# Acting Different(*-ly*): Bringing Derivational Morphology into Variationist Linguistics

Dan DeGenaro[*] [†] [‡]

**Abstract.** In some varieties of English, adjectives are generally employed rather than adverbs when modifying syntactic constituents that are noun-like, and adverbs are employed otherwise. Many English adjectives and adverbs arise in corresponding lexeme pairs due to the high productivity of the derivational morpheme *-ly,* which can produce de-adjectival adverbs (i.e. *different* → *differently*). Standard American English (SAE), along with many other varieties of English, generally requires adverbs to be expressed with *-ly* if possible. However, in African American English (AAE), as well as still more varieties of English, there is strong evidence that this suffix is not required in certain contexts, although such contexts may mandate its use in the first grouping of varieties. I present a corpus-based study of the derivational morpheme *-ly* as it is used in AAE, focusing on defining the contexts in which it may be obligatory, variable, or ungrammatical.

**Keywords.** -ly adverbs; zero adverbs; African American English; part of speech tagging

[*] Authors: Dan DeGenaro, University of Massachusetts Amherst (ddegenaro@umass.edu)

[†] The author wishes to acknowledge Lisa Green and Brian Dillon, both in the Department of Linguistics at the University of Massachusetts Amherst.

[‡] All code and data used in this study is available on Github.

- Are you sure you like it? 'Cause if you wanna be down with the Lethal Interjection crew, you got to really think it sounds good.
- I do.
- How good?
- Uh, real good?
- Okay, you passed.

Exchange between rapper Thugnificent and eight-year-old Riley Freeman on "The Boondocks"

## 1. Introduction.

1.1. DESCRIPTION OF TOPIC. In this thesis, I seek to investigate the behavior of the *-ly* adverbial suffix in African American English (AAE). The *-ly* adverbial suffix is a highly productive derivational morpheme, common to many varieties of English, especially those varieties with overt prestige, i.e. those used in many public-facing and formal contexts. This set of varieties includes formal Standard American English (SAE), the linguistic variety in which this paper is (approximately) written. While there are two related derivational morphemes which form de-adjectival adjectives (kind → kindly) and de-nominal adjectives (friend → friendly), the specific morpheme I am investigating has the function of transforming adjectives into de-adjectival adverbs (different → differently).

This morpheme is currently under-described in the literature on AAE, and in variationist linguistics in general. Most current literature on the morpheme deals with its historical evolution and the grammaticalization processes that have given it the properties that it currently has in varieties like formal SAE, and those varieties used in the southern United Kingdom (Nevalainen 2004, 2008; Killie 2015). The variationist literature available on the morpheme at present claims that it is possible for some varieties of English to omit or disallow *-ly* where other varieties prefer or even require it be used to denote de-adjectival adverbs (Schneider & Kortmann 2004; Nevalainen 2008; Wolfram & Schilling 2015). Schneider & Kortmann (2004), in particular, find that more varieties seem to omit it rather than prefer it. Among those that tend to omit *-ly* is AAE. In my study, I will be focusing on both lexemes and syntactic environments for which the *-ly* suffix may be obligatory, variable, or disallowed entirely in AAE. Both phenomena are more fully described in the sections that follow. To do this, I will be analyzing the Corpus of Regional African American Language (Kendall & Farrington 2021), or CORAAL for short, maintained by linguists at the University of Oregon. CORAAL consists of a set of seven subcorpora: ATL, DCA, DCB, LES, PRV, ROC, VLD.

I use statistical description of this corpus to make claims about whether the *-ly* suffix is obligatory, variable, or disallowed for particular lexemes when being used as adverbs, and whether there are more general syntactic environments that cause the *-ly* suffix to be obligatory, variable, or disallowed. Note that throughout this study, I will refer to adverb/adjective pairs such as *excited/ly* as single "lexemes."

An effective analysis of CORAAL involves developing methods to collect data not only

about *-ly*'s distribution in the corpus, but also the environments in the corpus in which it theoretically could occur in AAE. As previously noted, while some varieties tend to require *-ly* more frequently and strictly, in AAE, *-ly* may be comparatively absent, omitted, or suppressed (the appropriate description likely depends on what particular grammatical process warrants such behavior in the first place). Such environments where *-ly* is possibly absent must also be examined and included in all calculations and description done in this study. To accomplish this, I make use of an automated part-of-speech tagger in addition to a more traditional regular expression search process. In particular, I re-examine any items in my dataset built by regular expression search which were tagged differently by human annotation versus automated tagging. The hope is that this process will turn up missing or erroneously tagged uses of adverbs, culminating in a more thorough and accurate analysis of the corpus. In addition, this process provides an opportunity to make claims about the accuracy of an appropriately trained part-of-speech tagger on the task of disambiguating adjectival and adverbial uses.

1.2. SCIENTIFIC QUESTIONS. This research paper seeks to address some principal linguistic questions about the distribution of *-ly* in AAE, and as a sort of lemma, I seek also to determine how well the ARK Tagger (Gimpel et al. 2010; Owoputi et al. 2013), a conditional random field (CRF) part-of-speech tagger trained on non-standard English writing (namely tweets), disambiguates adjectives and adverbs in AAE. I define the adverbial morpheme *-ly* as a potential linguistic variable (Lavandera 1978), following the definition that its suppression or overt expression causes no concrete semantic difference, but rather serves as a stylistic choice in the variety of English at hand. I further define two variants. The overt adverbial (bound) morpheme (OAM) is that form that is expressed with *-ly,* and the zero adverbial morpheme (ZAM) is that form that is expressed without it. Formally, the questions I wish to answer in my thesis are:

1. On lexemes.
    (a) Which lexemes, if any, require *-ly* affixation (OAM) when being used as adverbs? For instance, is *he said it intentionally* grammatical while *he said it intentional∅* is ungrammatical?
    (b) Which lexemes, if any, disallow *-ly* affixation (ZAM) when being used as adverbs? For instance, is *she wrote it different∅* grammatical while *she wrote it differently* is ungrammatical?
    (c) Which lexemes, if any, have varying use of *-ly* affixation when being used as adverbs? For instance, are both *they built it solid∅* and *they built it solidly* grammatical?
2. On syntactic environments.
    (a) Which syntactic environments, if any, require *-ly* affixation (OAM) when being used as adverbs? For instance, might preverbal adverbs require *-ly* affixation? Is *he quickly ran* grammatical while *he quick∅ ran* is ungrammatical?
    (b) Which syntactic environments, if any, disallow *-ly* affixation (ZAM) when being used as adverbs? For instance, might preadjectival predicate-modifying adverbs disallow *-ly* affixation? Is *she ∅ real∅ smart* grammatical while *she ∅ really smart* is ungrammatical?

(c) Which syntactic environments, if any, have varying use of *-ly* affixation when being used as adverbs? For instance, might postverbal adverbs variably allow *-ly* affixation? Are both *they cheered happy*∅ and *they cheered happily* grammatical?

3. On part-of-speech tagging.

   (a) Can the ARK Tagger reliably recognize when a form like *different* is being used as an adjective or an adverb?

   (b) If not, what improvements could theoretically be introduced to capture this phenomenon?

1.3. ON LEXEMES. One way in which these environments may be specified is by the particular root adjective from which an adverb is being constructed. Particular lexemes such as *different/ly* may behave distinctly from ones like *intentional/ly*. For example, it may be the case that *she was doing it different* is grammatical, while *she was doing it intentional* may be ungrammatical. This may be due to a prosodic constraint (in.**ten**.tion.al vs. **diffe**.rent), a function of number of syllables (*different* has two while *intentional* has four), or frequency. A cursory regular expression search that did not take variation of case into account returned 1424 instances of the pattern `different`, of which 32 were actually `differently` ($\approx 2\%$), and only five matches for `intentional`, of which four were `intentionally` (80%). Note that a regular expression search is not advanced enough to suss out ZAM forms, only OAM forms, so these proportions cannot be considered relative adverb/adjective frequencies. However, it seems that the lexeme *intentional/ly* likely arises with this *-ly* suffix proportionally more frequently than does *different/ly*. It is possible that highly frequent words tend to be shortened more frequently, leading to the shorter form becoming more common, and especially so if *-ly* is variable or disallowed in certain environments.

1.4. ON SYNTACTIC ENVIRONMENTS. A second, more substantial constraint on the use of the *-ly* suffix may arise in the syntactic environment in which an adverb is being used. I identify five distinct (and mutually disjoint) syntactic environments that give rise to *-ly* adverbs:

1. Preverbal adverbs (`PREVB`) are those that linearly occur prior to the verb they are immediately modifying within a given utterance.

   (1) a. *He **quickly** ran.*
   b. ***Quickly,** he ran.*

2. Postverbal adverbs (`POSTVB`) are those that linearly occur after the verb they are immediately modifying within a given utterance.

   (2) a. *She ran (so) **quickly.***
   b. *How she ran was **quickly.***
   c. *How did she run? **Quickly.***

3. Preadjectival predicate-modifying adverbs (`PREJJ`) are those that linearly occur prior to the predicate adjective they are immediately modifying within a given utterance.

(3) a. *Their pace was **really** quick.*
b. ***Really** quick, their pace.*
c. *How was their pace? **Really** quick.*

4. Postadjectival predicate-modifying adverbs (POSTJJ) are those that linearly occur after the predicate adjective they are immediately modifying within a given utterance.

(4) *Their pace was quick, **really**.*

5. Non-predicate adjective-modifying adverbs (NPAM) are those that modify adjectives which are not predicate adjectives.

(5) *Their **really** quick pace was impressive.*

While I am not a native speaker of AAE, having it heard it for much of my life, I have some intuition about what might or might not be allowed. Therefore, I hypothesize, based solely on my own experience, that environments 1 and 4 do not allow ZAM forms, while environments 2, 3, and 5 do allow ZAM forms. Here are the above examples with ZAM forms (I use * for syntactic violations and ? for possible syntactic violations):

1. Preverbal adverbs (PREVB):

(6) a. **He **quick**∅ ran.*
b. ?***Quick**∅**,** he ran.*

2. Postverbal adverbs (POSTVB):

(7) a. *She ran (so) **quick**∅**.***
b. *How she ran was **quick**∅**.***
c. *How did she run? **Quick**∅**.***

3. Preadjectival predicate-modifying adverbs (PREJJ):

(8) a. *Their pace was **real**∅ quick.*
b. ***Real**∅ quick, their pace.*
c. *How was their pace? **Real**∅ quick.*

4. Postadjectival predicate-modifying adverb (POSTJJ):

(9) **Their pace was quick, **real**∅.*

5. Non-predicate adjective-modifying adverb (NPAM):

(10) *Their **real**∅ quick pace was impressive.*

1.5. SIGNIFICANCE. The *-ly* morpheme's properties in AAE are not thoroughly described in the literature on AAE, nor in the literature on other varieties of English, including both those varieties that seem to prefer it and those that tend to omit it where possible. This study attempts to fill in this gap, at least in the description of AAE, and begin to bring more derivational morphology into variationist linguistics, which at present tends to focus more on inflectional morphology, with little research having been done as of yet on variation in derivational morphology (Prieto 2005; Kortvelyessy & Stekauer 2014).

Other studies have focused on biases taken on by artificial intelligence (AI) systems, such as automated speech recognition (ASR) systems. Koenecke et al. (2020) found that state-of-the-art (at the time of the study) ASR systems developed by major tech companies (Amazon, Apple, Google, IBM, and Microsoft) had a word error rate of 0.35 on average when recognizing the speech of Black speakers versus 0.19 on average when recognizing the speech of white speakers. They define the word error rate as

$$WER = \frac{S + D + I}{N} \tag{1}$$

where $S$, $D$, and $I$ are the number of incorrectly substituted, deleted, or inserted words, and $N$ is the total number of words actually uttered by the speaker. As society begins to rely more heavily on automation, it is becoming increasingly important to turn towards equitable AI, which is useful and reliable for all its potential users. By examining the performance of the ARK Tagger on this task, I hope to nudge the AI community slightly in the direction of equitable systems.

As he utter the calm flow, don't talk about my moms, yo
Sometimes he rhyme quick, sometimes he rhyme slow
Or vice versa, whip up a slice of nice verse pie
Hit it on the first try, villain: the worst guy

"ALL CAPS," MF DOOM

## 2. Literature review.

2.1. AFRICAN AMERICAN ENGLISH. AAE is a broad set of linguistic varieties spoken throughout the United States and Canada, primarily by people of African American descent (Green 2002). It was first formally described by Labov et al. (1968), who sought to legitimate the speech of Black New Yorkers; up until that point, it was believed that the "improper speech" of many Black people in the United States was due to lack of education or cultural deficit. Labov and others showed that AAE is rule-based, like any other language. This was originally done by defining sets of features distinguishing AAE from SAE, such as Fasold & Wolfram (1972). However, recent literature has focused on defining patterns of usage rather than isolated features, and establishing AAE as a variety in its own right. Since the pioneering work in the 1960s, AAE has been the subject of extensive study within the discipline of variationist linguistics due to morphosyntactic patterns (Washington et al. 1998; Green 2002; Wolfram 2008; Green & Sistrunk 2015; Forrest & Wolfram 2019; Lightfoot & Havenhill 2019) as well as phonological characteristics (Green 2002; Craig et al. 2003; Thomas & Carter 2006; Thomas 2007; Charity 2008; Eberhardt 2008; Jones et al. 2019; Mufwene et al. 2021; Bailey & Thomas 2021) that distinguish it from other varieties of English. In this study, I focus on a particular morphosyntactic pattern: the ZAM and OAM variants of the adverbial *-ly* morpheme.

2.2. THE MORPHOSYNTAX OF AFRICAN AMERICAN ENGLISH.

2.2.1. ZERO WORDS. Given the content of this study, syntactic zero patterns are of particular interest. I first wish to establish the quantity, commonality, and productivity of such zero objects that have already been shown to exist in AAE. Green (2002), Wolfram (2008), and Green & Sistrunk (2015) note several of these patterns in word-like objects which are grammatical in AAE, but ungrammatical or varyingly grammatical in closely related varieties. Relative clauses may use the zero complementizer where some varieties cannot (*the man ∅ prepared the meal is a chef* is an example put forth by Green & Sistrunk (2015)). Questions may be formed without auxiliary *do* (*∅ she be working at the law firm?*). Left out of many discussions is the somewhat less frequent zero infinitival marker, which is generally overt in varieties such as SAE (*to*, as in *to boldly go where no one has gone before*). Washington et al. (1998) provide the example *people waitin' ∅ get on there.*

Most notable among zero words in AAE are the zero auxiliary and the zero copula[1]. For instance, *they ∅ walking* uses a zero auxiliary (connecting a subject and a verbal *-ing*), while *they ∅ good people* uses a zero copula (connecting a subject and some other form). Labov (1969)

[1] Note that Labov analyzes zero auxiliary and zero copula as deletion of contracted forms, as opposed to zero words.

was among the first studies of the zero copula, finding that AAE-speaking adolescents and pre-adolescents from Harlem used zero the copula up to 77% of the time in group settings, and to a nearly-as-great extent in individual conversations. Notably, Labov found that there was significant context-dependence on whether the zero copula tended to surface. Across two age groups of AAE-speakers from Harlem, the zero auxiliary surfaced more than 60% of the time preceding *-ing* verbs, and more than 80% of the time preceding the future marker *gon.* However, the zero copula surfaced preceding noun phrases less than one-third of the time, indicating that the first two environments are significantly more favorable for a zero form. Wolfram (1969) confirmed a similar analysis, finding that the copula in AAE speakers hailing from Detroit may be realized as a zero form greater than 50% of the time, particularly amongst speakers of working-class socioeconomic status. Wolfram also found that upper-middle class AAE speakers used the zero copula as much as 3-4% of the time on average.

2.2.2. ZERO MORPHEMES. In addition to such word-like zero objects, AAE also exhibits many highly productive, bound zero morphological objects, which behave similarly or identically to corresponding overt inflectional morphemes in varieties which do not allow the zero forms. Wolfram & Fasold (1974) noted these forms early on. Third-person singular *-s* vs. its corresponding zero form was shown to exhibit the most variation, and to surface as zero most frequently. Working-class AAE-speakers in Detroit were found to use the zero form as much as 70% of the time on average. While middle-class AAE-speakers from the same area still used the zero form, it surfaced less than 10% of the time on average. Analogously, possessive *-'s* surfaced as zero as much as 20% of the time or more among working-class speakers, and plural *-s* surfaced as zero around 5% of the time in the same group. Both zeros arose somewhat rarely among middle-class speakers. The same study also found that younger speakers used zero forms, particularly the third-person singular zero, more frequently.

A substantial portion of the literature on AAE's morphosyntax has focused on acquisition. AAE-speaking children may have less contact with SAE or other varieties than older speakers, and therefore are more likely to exhibit AAE-specific variants more frequently. This can allow for a clearer picture of what patterns may be specific to AAE, and which patterns may arise in AAE speakers due to bidialectalism. Studies have found that AAE-speaking children also do not produce third-person singular *-s* consistently, even if they begin to acquire it. Indeed, Reveron (1979) examined not only production of third-person singular *-s* by AAE-speaking children, but also plural *-s*, genitive marker *-'s* and past tense *-ed*. Her study found that none of the four inflectional morphemes studied were consistently produced (defined as $>$80% of possible contexts) by AAE-speaking children aged three to six from lower- and middle-class backgrounds. This can be taken as evidence that these morphemes are frequently zero in AAE, and that an ideal AAE speaker-hearer might never produce overt pronunciations of these morphemes. Similarly, Cole (1980) defines plural *-s*, genitive marker *-'s*, past tense *-ed*, and copula/auxiliary *be* as "non-obligatory" in AAE, including in adult AAE. She found that while zero plurals and possessives were not as common as overt ones, zero past tense, zero third-person singular, and zero copula/auxiliary *be* were often produced at criterion (i.e. these zero forms were acquired by the AAE-speaking children as opposed to the standard variants).

More recent studies such as Johnson (2005) have found evidence that AAE-speaking children do not comprehend third-person singular *-s*. Such a conclusion begs the claim that although overt third-person singular *-s* exists in some varieties of English, it may not exist in AAE. This

study also found that there was still considerable variation among participants in their ability to comprehend third-person singular *-s*, and the authors pose a few possibilities to address this, including the possibility that the children who seemed to interpret the zero form as ambiguous may have begun code-switching.

De Villiers & Johnson (2007) found that AAE-speaking children generally do not produce overt third-singular present *-s*, though SAE-speaking children reliably acquire it by age five. Newkirk-Turner & Green (2016) and Newkirk & Green (2021) present extensive descriptions of the the zero third-person singular marker in AAE (*he do∅ it all the time*), which, for the sake of comparison, is ungrammatical in formal SAE (in which third-person singular contexts are marked with the verbal suffix *-s*). This zero morpheme is readily produced and comprehended by AAE-speaking children, while the overt morpheme *-s* generally is not.

Other inflectional morphemes in AAE have been considered linguistic variables despite being non-zero. The alternative *-ing* allomorph, often notated *-in'* in popular literature and pronounced as /ɪn/. This last morpheme is not a zero, nor is it technically consonant cluster reduction, though it is a linguistically variable inflectional morpheme nonetheless. Forrest & Wolfram (2019) describe the variable pronunciation of *-ing* in AAE, which can be pronounced either /ɪn/ or /ɪŋ/. While such a linguistic variable is more phonological in nature, the techniques used are very much applicable to other patterns. The authors define phonological environments and collect descriptive statistics about the two variants in these environments. For instance, they find that a pause following a word ending with *-ing* was far more likely to result in the /ɪŋ/ pronunciation for speakers from the CORAAL DCB subcorpus (Kendall & Farrington 2021), whereas the same speakers favored /ɪn/ when a velar sound followed the *-ing* word. Unfortunately, there is scant literature concerning the properties of derivational morphology in AAE.

2.3. THE *-ly* MORPHEME AND ITS PROPERTIES IN OTHER VARIETIES. Adverbial *-ly* is rarely discussed in the context of AAE, but it has been thoroughly investigated in the extensive literature written about other English varieties. While it is important to note that *-ly* may behave differently in AAE as opposed to other varieties, understanding its properties in a closely related variety such as formal SAE (for which more theory-heavy literature exists) may shed light on its behavior in AAE. *-ly* is particularly noteworthy in the discussion of English morphology due to the lack of consensus on its status as a derivational morpheme. Zwicky (1995) presents evidence against its classification as inflectional despite its high productivity. He argues that classification of adverbial *-ly* as an inflectional morpheme is incorrect due to the notion that adjectives and adverbs are different grammatical classes, and an important property of inflectional morphemes is that affixing one to a given word must not change the word's grammatical class. This is considered the conventional and prevailing wisdom, that adverbial *-ly* is a derivational morpheme in English.

However, there are those who argue that it is in fact inflectional, or becoming inflectional by a grammaticalization process. One argument in favor of the inflectional status of *-ly* presented by Giegerich (2012) is that affixing *-ly* to an adjective to form an adverb bars further derivation: it is not possible to form words like **nicelier/quickliest/intenseliness*, though non-*ly* adverbs like *soon* can be further derived: *sooner/soonest* are grammatical. It is generally the case that inflectional morphemes are affixed to a root lexeme after derivational morphemes, which therefore disallows further affixation. A great many scholars have claimed in addition to Giegerich (2012) that *-ly* holds the status of an inflectional morpheme in English already (Lyons 1966; Emonds 1976; By-

bee 1985; Radford 1988), or is at least transitioning towards such behavior via a grammaticalization process (Nevalainen 2004, 2008; Killie 2015). Killie (2015) notes several aspects of grammaticalization that seem to apply to *-ly.* Though no specific criteria are yet firmly established for the definition of a grammaticalization process, likely contenders for such criteria include obligatorification and paradigmatization, both of which *-ly* exhibits in some varieties of English, though by no means all (Schneider & Kortmann 2004). It may just be the case that grammaticalization of *-ly* is taking place in some varieties, which tends to make *-ly* obligatory, but is not (yet) occurring in AAE. Indeed, Nevalainen (2008) finds that historically, *-ly* usage in English may have been stratified, for instance, by social class - she even states "If a grammaticalization process fails to apply to some significant sections of the language community, extralinguistic determinants can lead to dialect differentiation, and give rise to variety grammars with grammaticalization histories of their own." As for AAE, more historical data may be needed before this can be definitively claimed, though it appears, on the surface, to be a possible explanation for such variation.

*-ly*'s contentious status as either a derivational or inflectional morpheme makes it a natural starting place to begin investigating the rules governing derivational morphology in AAE. It is likely that no matter what category it best fits into, its properties, which align significantly with those of English inflectional morphemes, make techniques applied to the study of inflectional morphemes applicable to *-ly.* Furthermore, its extremely high productivity makes it possible to build large datasets of examples of its use, which in turn allows for accurate descriptive statistics to be collected. Such data may shed light on how the *-ly* affix behaves in AAE, whether that behavior differs significantly from other varieties, and if so, in precisely what ways, and for what underlying reasons.

2.4. ANECDOTAL EVIDENCE OF ZERO *-ly* BEHAVIOR IN AFRICAN AMERICAN ENGLISH. Anecdotally, it is possible in AAE to omit the adverbial *-ly* suffix on adverbs in contexts which mandate the suffix in other varieties. It may also be the case that there is simply no suffix in the first place, though further theorizing about the underlying structure of this phenomenon will likely rely on the results garnered in this study. I present a few examples of this phenomenon that occur with the form *different∅* found in the transcriptions of Kendall & Farrington (2021):

(11) a. *People might act **different∅.*** (ATL_se0_ag1_f_03_1)
b. *I performed **different∅. I operated different∅, and I got different results.*** (ATL_se0_ag1_m_03_1)
c. *Cali, all the houses are different, painted **different∅, they all have they own personalities.*** (ATL_se0_ag1_m_04_1)

Note the contrast between the bold and non-bold instances of *different. different* is the same adjective that is used in formal SAE, and as an adjective, it is used to modify nouns directly (*different middle schools, different results*) or as a predicate adjective (*the houses are different*). On the other hand, ***different∅*** is an adverb, whose formal SAE equivalent is the word *differently*. This word is not used to modify nouns. In the above examples, it modifies verbs (*act **different∅**, performed **different∅**, operated **different∅***) as well as adjectives (*painted **different∅***). This is in contrast to a resultative construction, such as *hammer the metal flat,* wherein the metal is what is flat, not the act of hammering. This construction still falls under adjectival use.

Scholars of English literature have also drawn attention to this phenomenon, without formally describing it as a linguistic system. Heller (2003) notes that in Toni Cade Bambara's short

story "The Lesson" Bambara (1972), the character Sylvia produces many utterances in which a word which clearly functions as an adverb is lacking the *-ly* suffix:

(12) a. *then spread out **gradual**∅ to breathe*
b. *he don't need it **bad**∅ as I do*
c. *"Can I steal?" Sugar asks very **serious**∅*
d. *sail two kittens across the pond if you strap them to the posts **tight**∅*
e. *wanted it **bad**∅*
f. *teased me so **bad**∅ I tied her up in the shower*

One might note the fact that these instances all fall under the category of postverbal adverbs (POSTVB), a category that I have hypothesized to consistently allow this construction. However, more formal linguistic description of this phenomenon is required to make effective claims. Importantly, Heller also claims that *I watched Miss Moore who is **steady** watchin us like she waitin for a sign* is an example of this phenomenon, but in this case, *steady* is not truly an adverb; this is a use case of the habituality/consistency/intensity marker *steady* (Green 2002). Heller further claims that *I talk about his mama something **ferocious*** is an example of this, though this is contentious; *ferocious* seems to modify *something*, which would make it an adjective. As this construction is not present in formal SAE, the overtly prestigious variety on which theory tends to fixate, it is difficult to make claims about the specific construction being used here, and how to correctly tag the word *ferocious* in this instance.

2.5. LEGITIMACY AS A SOCIOLINGUISTIC VARIABLE. Defining *-ly* as a sociolinguistic variable in AAE may be appropriate. Following Lavandera (1978) in her discussion of Labov, I define the OAM/ZAM forms jointly as the *-ly* variable. Lavandera introduces the criterion, following Labov, that a sociolinguistic variable should be some structure or phonation that a speaker can choose, in the sense that choosing one variant or another has no bearing on the concrete or abstract meaning being conveyed. That being said, both authors contend that a sociolinguistic variable carries stylistic or social meaning, regardless of the fact that its variants should not change the information being conveyed by the speaker outside of this.

"Code-switching," "dialect-shifting," or "style-shifting" are variously applied as monikers for a particular phenomenon in which a speaker adjusts stylistic elements (sociolinguistic variables) in response to their external environment (Craig 2019). One might postulate that the social meaning of the *-ly* variable is a signification of belonging within the AAE-speaking community. For instance, if the ZAM form is produced more readily by an AAE-speaker in conversation with an AAE-speaking interlocutor than in conversation with an SAE-speaking interlocutor, this might be evidence of such a social meaning. Macaulay (1995) found *-ly* to have a class-based social meaning and identification in Ayr, Scotland, and Tagliamonte & Ito (2002) found a similar result in York, England. Future research with AAE's behaviors with regard to *-ly* beyond this study should examine, to the extent possible, variable production of the ZAM and OAM morphemes in conversation with various interlocutors to confirm or deny a particular social meaning.

Such a result would fit neatly into the work of Eckert (2006), in which it is argued that people belong to particular communities whose linguistic practices are similar, due to both internal and external factors. Those people who align in many social categories such as race, gender, ethnicity, age, etc. are likely to align their speech as well. This behavior tends to create in-groups and out-groups, and one might define users of the ZAM form as in some sort of in-group, while

those who tend not to use it are in the out-group. Therefore, usage of the ZAM form might correlate with group membership, and thus encodes a social meaning of belonging.

2.6. REGULAR EXPRESSIONS. Pivoting to a discussion of the research tools I employ, I first introduce regular expressions. Regular expressions are a notation that can be used to specify regular languages – sets of strings (textual sequences) that can be recognized by a finite state machine (Kleene 1956). The first algorithm for searching for and matching strings in digital documents using regular expressions is attributed to Ken Thompson (Thompson 1968). I use Python's implementation of regular expressions (`re`), which functions similarly to the original one, but with added features for convenience. To design regular expressions, I rely a word list I derived from Princeton's WordNet (Miller 1995), using the interface provided by the Natural Language Toolkit (Loper & Bird 2002).

2.7. PART-OF-SPEECH TAGGING. Part-of-speech tagging is defined as labeling a sequence of textual tokens with the appropriate part-of-speech categories (grammatical classes) from a predefined set of possible tags (Jurafsky 2000). Generative models such as Hidden Markov Models (HMMs - for instance, Thede & Harper (1999); Brants (2000)) and discriminative models such as conditional random fields (CRFs - for instance, Lafferty et al. (2001)) have been applied to this task, which falls under the umbrellas of artificial intelligence, machine learning, and natural language processing. More modern approaches generally make use of neural networks, and particularly those whose architectures lend themselves to processing sequential information effectively (RNNs, LSTMs, transformers). A bidirectional long short-term memory recurrent neural network (BiLSTM) was shown to be effective by Ling et al. (2015), and BERT (Devlin et al. 2019) has been effective on this task with fine-tuning.

The ARK Tagger (Gimpel et al. 2010; Owoputi et al. 2013), a CRF part-of-speech tagger trained on Twitter data, was shown to perform better on AAE-like language than part-of-speech taggers trained on formal SAE writing (Jrgensen et al. 2016). It has also been successfully employed in a dependency parsing study (Blodgett et al. 2018) of AAE-like language, and was even used to identify AAE-like constructions from tweets (Blodgett et al. 2016). In that study, the authors used topic modeling to produce an automatically annotated corpus of text split into African American-aligned text (likely written by an AAE speaker) and white-aligned text (likely written by a non-AAE speaker). Although the CRF is not the most modern algorithm, Manning (2011) notes that the best HMMs, CRFs, and neural algorithms all perform at around 97% accuracy, similar to accuracies reported for humans on the same tasks.

2.8. PROSODY IN AFRICAN AMERICAN ENGLISH. AAE is known to place primary stress on the first syllable of many words that do not have primary stress on the first syllable in other varieties. Wolfram & Fasold (1974) note the words *July* and *hotel* are pronounced with stress on the first syllable in many contexts. They also note that the first syllables of some words can be absent if the stress is not on the first syllable; *'bout*, *'cept*, and *'member* are common pronunciations of the words *about*, *except*, and *remember*. Smitherman (1977) provides *police* and *Detroit* as examples of words that variably have initial stress. Baugh (1983) also notes *detain*, *define*, and *polite*. Baugh found that this stress shift, termed "forestressing," is indeed a linguistic variable, and occurs only in certain social contexts (it is generally disfavored in formal contexts), and only on bisyllabic words. It is worth considering the co-occurrence of forestressing and the *-ly* variable as a first guess at explaining the data in Section 4.1.

We real cool. We
Left school. We
Lurk late. We
Strike straight. We
Sing sin. We
Thin gin. We
Jazz June. We
Die soon.

Gwendolyn Brooks, "We Real Cool"

## 3. Methods.

3.1. RESOURCES. I use the textual transcriptions of the sociolinguistic interviews in CORAAL (Kendall & Farrington 2021) as a source of data, as well as the ARK part-of-speech Tagger (Gimpel et al. 2010; Owoputi et al. 2013), the Natural Language Toolkit (*nltk*) Python implementation (Loper & Bird 2002), `nltk`'s implementation of the Princeton WordNet Miller (1995), and Python's built-in regular expression package (`re`). The two primary methods of inquiry were 1) searching for regular expressions within CORAAL and manually annotating them and 2) comparison of human annotation with the ARK Tagger's tags over CORAAL.

3.1.1. CORAAL. CORAAL contains seven subcorpora from different locales in the United States. Importantly, per Cukor-Avila & Balcazar (2019), there is significant variation among the subcorpora. While I do not attempt to analyze regional differences in this study, such an analysis would be a key future direction. A brief summary of the data is available in Table 1.

| Abbrev. | Locale | Year of Recording | Proportion of Subcorpus |
|---|---|---|---|
| ATL | Atlanta, Georgia | 2017 | 5.88% |
| DCA | Washington, D.C. | 1968 | 23.75% |
| DCB | Washington, D.C. | 2016 | 35.19% |
| LES | Lower East Side (Manhattan), New York | 2009 | 6.36% |
| PRV | Princeville, North Carolina | 2004 | 10.43% |
| ROC | Rochester, New York | 2016 | 10.09% |
| VLD | Valdosta, Georgia | 2017 | 8.30% |

Table 1. Table detailing the seven subcorpora of CORAAL. "Proportion of Subcorpus" refers to the space used by the files of each subcorpus in relation to the space used by the entire corpus.

Each subcorpus is a collection of sociolinguistic interviews with AAE speakers and corresponding transcriptions. I work with the transcriptions in this study. Although writing cannot perfectly capture all the details and intricacies of spoken language, the CORAAL transcriptions are of extremely high quality, and errors are few and far between. For word-level frequency-based analyses such as this study, text is also a far simpler medium to work with in terms of computational processing, annotation, and categorization.

3.1.2. COMPUTATIONAL RESOURCES. I use the ARK Tagger, a CRF part-of-speech tagger (Gimpel et al. 2010; Owoputi et al. 2013). This tagger achieved around 90% accuracy in tagging a corpus of tweets. It was trained specifically to work on informal and non-standard writing having been trained on tweets, meaning it is ideally more adaptable to transcriptions of non-standard speech than other out-of-the-box part-of-speech taggers available. I also make use of NLTK's implementation of Princeton's WordNet database as a resource for building a word list of English adjectives and adverbs. The list of adjectives and adverbs are derived from `nltk.corpus.wordnet.all_lemma_names`. Finally, I use Python's built-in regex implementation `re`. This is used to perform keyword searches while also capturing the immediate context surrounding each instance of any keyword.

3.2. PREPROCESSING. To build the dataset, I downloaded CORAAL version 2021.07 (`get_coraal.py`) and cleaned all 231 of the textgrid files automatically (`preprocess.py`). This involved deleting lines spoken by interviewers, removing markup such as brackets and slashes which denote the interlocutor's interruptions and muffled speech, replacing redacted items with simpler "generic words," (for instance, */RD-SCHOOL-4/* $\rightarrow$ */school/*), and stripping out extraneous whitespace between lines. This resulted in a new set of 231 cleaned text files, which were taken as the raw data for further analysis in this study.

3.3. REGULAR EXPRESSION SEARCH. The first approach I take involves searching over the entire corpus for regular expressions. To design a collection of regular expressions which could retrieve any and all words I might care to examine, I make use of WordNet (Miller 1995) via the Natural Language Toolkit (NLTK) Python library (Loper & Bird 2002).

I build a set of target words by extracting all adjectival and adverbial lemma forms from NLTK's WordNet interface. I then filter these in three ways. First, for any given adjective I look at, there should be an *-ly* form. If I take an adjectival lemma and affix *-ly* according to the correct rule, I should find that the result is in the set of adverbial lemmas. I can then take both the adjective and adverb to be candidate words to search for in the corpus.

Next, I exclude any words that are in a pre-existing list of "ill-behaved" lexemes which I manually curated. This list includes pairs such as *like/ly*, *just/ly*, and *hard/ly* where the *-ly* form is used for a qualitatively different meaning most of the time. For example, *hard* as an adjective means something like "durable, difficult to break," etc., while *hardly* means something like "rarely, not very much at all." The list also includes the words *atypical/ly* and *unintelligible/y* due to the adjectival forms' use as markup in CORAAL.

Finally, I search for each candidate word via regular expression. If I find just one instance of the candidate word, I retain it as an official target word for this study. If I do not find it, it will be ignored beyond this point. This filtering process produces a list of approximately 1300 target words. The process is detailed in Figure 1.

I then design a regular expression for each target word. The expression consists of the target word itself, ignoring case, with up to 30 characters of context on either side of the word. This amount of surrounding character-level context was sufficient to determine the syntactic context of use of the target word the overwhelming majority of the time. I search the entire corpus for each regular expression (target word + context), and save all the matches to a spreadsheet. The unannotated data consists of over 70,000 examples, a portion of which are excluded from the final cut due to expansion of the list of "ill-behaved" lexemes which occurred during the annotation process as necessary.

```
procedure WORDLIST
    A ← set of all adjective lemmas from WordNet
    R ← set of all adverb lemmas from WordNet
    P ← {"good", "well"}                          ▷ Set of candidate adverbs.
    X ← words to be ignored
    for (a ∈ A) do                                ▷ For each adjective,
        r ← AFFIXLY(a)                            ▷ produce an -ly form.
        if (r ∈ R) then                           ▷ If such a form is an actual word,
            if (r ∉ X) ∧ (a ∉ X) then             ▷ and I don't wish to exclude it,
                P ← P ∪ {a, r}                    ▷ both the adjective and adverb are candidates.
            end if
        end if
    end for
    F ← {}                                        ▷ Set of found candidates.
    C ← CORAAL corpus
    for (p ∈ P) do                                ▷ For each candidate,
        e ← REGEX(\W + p + \W, ignore case)       ▷ wrap the word in non-letters.
        if (e ∈ C) then                           ▷ If this expression is found in the corpus,
            F ← F ∪ {p}                           ▷ it is a target word.
        end if
    end for
    return F                                      ▷ Final list of target words.
end procedure
procedure AFFIXLY(w)                              ▷ Helper function that generates an -ly form.
    if w ends with "y" then
        return w[: −1] + "ily"
    else if w ends with "le" then
        return w[: −1] + "y"
    else if w ends with "ic" then
        return w + "ally"
    else if w ends with "ue" then
        return w[: −1] + "ly"
    else
        return w + "ly"
    end if
end procedure
```

Figure 1. Word list build process. Note that although I ultimately decided to keep *good/well*, it is an irregular example of the adjective/adverb paradigm and should be considered separately.

I annotate each example with one of seven categorical markings: `A` for adjectives, `PREVB`, `POSTVB`, `PREJJ`, `POSTJJ`, and `NPAM` for the five syntactic contexts I defined for adverbs under 1.4, and `E` for "exclude." Exclusion of data was done when the context captured by the regular expression was insufficient to determine syntactic context, or when the context was ambiguous.

I also introduce a "count" for each example, which automatically counts the number of instances of use within a match. For example, if a participant said *I really think that, I really do,* this would count as **two** instances of *really*, despite the fact that the regular expression search would capture them together as a single example due to the window size of 30 characters on either side; importantly, regular expression matches are always non-overlapping. Such a process would also count *I really, really think that,* as two separate instances, which is more dubious, but also not unreasonable; the word is indeed used twice in the given fashion. One more issue with this process is the potential to miscategorize one of the instances by applying a single label to both as in *I really think that. I do, really.* The first instance would be considered `PREVB`, whereas the latter would be `POSTVB`. However, as I annotated the data, I noted such instances were so rare that it was sufficient to just exclude them and retain almost all the other data. The benefits of the automated count significantly outweighed the disadvantages in this regard.

The final product is a collection of data for which quantitative analysis is possible. I can break a lexeme down by its use as an adjective versus as an adverb, by its syntactic contexts as an adverb, and by its overall frequency in the corpus. The most interesting and significant findings are presented in section 4.

3.4. TAGGER-ASSISTED RE-ANALYSIS. After annotating the data by hand, I run the ARK Tagger on each example as well. The ARK Tagger produces a space-separated list of tokens and a corresponding list of space-separated part-of-speech tags. I automatically locate the tag corresponding to the target word for each example to determine the tagger's accuracy with respect to the human annotation. In particular, I calculate a confusion matrix (true positives, false positives, true negatives, false negatives). I define the four categories as follows. A true positive occurs when the human tag is one of the five adverb categories (`PREVB`, `POSTVB`, `PREJJ`, `POSTJJ`, `NPAM`) and the tagger tag is `R` (the adverb tag). A false positive occurs when the human tag is `A` (adjective) while the tagger tag is `R`. A true negative occurs when both the human tag and the tagger tag are `A`. A false negative occurs when the human tag is one of the five adverb categories and the tagger tag is `A`. In this way, I exclude from consideration any situation where other parts of speech are potentially involved. This ensures that the tagger is being solely evaluated on its ability to disambiguate adjectives and adverbs.

Once the true positives, true negatives, false positives, and false negatives are counted, I can calculate a few key metrics regarding the tagger's performance with respect to the human labels. On binary tasks such as adjective/adverb disambiguation, it is common to calculate three metrics: precision, recall, and $F_1$. Precision is defined as:

$$\text{precision} = \frac{\text{tp}}{\text{tp} + \text{fp}} \in [0, 1] \tag{2}$$

where a perfect score of 1 indicates that the tagger never incorrectly labeled an adjective as an adverb. Recall is defined as:

$$\text{recall} = \frac{\text{tp}}{\text{tp} + \text{fn}} \in [0, 1] \tag{3}$$

where a perfect score of 1 indicates that the tagger never incorrectly labeled an adverb as an adjective. $F_1$, the geometric mean of precision and recall, is a way to combine the two scores into an overall performance metric:

$$F_1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \in [0, 1] \tag{4}$$

where a score of 1 indicates that the tagger correctly tagged every adverb in the corpus (and did not tag any adjectives or other words as adverbs), and a score of 0 indicates that it didn't tag a single adverb correctly. I can expect the ARK Tagger's $F_1$ to be above 90%, given the high rates of accuracy on Twitter data reported by its creators.

3.5. VERIFICATION OF DATA QUALITY. I listened to a few of the sociolinguistic interviews in CORAAL to ensure the transcriptions are mostly accurate. I sample interviews proportionally (for instance, if ATL is 10% of the corpus by quantity of data, I select 10% of the transcriptions that I check from ATL) from each division of the corpus and check that the transcriptions don't contain many errors. I also tried to balance the amount of sampling from male and female interviewees' speech to the extent this was possible. I find that the word error rate of the human transcribers is negligible: on the order of a few words, perhaps ten or fewer per transcription out of several thousand. The data I listened to for this purpose is detailed in Table 2.

| Subcorpus | Files | % Female | Proportion of Listening |
|---|---|---|---|
| ATL | ATL_se0_ag1_m_04_1 | 66.2% | 5.9% |
| | ATL_se0_ag2_f_02_1 | | |
| DCA | DCA_se1_ag1_f_04_1 | 50.0% | 23.7% |
| | DCA_se1_ag1_m_08_1 | | |
| | DCA_se1_ag4_m_02_1 | | |
| | DCA_se2_ag1_f_03_1 | | |
| | DCA_se2_ag3_m_03_1 | | |
| | DCA_se3_ag1_f_02_1 | | |
| DCB | DCB_se1_ag1_m_02_1 | 50.1% | 35.1% |
| | DCB_se2_ag2_f_02_1 | | |
| | DCB_se3_ag3_f_01_1 | | |
| | DCB_se3_ag3_m_02_1 | | |
| LES | LES_se0_ag3_m_01_1 | 25.0% | 6.7% |
| | LES_se0_ag4_f_01_1 | | |
| PRV | PRV_se0_ag1_f_02_1 | 49.8% | 10.5% |
| | PRV_se0_ag1_m_01_1 | | |
| ROC | ROC_se0_ag2_m_01_1 | 50.0% | 10.5% |
| | ROC_se0_ag2_m_01_3 | | |
| | ROC_se0_ag2_m_01_4 | | |
| | ROC_se0_ag3_f_01_1 | | |
| VLD | VLD_se0_ag3_f_01_1 | 50.0% | 10.0% |
| | VLD_se0_ag3_m_02_1 | | |

Table 2. Table detailing the interviews I listened to for the purpose of checking textual transcriptions for accuracy. Proportion of listening is calculated with regard to space used by files.

3.6. JUSTIFICATION OF APPROACH. While the ARK Tagger performs well on the task of identifying adverbs in the corpus, there is reason to be concerned that the ZAM forms may slip through the cracks. Many part-of-speech taggers rely on estimating a probability of some particular tag for each token encountered, distributed over the grammatical classes defined in the tagset. However, if the model never, or rarely, encountered a form like *different* with the tag "adverb" in its training set, it is likely to assign a near-zero probability to the "adverb" category. Such a situation is likely to arise for two reasons. First, it is possible that the ZAM construction is rare and simply did not surface in the training set at any point, or surfaced rarely enough that the tagger did not learn it well. Second, it is possible that human annotators were unaware of this construction, and mistakenly tagged the word *different* as an adjective, regardless of the specific context and use of the form.

Furthermore, the tagger cannot tell us anything about the syntactic contexts of use of any adverbs it tags. Thus, we indeed require human annotations. It is for these reasons that I make use of the regular expression search to identify all potential adverbs in the corpus. The search iterated through every lexeme in my carefully designed word list, which contains every adjective with a corresponding *-ly* form, as well as the *-ly* form itself, that appears in the CORAAL transcriptions. The regular expression search was also designed to capture a context window, ideal for determining syntactic context, as well as providing sufficient context for the part-of-speech tagger to reliably tag words.

The purpose of also performing an analysis of the tagger's function on the human-annotated dataset is twofold; firstly, I wish to re-analyze any potentially problematic data that may have been passed over initially. After recording all instances of disagreement between the human annotations and the tagger tags (false positives and negatives), I analyze these examples further with the goal of finding mistakes in the human annotations as well as potential points of improvement for the ARK Tagger. A second purpose is a sort of lemma to this study. I determine how well the ARK Tagger performs on adjective/adverb disambiguation, opening the door for potential improvements of part-of-speech taggers that are intended for use on English varieties with ZAM forms.

I grieve different.

Kendrick Lamar, "United in Grief"

## 4. Results.

4.1. REGULAR EXPRESSION SEARCH. I examine the lexemes with the most interesting distributional properties based solely on the regular expression search. Results are presented both as raw counts and as %ZAM in Table 3. %ZAM is defined as the proportion of adverbial uses (excluding the data in the A column, then) which are without *-ly* divided by the total number of adverbial uses of a lexeme. For instance, with *bad/ly*, one obtains:

$$\%\text{ZAM} = \frac{\text{count of adverbial uses of } \textit{bad}}{\text{count of adverbial uses of } \textit{bad} \text{ or } \textit{badly}} = \frac{3+45+1}{3+45+3+1} \approx 94\% \tag{5}$$

It seems that some lexemes do exhibit a sort of preference for *-ly* or the lack thereof. Particularly noteworthy are *good*, *real*, and *serious*, which all exhibit a preference for *ly* with relatively high frequency counts, while *bad*, *different*, and *quick* show the reverse. And yet despite the clear preference for *-ly, good*, *real*, and *serious* all clearly allow for use of the ZAM form. Indeed, they were some of the most frequent ZAM adverbs. Histograms for the most common ZAM lexemes and their corresponding OAM variants are available in Figures 2 and 3.

The table also shows that some lexemes even exhibit preferences for ZAM or OAM in certain contexts, though it is difficult to make claims about this without understanding the more general distribution of these words. *real* is notably the only lexeme that appeared as both OAM and ZAM in every context, and may be a good candidate for broader claims. As an adverb, its ZAM variant seemed to prefer NPAM and PREJJ over POSTVB, and strongly dispreferred the PREVB and POSTJJ environments. The frequency counts for this lexeme lend themselves as evidence to the claim that ZAM adverbs are possibly ungrammatical in the PREVB context, and are acceptable, though variable, in the NPAM, PREJJ, and POSTVB contexts. On the other hand, some lexemes never arose as ZAM adverbs in any environment. The most significant results from the data are in table 4.

A breakdown of all instances across the corpus shows that while ZAM forms are rarely adverbs in general, they still do arise at a non-trivial rate. POSTVB is the most common syntactic environment for ZAM forms to arise, while NPAM is the most preferred environment from ZAMs to arise with respect to OAM form in the same environment. The PREJJ environment also tends to allow ZAM forms, while the PREVB environment seems to generally disprefer ZAM forms. Indeed, counts for ZAM and OAM forms are of comparable magnitude for NPAM and PREJJ, as well as POSTVB to a lesser extent. As for PREVB, it is clear that the *-ly* form is strongly preferred. It is difficult to say anything about the POSTJJ category definitively due to its low overall frequency, but the count for *-ly* forms in that category is quite a bit larger, relatively speaking. A summary is available in Table 5.

A $\chi^2$ contingency-table analysis of the counts column in Table 5 shows that ZAM/OAM distributions are clearly correlated with syntactic environment across all six environments shown ($p$-value $\approx 0$, $\chi^2$ statistic $\approx 29649$, 5 degrees of freedom). The expected counts are in Table 6. A per-category visualization of OAM preference and confidence intervals for such preferences is shown in Figure 4.

| **lexeme** | A | PREVB | POSTVB | PREJJ | POSTJJ | NPAM | **%ZAM** |
|---|---|---|---|---|---|---|---|
| *funny/ly* | 231 / | | 8 / | | | | 100 |
| *healthy/ly* | 11 / | | 6 / | | | | 100 |
| *last/ly* | 579 / | | 5 / | | | | 100 |
| *tight/ly* | 79 / | | 8 / | | 1 / | | 100 |
| *bad/ly* | 422 / | 3 / | 45 / 3 | | 1 / | | 94 |
| *sure/ly* | 415 / | 96 / 6 | 1 / 4 | 1 / | | | 91 |
| *crazy/ly* | 340 / | | 6 / | 3 / | | / 1 | 90 |
| *loud/ly* | 30 / | / 1 | 14 / 1 | | | | 88 |
| *nice/ly* | 485 / | 2 / | 15 / 3 | 2 / | 2 / | | 88 |
| *loose/ly* | 7 / | | 5 / 1 | | | | 83 |
| *proper/ly* | 21 / | | 8 / 2 | | 1 | | 82 |
| *heavy/ly* | 30 / | | 4 / 5 | 1 / 2 | | | 71 |
| *quick/ly* | 12 / | 1 / 4 | 30 / 11 | | / 1 | | 66 |
| *close/ly* | 285 / | | 2 / 7 | 6 / 1 | | 6 / | 64 |
| *great/ly* | 299 / | | 4 / 3 | | | | 57 |
| *different/ly* | 1341 / | | 41 / 32 | | | | 56 |
| *serious/ly* | 65 / | / 8 | 7 / 23 | / 3 | / 2 | | 16 |
| *slow/ly* | 33 / | / 15 | 4 / 7 | | | | 15 |
| *real/ly* | 294 / 1 | 1 / 1982 | 61 / 569 | 219 / 540 | 1 / 6 | 96 / 138 | 10 |
| **sum** | 6564 / 3 | 105 / 5528 | 352 / 973 | 232 / 627 | 7 / 9 | 107 / 152 | 9 |
| *total/ly* | 3 / | / 11 | / 8 | / 50 | | 5 / 11 | 6 |
| *good/well* | 1582 / 2 | 2 / 3501 | 78 / 294 | / 31 | 1 / | / 2 | 2 |

Table 3. A few key results of the regular expression search. The slashes separate the counts for the ZAM / OAM forms. The last column is the percent of instances tagged as adverbs (ignoring A tags) which were ZAM. Raw counts are presented as well because lexemes like *real/ly* were found in ZAM form many times, though these instances comprise only 10% of adverbial instances of the lexeme in general. Criterion for inclusion in the table: the ZAM form arose in the same adverbial environment at least 4 times.

| **lexeme** | PREVB | POSTVB | PREJJ | POSTJJ | NPAM | **sum** |
|---|---|---|---|---|---|---|
| actually | 637 | 190 | 34 | 7 | | 868 |
| probably | 443 | 293 | 22 | 1 | | 759 |
| definitely | 197 | 122 | 25 | | 1 | 345 |
| basically | 198 | 125 | 10 | | | 333 |
| especially | | 209 | | | | 209 |
| finally | 140 | 5 | | | | 145 |
| literally | 95 | 44 | 3 | | | 142 |
| eventually | 94 | 12 | | | | 106 |
| honestly | 62 | 14 | 1 | | | 77 |
| necessarily | 38 | 27 | 10 | | | 75 |
| recently | 33 | 31 | | | | 64 |
| completely | 9 | 20 | 26 | 1 | 1 | 57 |
| obviously | 35 | 18 | 2 | | | 55 |
| normally | 42 | 6 | 4 | | | 52 |
| **sum** | 2023 | 1116 | 137 | 9 | 2 | 3287 |

Table 4. A breakdown of the most frequent lexemes that were strictly OAM. Criterion for inclusion in the table: used as an adverb at least 50 times, but never used as a ZAM adverb.

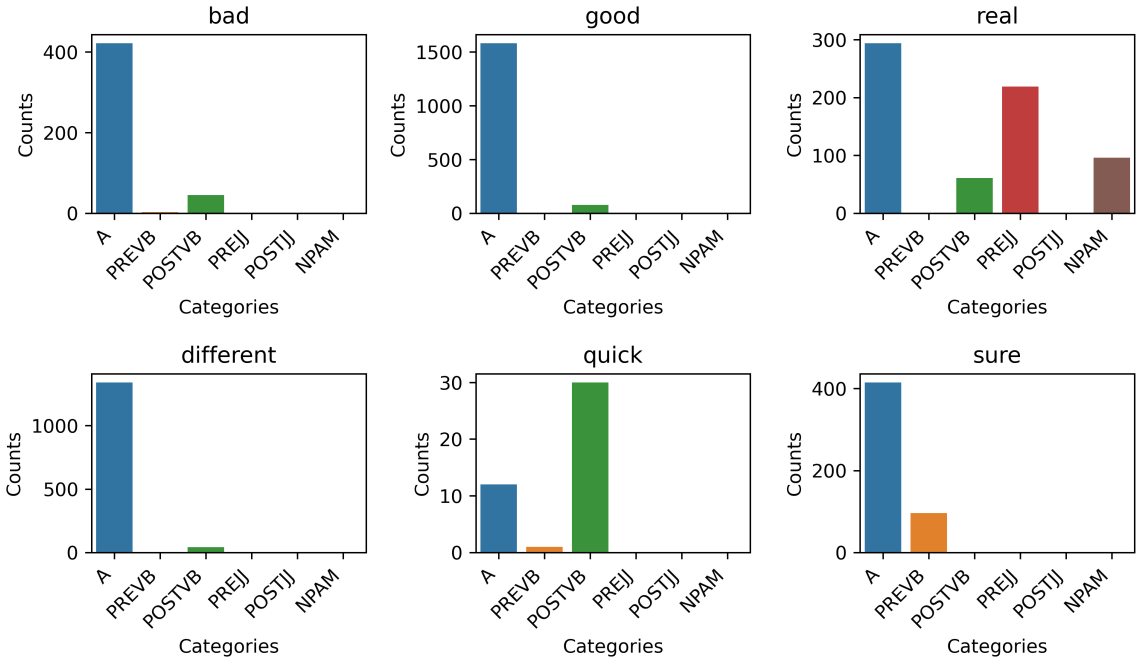


Figure 2. Histograms of common ZAM forms by environment.

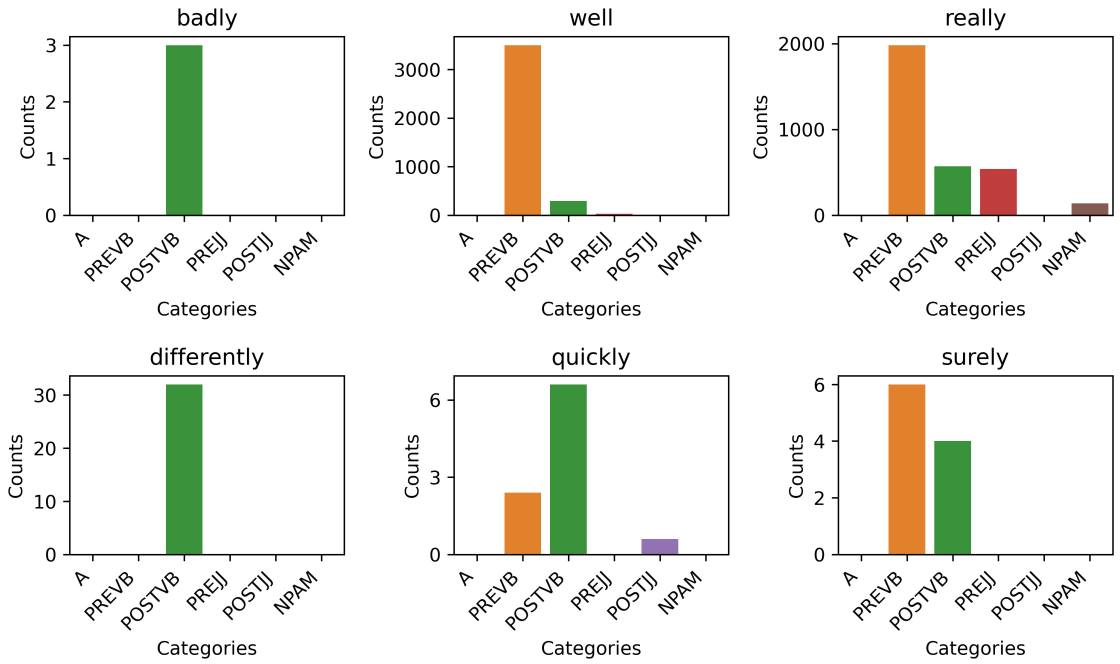


Figure 3. Histograms of common OAM forms by environment.

| **category** | **counts** (ZAM / OAM) | **%** (ZAM / OAM) | **combined counts** | **combined %** |
|---|---|---|---|---|
| A | 19372 / 14 | 95.6 / 0.1 | 19386 | 58.9 |
| PREVB | 115 / 8503 | 0.6 / 67.4 | 8619 | 26.2 |
| POSTVB | 418 / 2919 | 2.1 / 23.1 | 3339 | 10.2 |
| PREJJ | 240 / 960 | 1.2 / 7.6 | 1204 | 3.7 |
| POSTJJ | 8 / 33 | 0 / 0.3 | 41 | 0.1 |
| NPAM | 110 / 190 | 0.5 / 1.5 | 300 | 0.9 |
| **sum (excluding A)** | 891 / 12605 | 4.4 / 99.9 | 13503 | 41.1 |

Table 5. A summary of all frequency counts across the entire corpus for both OAM and ZAM forms. Each percentage refers to the proportion of all ZAM and OAM forms in the corpus which fall into each category.

| **category** | **expected counts** (ZAM / OAM) |
|---|---|
| A | 11948 / 7438 |
| PREVB | 742 / 3307 |
| POSTVB | 2058 / 1281 |
| PREJJ | 742 / 462 |
| POSTJJ | 25 / 16 |
| NPAM | 185 / 115 |
| **sum (excluding A)** | 3752 / 5181 |

Table 6. A summary of $\chi^2$ expected frequency counts across the entire corpus for both OAM and ZAM forms.

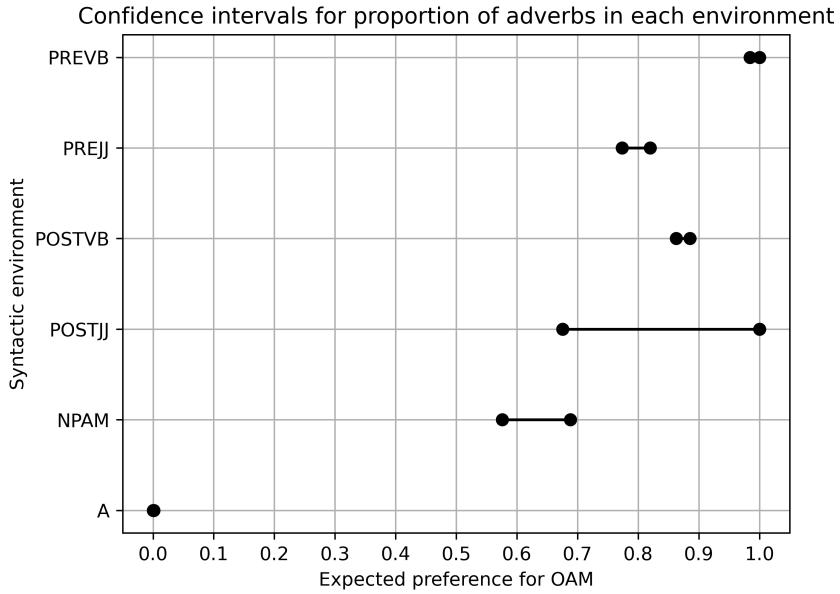


Figure 4. 95% confidence intervals for binomial tests of the expected proportion of OAM forms by category. A (adjectival use), which is something of a control group, shows a nearly total preference for ZAM, while PREVB shows a nearly total preference for OAM. All $p$-values were beneath $10^{-4}$.

A key fact is that all 898 instances of ZAM adverbs found in the corpus, save a select few, are stress-initial words (in both OAM/ZAM form). The only exceptions are detailed in Table 7. If we examine only non-monosyllabic ZAM adverbs, we find 126 with initial stress versus the 13 instances of non-initial stress detailed there. It is difficult to establish whether number of syllables or frequency count are relevant factors, or whether they interact with or make sharper the commonality of stress-initial ZAM forms. This would be an interesting direction for further research.

4.2. TAGGER-ASSISTED RE-ANALYSIS. An initial run of the ARK Tagger on the human-annotated data turned up several mistakes in the human annotations, as well as a small, contiguous section of data that was not annotated at all. As such, bringing automation into the annotation process as

| **category** | **# exceptions** | **category rate** (%) | **lexemes** |
|---|---|---|---|
| PREVB | 2 | 1.7 | *original*, *particular* |
| POSTVB | 5 | 1.2 | *aggressive*, *consistent*, *direct* $\times 2$, *naive* |
| PREJJ | 5 | 2.0 | *academic* $\times 2$, *automatic*, *exact*, *mechanical* |
| POSTJJ | 0 | 0 | |
| NPAM | 1 | 0.9 | *absolute* |
| **sum** | 13 | 1.4 | |

Table 7. ZAM forms with non-initial stress. "Number of exceptions" is the count of ZAM forms in each category with this attribute. "Category rate" is the number of exceptions divided by the total number of occurrences of ZAM in each category.

a second annotator seems to be beneficial in regard to producing more accurate human-annotated datasets. All data presented in this study is presented after making corrections to the human annotations.

After correcting these errors, the tagger produced 8889 true positives, 35 false positives, 17338 true negatives, and 723 false negatives. This yields a precision of 0.9961, a recall of 0.9247, and an $F_1$ of .9591. These are excellent scores, and they near total agreement between the human annotations and the tagger-generated tags. Since the precision is slightly better than the recall, the tagger has a slight tendency to label adverbs as adjectives, rather than the other way around. This is key, because it points to ZAM forms as a possible issue for the tagger.

Looking at this phenomenon on a per-lexeme basis, a striking resemblance to Table 3 emerges, in terms of the most interesting behavior. We note the most commonly mis-tagged words in Table 8. All of these words are the most common ZAM adverbs from Table 3. Coupled with the comparatively low recall (as compared with precision), it is clear that the tagger is struggling to perceive ZAM forms as adverbs. Despite the fact that ZAM forms occur in the same syntactic environments where OAM forms might, they are not correctly being tagged as adverbs due to their morphology. One possible reason for this is that the annotators of the data that was used to train the ARK Tagger were unaware of this construction, and did not tag ZAM forms as adverbs when and if they were encountered in the training set (or those annotators disagree with me about this phenomenon, or how to appropriately tag it). Another possibility is that ZAM forms did not occur very often in the training set, and thus, the tagger is relying mostly or entirely on the form of the word to determine its tag.

To correct this behavior, something like what is suggested by Linzen et al. (2016) might be effective. In trying to get a long short-term memory (LSTM) network to learn long-range number agreements, they claim, "these results suggest that oversampling difficult training cases may be beneficial." This could be an effective approach for a part-of-speech tagger, as both tasks (number agreement and part-of-speech tagging) rely on local, sequential information. Further research might involve training the ARK Tagger specifically on the task of adjective/adverb disambiguation using the dataset I have developed here. This may improve the ARK Tagger's ability to handle ZAM forms. Given that the ARK Tagger is a CRF model, the simplest approach would be to present it with appropriately tagged examples of ZAM adverbs in context.

| **word** | **# of incorrect tags** | ARK tag (wrong) |
|---|---|---|
| *bad* | 49 | A |
| *close* | 14 / 5 | A / R |
| *crazy* | 9 | A |
| *different* | 39 | A |
| *funny* | 7 | A |
| *good* | 77 | A |
| *great* | 4 | A |
| *healthy* | 6 | A |
| *heavy* | 5 | A |
| *last* | 5 / 2 | A / R |
| *loud* | 12 | A |
| *nice* | 19 | A |
| *proper* | 9 | A |
| *quick* | 31 | A |
| *real* | 182 / 3 | A / R |
| *serious* | 7 | A |
| *sure* | 70 / 3 | A / R |
| *tight* | 9 | A |
| *total* | 5 | A |
| *well* | 67 | A |

Table 8. Words (**not** lexeme pairs) tagged incorrectly by the ARK Tagger, and the incorrect tag given. A is adjective, R is adverb. Criterion for inclusion in the table: the word was tagged incorrectly the same way at least 4 times in ZAM form.

- You don't wanna do nothing different?
- No, man. Look, I got, I got a photo shoot coming up, man. So I just want to look good.
- Oh! Oh, you got a photo shoot coming up.
- Yeah, man.
- I mean - Oh, you Hollywood now, huh?
- Nah, man. It's just a write-up for me in a magazine.
- Yeah, my boy done went all Hollywood.

Exchange between Paper Boi and his barber, Bibby, on "Atlanta"

**5. Discussion.** This study is the first detailed analysis of ZAM and OAM adverbs in any English dialect to the best of the author's knowledge. Therefore, rather than drawing strict conclusions in this section, I intend to present a select few claims based on the evidence presented here, while still leaving room for alternative interpretations, as well as future study of this phenomenon with even more nuance and care.

It seems that there is some merit to the idea that certain lexemes prefer ZAM or OAM overall. As can be seen in Table 3, among all lexemes that seem to allow ZAM adverbs with some consistency, the % ZAM metric ranges from 2-100%, with lots of intermediate values represented (evidence of lexemes which allow some variation). While it is hard to draw conclusions due to the rarity of these constructions more generally, one might argue that if CORAAL is a reasonably accurate depiction of AAE, the % ZAM metric is a reasonably accurate determinant of a lexeme's preference to surface as a ZAM adverb in AAE more generally (i.e. outside of this corpus).

It may be the case that a lexeme with 100% ZAM adverbs (in this corpus: *funny/ly*, *healthy/ly*, *last/ly*, *tight/ly*) disallows OAM forms, but none of these words were particularly frequent overall. More data, such as might be garnered via an elicitation experiment, would be useful to address this more carefully. Notably, these four lexemes appear almost exclusively in the `POSTVB` environment.

On the other hand, Table 4 shows that there are many lexemes that arise frequently and always in OAM form, even in syntactic contexts which allow for variation more generally. Indeed, there are more lexemes that fit this pattern that are not shown in the table. Making a similar argument, one might say that a lexeme with 100% OAM adverbs, such as any lexeme from that table, might disallow ZAM forms. Some of these words were quite frequent, though confirmation of this finding via an elicitation experiment would be useful.

It is difficult to disentangle the various factors that might unite those lexemes with ZAM adverb preference. However, primary stress (or some variable heavily correlated with it) seems to be at play. Indeed, out of 898 instances of ZAM adverbs, including 139 non-monosyllabic instances, only 13 did not have initial stress. A more nuanced analysis of lexeme types might include comparisons of overall frequencies of use in the corpus, other prosodic factors such as intonation, phonological contexts of use, and social connotations that lexemes may carry. These

factors may or may not be correlated with primary stress, though they may reveal interesting patterns or deeper structural reasons for ZAM/OAM contrast in AAE regardless. Questions that could be answered about these factors might be:

1. Are more frequent lexemes more likely to surface as ZAM when used as adverbs?
2. Do certain lexemes tend to be used more often with certain intonation, such as due to their contexts of use? Do particular intonation patterns or contexts correlate with ZAM frequency?
3. Do certain lexemes tend to surface as OAM or ZAM depending on the phoneme or phonemes that immediately follow their use?
4. Do certain lexemes possibly carry social information? For instance, a lexeme such as *methodical/ly* may sound more prestigious than *careful/ly* - does one tend to surface as OAM or ZAM more frequently than the other?

A final point about individual lexemes' preferences lies in the behavior exhibited by *real/ly*. *real/ly* was the only lexeme to appear in every environment in both OAM and ZAM form. However, the ZAM form seems to be allowed in certain environments (`POSTVB`, `PREJJ`, `NPAM`) and disallowed in others (`PREVB`, `POSTJJ`). The single use of the ZAM form in each of these last two environments could be due to speech errors. The `PREVB` instance was in the context "...guys, the teenagers, don't real appreciate the roots and the..." (`PRV_se0_ag2_m_02_1`). The speaker makes no attempt to correct themself, but this is the only instance in the entire corpus in which *real* is used in this way. More notably, the `POSTJJ` instance was in the context "...It was- it was- it was- good, real. They, uh, basically- They..." (`DCB_se3_ag4_f_02_1`). The speaker was stuttering quite a lot in this instance's context. It is difficult to make claims about whether the pattern that *real/ly* exhibits is generalizable due to insufficient data, but it does conform with the overall results by syntactic category. Future research might focus specifically on this lexeme due to its commonality and productivity, as well its use as an intensifier, which was not distinguished from its use as an adverb in this study.

Indeed, the binomial confidence intervals paint a picture largely confirming my hypotheses about syntactic environments. It seems that the `PREVB` environment almost categorically refuses ZAM adverbs across all lexemes (this is true even if we exclude the admittedly ill-behaved lexemes *real/ly* and *good/well*, which together account for a slight majority of `PREVB` instances). There is a non-zero chance that `POSTJJ` does as well, though a lack of data about that environment is evident. Conversely, the `POSTVB`, `PREJJ`, and `NPAM`, in order of increasing preference for ZAM adverbs, all clearly exhibit some degree of variation. It is clear that the claim made by Schneider & Kortmann (2004), that AAE generally disprefers *-ly,* is not strictly true. The syntactic environment is key, and in particular, AAE strongly prefers *-ly* in preverbal contexts, while generally preferring it across all adverbial contexts.

As for the ARK Tagger, it is abundantly clear that ZAM adverbs are mostly tagged incorrectly as adjectives. It's unclear at present why this is. It could have to do with the way its training data was labeled by human annotators. The ZAM adverb may also simply be too rare in the the ARK Tagger's training data, such that the Tagger does not "understand" it. Both are probably true to some extent. To remedy this, it might be worth training the ARK Tagger additionally on the examples in the dataset I have built for this study, with correct adjective and adverb

tags where applicable. Linzen et al. (2016) found that neural language models might benefit from such task-specific training when shortcomings are evident. Although the ARK Tagger uses a CRF model, this idea may be applicable. In the future, it will be important to make sure that any part-of-speech tagger that may receive AAE input (or input from other varieties of English in which this construction is prominent) is aware of this phenomenon and has been trained on appropriately labeled data.

Other future directions for research should involve further breakdowns and statistical analysis of the dataset presented here. No analysis has been done on the bases of region or gender, each of which may yield interesting results regarding this phenomenon. Furthermore, some lexemes have not been given sufficient attention. The especially noteworthy properties that *real/ly* exhibits are now well-established, and this behavior probably warrants its own study, given also that this lexeme makes up more than 10% of the dataset. Another key lexeme that deserves further attention is *sure/ly*. The ZAM form *sure* appeared in preverbal contexts 96 times; there were only ten other instances of preverbal ZAM forms in the entire corpus. This construction arises in contexts such as "...it's a new car out, and I sure could use one. He said..." (`DCB_se2_ag4_f_05_1`). It is unclear what is so special about *sure* that allows it to be used in this way whereas other ZAM adverbs are generally forbidden in this context. Even more confusingly, it also seems to be acceptable to use *surely* in such contexts, such as in "...I believe so. I surely believe they is. Mmhm..." (`LES_se0_ag3_m_02_1`). Finally, the set of excluded items is large, and many of these, if not all, warrant some consideration that could not be given to them in this study. The excluded items are listed in Table 9 along with their reason for exclusion.

| **lexeme** | **excluded because...** |
|---|---|
| *atypical/ly* | used for markup |
| *elementary/ly* | overwhelmingly a noun |
| *even/ly* | semantic difference |
| *fine/ly* | semantic difference |
| *first/ly* | overwhelmingly ZAM and adverbial + time constraints |
| *hard/ly* | semantic difference |
| *high/ly* | semantic difference |
| *inaudible/ly* | used for markup |
| *just/ly* | semantic difference |
| *kind/ly* | OAM form is a separate adjective as well |
| *late/ly* | semantic difference |
| *laughing/ly* | used for markup |
| *like/ly* | semantic difference + part-of-speech issues |
| *mocking/ly* | used for markup |
| *most/ly* | ill-behaved + time constraints |
| *only* | lack of true ZAM form (one) |
| *over/ly* | semantic difference |
| *poor/ly* | semantic difference |
| *pretty/ly* | semantic difference |
| *right/ly* | semantic difference + time constraints |
| *short/ly* | semantic difference |
| *single/ly* | overwhelmingly ZAM and adjectival |
| *unintelligible/ly* | used for markup |
| *very/ly* | lack of true OAM form (verily) |

Table 9. The set of lexemes which were explicitly excluded from this study along with justification for doing so. It may be worth eventually analyzing some of these lexemes as well, though many are simply ill-behaved or not very interesting.

## References


Bailey, Guy & Erik Thomas. 2021. Some aspects of African-American vernacular English phonology. In *African-American English*, 93–118. Routledge.

Bambara, Toni Cade. 1972. *The Lesson*. New York, NY, USA: Bedford/St.Martin's.

Baugh, John. 1983. *Black Street Speech: Its History, Structure, and Survival*. University of Texas Press. Google-Books-ID: gzaRH7L0h84C.

Blodgett, Su Lin, Lisa Green & Brendan O'Connor. 2016. Demographic Dialectal Variation in Social Media: A Case Study of African-American English. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 1119–1130. Austin, Texas: Association for Computational Linguistics. https://doi.org/10.18653/v1/D16-1120. https://aclanthology.org/D16-1120.

Blodgett, Su Lin, Johnny Wei & Brendan O'Connor. 2018. Twitter Universal Dependency Parsing for African-American and Mainstream American English. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1415–1425. Melbourne, Australia: Association for Computational Linguistics. https://doi.org/10.18653/v1/P18-1131. https://aclanthology.org/P18-1131.

Brants, Thorsten. 2000. TnT-a statistical part-of-speech tagger. *arXiv preprint cs/0003055* .

Bybee, Joan L. 1985. *Morphology: A Study of the Relation Between Meaning and Form*. John Benjamins Publishing.

Charity, Anne Harper. 2008. African American English: An Overview. *Perspectives on Communication Disorders and Sciences in Culturally and Linguistically Diverse (CLD) Populations* 15(2). 33–42. https://doi.org/10.1044/cds15.2.33. https://pubs.asha.org/doi/abs/10.1044/cds15.2.33. Publisher: American Speech-Language-Hearing Association.

Cole, Lorraine Theresa. 1980. *A Developmental Analysis of Social Dialect Features in the Spontaneous Language of Preschool Black Children*. United States – Illinois: Northwestern University Ph.D. https://www.proquest.com/docview/288405893/abstract/EDEBF71EDAD34AA8PQ/1. ISBN: 9781083639059.

Craig, Holly. 2019. *: The Role of Dialectal Code Switching*. New York: Routledge. https://doi.org/10.4324/9781315813394.

Craig, Holly K., Connie A. Thompson, Julie A. Washington & Stephanie L. Potter. 2003. Phonological Features of Child African American English. *Journal of Speech, Language, and Hearing Research* 46(3). 623–635. https://doi.org/10.1044/1092-4388(2003/049). https://pubs.asha.org/doi/full/10.1044/1092-4388(2003/049). Publisher: American Speech-Language-Hearing Association.

Cukor-Avila, Patricia & Ashley Balcazar. 2019. Exploring Grammatical Variation in the Corpus of Regional African American Language. *American Speech* 94(1). 36–53. https://doi.org/10.1215/00031283-7321989. https://doi.org/10.1215/00031283-7321989.

De Villiers, Jill G. & Valerie E. Johnson. 2007. The information in third-person /s/: acquisition across dialects of American English. *Journal of Child Language* 34(1). 133–158. https://doi.org/10.1017/S0305000906007768. https://www.cambridge.org/core/journals/journal-of-child-language/article/

information-in-thirdperson-s-acquisition-across-dialects-of-american-eng 6D2DD06A823C997AEDCB1A948195A842. Publisher: Cambridge University Press.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee & Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. Minneapolis, Minnesota: Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423. https://aclanthology.org/N19-1423.

Eberhardt, Maeve. 2008. The Low-Back Merger in the Steel City: African American English in Pittsburgh. *American Speech* 83(3). 284–311. https://doi.org/10.1215/00031283-2008-021. https://doi.org/10.1215/00031283-2008-021.

Eckert, Penelope. 2006. Communities of practice. *CONCISE ENCYCLOPEDIA OF* 109.

Emonds, Joseph E. 1976. *A transformational approach to English syntax: Root, structure-preserving, and local transformations*. Academic press.

Fasold, Ralph W. & Walt Wolfram. 1972. Some linguistic features of Negro dialect. *Language, Speech, and Hearing Services in Schools* 3(4). 16–49. Publisher: ASHA.

Forrest, Jon & Walt Wolfram. 2019. The Status Of (ING) in African American Language: A Quantitative Analysis of Social Factors and Internal Constraints. *American Speech* 94(1). 72–90. https://doi.org/10.1215/00031283-7308049. https://doi.org/10.1215/00031283-7308049.

Giegerich, Heinz J. 2012. The morphology of -ly and the categorial status of adverbs in English. *English Language & Linguistics* 16(3). 341–359. https://doi.org/10.1017/S1360674312000147. https://www.cambridge.org/core/journals/english-language-and-linguistics/article/morphology-of-ly-and-the-categorial-status-of-adverbs-in-english1/B97CE3AC6AA6DECFD67EFA7525C3A6BA. Publisher: Cambridge University Press.

Gimpel, Kevin, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan & Noah A. Smith. 2010. Part-of-speech tagging for twitter: Annotation, features, and experiments. Tech. rep. Carnegie-Mellon Univ Pittsburgh Pa School of Computer Science.

Green, Lisa J. 2002. *African American English: A Linguistic Introduction*. Cambridge University Press.

Green, Lisa J. & Walter Sistrunk. 2015. Syntax and Semantics in African American English. In *The Oxford Handbook of African American Language* Oxford Handbooks, Oxford University Press. https://academic.oup.com/edited-volume/28056/chapter/212012532.

Heller, Janet Ruth. 2003. Toni Cade Bambara's Use of African American Vernacular English in "The Lesson". *Style* 37(3). 279–293. https://libproxy.uccs.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=f5h&AN=12677055&site=ehost-live. Publisher: Pennsylvania State University Press.

Johnson, Valerie E. 2005. Comprehension of Third Person Singular /s/ in AAE-Speaking Children. *Language, Speech, and Hearing Services in Schools* 36(2). 116–124. https://doi.org/10.1044/0161-1461(2005/011). https://pubs.asha.org/doi/full/10.1044/0161-1461%282005/011%29. Publisher: American Speech-Language-Hearing Association.

Jones, Taylor, Jessica Rose Kalbfeld, Ryan Hancock & Robin Clark. 2019. Testifying while black: An experimental study of court reporter accuracy in transcription of African American English. *Language* 95(2). e216–e252. https://doi.org/10.1353/lan.2019.0042. https://muse.jhu.edu/pub/24/article/727848. Publisher: Linguistic Society of America.

Jurafsky, Dan. 2000. *Speech & language processing*. Pearson Education India.

Jrgensen, Anna, Dirk Hovy & Anders Sgaard. 2016. Learning a POS tagger for AAVE-like language. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1115–1120. San Diego, California: Association for Computational Linguistics. https://doi.org/10.18653/v1/N16-1130. https://aclanthology.org/N16-1130.

Kendall, Tyler & Charlie Farrington. 2021. The Corpus of Regional African American Language. Version 2021.07. https://oraal.uoregon.edu/coraal.

Killie, Kristin. 2015. Secondary grammaticalization and the English adverbial -ly suffix. *Language Sciences* 47. 199–214. https://doi.org/10.1016/j.langsci.2014.10.003. https://www.sciencedirect.com/science/article/pii/S0388000114001296.

Kleene, Stephen C. 1956. Representation of events in nerve nets and finite automata. *Automata studies* 34. 3–41. Publisher: Princeton, NJ.

Koenecke, Allison, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R. Rickford, Dan Jurafsky & Sharad Goel. 2020. Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences* 117(14). 7684–7689. https://doi.org/10.1073/pnas.1915768117. https://www.pnas.org/doi/abs/10.1073/pnas.1915768117. Publisher: Proceedings of the National Academy of Sciences.

Kortvelyessy, Livia & Pavol Stekauer. 2014. Derivation in a Social Context. In Rochelle Lieber & Pavol Stekauer (eds.), *The Oxford Handbook of Derivational Morphology*, 0. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199641642.013.0023. https://doi.org/10.1093/oxfordhb/9780199641642.013.0023.

Labov, William. 1969. *Language in the inner city: Studies in the Black English vernacular* 3. University of Pennsylvania Press.

Labov, William, Paul Cohen, Clarence Robins & John Lewis. 1968. A Study of the Non-Standard English of Negro and Puerto Rican Speakers in New York City. Volume I: Phonological and Grammatical Analysis. Tech. rep. Columbia University New York, NY, USA. https://eric.ed.gov/?id=ED028423. ERIC Number: ED028423.

Lafferty, John, Andrew McCallum & Fernando CN Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *International Conference on Machine Learning*, .

Lavandera, Beatriz R. 1978. Where Does the Sociolinguistic Variable Stop? *Language in Society* 7(2). 171–182. https://www.jstor.org/stable/4166996. Publisher: Cambridge University Press.

Lightfoot, David W. & Jonathan Havenhill. 2019. *Variable Properties in Language: Their Nature and Acquisition*. Georgetown University Press. Google-Books-ID: KaWSDwAAQBAJ.

Ling, Wang, Chris Dyer, Alan W Black, Isabel Trancoso, Ramn Fermandez, Silvio Amir, Lus Marujo & Tiago Lus. 2015. Finding Function in Form: Compositional Character Models for Open Vocabulary Word Representation. In *Proceedings of the 2015 Conference on*

*Empirical Methods in Natural Language Processing*, 1520–1530. Lisbon, Portugal: Association for Computational Linguistics. https://doi.org/10.18653/v1/D15-1176. https://aclanthology.org/D15-1176.

Linzen, Tal, Emmanuel Dupoux & Yoav Goldberg. 2016. Assessing the Ability of LSTMs to Learn Syntax-Sensitive Dependencies. *Transactions of the Association for Computational Linguistics* 4. 521–535. https://doi.org/10.1162/tacl_a_00115. https://doi.org/10.1162/tacl_a_00115.

Loper, Edward & Steven Bird. 2002. Nltk: The natural language toolkit. *arXiv preprint cs/0205028* .

Lyons, John. 1966. Towards a notional theory of the parts of speech. *Journal of Linguistics* 2(2). 209–236. https://doi.org/10.1017/S0022226700001511. https://www.cambridge.org/core/journals/journal-of-linguistics/article/abs/towards-a-notional-theory-of-the-parts-of-speech/BC327CB92CB5E96BF35839A4E402DD74. Publisher: Cambridge University Press.

Macaulay, Ronald K. S. 1995. The Adverbs of Authority. *English World-Wide* 16(1). 37–60. https://doi.org/10.1075/eww.16.1.03mac. https://www.jbe-platform.com/content/journals/10.1075/eww.16.1.03mac. Publisher: John Benjamins.

Manning, Christopher D. 2011. Part-of-speech tagging from 97% to 100%: is it time for some linguistics? In *International conference on intelligent text processing and computational linguistics*, 171–189. Springer.

Miller, George A. 1995. WordNet: a lexical database for English. *Communications of the ACM* 38(11). 39–41. https://doi.org/10.1145/219717.219748. https://dl.acm.org/doi/10.1145/219717.219748.

Mufwene, Salikoko S., John R. Rickford, Guy Bailey & John Baugh. 2021. *African-American English: Structure, History, and Use*. Routledge. Google-Books-ID: dJZEEAAAQBAJ.

Nevalainen, Terttu. 2004. *Three perspectives on grammaticalization*. John Benjamins Publishing Company.

Nevalainen, Terttu. 2008. Social variation in intensifier use: constraint on -ly adverbialization in the past? *English Language & Linguistics* 12(2). 289–315. https://doi.org/10.1017/S1360674308002633. https://www.cambridge.org/core/journals/english-language-and-linguistics/article/social-variation-in-intensifier-use-constraint-on-ly-adverbialization-ir
5C51DC6155D913AE228449EECAD6DF46. Publisher: Cambridge University Press.

Newkirk, Turner Brandi L. & Lisa Green. 2021. Language Use and Development in Third-Person Singular Contexts: Assessment Implications. *Language, Speech, and Hearing Services in Schools* 52(1). 16–30. https://doi.org/10.1044/2020_LSHSS-19-00109. https://pubs.asha.org/doi/full/10.1044/2020_LSHSS-19-00109. Publisher: American Speech-Language-Hearing Association.

Newkirk-Turner, Brandi L. & Lisa Green. 2016. Third person singular-s and event marking in child African American English. *Linguistic Variation* 16(1). 103–130. Publisher: John Benjamins.

Owoputi, Olutobi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider & Noah A. Smith. 2013. Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 380–390. Atlanta,

Georgia: Association for Computational Linguistics. https://aclanthology.org/N13-1039.

Prieto, Victor Moises. 2005. *Spanish evaluative morphology: Pragmatic, sociolinguistic, and semantic issues*. United States – Florida: University of Florida Ph.D. https://www.proquest.com/docview/304995116/abstract/6843840C5A604B8APQ/1. ISBN: 9780542353734.

Radford, Andrew. 1988. *Transformational Grammar: A First Course*. Cambridge University Press. Google-Books-ID: yhkIh_uMwyIC.

Reveron, Wilhelmina Wright. 1979. *The Acquisition of Four Black English Morphological Rules by Black Preschool Children.* United States – Ohio: The Ohio State University Ph.D. https://www.proquest.com/docview/302951444/citation/2F4EA6F5D7624451PQ/1. ISBN: 9798661987853.

Schneider, E.W. & B. Kortmann. 2004. *A Handbook of Varieties of English: Morphology and syntax* A Handbook of Varieties of English: A Multimedia Reference Tool. Walter de Gruyter GmbH & Company KG. https://books.google.com/books?id=aI1iAAAAMAAJ.

Smitherman, Geneva. 1977. *Talkin and Testifyin: The Language of Black America*. Wayne State University Press. Google-Books-ID: HXD7pYv80bUC.

Tagliamonte, Sali & Rika Ito. 2002. Think really different: Continuity and specialization in the English dual form adverbs. *Journal of Sociolinguistics* 6(2). 236–266. https://doi.org/10.1111/1467-9481.00186. https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9481.00186. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9481.00186.

Thede, Scott M. & Mary P. Harper. 1999. A second-order Hidden Markov Model for part-of-speech tagging. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics* ACL '99, 175–182. USA: Association for Computational Linguistics. https://doi.org/10.3115/1034678.1034712. https://doi.org/10.3115/1034678.1034712.

Thomas, Erik R. 2007. Phonological and Phonetic Characteristics of African American Vernacular English. *Language and Linguistics Compass* 1(5). 450–475. https://doi.org/10.1111/j.1749-818X.2007.00029.x. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-818X.2007.00029.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-818X.2007.00029.x.

Thomas, Erik R. & Phillip M. Carter. 2006. Prosodic rhythm and African American English. *English World-Wide* 27(3). 331–355. https://doi.org/10.1075/eww.27.3.06tho. https://www.jbe-platform.com/content/journals/10.1075/eww.27.3.06tho. Publisher: John Benjamins.

Thompson, Ken. 1968. Programming techniques: Regular expression search algorithm. *Communications of the ACM* 11(6). 419–422. Publisher: ACM New York, NY, USA.

Washington, Julie A., Holly K. Craig & Amy J. Kushmaul. 1998. Variable Use of African American English Across Two Language Sampling Contexts. *Journal of Speech, Language, and Hearing Research* 41(5). 1115–1124. https://doi.org/10.1044/jslhr.4105.1115. https://pubs.asha.org/doi/full/10.1044/jslhr.4105.1115. Publisher: American Speech-Language-Hearing Association.

Wolfram, Walt. 2008. Urban African American Vernacular English: morphology and syntax. In *Urban African American Vernacular English: morphology and syntax*, 1511–1532. De Gruyter Mouton. https://doi.org/10.1515/9783110197181-097. https://www.degruyter.com/document/doi/10.1515/9783110197181-097/html?lang=en.

Wolfram, Walt & Ralph W. Fasold. 1974. *The Study of Social Dialects in American English*. Prentice-Hall. Google-Books-ID: iNN5AAAAIAAJ.

Wolfram, Walt & Natalie Schilling. 2015. *American English: dialects and variation*. John Wiley & Sons.

Wolfram, Walter A. 1969. *A Sociolinguistic Description of Detroit Negro Speech* (Urban Language Series 5). Washington, D.C.: Center for Applied Linguistics.

Zwicky, Arnold M. 1995. Why English adverbial-ly is not inflectional. In *Papers from the 31st regional meeting of the Chicago Linguistic Society*, 513–535.