How allomorphic is English article allomorphy?

Marjorie Pak
mgpak@emory.edu
Emory University

July 11, 2014

1.   The *a/an* paradox

English *a/an* is at first sight a very simple alternation: use *an* before a vowel, otherwise *a* (1). But is this alternation **phonological** (2a) or **allomorphic** (2b)?

1) **an** apple, **an** interesting book   *vs.*   **a** book, **a** very red apple

2) a. PHONOLOGICAL: $\emptyset \rightarrow$ n / ə __#V (*n*-insertion) *or* n $\rightarrow \emptyset$ / ə __#C (*n*-elision)
        (Hurford 1972/1974; Perlmutter 1970; Venneman 1974)

    b. ALLOMORPHIC: D[-def] $\leftrightarrow$ an / __⌢V...
                            $\leftrightarrow$ a / elsewhere
        (Hayes 1990, Kaisse 1985, Lee 2009, Mascaró 1996b, Nevins 2011)

The obvious problem with the phonological approach is its lexical restrictedness: *a/an* is the only context where /n/ alternates with $\emptyset$ in English (3). So it seems simpler just to store *a* and *an* as suppletive allomorphs than to posit such a highly specialized phonological rule.

3) a. No general *n*-insertion: * my idea**n** is, *Santa**n** and Rudolph
    b. No general *n*-elision: * fountai(**Ø**) pen, * ope(**Ø**) door

On the other hand, <u>allomorphy is typically a word-internal phenomenon, but *a/an* needs to 'see' across the word boundary</u>. Once we admit **phrasal** or **external allomorphy** into the grammar, we need to ask: What are its implications and restrictions?

According to Mascaró (1996a, 1996b, 2007), external allomorphy provides a special context where **phonologically optimizing** effects may emerge (TETU). In his OT-based analysis, both *a* and *an* are listed as (suppletive) allomorphs of D[-def]; since they are equally faithful, the choice between them is determined by hiatus-avoiding markedness constraints (Mascaró 1996b: 517):

4)

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

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

5) Other proposed cases of allomorphy as TETU   (Mascaró 1996a, 1996b, 2007; Lee 2009)
   a. French **bo** __C *vs.* **bɛl** __V (*beau* mari, *bel* enfant 'good-looking husband/child')
      (also *nouveau/nouvel* 'new', *ce/cet* 'this', *ma/mon* 'my', *vieux/vieil* 'old', etc.)
   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* ʎiβre, *ést* ɔme 'this book/man')
   e. Moroccan Arabic C__ **-u** *vs.* V__ **-h** (ktab-*u*, x^{tˤ}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** (argi-*tu* 'clear up', ilun-*du* 'darken') (also *dar/tar, ko/go, tik/dik*)

Here I look at *a/an* alongside a strikingly similar but less widely recognized phenomenon – the alternation between /ði(j)/ and /ðə/ in the English definite article (henceforth *THE*). The distribution of the alternants is almost identical to that found with *a/an*: use /ði(j)/ if the immediately following word starts with a vowel; otherwise use /ðə/ (Ladefoged 1975:91-92).

6) a. /ðij/ **a**pple, /ðij/ **i**nteresting book
   b. /ðə/ **b**ook, /ðə/ **v**ery red apple

Since tense vowels are diphthongized in English (Chomsky & Halle 1968:183ff), THE could be easily incorporated into Mascaró's framework, and might even be viewed as an additional source of support for the idea that external allomorphy is always phonologically optimizing.

7) TETU analysis of THE *(to be rejected)*

{ ðə, ðij} **book**	ONSET	NO-CODA
☞ ðə.book		*
ðij.book		**!

{ ðə, ðij} **egg**	ONSET	NO-CODA
ðə.egg	*!	*
☞ ði.j egg		*

Nevertheless, **I argue that THE is phonological, not allomorphic, in nature**.

8) PROPOSAL:
   /ðə/ is derived from /ði/ by a structurally restricted phonological rule (vowel-reduction).

➢ I adopt a serialist architecture with allomorphy preceding various kinds of ordered phonological rules, including morphosyntactically restricted phonological rules.
➢ I show that this model provides a more complete account of the distribution of *a/an* and THE, including the fact that their surface forms are *not* always phonologically optimizing.

2.   Criteria for distinguishing allomorphy from phonology

Allomorphy is generally assumed to be *suppletive* in the sense that one alternant completely replaces the other in its designated contexts. Suppletive allomorphy is most clearly at work when the alternants in question have very different pronunciations. Consider Korean *-i*/*ka* (5f):

9) a. kae-<u>ka</u> 'dog-NOM', so-<u>ka</u> 'cow-NOM'

b. chaek-<u>i</u> 'book-NOM', sok-<u>i</u> 'inside-NOM'

Since no plausible phonological rule(s) could derive *-i* from *-ka* or vice versa, it is assumed instead that *-i* and *-ka* are (suppletive) allomorphs of the [NOMINATIVE] morpheme.

So in deciding whether an alternation is phonological or allomorphic, we consider (*inter alia*):

10) ALLOMORPHY OR PHONOLOGY?

Criteria	*Allomorphic*	*Phonological*
A. Degree of phonological resemblance	No resemblance (e.g. Korean *i/ka*)	Very close resemblance (e.g. English flapping: t/d/ɾ)
B. Degree of lexical/ structural restrictedness	Restricted to one morpheme (e.g. Korean *i/ka*)	Potentially across the board (e.g. Eng. flapping: *a<u>t</u>om, a<u>t</u> 'em,...*)

These are both **gradient** criteria, so the clearest cases will be at the extremes. Korean *i/ka* is clearly allomorphic because (i) *i* and *ka* are so dissimilar that neither can be plausibly derived from the other phonologically; and (ii) the alternation is restricted to a single morpheme.

What about THE?

- Criterion A: /ði/ and /ðə/ are phonologically identical except that one has a full vowel /i/ where the other has /ə/.

- Criterion B: V~ə alternations occur in many contexts in English, e.g.:

11) a. *Stylistic variation:* b<u>e</u>lieve, b<u>e</u>have, r<u>e</u>lax, <u>e</u>mergency, <u>e</u>raser, <u>e</u>normous, <u>e</u>leven
   b. *Word-formation*: beaut**/i/** ~ beaut/ə/ful, expl/**e**/n ~ expl/ə/nation, s**o**lid ~ s/ə/lidity
   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/**ɔ́**/r it, not against it. ~ I voted f/ə/r Jóhn.
      iii. John wrote th/**í**/ paper on parentheticals. ~ John wrote th/ə**/** páper**.**

THE can be phonologically derived by a relatively natural rule – one of *unstressed-vowel reduction*, a cross-linguistically well-precedented phenomenon (Crosswhite 2004).

➤ While *a/an* and THE have many properties in common, <u>THE is not idiosyncratic or arbitrary to the same degree as *a/an*</u>. If we treated THE as suppletive allomorphy, the phonological similarity between /**ði**/ and /**ðə**/ would have to be viewed as **accidental**, as would its parallels to other V~ə alternations.

Our CHILDES corpus study shows that:
(MacWhinney 2000; see Appendix)

i) Children don't use prevocalic *an*
    or /ði/ consistently until after age 5.
    (Until they do, they generally insert a
    glottal stop between /ə/ or /ðə/ and the
    following word, e.g. *a ʔelevator*.)

ii) Adults are less consistent with THE
    than with *a/an* (an adult who never
    says *a apple* may sometimes say
    /ðə/ *apple*) (*p* < .0001).

iii) Despite this less-consistent input,
     children acquire THE slightly *earlier*
     than *a/an*.[1]

Conceivably, children are noticing
parallels between THE and other V~ə
alternations in English and are therefore
able to acquire THE more easily than the
unprecedented and arbitrary *a/an*.

Figure 1. Frequency of *an* (vs. *a*) and *thi* (vs. *thə*) in pre-vocalic contexts
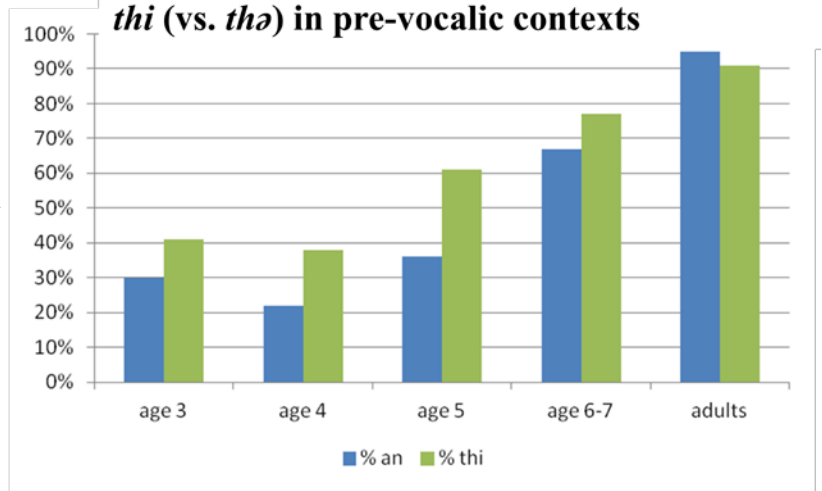
	% an	AN / A+AN	*% ði*	ði / ðə+ði
age 3	**30%**	(166/561)	**41%**	(160/388)
age 4	**22%**	(73/326)	**38%**	(109/289)
age 5	**36%**	(32/90)	**61%**	(115/187)
age 6-7	**67%**	(59/88)	**77%**	(94/122)
age 8-9	**74%**	(14/19)		
age 10-11	**95%**	(42/44)		
adults	**95%**	(2883/3019)	**90%**	(773/863)

12) JO (5;2): if you don't want me to take th[ə **ʔ**]elephant, let me have this one.
    E (3;6):   I want you to take th[ə **ʔ**]el[**o**]phant. This elephant.
    JO: I don't want th[**i**] elephant. I wanna have the horse.   (Sawyer 2-28-92)

3. Analysis

In laying out a phonological analysis of THE, the challenge is to:

➢ represent the parallels between ði~ðə, *beau<u>t</u>y~beautiful*, etc. as **non-accidental** – e.g. by
  deriving all these alternations from a single vowel-reduction rule (V[-stress] → ə /__C)
➢ but avoid an analysis that overgenerates. <u>THE is an exceptional case: [-low] vowels *don't*
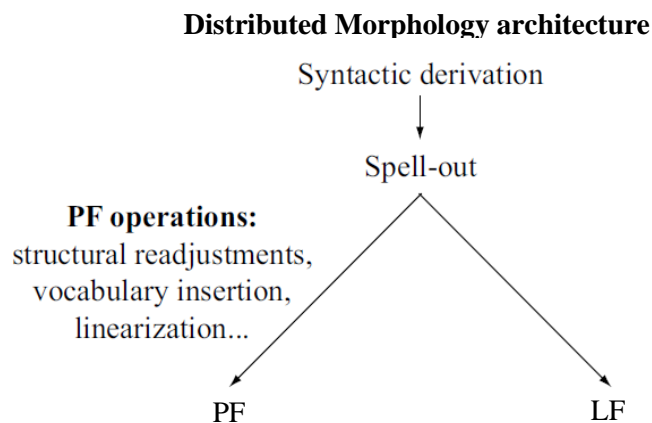 usually reduce word-finally</u> (Chomsky & Halle 1968:111):

13) *carr/ə/ me, *craz/ə/ kids, *beaut/ə/ within

Informally, <u>*the* behaves like 'part of the following word' for the purposes of vowel reduction.</u>

---

[1] Newton & Wells 1999 report similar results in a study of spontaneous speech, sentence repetitions and
   story repetitions by 94 British 3- to 7-year-olds. The frequency of adult-like *an* lagged behind the
   frequency of adult-like /ði/ by approximately 5-15 percentage points at every age (e.g. 27% vs. 35% at
   age 3, 80% vs. 89% at age 7).

4

BACKGROUND ASSUMPTIONS:

i) A serialist model of PF in which (morpho-) syntactic structures are converted to phonetic strings by a series of PF operations, e.g.:
  – *Linearization* (word-internal and phrasal)
  – *Vocabulary insertion* (exponence of functional heads, including allomorphy)
  – Limited structural *readjustments* (e.g. 'cliticization' or local dislocation)
  – *Phonological rules* of various kinds

Distributed Morphology architecture

Syntactic derivation

↓

Spell-out

PF operations:
structural readjustments,
vocabulary insertion,
linearization...

PF   LF

ii) Syntactic structures – including internally complex words – are spelled out in **phases** instead of all at once (Chomsky 2000 et seq., Marvin 2002, Embick 2010).

iii) <u>Phonological rules apply at different stages in PF and thus have access to different kinds of information</u> (Kaisse 1985, Seidl 1999, etc.). Specifically, phonological rules apply as phases of increasing size are spelled out and linearized (Pak 2008).

INGREDIENTS:

To explain the affix-like behavior of *the*, I propose that English has an *Article Cliticization* operation that adjoins D[±def] to the following word (a type of Local Dislocation (Embick & Noyer 2001)). This is similar to the Article Cliticization rule that produces the contracted *l'* form of the French definite article, as argued in Embick 2003:328ff, 2010:87ff.

14) English Article Cliticization: D[±def] ⌢ [X...] → [D[±def] [X...]]

15) Effects of Article Cliticization in French:
  a. l'arbre 'the tree' (*<u>le</u> arbre), l'école 'the school' (*<u>la</u> école)
  b. *cf.* <u>le</u> chien 'the.MASC dog', <u>la</u> fille 'the.FEM girl'

At Vocabulary Insertion, D[-def] is spelled out as /ði/ (16). The alternation between /ði/ and /ðə/ is then produced by a phonological **vowel reduction** rule (VR), which is strictly **word-internal**.

16) Vocabulary Insertion: D[-def] ↔ ði

17) Vowel Reduction (VR):   V[-stress] → ə /__C   *(to be revised)*

SAMPLE DERIVATIONS:

First consider the individual words **beauty** and **beautiful**. *Beauty* is composed of the root √BEAUTY plus a null category-defining *n* head, while *beautiful* is composed of the root √BEAUTY attached to the adjectival suffix *-ful*:

18) a. [$_n$ √BEAUTY [$_n$ Ø]]
   b. [$_a$ √BEAUTY [$_a$ -ful]]

After each of these words is spelled out, it is subjected to the Vowel-Reduction (VR) rule in (17). VR fails to apply in *beauty* because its context is not met; there is no unstressed vowel followed by a consonant *within* that word. In *beautiful*, however, the context for VR is met – the unstressed /i/ in the root √BEAUTY is followed by the /f/ in *-ful*, so it can reduce to schwa.

Now let's consider the DP **the crazy kid**. Why does the /i/ get reduced in *the* but not in *crazy*?

1. Within the nP...

   i) After each M-word ([$_a$ √CRAZY [$_a$ Ø]] and [$_n$ √KID [$_n$ Ø]]) has been spelled out, it is subjected to the Vowel-Reduction rule in (17). <u>But VR has no effect because it is strictly word-internal</u>: its context ('__C') is never met *within* either word.

   ii) The M-words [$_a$ √CRAZY [$_a$ Ø]] and [$_n$ √KID [$_n$ Ø]] are linearized (concatenated): [$_a$ √CRAZY [$_a$ Ø]] ⌢ [$_n$ √KID [$_n$ Ø]]
   Now the /i/ in *crazy* is followed by a consonant, but VR has already applied and cannot have any effects at this phrasal level.
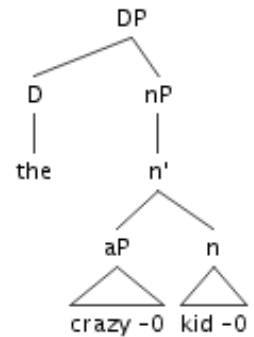
2. At the next cycle, the DP layer of structure is introduced.

   i) *Concatenation*: D[+def] is concatenated with the following material: D[+def] ⌢ [$_a$ √CRAZY [$_a$ Ø]] ...

   ii) *Article Cliticization* (see (14)): D[+def] ⌢ [$_a$ √CRAZY [$_a$ Ø]] → [D[+def] [$_a$ √CRAZY [$_a$ Ø]]]

   iii) *Vocabulary Insertion* (see (16)): D[+def] ↔ **ði**

   iv) *Vowel Reduction:* applies within the M-word [D[+def] [$_a$ √CRAZY [$_a$ Ø]]]; its context is met by the /k/ in *crazy*, so **ði → ðə**.

Now let's address the question: why does the /i/ reduce in **beautiful** but not in **happiness**?

− In the spirit of Marvin 2002, I take this 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)b),[2] while *happiness* is spelled out in two: first the root combines with a null adjectival head, then this adjective combines with the suffix *-ness*:

---

[2] 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

6

19) *Phase 1:* [$_a$ √HAPPY [$_a$ Ø]]
  Phase 2: [$_n$ [$_a$ √HAPPY [$_a$ Ø]] [$_n$ -ness]]

− So we could try solving the *beautiful/happiness* problem by making our Vowel-Reduction rule *cyclic*. The idea would be that VR doesn't apply in *happiness* because /i/ is not followed by a C *within the same cycle*.

− But the problem with making VR cyclic, as pointed out by Chomsky & Halle (1968:113), is that we don't want VR to apply to inner cycles in words like *sol<u>i</u>d, brut<u>a</u>l, presid<u>e</u>nt* and then have no way to 'recover' the full vowels when stress-shifting affixes are added on later cycles (*sol<u>i</u>d-ify, brut<u>a</u>l-ity, presid<u>e</u>nt-ial*).

− It seems like we need something along the lines of what Chomsky & Halle 1968 propose: an additional feature (±tense) that plays a role in conditioning VR. Then we could have (i) a cyclic tensing rule, preceding (and bleeding) (ii) a non-cyclic VR rule that targets only [-stress] *and* [-tense] vowels:

20) a. Tensing (*cyclic*):   V[-low -stress] → [+tense] / __#
   b. VR (*non-cyclic*):   V[-stress **-tense**] → ə /__C   *(revised)*

21) Derivation of *happiness*:
   Phase 1: [$_a$ √HAPPY [$_a$ Ø]] /**hǽpɪ**/
     Tensing: /hǽpɪ/ → /hǽp**i**/
   Phase 2: [$_n$ [$_a$ √HAPPY [$_a$ Ø]] [$_n$ -ness]] /**hǽpinɛs**/
     Tensing: NA (no V before word boundary)
     VR: /hǽpinɛs/ → /hǽpinəs/ (/i/ is unaffected because it is [+tense])

− My analysis can be modified in this way without any problems that I am aware of. But purely for simplicity of exposition, I will show only VR and leave out the Tensing rule in subsequent examples.

VOWEL-REDUCTION IN *a/an*. One advantage of viewing THE as a phonological alternation that it allows us to understand certain aspects of *a/an* as well.

➢ *A/an* is often implicitly assumed to be a two-way alternation, but many adults actually have four forms, distributed as in (22) (see also Clark & Fox Tree 2002:102, Jurafsky et al. 1998).

➢ Notice that <u>the 'strong' forms /e(j)/ and /æn/ have full vowels where their 'weak' counterparts have /ə/</u>. In other words, *a* and *an* behave exactly like other monosyllabic function words that have full vowels when stressed and /ə/ when stressless (23).

---

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/.

22) Four variants of D[-def]

|	__V	__{C/Ø}
[+stress]	/**æn**/ I want /**æn**/ apple, not two apples.	/**e**(j)/ I want /**e**(j)/ book, not two books.
[-stress]	/**ən**/ I want /ən/ apple.	/**ə**/ I want /ə/ book.

23) i.   You c/**ǽ**/n finish early, but you won't. ~ You c/ə/n dó it.
    ii. I voted f/**ɔ́**/r it, not against it. ~ I voted f/ə/r Jóhn.
    iii. John wrote th/**í**/ paper on parentheticals. ~ John wrote th/ə/ páper**.**

We can use our vowel-reduction rule to derive all the 'weak' forms of *a/an*:

24) Derivation of *a book* (DP cycle) [3]
    i.   *Concatenation:* D[-def] and [*n* √BOOK [*n* Ø]] are linearized: D[-def] ⌐ [*n* √BOOK [*n* Ø]]
    ii. *Article Cliticization:* D[-def] ⌐ [*n* √BOOK [*n* Ø]] → [D[-def] [*n* √BOOK [*n* Ø]]]
    iii. *Vocabulary Insertion:*   D[-def] ↔ æn /__V
                                          ↔ e elsewhere
         (/e/ is inserted here because the following segment is the consonant /b/.)
    iv. *Vowel Reduction:* applies to the material spelled out on the current cycle /e/;
         its context is met by the /b/ in *book*, so **e → ə**

➤ This analysis captures the observation that /ej/ and /ə/ are similar to each other in the same way as /æn/~/ən/ *and* /ði/~/ðə/. Each /ə/ form is derived from its strong counterpart by VR.

➤ At the same time, this analysis captures an important difference between *a/an* and THE: While THE is a two-way alternation that can be attributed to VR alone, *a/an* is a four-way alternation that involves *both* allomorphy *and* VR.[4]

25)

---
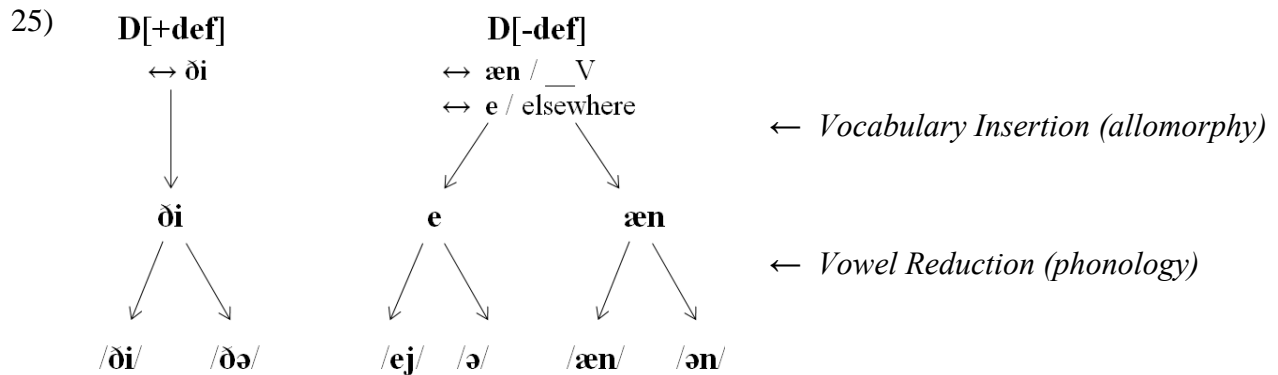
3   Diphthongization applies after step (iv), inserting /j/ after a tense front vowel (ði → ðij, e → ej).
4   The initial split between /e/ and /æn/ appears to be allomorphic rather than phonological in nature by both Criterion A (almost no phonological resemblance between /e/ and /æn/) and Criterion B (restricted to a single morpheme). A phonological analysis is possible, but it would require *two* highly idiosyncratic rules: *n*-insertion (2a) and vowel lowering/laxing (e → æ).

➢ Under a uniformly allomorphic treatment of *a/an* and THE, we would have to assume 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.*

26) Fully suppletive indefinite-article allomorphy *(rejected)*

D[-def] ↔ æn / [+stress] _V
           e / [+stress]
           ən / __V
           ə

D[-def]
/ej/   /ə/   /æn/   /ən/
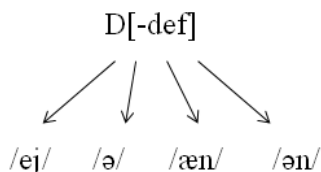
4.   Rule-ordering effects

4.1   Emphatic glottal stops and non-optimal syllables

As noted in §2, when children use *a/*ðə prevocalically, they frequently insert /ʔ/. Adult speakers with variable *a/an* and THE also use /ʔ/ in this context (Gabrielatos et al. 2010, Britain & Fox 2009). /ʔ/ here seems to be a hiatus-breaking mechanism, which is eventually replaced (variably or categorically) by /n/ in D[-def] and /j/ in D[+def].

27) a.   he has [ə ʔ]allergy   (Braunwald 3-05-28b)
      b.   if you don't want me to take [ðə ʔ]elephant   (Sawyer 2-28-92)

However, /ʔ/ has another function in many varieties of English: it is frequently found at the beginning of a prominent (e.g. pitch-accented) V-initial syllable (Garellek 2012).

28) a. He'll fall asleep /ʔ/ánywhere.
      b. I haven't seen John in for/ʔ/éver.

Notably, this '**emphatic /ʔ/**' can be used in *non*-hiatus contexts – <u>including after *an* or ði</u>. Keating et al. (1994:137) report /ʔ/ after approx. 30% of prevocalic /ði/ in the TIMIT corpus; and we found /ʔ/ after 21% of adults' prevocalic /ði/ (161/773) in a subset of the CHILDES corpus.

29) a.   That's an /ʔ/éxcellent idea.
      b.   What an /ʔ/ídiot.
30) a.   That was /ði ʔ/óther guy.
      b.   She's got the knobs for /ði ʔ/áir conditioner.   (Braunwald 1-05-10)

Under Mascaró's (1996) analysis of *a/an* as TETU (§1), such utterances are problematic. [nʔ] is not a possible onset in English, so *an ʔídiot* must be syllabified as /an.ʔidiot/. But /an.ʔidiot/

should always be beaten by either /a.ʔidiot/ or /a.n idiot/, which have fewer NO-CODA violations:[5]

31)

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

In the current model, Emphatic /ʔ/ Insertion can be analyzed as a relatively late, optional phrasal phonological rule, applying well after both Vocabulary Insertion and Vowel Reduction.

32) a. Vocabulary insertion   /æn/ idiot   /ði/ idiot
      b. Vowel reduction   /ən/ idiot   /ði/ idiot
      c. Emphatic /ʔ/ Insertion *(optional)*   /ən ʔ/idiot   /ði ʔ/idiot

Since /ʔ/ is not yet present at the stage when Vocabulary Insertion and VR apply, it does not 'count' as a consonant for the purpose of these early word-internal rules.

On the other hand, emphatic /ʔ/ is visible for *later* phrasal phonological rules – e.g. Flapping. In the current model, we can view this contrast as a reflex of rule-ordering: the emphatic /ʔ/ is inserted after *a/an* allomorphy and before Flapping.

33) a. That's Fa[ɾ] Albert.
      b. That's Fa[t̚, *ɾ] ʔÁlbert, not Flat Stanley.

A similar solution can be applied to data from Hurford 1971: older Cockney English speakers who otherwise have categorical pre-V *an* use *a* iff the following word starts with 'dropped' /h/:

34) a. a half [əɑːf], a heart [əɑːʔ]
      b. an artist [*əɑːtɪst], an office [*əɔfɪs]

Suppose this dialect has a phrasal rule of /h/-deletion that applies well after Vocabulary Insertion. Since the /h/ in *half* is still present when Vocabulary Insertion applies, *a* is inserted rather than *an*. Later, /h/ deletion applies, producing the non-optimal forms.

35) a.   Vocabulary insertion   /e hɑːf/
      b.   Vowel reduction   /ə hɑːf/
      c.   /h/ deletion   /ə ɑːf/

---

[5] A similar problem for Korean *i/ka* is described by Lee (2009): *-i* is chosen after roots ending with /ŋ/ (waŋ-i 'king-NOM'), but /ŋ/ is not a possible onset in Korean (💣/wa.ŋi/). To solve this problem, Lee proposes a DEFAULT constraint, which identifies the phonologically simpler form as preferred. This solution will not work for *a/an,* however, because the unexpected form is *an* rather than the simpler *a,* and because there are independent reasons to treat *a* as the default (see Rotenberg 1978).

We have seen that *a/an* and THE, while they may play a role in creating optimal syllables, do not operate on surface phonetic strings in the same way as Flapping. In the current model, this is because Vocabulary Insertion operates on whatever information is available early in PF; later phonological processes may then add, delete, or modify segments.

4.2   Pause-fillers

As further support for a derivational approach, notice that the 'strong' form of the indefinite article /ej/ shows up not only when stressed and pre-consonantal but also before the pause-fillers *uh/um* – without an intervening silence. This is also a context where /**ði**/ is used.

36)   I want /**éj**/ book, not two books.

37)   a.   I'd like /**ej**/ um... a large coffee and a croissant.
       b.   This is /**ej**/ uh… part of a trailer truck.   (Braunwald ale33)

38)   And from the-uh /**ði.jə**/ spectator point of view it looks like airplanes going in all
       directions. (Clark & Fox Tree 2002: 103)

Unlike other V-initial words, *uh* and *um* do **not** trigger insertion of ***an*** in D[-def]. Why not?

39)   a.   I'd like {/ej/, ?*an} um...
       b.   I'd like {*/ej/, an} umbrella.

Suppose that *uh* and *um* are not present in the syntax at all, but are inserted post-syntactically, in PF (see Kaisse 1985, Rotenberg 1978 for precedent for this idea). Crucially, *uh* and *um* are not present at the stage when vocabulary insertion applies to D[-def], so the 'elsewhere' /e/ is selected.[6]

40)   Derivation of *I'd like /ej/ um*...
       a.   Linearization of DP: **D[-def]**   *(nothing follows D[-def] at this stage)*
       b.   Vocab insertion: **D[-def]** ↔ **e** (__V *context for **æn** isn't met; so **e** is inserted)*
       c.   Pause-filler insertion: **e ʌm**
       d.   Vowel reduction: **e ʌm**   *(NA because __C context isn't met)*
       e.   Glide insertion: **ej ʌm**

On the other hand, pause-fillers are not completely 'outside the grammar' – as noted by Clark & Fox Tree (2002), *uh* and *um* act like ordinary words in many respects. They are visible for at least some phonological rules, e.g. Flapping:

41)   Bu/ɾ/ uh ... we think tha/ɾ/ uh ...

---

[6] Cf. Rotenberg's (1978) treatment of e.g. *This is a(*n), although I hate to admit it, very silly idea*.

Like the emphatic /ʔ/, pause-fillers seem to be invisible for some grammatical process (Vocabulary Insertion) but visible for others (Flapping). Again, this contrast is taken to be a rule-ordering effect: pause-fillers are inserted after Vocabulary Insertion but before Flapping.

5   Conclusions

I have argued that <u>English THE is not external allomorphy</u>, but is derived by:

(i)   a morphological restructuring rule that makes D[+def] 'part of the same word' as the following segment (Article Cliticization)
(ii)   a structurally restricted phonological rule (Vowel Reduction)

It is an open question whether a similar analysis could be applied to other cases of apparent external allomorphy (see §1), or even to the English *to* alternation:

42)   a.   I'm flying t/**u**/ Atlanta.   (/u/ before V)
       b.   I'm flying t/ə/ Dallas.   (/ə/ before C)

The current model allows for a wide range of types of phonological rules. Some rules apply word-internally (like Vowel Reduction) while others apply across utterances (like Flapping), and still others apply at various intermediate stages (see Pak 2008).

What this means is that <u>we can view both phonetic naturalness (Criterion A) and structural restrictedness (Criterion B) as gradient, rather than binary, measures</u>, and we are not necessarily forced to adopt an allomorphic treatment of an alternation just because it is not a 'low-level phonetic' or 'across-the-board' rule. (See Pak 2008:ch6 for similar arguments in favor of treating French liaison phonologically rather than allomorphically.)

Opening up the possibility of a phonological treatment here allowed us to …

−   recognize parallels between THE and other V~ə alternations
−   account for the distribution of the indefinite-article forms /**e**/ and /**æn**/ (as well as **ə/ən***).*
−   explain various rule-ordering effects observed with *a/an*, THE, pause-fillers and /ʔ/, including the fact that *a/an* and THE are *not* always phonologically optimizing.

Appendix 1: Variation

The analysis laid out here can be adapted to account for inter- and intraspeaker variation in *a/an* and THE. The mini-grammar DEF1 below represents a hypothetical speaker with categorical prevocalic /ði/. DEF2 represents a hypothetical speaker with categorical /ðə/ (i.e. no alternation).

43) *Grammar DEF1*
    a. Article Cliticization
    b. Vocabulary insertion: D[+def] ↔ ði
    c. Vowel reduction: V[-stress] → ə /__C   →   /ðə/ book, /ði/ apple

44) *Grammar DEF2*
    a. Vocabulary insertion: D[+def] ↔ ðə   →   /ðə/ book, /ðə/ apple[7]

DEF1 produces 100% prevocalic /ði/ while DEF2 produces 0% prevocalic /ði/. Speakers with intermediate rates of prevocalic /ði/ –i.e. most speakers of standard English – can be assumed to have access to *both* DEF1 and DEF2, and to go back and forth between these competing grammars depending on register, style, carefulness, and other factors that remain to be explored (Kroch 1994, Embick 2008).

During acquisition, children are assumed to start out favoring the simpler grammar in DEF2. Over time, they learn to use DEF1 more and more frequently until they reach the adult pattern for their particular variety of English.[8]

For the indefinite article, grammar INDEF1 represents a speaker with four variants (/e/, /æn/, /ə/, /ən/). This grammar can be assumed to exist alongside a 'non-alternating' grammar with a single invariant form /ə/ (INDEF2).

45) *Grammar INDEF1*
    a. Article Cliticization
    b. Vocab. insertion: D[-def] ↔ æn /__V
                        ↔ e elsewhere
    c. Vowel reduction: V[-stress] → ə /__C   → /ə/ bóok, /ən/ ápple, /é/ book, /ǽn/ apple

46) *Grammar INDEF2*
    a. Vocab. insertion: D[-def] ↔ ə   → /ə/ bóok, /ə/ ápple, /ə́/ book, /ə́/ apple

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 /æn/.

---

[7]   I also assume a rule of glottal-stop insertion that adds /ʔ/ between *a*/ðə and a following vowel.
[8]   It is possible, of course, that children acquire additional grammars beyond DEF1 and DEF2, and that some of these grammars are eventually abandoned. One possibility would be a grammar with an allomorphy rule inserting /ði/ before a memorized list of words (e.g. *end, other*) and /ðə/ elsewhere. Another would be a grammar where /ði/ and /ðə/ are (realizations of) different morphemes.

47) *Grammar INDEF3*
    a. Article Cliticization
    b. Vocab. insertion: D[-def] ↔ ən /__V
                             ↔ ə elsewhere → /ə/ bóok, /ən/ ápple, /ə́/ book, /ə́n/ apple

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 and/or INDEF1 until they achieve the pattern for their variety of English.

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

48) a. This is {ðí/ðə́} book to read on global warming.
      b. I said I wanted {éj/ə́} 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.

Appendix 2: CHILDES corpus information

Corpus	*Children (N)*	*Adults (N)*	*# a/an tokens*	*# THE tokens*
Bliss, L. (1988). The development of modals. *The Journal of Applied Developmental Psychology, 9*, 253–261.	3 (5;4, 4;6, 6;1)	1	7	--[9]
Braunwald, S. R. 1993. Differences in two sisters' acquisition of first verbs. ERIC Document Reproduction Service.	2 (1;0-6;0, 4;0-7;0)	3	171	319
Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.	2 (2;3-4;10, 2;3-5;1)	10	794	--
Carterette, E. C., & Jones, M. H. (1974). *Informal Speech: Alphabetic and Phonemic texts with statistical analyses and tables*. Berkeley: University of California Press. Jones, M. H., & Carterette, E. C. (1963). Redundancy in children's free-reading choices. *Journal of Verbal Learning and Verbal Behavior, 2,* 489-493.	54 1st graders, 48 3rd graders, 48 5th graders	24	75	--
Ervin-Tripp	12	9	80	53
Evans	16 dyads of 1st-graders	0	13	--
Garvey, C. (1979). An approach to the study of children's role play. *Quarterly Newsletter of the Laboratory of Comparative Human Cognition, 12*. Garvey, C., & Hogan, R. (1973). Social speech and social interaction: Egocentrism revisited. *Child Development, 44*, 562–568.	10 triads (2;10-5;7)	0	44	--
Gathercole, V. (1980). *Birdies like birdseed the bester than buns: A study of relational comparatives and their acquisition*. Unpublished doctoral dissertation. University of Kansas.	6 (3-6 yrs)	0	8	--

---

[9] Tokens of THE were retrieved only from corpora with audio recordings.

Gleason, J. B., & Greif, E. (1983). Men's speech to young children. In B. Thorne, C. Kramerae, & N. Henley (Eds.), *Language, Gender and Society*. Rowley, MA: Newbury.	24 (2;1-5;2)	48	128	--
Kuczaj, S. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior, 16,* 589-600.	1 (2;4-5;0)	3	336	--
MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum.	2 (0;6-8;0, 0;7-5;6)	2	365	543
Nelson, K. (Ed.) (1989). *Narratives from the crib*. Cambridge, MA: Harvard University Press.	1 (2 yrs)	2	18	6
Providence. Demuth, K., Culbertson, J. & Alter, J. 2006. Word-minimality, epenthesis, and coda licensing in the acquisition of English. *Language & Speech*, 49, 137-174.	5 (1-4 yrs)	~12	1860	898
Sachs, J. (1983). Talking about the there and then: The emergence of displaced reference in parent–child discourse. In K. E. Nelson (Ed.), *Children's language, Vol. 4,* Hillsdale, NJ: Lawrence Erlbaum.	1 (1;2-4;9)	3	76	--
Sawyer	24 (3-6 yrs)	6	44	243
Snow (MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum.)	1 (2;5-3;9)	2	241	351
Warren-Leubecker, A. (1982). *Sex differences in speech to children.* Unpublished doctoral dissertation. Georgia Institute of Technology. Warren-Leubecker, A., & Bohannon, J. N. (1984). Intonation patterns in child-directed speech: Mother-father speech. *Child Development, 55,* 1379–1385.	6 (5-6 yrs)	8	24	--
Weist, R. M., Pawlak, A., & Hoffman, K. 2009. Finiteness systems and lexical aspect in child Polish and English. *Linguistics 47*, 1321-1350. Weist, R. M. & Zevenbergen, A. (2008). Autobiographical memory and past time reference. *Language Learning and Development*, 4 (4), 291 – 308.	4 (2-4 yrs)	16	344	--

References

Britain, David & Fox, Sue. 2009. The regularisation of hiatus resolution in British English. In Filppula, M., Klemola, J., & Paulasto, H. (eds.), *Vernacular universals and language contacts* (pp. 177-205). Routledge.

Chomsky, Noam & Halle, Morris. 1968. *The sound pattern of English*. New York: Harper & Row.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A life in language* (pp. 1-52). Cambridge: MIT Press.

Clark, Herbert H. & Fox Tree, Jean E. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition 84*, 73-111.

Crosswhite, Katherine. 2004. Vowel reduction. In Hayes, B. & Kirchner, R. (eds.), *Phonetically based phonology* (pp. 191-231). Cambridge University Press.

Embick, David. 2008. Variation and morphosyntactic theory: competition fractionated. *Language and Linguistics Compass 2*(1), 59-78.

Embick, David. 2010. *Localism versus globalism in morphology and phonology*. Cambridge, MA: MIT Press.

Embick, David & Noyer, Rolf. 2001. Movement operations after syntax. *Linguistic Inquiry 32*, 555-595.

Fox Tree, Jean E. & Clark, Herbert H. 1997. Pronouncing 'the' as 'thee' to signal problems in speaking. *Cognition 16*, 151-167.

Gabrielatos, Costas, Torgersen, Eivind Nessa, Hoffmann, Sebastian, & Fox, Susan. 2010. A corpus-based sociolinguistic study of indefinite article forms in London English. *Journal of English Linguistics, 38*(4), 297-334.

Garellek, Marc. 2012. Glottal stops before word-initial vowels in American English: distribution and acoustic characteristics. *UCLA Working Papers in Phonetics, 110,* 1-23.

Hayes, Bruce. 1990. Precompiled phrasal phonology. In Sharon Inkelas & Draga Zec (eds.), *The phonology-syntax connection* (pp. 85-108). Chicago: University of Chicago Press.

Healy, Alice F. et al. 1998. Toward the improvement of training in foreign languages. In A.F. Healy & L.E. Bourne, Jr. (eds.), *Foreign language learning: psycholinguistic studies on training and retention* (pp. 3-53). Mahwah, NJ: Lawrence Erlbaum.

Hurford, James R. 1972. The diachronic reordering of phonological rules. *Journal of Linguistics 8*, 293-295.

Hurford, James R. 1974. The base form of English *a/an*: a reply. *Lingua 33*, 129-135.

Joseph, Bryan. 1997. On the linguistics of marginality: the centrality of the periphery. *Papers from the 33rd Regional Meeting of the Chicago Linguistics Society*, 197-213.

Jurafsky, Daniel, Bell, Alan, Fosler-Lussier, Eric, Girand, Cynthia and Raymond, William. 1998. Reduction of English function words in Switchboard. In Proceedings of ICSLP-98.

Kaisse, Ellen M. 1985. *Connected speech: the interaction of syntax and phonology*. Orlando: Academic Press.

Keating, P., Byrd, D., Flemming, E., & Todaka, Y. 1994. Phonetic analysis of word and segment variation using the TIMIT corpus of American English. *Speech Communciations 14*, 131-142.

Kroch, Anthony. 1994. Morphosyntactic variation. *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parassession of Variation and Lingsuitics Theory* (pp. 180-201). Chicago: Chicago Linguistics Society.

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 et al. (eds.). Current Issues in Linguistic Interfaces: Proceedings of The 2009 Seoul International Conference on Linguistic Interfaces, Volume 2 (pp. 417-434). Seoul: Hankookmunhwasa.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Marvin, Tatjana. 2002. Topics in the stress and syntax of words. Doctoral dissertation, MIT.

Mascaró, Joan. 1996a. External allomorphy and contractions in Romance. *Probus 8*, 181-205.

Mascaró, Joan. 1996b. External allomorphy as emergence of the unmarked. In Jacques Durand & Bernard Laks (eds.), *Current trends in phonology: models and methods* (pp. 473-483). Salford, Manchester: European Studies Research Institute, University of Salford.

Mascaró, Joan. 2007. External allomorphy and lexical representation. *Linguistic Inquiry, 38*(4), 715-735.

Nevins, Andrew. 2011. Phonologically conditioned allomorph selection. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.) *The Blackwell companion to phonology* (ch. 99). Blackwell.

Newton, Caroline & Wells, Bill. 1999. The development of between-word processes in the connected speech of children aged between three and seven years. In Maassen, B. & Groenen, P (ed). *Pathologies of Speech and language: Advances in Clinical Phonetics and Linguistics* (pp. 67-75). Whurr Publishers.

Pak, Marjorie. 2008. *The postsyntactic derivation and its phonological reflexes*. Doctoral dissertation, U Penn.

Perlmutter, David. 1970. On the article in English. In Bierwisch, M. & Heidolph, K. (eds.), *Progress in linguistics* (pp. 233-248). The Hague: Mouton.

Rotenberg, Joel. 1978. *The syntax of phonology*. Doctoral dissertation, Massachusetts Institute of Technology

Seidl, Amanda. 2001. *Minimal indirect reference: a theory of the syntax-phonology interface*. New York: Routledge.

Vennemann, Theo. 1974. Restructuring. *Lingua 33*, 137-156.