A unified constraint-based account of the English indefinite article

Brian W. Smith • brianwsmith@ucla.edu

University of California Los Angeles

Overview and goals

The basic pattern: *a* before consonants, *an* before vowels.

a	*an*
a pear	an apple
a minute	an hour
a morpheme	an allomorph

Phonologically Conditioned Suppletive Allomorphy (PCSA): there are listed allomorphs; the choice between allomorphs is conditioned by phonology; often one allomorph is the default.

Listedness	/ej/ ~ /æn/
Phonology	/æn/ before vowels
Defaults	/ej/ elsewhere

Three questions about PCSA:
- Is it driven by markedness constraints or lexical subcategorization?
- What is the evidence for defaulthood? How to model it?
- Can putative cases of PCSA be reanalyzed with minor rules, or in OT, with morpheme-specific phonology?

Answers from *a* and *an*:
- Markedness constraints condition PCSA. Diverse repairs respect *ə.V:
  - *a ~ an* allomorphy   • [ʔ]-epenthesis   • *the* reduction
- Defaulthood is observed in production, and modeled with UR constraints.
- In OT, an analysis with morpheme-specific phonology is difficult at best.

A is the default

The allomorph *a* is used in the elsewhere environment. Counts below from the Switchboard Corpus of American English (Godfrey & McDaniel 1992).

Disfluencies. The allomorph *a* is more likely before disfluencies. This holds regardless of the word following the disfluency.

	uh	*um*	*uh/um* V…	*uh/um* C…
a	799	147	19	250
an	17	5	2	2
% *a*	98%	97%	90%	99%

Before disfluencies, function words are unreduced (Foxtree & Clark 1997). [ej] *um apple* * [ə] *um apple*

Speech errors. 95% of errors in Switchboard are *an → a*.
Phrases like *a apple* are about 17x more likely than *an pear*.

	V-initial N/Adj	C-initial N/Adj
a	230	51778
an	4080	13
% *a*	5%	99.9%

*ə.V across English

*ə.V   Assign one ✳ for every ə.V sequence in the surface representation.

This constraint has effects across English, driving and blocking a number of different phonological processes. This conspiracy was noticed as early as Stene (1954).

Phonotactics. No hiatus in English where left vowel is lax (Chomsky & Halle 1965)
- [mi.ə]   [redi.o]   *[mə.i]   *[redə.o].
- <10 counterexamples in the Carnegie Mellon University pronouncing dictionary.
- >2,500 examples of hiatus where the left vowel is tense.

[ʔ]-epenthesis. Epenthesis is nearly categorical in the context ðə_V (Keating et al 1994)
- Also within words (Plag 1999)
  mora[ʔ]-ize , samba[ʔ]-ing, dada[ʔ]-ist
- Children up to 10 yrs: a [ʔ] apple (Pak 2014)

Reduction. English speakers are less likely to reduce function words before vowels (Keating et al 1994).
- [ði] apple   [ðə] pear   *[ðə] apple
- Same results for *to* [tu]~[tə], shown in graph to the right. These data show the proportion of the strong (unreduced) form in the Buckeye corpus of conversational English (Buckeye: Pitt et al. 2007). Each box is a speaker in Buckeye.
- Speakers who use prevocalic [ðəʔ] or [təʔ] are still more likely to reduce before consonants.

Strong forms of function words across speakers
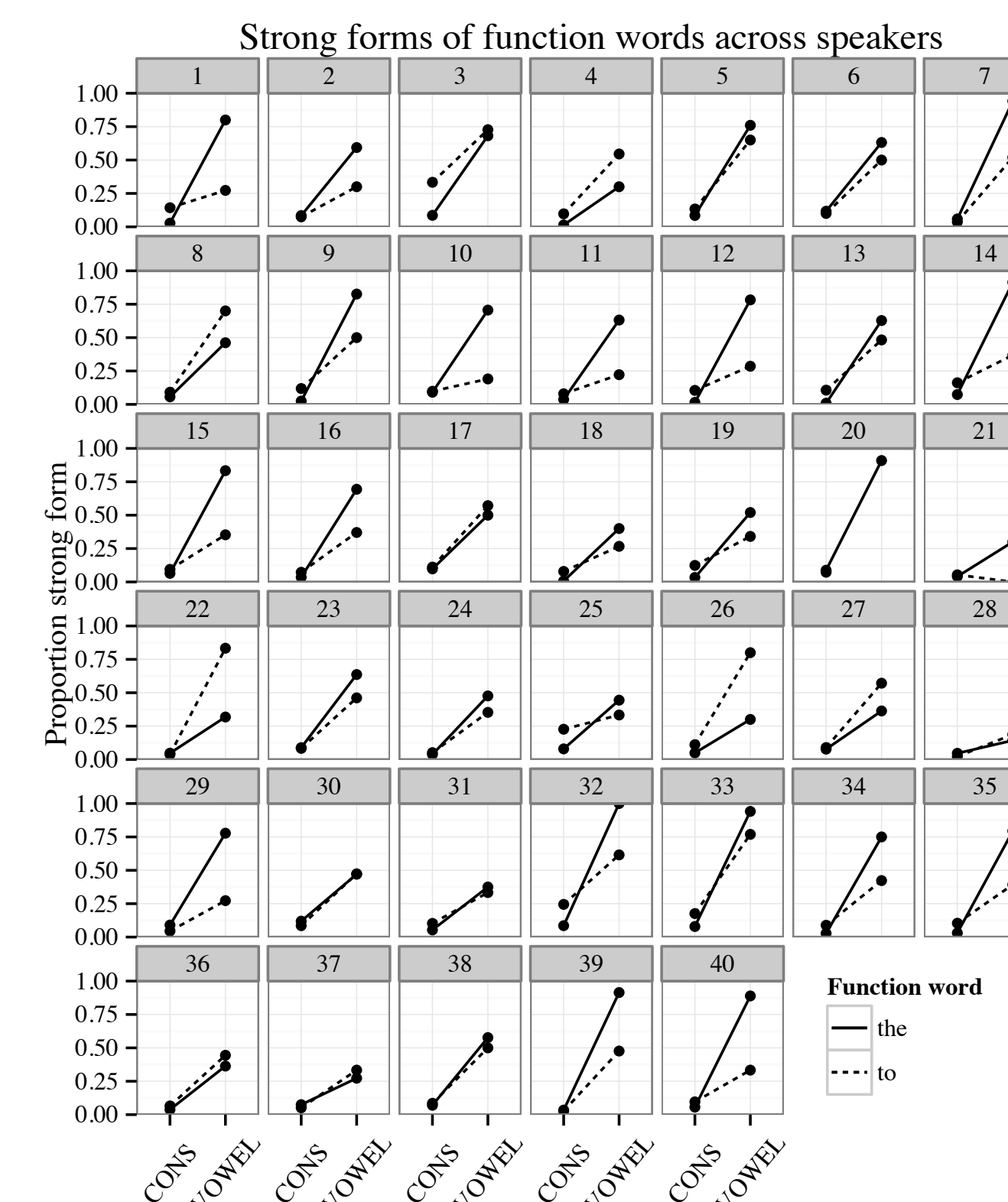
Account with UR constraints

The input to phonology consists solely of morphosyntactic features, abbreviated here as FS for 'Feature Structure' (Russell 1995, Boersma 1998, Zuraw 2000, Wolf 2008).

UR selection happens in phonology (Tranel 1996, Kager 1996, Wolf 2013 for overview).

FS–UR correspondences are encoded as violable UR constraints (Pater et al. 2012, see also Kager 1996, Zuraw 2000, Boersma 2001).

UR = AN   Assign one ✳ for every FS {Indef, Det} that does not correspond to /æn/

UR = A   Assign one ✳ for every FS {Indef, Det} that does not correspond to /ej/

*ə.V   Assign one ✳ for every ə.V sequence in the surface representation

REDUCE   Assign one ✳ for every function word with an unreduced vowel in a weak prosodic position (based on Selkirk 1972)

Ranking

Defaults and reduction

UR = A ≫ UR = AN obtains default.

REDUCE ≫ IDENT obtains reduction.

		*ə.V	DEP	REDUCE	UR=A	UR=AN	IDENT
a	☞ [ə] pear					✳	✳
b	[ən] pear				✳W	L	✳
c	[e] pear			✳W		✳	L
d	[ðə] pear						✳
e	[ði] pear			✳W			L

Three repairs for hiatus

*ə.V ≫ UR = A Default isn't used if it creates hiatus.

*ə.V ≫ REDUCE blocks reduction. Reverse ranking for speakers with pre-vocalic [ðəʔ].

Epenthesis is still possible, even though it isn't used for *a* or *the*.

		*ə.V	DEP	REDUCE	UR=A	UR=AN	IDENT
a	☞ [ən] apple					✳	
b	[e] apple			✳W	L	✳W	L
c	[əʔ] apple		✳W			L	✳W	✳
d	[ə] apple	✳W				L	✳W	✳
e	☞ [ði] apple				✳		
f	[ðəʔ] apple		✳W		L		✳
g	[ðə] apple	✳W			L		✳
h	☞ mora [ʔ] ize		✳				
i	mora ize	✳W			L		

Lexically indexed constraints

An alternative to PCSA for *a* and *an*: morpheme-specific phonology. This is implemented as a minor rule of n-epenthesis in Rotenberg (1974).

A possible implementation in OT is lexically indexed constraints, either faithfulness (Fukazawa 1997) or markedness (Pater 2010). Indexed markedness misses the conspiracy. Indexed faithfulness can capture both n-epenthesis and [ʔ]-epenthesis, although it cannot capture reduction.

Dep-ʔ_A   Assign one ✳ for every [ʔ] in the output without an input correspondent, if the [ʔ] follows the morpheme *a*.

Indexed faith

* əV, DEP-n ≫ DEP [ʔ]-epenthesis to repair hiatus, not n-epenthesis.

D [ʔ]-epenthesis blocked for *a*, n-epenthesis instead.

No epenthesis before a consonant. REDUCE ≫ IDENT gets reduction.

		*ə.V	D	REDUCE	DEP-n	DEP	IDENT
h	☞ mora [ʔ] ize					✳	
i	mora [ʔ] ize				✳W	✳	
	mora ize	✳W				L	
d	☞ [ən] apple					✳	✳	✳
e	[e] apple			✳W		L	L	L
f	[əʔ] apple		✳W		L	✳	✳
g	[ə] apple	✳W			L	L	✳
a	☞ [ə] pear						✳
b	[ən] pear					✳W	✳W	✳
c	[e] pear			✳W			L

Reduction in *the* introduces a ranking paradox. Why not [e] apple?

Incompatible with *the*

If n-epenthesis and reduction occur for *a*, they should also occur for *the*. If reduction is blocked for *the*, it should also be blocked for *a*.

		*ə.V	D	REDUCE	DEP-n	DEP	IDENT
a	☞ [ði] apple			✳			
b	[ðən] apple			L	✳W	✳W	✳W
c	[ðəʔ] apple			L		✳W	✳W
d	☞ [ən] apple				✳	✳	✳
e	[e] apple			✳W	L	L	L

The problem: the choice of repair is determined solely by the ranking of constraints. Blocking [ðən] and [ðəʔ] requires even more indexation. In the PCSA analysis, [ðən] isn't an option because /ðin/ isn't a listed allomorph.