a/an* is blind to it. This contrast is unexpected under Mascaró's (1996b) analysis of *a/an*; if allomorph choice were directly guided by *surface* syllable well-formedness, then *an* should be blocked by an emphatic /ʔ/ just as Flapping is. In the current model, however, this contrast follows automatically from the way the relevant operations are ordered: emphatic /ʔ/ is inserted after the early rules of allomorphy and Tensing/Vowel Reduction, but before the late rules of resyllabification and Flapping.

#### 4.2 /h/ dropping

A similar solution can be applied to data described in Hurford (1972; 1974).<sup>27</sup> Hurford reports that older Cockney speakers who otherwise have categorical prevocalic *an* will use *a* if the following word starts with a “dropped” /h/. The resulting forms have

<sup>26</sup> Flapping has independently been characterized as “late” primarily because it applies nearly across-the-board, crossing word and phrase boundaries of various types (e.g. *Bring your jacke[r]*, *it's cold outside*), and most serialist or stratal models assume that phonological domains increase in size as the derivation proceeds (e.g. Kaisse 1985; Bermúdez-Otero 2007). Independent evidence that Flapping is a late-stage phenomenon comes from its interaction with Canadian Raising (a classic opacity effect): Canadian Raising applies in *writer* [ɹɪjɹɛɪ] but not *riɹer* [ɹɹɹɛɪ], even though the segment following the diphthong is identical (a flap) on the surface. The solution adopted by Bermúdez-Otero (2004; 2007) and others is to assume that Flapping follows and counterbleeds Raising (consistent with the idea that Raising is word-bounded and early while Flapping is phrasal and late). Raising applies to the stem *write* at the stage when its final segment is a voiceless [t̚]; this [t̚] is subsequently resyllabified and flapped at the phrase level. Thanks to an anonymous reviewer for referring me to Bermúdez-Otero's work.

<sup>27</sup> As pointed out by a reviewer, the pattern in (42) qualifies as evidence against a TETU-based analysis of *a/an* only for this particular dialect.



non-optimal (hiatus) syllable structures, so it is unclear how they would surface in a TETU-style approach:

- (42) a. a half [əɑ:f], a heart [əɑ:ʔ]  
 b. an artist [\*əɑ:tɪst], an office [\*əɔfɪs]

In the current model, we can propose that this dialect has a rule of /h/-Deletion that applies after Vocabulary Insertion. Since the /h/ in *half* is still present when Vocabulary Insertion applies to D[-def], *a* is inserted rather than *an*. Later, /h/-Deletion applies, producing the forms in (42)a.

- (43) a. Vocabulary Insertion            /ɛ hɑ:f/  
 b. Tensing / Vowel Reduction       /ə hɑ:f/  
 c. /h/-Deletion                        /ə ɑ:f/

### 4.3 Pause-fillers

Recall that *a/an* is actually a four-way alternation for many speakers, with the full-vowel strong variants /e(j)/ and /æn/ as well as their reduced-vowel weak counterparts /ə/ and /ən/ (rsp.). In §3.2 I argued that the default allomorph of D[-def] is /ɛ/, and that /ɛ/ either becomes tense and diphthongized /ej/ (when stressed) or is reduced to /ə/ (otherwise).

Interestingly, D[-def] surfaces as /ej/ not only in [+stress] contexts like *I want /éj/ book, not two books*, but also in contexts like (44) – with little or no stress, before the vowel-initial pause-fillers *uh/um*, often with no intervening silence. This is also a context where /ði(j)/ is used (Fox Tree & Clark 1997; Clark & Fox Tree 2002).

- (44) a. I'd like /ej/ um... a large coffee and a croissant.  
 b. This is /ej/ uh... part of a trailer truck. (Braunwald ale33)
- (45) And from the-uh /ði(j)ə/ spectator point of view it looks like airplanes going in all directions...We-um have a-uh /ejə/ pyro-techniques team. (Clark & Fox Tree 2002: 103)

If allomorph choice were driven by surface syllable well-formedness, the question would be why we find /ej/, rather than *an*, in contexts like (44)–(45). Put slightly differently, why would *an* be chosen before e.g. *umbrella* but not before *uh/um*?<sup>28</sup>

- (46) a. I'd like {/ej/, ?\*an} um...  
 b. I'd like {\*/ej/, an} umbrella.

In a serialist model like the one assumed here, on the other hand, it is possible to follow the intuition that pause fillers like *uh* and *um* are *structurally* exceptional, with a fundamentally different status in the grammar from words like *umbrella*. Arguably, pause-fillers are not present in the syntax at all, but are inserted post-syntactically during the PF

<sup>28</sup> In the CHILDES corpora examined here (see Appendix), we found one instance of *an um* (uttered by a 3-year-old) and no instances of *an uh*, compared to 38 instances of *a um/uh*. Within the analysis laid out here, it is possible for the occasional instance of *an uh/um* to be derived as follows: a speaker prepares to say e.g. *an elephant*, then suddenly changes their mind just after uttering *an* and replaces *elephant* with a pause-filler. I believe that such scenarios, while possible, are somewhat exceptional, and that the usual situation where speakers insert pause-fillers after determiners is when they have not yet figured out which particular word they want to utter next (and correspondingly insert the default form *a*). This more common scenario is the one assumed in the derivation in (47). Rotenberg (1978: 40–41) makes a similar distinction between planned parentheticals and “last-minute performance effects” (e.g. a cough or exclamation like *Hey, did you see that?!*), the latter of which might intervene between *an* and a vowel-initial complement.

derivation (see Rotenberg 1978; Kaisse 1985 for precedent for this idea). If this proposal is on the right track, it suggests an explanation for the contrast in (46).

Recall that Article Local Dislocation can apply only if there is something right-adjacent to D[±def] for D[±def] to cliticize onto. If pause-fillers are not yet present, this condition will not be met (there will be no statement  $D[\pm\text{def}] \widehat{[X\dots]}$  to provide the necessary input for rule (17)) and D[±def] will remain an independent word, triggering insertion of the elsewhere allomorph.

- (47) Derivation of *I'd like /ej/ um...* (DP cycle):
- a. *Article Local Dislocation* (11): NA; nothing follows D[-def] at this stage
  - b. *Vocabulary Insertion*: D[-def] ↔ æn /\_V  
↔ ε elsewhere  
(/ε/ inserted here because nothing follows D[-def] word-internally)
  - c. *Tensing* (15)a: ε → e (because nothing follows D[-def] word-internally)
  - d. *Vowel Reduction* (15)b: NA because /e/ is [+tense]

For the purposes of this analysis, the pause-filler could be inserted at any point in PF after step (a) Article Local Dislocation. Once Article Local Dislocation fails, nothing outside of D[±def] can be visible for the word-bounded rules of Vocabulary Insertion, Tensing, or Vowel Reduction, so D[-def] will surface as /ej/ (after Diphthongization). The larger question of when and how exactly pause-fillers are inserted in PF – if, for example, different kinds of pause-fillers might be inserted at different points or by different mechanisms – remains open for future investigation.

The derivation of /ði/ *uh/um* in the definite article works exactly as in (47), except that there is no allomorphy in step (b):<sup>29</sup>

- (48) Derivation of *I'd like /ði/ um...* (DP cycle):
- a. *Article Local Dislocation* (11):  
NA; nothing follows D[+def] at this stage
  - b. *Vocabulary Insertion*: D[+def] ↔ ðɪ
  - c. *Tensing* (15)a: ɪ → i (nothing follows D[+def] word-internally)
  - d. *Vowel Reduction* (15)b: NA because /i/ is [+tense]

It is important to recognize that like emphatic glottal stops, pause-fillers are not completely “outside the grammar” even though they are inserted late. Whether or not they count as *bona fide* words in other respects (see Clark & Fox Tree 2002 for discussion), *uh* and *um* are clearly visible for the apparently late-stage phonological rule of Flapping:

- (49) Bu/ɾ/ *uh* ... we think tha/ɾ/ *uh* ...

Again, the fact that pause-fillers are invisible for some rules (allomorphy) but visible for others (Flapping) is taken to be a rule-ordering effect: pause-fillers are inserted after Article Local Dislocation but before Flapping. If *a/an* were determined by surface syllable well-formedness constraints, however, we would not expect *a/an* to behave any differently from Flapping with respect to whether its alternating segment could syllabify as an onset onto *uh/um*.

<sup>29</sup> As a reviewer points out, both /ði/ and /ðə/ are possible in e.g. *I'd like the, um, I don't know...* In our CHILDES corpus study, we found that *the* was pronounced as /ði/ 75% of the time when it was immediately followed by a pause-filler *uh* or *um* (69 out of 92 adult utterances of *the uh/um* from the Braunwald, Ervin-Tripp, MacWhinney, Nelson, Providence, Sawyer, and Snow corpora). Josef Fruehwald (p.c.) reports 86% /ði/ in utterances of *the uh/um* from the Philadelphia Neighborhood Corpus. To explain why /ðə/ is sometimes used before *uh/um*, I propose in §5 that speakers may have access to two competing grammars – one with the ði/ðə alternation and one with invariant /ðə/. At this time I do not have an explanation for why our CHILDES study adults used the ði/ðə grammar less frequently before pause-fillers than in other prevocalic contexts (75% vs. 90%,  $p < .001$ ) – although the ði/ðə grammar was clearly the preferred grammar in both contexts.

#### 4.4 Interim discussion

What I have shown in the preceding subsections is that neither *a/an* nor *the* can be analyzed as allomorphy as TETU,<sup>30</sup> at least not for varieties of English where any of the phenomena discussed here (emphatic glottal stop, /h/-dropping, pause-filler insertion) apply as described. While *a/an* plays an indirect role in creating many optimally syllabified strings (V.CV), it can also contribute segments that are ultimately syllabified in a non-optimal way (VC.CV). In the current proposal, this is because Vocabulary Insertion operates on whatever information is available early in PF; later phonological processes may then add, delete, or modify segments.

As pointed out by a reviewer, the theory of allomorphy as TETU is not necessarily threatened just because it turns out to be inappropriate for English *a/an*. Mascaró (2007) does not claim that *all* phonologically conditioned allomorphy (PCA) is optimizing; rather, he distinguishes optimizing (externally conditioned, regular) PCA from non-optimizing (internally conditioned, lexical, arbitrary) PCA (see also Bonet et al. 2007).<sup>31</sup> It could be that English *a/an*, despite initial appearances and *contra* Mascaró 1996b, is an example of arbitrary PCA and thus would not be expected to yield to a TETU analysis.

However, earlier in the paper I showed that *a/an* is best analyzed in a serialist architecture where allomorphy precedes phonology; this approach allows us to account for all four variants (/ej/, /æn/, /ə/, /ən/) in a way that captures the parallels between the strong/weak forms here and in other function words, including *the*. As noted at the beginning of §4, it is impossible in this model for allomorphy to have access to surface phonology, since Vocabulary Insertion strictly precedes phonology. To introduce a new kind of late-stage, post-phonological, surface-sensitive allomorphy into this model would represent a significant addition. Is this addition necessary? Or, more generally: Do we really need two kinds of allomorphy, or can we get by with just one?

The main argument for TETU-based allomorphy that appears in the literature is that it captures an important cross-linguistic generalization: “[T]he linguistic generalization that the allomorph is chosen because it yields an unmarked structure should be incorporated into grammatical theory, since it rests on an extensive empirical base” (Mascaró 2007: 716). A key question for future research is whether the many proposed cases of optimizing allomorphy from the literature, including those in (30), really are consistently optimizing *on the surface*. Surface-optimization could be established by testing whether the given alternation interacts with pause-fillers and the output of late-stage phrasal phonology in the expected way (as demonstrated in §§4.1–4.3).<sup>32</sup> We have seen that English *a/an* is not surface-optimizing, even though at first sight it seems like a well-behaved textbook case of allomorphy as TETU.

TETU-based proposals make a strong prediction that the given alternation should “see” exactly what is there on the surface, including pause-fillers and epenthetic segments, much like American English Flapping. What I hope to have shown is that *apparent* phonological

<sup>30</sup> The data in this section present a problem for any account of *a/an* that requires reference to surface syllable well-formedness, independent of whether *a/an* is treated allomorphically or phonologically. For example, Yang’s (2004) modified analysis of *a/an*, where there is no allomorphy but rather a single exponent  $a < n >$  with a ‘ghost’ /n/ that is realized iff it surfaces as a syllable onset, runs afoul of the same problems as Mascaró (1996b).

<sup>31</sup> Two examples of arbitrary PCA from the literature are the Tzeltal perfective (-eh after polysyllabic stems; -oh after monosyllabic stems) (Mascaró 2007: 715–716) and the Turkish causative (-t after polysyllabic stems ending in V, l, r; -dir elsewhere) (Bonet et al. 2007: 904).

<sup>32</sup> For some case studies there may be no available evidence of this kind, due to the phonological shape and position of the alternating segment and its surface-adjacent material, or due to the absence of phenomena like emphatic glottal stop or Flapping that could be used as diagnostics in the given language. I would view these cases as ambiguous – i.e., not clearly supporting an argument for or against allomorphy as TETU.

optimization does not always correspond to *surface* phonological optimization across contexts.

## 5 Accounting for variation

As noted earlier, neither *a/an* nor *the* is a categorical alternation. The goal of this section is to give a slightly fuller picture of the inter- and intraspeaker variation with *a/an* and *the*, and to show how the analysis presented in §3 can be incorporated into a competing-grammars approach to account for this variation. I do not make any internal changes to the analysis from §3 here.

Prevocalic *a* is common in many varieties of British and American English; in fact, some of these varieties have invariant *a*, prevocalically and preconsonantly (see Gabrielatos et al. 2010 and references cited there). Prevocalic /ðə/ is also a feature of at least some of these dialects (see e.g. Britain & Fox 2009). In fact, prevocalic /ðə/ occurs even in varieties that generally do not have prevocalic *a*, e.g. “standard” American English (Healy et al. 1998; Raymond et al. 2002), and appears to be becoming more common in younger generations in some regions (Todaka 1992; Keating et al. 1994: 136–138).

Prevocalic *a* and /ðə/ are also a well-known feature of children’s speech. As shown earlier (Table 1, §2), children in our CHILDES corpus study use the “standard” prevocalic forms *an* and /ði/ far less frequently than their adult caregivers. Table 2, which includes data from more corpora and breaks down the children’s data by age,<sup>33</sup> shows that children’s use of prevocalic *an* and /ði/ does not reach even 65% frequency until age 6 (see also Newton & Wells 1999).

	% <i>an</i>	<i>an</i> /( <i>a+an</i> )	% ði	ði/(ðə+ði)
age 3	30%	(166/561)	41%	(160/388)
age 4	22%	(73/326)	38%	(109/289)
age 5	36%	(32/90)	61%	(115/187)
age 6–7	67%	(59/88)	77%	(94/122)
age 8–9	74%	(14/19)	(no audio data)	
age 10–11	95%	(42/44)	(no audio data)	
adults	95%	(2883/3019)	90%	(773/863)

**Table 2:** Frequency of prevocalic *an* and /ði/ in North American children and adults in CHILDES.<sup>34</sup>

Examples of prevocalic *a* and /ðə/ in a 5-year-old’s speech are given in (50). Notice that a /ʔ/ is inserted after prevocalic *a*/ðə. This /ʔ/ is also a feature of adult speech (Todaka 1992; Britain & Fox 2009); in our corpus study, for example, 80% (72/90) of adults’ prevocalic /ðə/ had a /ʔ/ between /ðə/ and the following vowel.

- (50) a. Pretend this was a [əʔ] *elevator*. (Sawyer 2-26-92)  
 b. if you don’t want me to take *the* [ðəʔ] *elephant* (Sawyer 2-28-92)

The analysis laid out in §3 can be adapted to account for inter- and intraspeaker variation in *a/an* and *the*. To see how this might work, consider the mini-grammars in (51)–(52).

<sup>33</sup> Since the /ði~/~ðə/ distinction is not reflected in transcriptions, it can only be observed by listening to audio recordings. Table 1 in §2 includes data only from corpora with audio recordings, allowing for a more direct comparison of *a/an* to *the*. Table 2, in contrast, combines data from corpora with and without audio recordings, meaning that more speakers are represented under the *a/an* columns than under the *the* columns (e.g. 306 children for *a/an* vs. 48 children for *the*).

<sup>34</sup> See Appendix for a full list of corpora included in Table 2; see MacWhinney (2000) for more information about individual corpora.



It is likely that many speakers also have a third “intermediate” grammar: one that has the basic /n/ ~ Ø alternation but lacks the full-vowel forms /e/ and /æɪn/.

- (55) Grammar INDEF3 (/ə/ bóok, /ən/ ápple, /ə/ book, /əɪn/ apple)
- a. Article Local Dislocation
  - b. Vocab. insertion: D[-def] ↔ ən / \_V  
↔ ə elsewhere

As with the definite article, I assume that children initially favor the simple grammar that inserts /ə/ categorically (INDEF2). Over time, they increase their use of INDEF3 (with allomorphy) and/or INDEF1 (with allomorphy and Tensing/Vowel Reduction) until they achieve the adult pattern for their given variety of English. Some adult speakers may alternate between INDEF2 and INDEF3, some between INDEF1 and INDEF3, and some among all three grammars. Again, it is possible that children (and adults) use additional grammars beyond those sketched here.

Among other things, this approach explains why there is intraspeaker variability in the pronunciation of pitch-accented articles:

- (56) a. This is {ðí/ðə} book to read on global warming.  
b. I said I wanted {éj/ə} croissant, not two croissants.

When the full-vowel form is chosen, the speaker is using grammar (IN)DEF1. When the /ə/ form is chosen, the speaker is using grammar DEF2, INDEF2 or INDEF3.

## 6 Concluding thoughts

The question posed in the title of this paper is “How allomorphic is English article allomorphy?” I have answered this question as follows:

- i. *the* is *not* allomorphic. It is derived by the same phonological rules – Tensing and Vowel Reduction – as other V ~ ə alternations in English, e.g. *beauty* ~ *beautiful*, /kæɪn/ ~ /kən/ in the function word *can*.
- ii. *A/an* is *partly* allomorphic. For speakers with the strong forms /e/ and /æɪn/ as well as /ə/ and /ən/, *a/an* is best analyzed as involving both allomorphy and phonology: first allomorphy establishes a basic split between /æɪn/ (before vowels) and /e/ (elsewhere); then the *same phonological rules involved in the* – Tensing and Vowel Reduction – derive the variants /ə/ and /ən/ in their designated contexts.

Recall that children do not reach adultlike patterns with *a/an* and *the* until age 6 or later (Table 2; Newton & Wells 1999). In this respect, *a/an* and *the* are very different from some of the other “between-word processes” that have been examined in the acquisition literature, e.g. cluster simplification and assimilation (Newton & Wells 1999):

- (57) a. Cluster simplification: just like [ɔ̃ɪslak]  
b. Assimilation: one cloud [wʌŋklaud]

Cluster simplification and assimilation show no clear developmental trend: the 3-year-olds in Newton & Wells (1999) apply them at roughly the same rates as the 7-year-olds. Newton & Wells speculate that *a/an* and *the*, unlike cluster simplification and assimilation, must be gradually learned because they are language-specific and relatively “unnatural” from a phonetic perspective (1999: 74). In a follow-up study, Newton & Wells (2000) look at another between-word process, /r/-liaison in British English (e.g. *saw a* [sɔ̃ɪə]), and show that it is also gradually acquired. This is as expected, since the glide-like behavior of /r/ is a language-specific rather than cross-linguistic phenomenon.

These findings reinforce the point made in §2 that “phonological naturalness” involves *degrees on a scale*. While I have argued for a phonological treatment of *the* on the grounds that *the* is less arbitrary and idiosyncratic than *a/an*, I have not suggested that vowel reduction is an “automatic” or “low-level phonetic” rule like cluster simplification or assimilation. Some languages do not have vowel reduction at all, while others (like English) impose language-specific constraints on vowel reduction (e.g. word-internal, sensitive to tense/lax distinction), so that children have to learn the rule itself as well as figuring out exactly when it applies.

The current model allows for a wide range of types of phonological rules. Some rules tend towards ease of articulation, and it is expected that these will be easier to acquire than less natural ones (all else being equal). Furthermore, phonological rules are structurally restricted depending on when they apply in PF, so that some rules apply cyclically during word-formation (like Tensing) while others apply over entire utterances (like Flapping), and still others apply at various intermediate stages (see Pak 2008). Under this type of approach, we can view both phonetic similarity (Criterion A) and structural restrictedness (Criterion B) as gradient, rather than binary, measures, and we are not necessarily forced to adopt an allomorphic treatment of an alternation just because it is not a “low-level” or “across-the-board” rule.

Another contribution of this paper has been to call attention to the question of what it means for allomorphy to be phonologically optimizing, and how (or whether) phonological optimization should be explained in the grammar. In §4 I used evidence from emphatic glottal stops, /h/-dropping and pause-fillers to show that despite initial appearances, English *a/an* is not always phonologically optimizing on the surface (e.g. *an* /ʔ/ápple, *I want a um...*). In the model I adopted in §3 to analyze *a/an* and *the*, this result is unsurprising: since allomorphy strictly precedes phonology, it is not expected to be able to “see” the surface syllable structure. The question I posed at the end of §4 is whether other proposed cases of allomorphy as TETU really are demonstrably *surface-optimizing*, using diagnostics similar to those I use in §§4.1–4.3.

It has not been my intent to argue for a phonological treatment for every ambiguous case that has been cited in an “allomorphy vs. phonology” debate. I have, however, laid out an analysis of English *a/an* and *the* whose ingredients may be involved in many of these other cases. As we have seen, opening up the possibility for a phonological treatment allowed us to recognize a number of important similarities *and differences* between *a/an* and *the*, which would have gone unexplained in a uniformly allomorphic treatment.

## Supplementary Files

The supplementary files for this article can be found as follows:

- **Supplementary File 1:** Appendix. <http://dx.doi.org/10.5334/gjgl.62.s1>

## Acknowledgements

Thanks to David Embick, Kim Edmunds, Ian Kirby, Chris Naber, Don Tuten, and audiences at the Emory Linguistics Colloquium, the 2014 LSA Annual Meeting, SECOL 81, and the University of Pennsylvania F-MART group for helpful feedback and discussion. Thanks to Kim Edmunds, Ian Kirby, Chris Naber, Greg Tracy and Denton Williams for help with CHILDES corpus data retrieval and coding, and to the Emory Program in Linguistics for support for this project. Finally, thanks to the CHILDES database contributors (see Appendix) for the corpus data presented here. Any errors are of course my own.

## Competing Interests

The author declares that she has no competing interests.

## References

- Asudeh, Ash & Ewan Klein. 2002. Shape conditions and phonological context. In Frank Van Eynde, Lars Hellan & Dorothee Beermann (eds.), *Proceedings of the 8<sup>th</sup> International HSPG Conference*, 20–30. Stanford: CSLI Publications.
- Bermúdez-Otero, Ricardo. 2004. Raising and flapping in Canadian English: Grammar and acquisition. (Handout of paper presented at the CASTL Colloquium, University of Tromsø, 2 November 2004.)
- Bermúdez-Otero, Ricardo. 2007. Word-final prevocalic consonants in English: Representation vs. derivation. (Handout of paper presented at the Old World Conference in Phonology 4, Rhodes, 20 January 2007.)
- Bloomfield, Leonard. 1935. *Language*. London: George Allen & Unwin.
- Bonet, Eulàlia, Maria-Rosa Lloret & Joan Mascaró. 2007. Allomorph selection and lexical preferences: Two case studies. *Lingua* 117. 903–927. DOI: <http://dx.doi.org/10.1016/j.lingua.2006.04.009>
- Boroff, Marianne L. 2007. *A Landmark Underspecification account of the patterning of glottal stop*. Stony Brook, NY: Stony Brook University dissertation.
- Britain, David & Sue Fox. 2009. The regularisation of hiatus resolution in British English. In Markku Filppula, Juhani Klemola & Heli Paulasto (eds.), *Vernacular universals and language contacts*, 177–205. New York: Routledge.
- Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.
- Clark, Herbert H. & Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84. 73–111. DOI: [http://dx.doi.org/10.1016/S0010-0277\(02\)00017-3](http://dx.doi.org/10.1016/S0010-0277(02)00017-3)
- Crisma, Paola. 2009. Word-initial h- in Middle and Early Modern English. In Donka Minkova (ed.), *Phonological weakness in English: From Old to Present-Day English*, 130–167. Palgrave Macmillan.
- Crosswhite, Katherine. 2004. Vowel reduction. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 191–231. Cambridge: Cambridge University Press. DOI: <http://dx.doi.org/10.1017/CBO9780511486401.007>
- Elordieta, Gorka. 2008. An overview of theories of the syntax-phonology interface. *Journal of Basque Linguistics and Philology* 42. 209–286.
- Embick, David. 2007. Linearization and Local Dislocation: Derivational mechanics and interactions. *Linguistic Analysis* 33. 303–336.
- Embick, David. 2008. Variation and morphosyntactic theory: Competition fractionated. *Language and Linguistics Compass* 2(1). 59–78. DOI: <http://dx.doi.org/10.1111/j.1749-818X.2007.00038.x>
- Embick, David. 2010. *Localism versus globalism in morphology and phonology*. Cambridge, MA: MIT Press. DOI: <http://dx.doi.org/10.7551/mitpress/9780262014229.001.0001>
- Embick, David & Rolf Noyer. 2001. Movement operations after syntax. *Linguistic Inquiry* 32. 555–595. DOI: <http://dx.doi.org/10.1162/002438901753373005>
- Fox Tree, Jean E. & Herbert H. Clark. 1997. Pronouncing ‘the’ as ‘thee’ to signal problems in speaking. *Cognition* 16. 151–167. DOI: [http://dx.doi.org/10.1016/S0010-0277\(96\)00781-0](http://dx.doi.org/10.1016/S0010-0277(96)00781-0)
- Gabrielatos, Costas, Eivind Nessa Torgersen, Sebastian Hoffmann & Susan Fox. 2010. A corpus-based sociolinguistic study of indefinite article forms in London English. *Journal of English Linguistics* 38(4). 297–334. DOI: <http://dx.doi.org/10.1177/0075424209352729>
- Garellek, Marc. 2012. Glottal stops before word-initial vowels in American English: Distribution and acoustic characteristics. *UCLA Working Papers in Phonetics* 110. 1–23.
- Garellek, Marc. 2013. *Production and perception of glottal stops*. Los Angeles: University of California dissertation.
- Gramley, Stephan. 2012. *The history of English: An introduction*. Routledge.



- Halle, Morris & Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In Kenneth Hale and S. Jay Keyser (eds.), *The view from Building 20*, 111–176. Cambridge, MA: MIT Press.
- Haugen, Jason & Daniel Siddiqi. 2013. On double marking and containment in realization theory. Unpublished manuscript, Oberlin College and Carleton University.
- Hayes, Bruce. 1990. Precompiled phrasal phonology. In Sharon Inkelas & Draga Zec (eds.), *The phonology-syntax connection*, 85–108. Chicago: University of Chicago Press.
- Healy, Alice F., Immanuel Barshl, Robert Crutcher, Liang Tao & Timothy Rickard. 1998. Toward the improvement of training in foreign languages. In Alice F. Healy & Lyle E. Bourne, Jr. (eds.), *Foreign language learning: Psycholinguistic studies on training and retention*, 3–53. Mahwah, NJ: Lawrence Erlbaum.
- Hopper, Paul & Janice Martin. 1985. Structuralism and diachrony: The development of the indefinite article in English. In Anna Giacalone Ramat, Onofrio Carruba & Giuliano Bernini (eds.), *Papers from the 7<sup>th</sup> International Conference on Historical Linguistics*, 295–304. Amsterdam: John Benjamins.
- Hurford, James R. 1972. The diachronic reordering of phonological rules. *Journal of Linguistics* 8. 293–295. DOI: <http://dx.doi.org/10.1017/S0022226700003339>
- Hurford, James R. 1974. The base form of English *a/an*: A reply. *Lingua* 33. 129–135. DOI: [http://dx.doi.org/10.1016/0024-3841\(74\)90031-X](http://dx.doi.org/10.1016/0024-3841(74)90031-X)
- Joseph, Bryan. 1997. On the linguistics of marginality: The centrality of the periphery. *Chicago Linguistic Society* 33. 197–213.
- Jurafsky, Daniel, Alan Bell, Eric Fosler-Lussier, Cynthia Girand & William Raymond. 1998. Reduction of English function words in Switchboard. In Robert H. Mannell & Jordi Robert-Ribes (eds.), *Proceedings of the 5<sup>th</sup> International Conference on Spoken Language Processing*, vol. 7, 3111–3114. Canberra City: Australian Speech Science and Technology Association, Inc.
- Kaisse, Ellen M. 1985. *Connected speech: The interaction of syntax and phonology*. Orlando: Academic Press.
- Keating, Patricia A., Dani Byrd, Edward Flemming & Yuichi Todaka. 1994. Phonetic analysis of word and segment variation using the TIMIT corpus of American English. *Speech Communication* 14. 131–142. DOI: [http://dx.doi.org/10.1016/0167-6393\(94\)90004-3](http://dx.doi.org/10.1016/0167-6393(94)90004-3)
- Kroch, Anthony. 1994. Morphosyntactic variation. *Chicago Linguistic Society* 30. 180–201.
- Ladefoged, Peter. 1975. *A course in phonetics*. New York: Harcourt Grace Jovanovich Inc.
- Lee, Yongsung. 2009. Allomorphy selection: Universal and morpheme-specific constraints. In Young-Se Kang (ed.), *Current issues in linguistic interfaces*, 417–434. Seoul: Hankook Munhwasa.
- Levelt, Clara. 2008. Phonology and phonetics in the development of schwa in Dutch child language. *Lingua* 118. 1344–1361. DOI: <http://dx.doi.org/10.1016/j.lingua.2007.09.010>
- Mackenzie, Laurel. 2012. *Locating variation above the phonology*. Philadelphia, PA: University of Pennsylvania dissertation.
- MacWhinney, Brian. 2000. *The CHILDES Project: Tools for analyzing talk*, 3<sup>rd</sup> edition. Mahwah, NJ: Lawrence Erlbaum Associates.
- Marvin, Tatjana. 2002. *Topics in the stress and syntax of words*. Cambridge, MA: MIT dissertation.
- Mascaró, Joan. 1996a. External allomorphy and contractions in Romance. *Probus* 8. 181–205. DOI: <http://dx.doi.org/10.1515/prbs.1996.8.2.181>
- Mascaró, Joan. 1996b. External allomorphy as emergence of the unmarked. In Jacques Durand & Bernard Laks (eds.), *Current trends in phonology: Models and methods*, 473–483. Salford, Manchester: European Studies Research Institute, University of Salford.

- Mascaró, Joan. 2007. External allomorphy and lexical representation. *Linguistic Inquiry* 38. 715–735. DOI: <http://dx.doi.org/10.1162/ling.2007.38.4.715>
- Nádasdy, Ádám. 2013. 'PRESUME-tensing' and the status of weak /i/ in RP. In Péter Szigetvári (ed.), *Papers in linguistics presented to László Varga on his 70<sup>th</sup> birthday*, 363–382. Budapest: Tinta Publishing House.
- Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.
- Nevins, Andrew. 2011. Phonologically conditioned allomorph selection. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, 2357–2382. Chichester: Wiley-Blackwell.
- Newton, Caroline & Bill Wells. 1999. The development of between-word processes in the connected speech of children aged between three and seven years. In Ben Maassen & Paul Groenen (eds.), *Pathologies of speech and language: Advances in clinical phonetics and linguistics*, 67–75. London: Whurr Publishers Inc.
- Pak, Marjorie. 2008. *The postsyntactic derivation and its phonological reflexes*. Philadelphia, PA: University of Pennsylvania dissertation.
- Perlmutter, David. 1970. On the article in English. In Manfred Bierwisch & Karl Erich Heidolph (eds.), *Progress in linguistics*, 233–248. The Hague: Mouton. DOI: <http://dx.doi.org/10.1515/9783111350219.233>
- Raymond, William D., Julia A. Fisher & Alice F. Healy. 2002. Linguistic knowledge and language performance in English article variant performance. *Language and Cognitive Processes* 17. 613–662. DOI: <http://dx.doi.org/10.1080/01690960143000380>
- Rhodes, Richard A. 1996. English reduced vowels and the nature of natural processes. In Bernard Hurch & Richard A. Rhodes (eds.), *Natural phonology: The state of the art*, 239–259. Berlin: Mouton de Gruyter. DOI: <http://dx.doi.org/10.1515/9783110908992.239>
- Rotenberg, Joel. 1978. *The syntax of phonology*. Cambridge, MA: MIT dissertation.
- Selkirk, Elisabeth. 1995. The prosodic structure of function words. In James A. Morgan & Katherine Demuth (eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 187–214. New York: Lawrence Erlbaum.
- Spencer, Andrew. 1991. *Morphological theory: An introduction to word structure in generative grammar*. Oxford: Blackwell.
- Todaka, Yuichi. 1992. Phonetic variants of the determiner 'the.' *UCLA Working Papers in Phonetics* 81. 39–47.
- Tranel, Bernard. 1990. On suppletion and French liaison. *Probus* 2.2. 169–208. DOI: <http://dx.doi.org/10.1515/prbs.1990.2.2.169>
- Vennemann, Theo. 1974. Restructuring. *Lingua* 33. 137–156. DOI: [http://dx.doi.org/10.1016/0024-3841\(74\)90032-1](http://dx.doi.org/10.1016/0024-3841(74)90032-1)
- Wells, John C. 1982. *Accents of English*. Cambridge University Press. DOI: <http://dx.doi.org/10.1017/CBO9780511611759>
- Wells, John C. 2008. *Longman pronunciation dictionary*, 3<sup>rd</sup> edition. Harlow: Pearson Education.
- Yang, So-Young. 2004. Latent segments in the English indefinite article. *Language and Information Society* 6. 68–83.
- Zwicky, Arnold. 1986. The general case: Basic form versus default form. *Proceedings of BLS* 12. 305–314. DOI: <http://dx.doi.org/10.3765/bls.v12i0.1875>

**How to cite this article:** Pak, Marjorie. 2016. How allomorphic is English article allomorphy? *Glossa: a journal of general linguistics* 1(1): 20. 1–27, DOI: <http://dx.doi.org/10.5334/gjgl.62>

**Published:** 15 July 2016

**Copyright:** © 2016 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

]u[ *Glossa: a journal of general linguistics* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 