This printout has been approved by me, the author. Any mistakes in this printout will not be fixed by the publisher. Here is my signature and the date: _____

Stress in *–ative* and the Role of Pre-*at-* Segments: Evidence from Dictionary Studies

Juliet Stanton

1. Overview

Nanni (1977) claims that, in American English *–ative* forms, whether or not *–at-* bears a secondary stress depends on several factors. One is rhythmic: *–at-* is not stressed if the syllable before it is stressed (*affírmative*, **affírmàtive*). Another is segmental: in contexts where *–at-* can bear stress, it does so when preceded by an obstruent or cluster (*invéstigàtive*, *législàtive*) but not a vowel or sonorant (*pálliative*, *spéculative*). This short paper provides dictionary data in support of the claim that segmental factors impact *–at-* stress and shows that the effect cannot be reduced to other considerations, like the frequency of the *–ative* derivative or a related form. §2 presents results from the Oxford English Dictionary. §3 presents results from four other dictionaries (Merriam Webster, American Heritage, Kenyon & Knott 1944, the CMU Pronouncing Dictionary), which are less informative due to corpus size and/or editorial convention. §4 summarizes, and briefly discusses the status of dictionary data as phonological evidence.

2. Oxford English Dictionary

The corpus of *–ative* forms from the Oxford English Dictionary (OED) included all non-obsolete forms in the dictionary as of 7/17 that satisfied several criteria. First, an IPA transcription and frequency information had to be present. Second, the *–ative* form's stem had to end in a trochee (as in *législative*, *σ́σ-σσ*), so as to control for potential effects of rhythmic context. In total, 334 *–ative* forms satisfied these criteria. The suffix *–at-* is "stressed" if the vowel is transcribed as an [eɪ], variably or invariably; it is "stressless" if the vowel is always a [ə]. For example: *mollificative* (/məˈlɪfəkədɪv/) is "stressless", and *communicative* (kəˈmjunəˌkeɪdɪv, /kəˈmjunəkədɪv/) and *motivative* (/ˈmoʊdəˌveɪdɪv/) are "stressed". (The choice to group variable and consistent *–at-* stress together was arbitrary but made to simplify the analysis. Results of the analyses largely do not differ according to this choice; the few cases in which the results change when the variable cases are grouped with the stressless cases are flagged.)

The data provide support for Nanni's (1977) claim that segmental identity influences *–at-* stress (1). They also reveal additional distinctions among segment types, as well as quite a bit of variability: it is not the case that stress on *–at-* categorically depends on the identity of the preceding segment(s).

(1) Role of identity of preceding segment(s) in *–at-* stress

Segment(s)	Stressed *–at-*	Stressless *–at-*	% stressed	Total
Vowel	22, e.g. *annunc**i**ative*	22, e.g. *enunc**i**ative*	50%	44
Sonorant	88, e.g. *muti**l**ative*	64, e.g. *specu**l**ative*	58%	152
Obstruent	92, e.g. *deprec**a**tive*	18, e.g. *dubi**t**ative*	84%	110
Cluster	27, e.g. *legi**sl**ative*	1, e.g. *ade**qu**ative*	96%	28

To ensure that this effect cannot be attributed to other unrelated factors, a logistic regression was fit to the data in (1). The dependent variable had a value of 0 if *–at-* was stressless, and a value of 1 if *–at-* was (variably or consistently) stressed. The role of segmental information, along with several other potentially relevant factors, were included as independent variables (all are described below).

* J. Stanton, New York University. My thanks to the audience members whose questions helped shape this material.

Cascadilla Proceedings Project
Completed June 22, 2018

- ***Identity of pre-ative segments*** *(V/R/O/CC; continuous variable)*

  The segmental information represented in (1) was encoded as a continuous variable, where V was assigned a value of 0, R was assigned a value of 1, O was assigned a value of 2, and CC was assigned a value of 3. This predictor was included so as to verify the version of Nanni's (1977) claim apparent in (1): the identity of the pre-*ative* segment or segments affects the rate at which *–at-* bears stress.

- ***Frequency of the –ative form*** *($Freq_{ative}$; continuous variable)*

  The frequency of the *–ative* form was encoded as a continuous variable. The lexical frequency information was taken from the OED, which divides words into one of eight frequency "bands", where lower bands indicate lower frequency.[1] This information was included to evaluate Kenyon & Knott's (1944) claim that more frequent *–ative* derivatives are more likely to bear stress on *–at-*.

- ***Frequency of related –ate and –ation forms*** *($Freq_{ate}$ and $Freq_{ation}$; continuous variables)*

  For many *–ative* derivatives, there is a related *–ate* and/or *–ation* form with the same stem (e.g. *legislative*, *legislate*, *legislation*). It is possible that the existence of these forms, in which *–at-* is always stressed, could influence the *–ative* form. Specifically, the more frequent the *–ate* or *–ation* form is, the more likely the speaker might be to stress *–at-* in the corresponding *–ative* form. Frequency information was from the OED; in the case that there was no related form, or the frequency was unavailable, it was marked as 0.

The logistic regression was fit using the glm function of R's lme4 package (Bates et al. 2015). Effects were considered significant if $p \leq .05$ (roughly, if the z-statistic $\geq |2|$), as assessed by the Wald test. A model including all factors indicated a significant effect of only V/R/O/CC. A likelihood ratio test (LRT) was then performed, comparing a model that included all four predictors to one that included only V/R/O/CC. The LRT indicated that the model including all predictors is not a significantly better fit to the data (χ^2 (3) = 3.28, p = .35), and thus the simpler model is preferable. The output of this simpler model is summarized in (2). The positive coefficient indicates that as the pre-*ative* material changes from a vowel to a sonorant to an obstruent to a cluster, *–at-* becomes significantly more likely to bear stress.

(2) Model results

	Estimate	***z* value**	**Significant?**
Intercept	-0.42	–	–
V/R/O/CC	0.96	5.56	Yes ($p < .001$)

I address next the possibility of finer-grained distinctions within the R, O, and CC categories (§2.1) and the extent to which the OED provides an accurate picture of stress in American English *–ative* (§2.2).

2.1. Further distinctions among consonant types

The division of (1) into four coarse categories means that distinctions within them have been glossed over. In (3), we see that subdivision of the R/O/CC categories yields some additional distinctions.

(3) Role of segmental identity in *–at-* stress, by natural class

Segment(s)	**Class**	**Stressed *–at-***	**Stressless *–at-***	**% stressed**	**Total**
R	Approximant	60, e.g. *muti**l**ative*	49, e.g. *gestic**ul**ative*	55%	109
	Nasal	28, e.g. *aggluti**n**ative*	15, e.g. *imagi**n**ative*	65%	43
O	Voiceless stop	69, e.g. *expli**c**ative*	16, e.g. *eradi**c**ative*	81%	85
	Voiced fricative	5, e.g. *inno**v**ative*	1, e.g. *aggra**v**ative*	83%	6
	Voiced stop	18, e.g. *propa**g**ative*	1, e.g. *eluci**d**ative*	95%	19
CC	CC	21, e.g. *arbi**tr**ative*	1, e.g. *ade**qu**ative*	95%	22
	CCC+	6, e.g. *illu**str**ative*	–	100%	6

These distinctions, however, are based on too few forms to be meaningful: within-category comparisons suggest that the observed differences are not statistically significant (results of Fisher's Exact Tests are p

[1] On the OED's frequency data, see http://public.oed.com/how-to-use-the-oed/key-to-frequency/.

$>$ .05 for all). Thus while the OED data provide support for a general effect of segmental identity on *–at-* stress, they do not provide enough power to draw additional distinctions within these larger categories.

2.1.1. Sources of transcriptions and dialect mixing

A potential worry is that the OED contains transcriptions from a number of different varieties of English. The possibility of dialect-specific differences must be taken into account, to ensure that the result in (1) is not a consequence of dialect mixing. To make this concern concrete: assume that *–at-* is stressed at higher rates in American than in British English. If transcriptions of *–ative* forms where *–at-* is preceded by an obstruent or cluster come mainly from American English, and the rest come mainly from British English, the apparent effect of segmental identity may be due how the OED was compiled.

The 334 *–ative* forms can be divided into two groups, according to the transcription's citation. For 181, the OED provides a transcription that is marked as coming from American English (AE); these transcriptions were used when available. For the rest, the OED does not remark on the transcription's source; I call these 151 forms "undifferentiated" (or UD). In (4) we see that when these forms are separated into groups, they display somewhat different properties. For AE, the overall rate of *–at-* stressing (at 85%) is higher, and the amount of cross-category variation is smaller; for UD, the overall rate of *–at-* stressing (at 48%) is lower, but the amount of cross-category variation is larger.

(4) Role of identity of preceding segments, separated by dialect

Segment(s)	Source	Stressed *–at-*	Stressless *–at-*	% stressed	Total
Vowel	AE	21, e.g. *perpet**u**ative*	6, e.g. *individ**u**ative*	78%	27
	UD	1, e.g. *insin**u**ative*	16, e.g. *eval**u**ative*	6%	17
Sonorant	AE	65, e.g. *allite**r**ative*	14, e.g. *imagi**n**ative*	84%	79
	UD	23, e.g. *extermi**n**ative*	50, e.g. *figu**r**ative*	32%	73
Obstruent	AE	54, e.g. *regurgi**t**ative*	7, e.g. *mollifi**c**ative*	89%	61
	UD	38, e.g. *depu**t**ative*	11, e.g. *habi**t**ative*	78%	49
Cluster	AE	15, e.g. *nome**ncl**ative*	1, e.g. *ade**qu**ative*	94%	16
	UD	12, e.g. *facul**t**ative*	–	100%	12

Logistic regressions fit to both datasets (identical in form to (2)) indicate that the effect of segmental identity is significant for UD ($p < .001$) and either trending ($p = .08$) or significant ($p < .001$) for AE, depending on whether the forms with variable *–at-* stress are grouped with the consistently stressed or consistently stressless forms, respectively (see discussion at the beginning of Sec. 2).[2]

These results suggest that segmental information is important in the AE and UD groups, but it raises the question of what the relationship between them is. The UD transcriptions are of British English (BE): they make use of BE vowels (e.g. [əʊ]), exhibit features that are typical of BE (e.g. palatalization of coronal stops, as in *dubitative* [djuːbɪteɪtɪv]), and are non-rhotic (e.g. *dissertative*, [dɪsəteɪtɪv]). The primary difference between the UD and the AE forms has to do with whether or not the form's entry has been fully revised for the OED's 3rd Edition: while the majority of entries that provide an AE form have been fully revised for the 3rd Edition (167/183), no entries that provide a UD form have been (0/151).[3]

What determines which words have been revised for the 3rd Edition? Catherine Sangster, Head of Pronunciation at Oxford Dictionaries, reports (p.c. to D. Steriade) that revisions to the 2nd Edition first concentrated on words beginning with *n*, *o*, and *p*. After this, revisions occurred throughout. This initial focus on the middle part of the alphabet is evident in the OED's entries for *–ative* forms: entries revised from 2000-2007 begin with *m/n/o/p/q*, and revisions occurred in alphabetical order (*m* in 2000-2003, *n* in 2003, *o* in 2004-2005, *p* in 2005-2007, and *q* in 2007; after this, the beginning letters are more heterogeneous, but entries beginning with *r* are disproportionately common until 2011). As for why the later revisions targeted the forms that they did, lexical frequency may be relevant: excluding forms that begin with *m-r*, the average frequency bin of words with revised entries is 3.1, compared to 2.8 for

[2] The results of LRTs, comparing these AE models to those where there is no dependent variable, similarly vary according to categorization of the variable cases. If variable is grouped with consistently stressed, the LRT does not find that including V/R/O/CC results in a better fit (χ^2 (1) = 3.10, $p = .08$). If variable is grouped with consistently stressless, the LRT finds that including V/R/O/CC results in a significantly better fit (χ^2 (1) = 12.37, $p < .001$).

[3] For more on the 3rd Edition, see http://public.oed.com/the-oed-today/guide-to-the-third-edition-of-the-oed/.

words without revised entries. Whether or not a form has been revised for the 3rd Edition thus appears to depend entirely on non-phonological factors like its beginning segment and its lexical frequency.

Is it justified to treat these unrevised UD forms as representative of American English speech? The AE forms definitely are: the process by which a pronunciation gets listed in the OED is lengthy and involves input from multiple native speakers of American English, including pronunciation editors, their consultants, and recording artists (C. Sangster, again p.c. to D. Steriade). But the source of the UD forms' transcriptions is unclear: many (107/151) were added in the 19th century. Thus while the OED data as a whole indicate that segmental identity is relevant, the conclusion is slightly weakened by the observation that the effect is strongest in a set of forms whose transcriptions' provenance is unclear.

3. Other dictionaries

To see if the OED trends replicate, I consulted 4 additional dictionaries: Merriam Webster, American Heritage, Kenyon & Knott 1944, and the CMU Pronouncing Dictionary (CMU). In all four cases, the transcriptions are claimed to come exclusively from American English. In all but CMU, however, any possible effect of segmental identity is muted by editorial convention.

3.1. Merriam Webster

To build a corpus of *–ative* forms from Merriam Webster (MW), I attempted to find transcription information for each form included in the OED study. Unlike the OED, MW does not provide a way to search for forms ending in *–ative*, so finding all of the *–ative* forms in the dictionary would have required a search of its entire contents. In addition, because MW does not provide usable frequency information, any analysis of the MW data that includes frequency needs to reference the OED's frequency bins. Thus, any analysis of the MW data needs to include only those forms that are also present in the OED.

Of the 334 forms under consideration here, Merriam Webster contains transcriptions for only 187. As shown in (5), in the vast majority of cases (182/187) these forms are transcribed with *–at-* stress. The five forms where *–at-* is stressless pattern as one would expect, given the OED findings: in two *–at-* is preceded by a vowel, in two *–at-* is preceded by a sonorant, and in one *–at-* is preceded by obstruent.

(5) Rates of *–at-* stress by segmental context (Merriam Webster)

Segment(s)	Stressed ***–at-***	Stressless ***–at-***	% stressed	Total
V	28, e.g. *recr**e**ative*	2, e.g. *humil**i**ative*	93%	30
R	79, e.g. *combi**n**ative*	2, e.g. *configu**r**ative*	98%	81
O	56, e.g. *resusci**t**ative*	1, e.g. *reci**t**ative*	98%	57
CC	19, e.g. *legi**sl**ative*	0	100%	19

Segmental identity is not a significant predictor of *–at-* stress ($p = .11$), nor is the frequency of the *–ative* derivative ($p = .55$) or the frequency of a related *–ate* or *–ation* form ($p = .06$, $p = .26$). (The regressions fit in this section are in all relevant respects identical to the logistic regression in Sec. 2.1 for the OED.)

The high rate of *–at-* stress can likely be linked to editorial policy. While OED transcriptions are informed by native speaker judgments, Joshua Guenter, a MW pronunciation editor, notes that in the cases of "Latinate words which are infrequently (if ever) used in speech [...] analogy is the determining factor", and "pronunciations are derived from the morphology of the word and analogy to words with similar roots and affixes" (p.c. to D. Steriade). Based on this, it seems reasonable to assume that the high rate of *–at-* stress in *–ative* forms comes from a judgment, on the part of the editors, that *–at-* ought to bear stress in *–ative* forms (perhaps because it always does in the related *–ate* and *–ation* forms) The generally high rate of *–at-* stress found in MW may then reflect entirely on editorial convention.

3.2. American Heritage

The corpus considered next includes those forms from the OED (n=208) that are also listed in American Heritage (AH). As AH does not include frequency information in its entries, this was necessary in order to account for any potential role of lexical frequency. The results for AH are more or less identical to the results for MW: the rate of *–at-* stress is high overall (194/201, or 97%), but the distribution of the small number of stressless *–at-* forms is roughly consistent with the OED data.

(6) Rates of *–at-* stress by segmental context (American Heritage)

Segment(s)	Stressed *–at-*	Stressless *–at-*	% stressed	Total
V	23, e.g. *recr**e**ative*	1, e.g. *eval**u**ative*	96%	24
R	99, e.g. *acce**l**erative*	5, e.g. *figu**r**ative*	95%	104
O	67, e.g. *precipi**t**ative*	1, e.g. *regurgi**t**ative*	99%	68
CC	12, e.g. *admini**str**ative*	0	100%	12

Effects of segmental identity and derivative frequency are not significant if the cases of variable *–at-* stress are grouped with the consistently stressed cases (as shown in (6); $p = .14$ and $p = .052$, respectively) but both are significant if the variable cases are grouped with the stressless cases ($p < .001$ for both). Frequencies of related *–ate* and *–ation* forms are not significant predictors under either assumption.

The existence of this frequency effect and the rarity of stressless *–at-* are both likely due to editorial convention. Evidence for this comes from discrepancies in the two different methods of recording stress. The first is a transcription: of the 208 *–ative* forms, AH provides transcriptions for 93. The entries for the remaining 115 are noted only with stress marks: extirpative, for example, is transcribed as ex′tir·pa′tive (in AH, stress marks follow stressed syllables). Whether or not AH provides a transcription for an *–ative* form is correlated with its frequency (7); this correlation is significant ($p < .001$, logistic regression).

(7) Rate of *–ative* transcription by frequency bin (American Heritage)

OED frequency bin	Transcription	No transcription	% transcription	Total
1	0	3	0%	3
2	9	61	13%	70
3	29	28	51%	57
4	28	18	61%	46
5	22	5	81%	27
6	5	0	100%	5

The trend in (7) suggests an editorial strategy: frequent words are more likely to be transcribed. For the words that are not transcribed, the vast majority have a stress on *–at-*, but there are a couple of exceptions (e.g. *iterative* is it′er·a·tive). Interestingly, however, when stress marks and a transcription co-exist, they can disagree. *dissociative*, for example, is dis·so′ci·a′tive, but immediately next to this transcription, the dictionary editors indicate with the parenthetical (-ə-tĭv) that *–at-* is stressless. *obliterative* is o·blit′er·a′tive, but the dictionary editors indicate with the parenthetical (-ə-rā′tĭv, -ər-ə-tĭv) that the stress of this form is actually variable. These disagreements are unidirectional: in no case does an *–ative* form's stress marks indicate that it is stressless (as in it′er·a·tive) but a transcription indicate that it bears stress.

Based on this, it is plausible to assume that the editors of American Heritage employ the following strategy for forms ending in *–ative*: transcribe with a stress on *–at-*, unless there is enough evidence from available native speaker pronunciations that *–at-* ought to be transcribed as stressless.

3.3. Kenyon & Knott (1944)

Kenyon & Knott (KK)'s discussion of *–ative* forms (p. 31) indicates that they assume the default pronunciation to involve *–at-* stress: they write that "when the main accent is on the second syllable before *–ative*, there is usually a secondary accent on the *–a-* of *–ative* [. . .] but in familiar words some speakers often pronounce *–ative* without accent [. . .] when only –ˌetɪv is given, –ətɪv is to be understood as a possible variant." Of the 35 *–ative* forms in KK, 34 are transcribed as bearing stress on *–at-* (8).

(8) Rates of *–at-* stress by segmental context (Kenyon & Knott)

Segment(s)	Stressed *–at-*	Stressless *–at-*	% stressed	Total
V	7, e.g. *pall**i**ative*	0	100%	7
R	13, e.g. *acce**l**erative*	1, e.g. *figu**r**ative*	93%	14
O	11, e.g. *vindi**c**ative*	0	100%	11
CC	3, e.g. *legi**sl**ative*	0	100%	3

The fact that *figurative* is the only stressless *–at-* form is consistent with the OED results, but given KK's explicit policy for the transcription of *–ative* forms, it is also possible that this transcription is a mistake.

3.4. The CMU Pronouncing Dictionary

Of the 334 forms in the OED *–ative* set, 43 are transcribed in the CMU Pronouncing Dictionary (CMU). As seen in (9), segmental identity plays a role in *–at-* stress.

(9) Rates of *–at-* stress by segmental context (CMU Pronouncing Dictionary)

Segment(s)	Stressed *–at-*	Stressless *–at-*	% stressed	Total
V	1, e.g. *appreci**a**tive*	1, e.g. *palli**a**tive*	50%	2
R	7, e.g. *stimu**l**ative*	17, e.g. *deco**r**ative*	29%	24
O	12, e.g. *medi**t**ative*	2, e.g. *communi**c**ative*	86%	11
CC	3, e.g. *legis**l**ative*	0	100%	3

A logistic regression finds that segmental identity ($p < .01$) but no frequency-related factor ($p = .86$ for *–ative*, $p = .33$ for *–ate*, and $p = .56$ for *–ation*), predicts *–at-* stress. What makes CMU so different from the others in this section? I suspect that the transcriptions in CMU are based on the productions of native American English speakers, and that its editors do not have a bias for *–at-* stress. If this is correct, then CMU provides us with an accurate picture of how *–ative* forms are pronounced in American English. Unfortunately, I have not been able to get in touch with the dictionary's maintainer to verify this.

4. Summary

The results of this study (10) support Nanni's (1977) claim: *–at-* is more likely to be stressed when preceded by V or R than when preceded by O or CC. In addition, the results of this study suggest that *–at-* stressing is more variable and sensitive to more distinctions among segment types than we knew. Nanni's original claim was that *–at-* stressing is a categorical effect: it is always stressless when preceded by a vowel or sonorant, and always stressed when preceded by an obstruent or a cluster. Given that Nanni does not cite a source for her claim, and that this categorical effect was not found in any of the dictionaries consulted, it seems likely that Nanni's claim reflects her intuitions or possibly a simplification thereof.

(10) Rates of *–at-* stress by segmental context (all dictionaries)

Dictionary	V	R	O	CC	Total
OED (AE)	21/27 (78%)	65/79 (83%)	54/61 (90%)	15/16 (94%)	153/181 (85%)
OED (UD)	1/17 (6%)	23/73 (32%)	38/49 (78%)	12/12 (100%)	74/151 (49%)
MW	28/30 (90%)	79/81 (98%)	56/57 (98%)	19/19 (100%)	182/187 (97%)
AH	23/24 (96%)	99/104 (95%)	67/68 (99%)	12/12 (100%)	194/201 (97%)
KK	7/7 (100%)	13/14 (93%)	11/11 (100%)	3/3 (100%)	34/35 (97%)
CMU	1/1 (50%)	7/24 (29%)	12/14 (86%)	3/3 (100%)	23/43 (53%)
Total	**81/106 (76%)**	**284/373 (76%)**	**238/260 (92%)**	**64/65 (98%)**	**660/798 (83%)**

These results also suggest that dictionaries are not equally reflective of native speaker judgments in their transcriptions of rare Latinate words. At one end of the spectrum is MW, which transcribes rare Latinate words by taking into account only the pronunciation of their relatives. At the other is the OED, where each word, no matter its frequency, is transcribed based on native speaker judgments (C. Sangster, p.c. to D. Steriade). This finding has several implications. First, of the dictionaries that list large numbers of infrequent forms, the OED is likely the most reflective of native speaker judgments. Second, the source of dictionary data must be taken into account when it is used as evidence for phonological generalizations.

References

(2017). Merriam-Webster.com. Online resource (accessed November 2017).

(2017). OED Online. Oxford University Press. Online resource (accessed July 2017).

(2018). The American Heritage dictionary of the English Language. Online resource (accessed February 2018).

(2018). The CMU Pronouncing Dictionary. Online resource (accessed February 2018).

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1–48.

Kenyon, John Samuel & Thomas Albert Knott (1944). *A Pronouncing Dictionary of American English*. Merriam, Springfield, Massachusetts.

Nanni, Debbie L. (1977). Stressing Words in -Ative. *Linguistic Inquiry* 8, 752–763.