

Some notes on “infinitival perfects” in Appalachian English

Christina Tortora, CUNY (CSI and The Graduate Center)

[work with Beatrice Santorini & Greg Johnson]

1. The Audio-Aligned and Parsed Corpus of Appalachian English (<http://csivc.csi.cuny.edu/aapcappE/>) Tortora, C., B. Santorini, F. Blanchette, & C.E.A Diertani (to appear)

The AAPCAppE is a publicly available corpus of vernacular speech which is:

- Available online by the end of 2016!
- 1,024,043 words / 127,375 sentence tokens
- syntactically annotated according to PPCHE method (and searchable by any standard tree query language e.g., *CorpusSearch*, Randall 2009)
- accompanied by a full set of digitized recordings of the underlying speech signal, in the form of *.wav* files, text-searchable using Praat (Boersma and Weenink 2011)

Underlying speech signal: The AAPCAppE is based on the speech from oral history project recordings housed at various colleges and institutions in the Appalachian region:

I. Dante Oral History Project (DOHP). Collection of interviews on cassette tape with residents of Dante, VA (recorded 1997-98). Recordings are housed at, and curated by, the Archives of Appalachia at ETSU.

II. Joseph Hall Collection (JHall). Interviews with residents of the Great Smoky Mountains in Tennessee and North Carolina (1939); collector: Joseph Hall.

III. Appalachian Oral History Project (AOHP_I) at Alice Lloyd College, in Pippa Passes, KY. This history project was conducted from 1971-75 and its materials are housed in the library at Alice Lloyd College, Pippa Passes, Kentucky.

IV. Appalachian Oral History Project (AOHP_II) at Appalachian State University, in Boone, NC. This history project was conducted from the 1960s through the 1980s, and its materials are housed in the library at Appalachian State, in Boone, NC.

V. The Appalachian Archive (SKCTC) at Southeast Kentucky Community and Technical College, in Cumberland, KY. This history project was conducted from the 1960s through the 1980s, and its materials are housed in the library at Southeast Kentucky Community and Technical College, in Cumberland, KY.

2. The infinitival perfect in the AAPCAppE

2.1 Infinitival *to*

The AAPCAppE exhibits structures with infinitival *to* followed by what appears to be a participial verb form, instead of an infinitival verb (e.g., *to go*), such as *to went*, *to been*, *to had*, *to got*, *to gone*, *to done*, *to slipped off*, *to worked*, *to told*, *to killed*, *to gave*,

- (1) a. And them mules went right on up there, just where they was supposed **to went**.
cf. *...just where they was supposed to go*.
- b. But the men should've refused **to went** in. (> into the mines)
cf. *The men should've refused to go in*.
- c. Course, if it'd've lasted much longer, he would've had **to went**. (> off to war)
cf. *If it'd've lasted much longer, he would've had to go off to war*.

- d. That was supposed **to been** a rare seed.
cf. *That was supposed **to be** a rare seed*
- e. She continued to work as long as she was able to work, which she had **to been** up in her sixties...
cf. *...she had **to be** up in her sixties (..when she quit working)*
- f. I'd've loved **to had** them back. (> about her grown children)
cf. *I'd've loved **to have** them back*
- g. I'[\u0259z] supposed **to got** one. (= an award)
cf. *I was supposed **to get** an award.*
- h. I would've really loved **to got** that baseball. (> foul ball in the stands at a game)
cf. *I would've really loved **to get** that baseball.*
- i. They wouldn't knowed what **to done** with the underclothes they've got today.
cf. *They wouldn't have known what **to do** with the underclothes they've got today.*

Is this none other than the **infinitival to-perfect**, with “silent HAVE,” as in (1)?

- (1') a. And them mules went right on up there, just where they was supposed **to HAVE went**.
b. But the men should've refused **to HAVE went** in. (> into the mines)
c. Course, if it'd've lasted much longer, he would've had **to HAVE went**. (> off to war)
d. That was supposed **to HAVE been** a rare seed. etc.

What suggests this? Like infinitival *to* in (1), modals in the AAPCAppE can also be immediately followed by a participial verb form, without an intervening infinitival auxiliary *have*:

2.2 Modals

- (2) a. I said, No, you the one **should said** something, cuz I wasn't going to say nothing...
b. But he **shouldn't** never **told** me like he did.
c. You **could heard** a pin drop.
d. I **must been** four or five years old.
e. I don't know what in the world I **would done** if it wasn't for Ginny.
f. They **wouldn't knowed** what to done with the underclothes they've got today.

As with (1) above, it seems reasonable to posit that these structures too are hidden **infinitival perfects**, with a “silent HAVE”:

- (2') a. I said, No, you the one **should HAVE said** something, cuz I wasn't going to say nothing...
b. But he **shouldn't** never **HAVE told** me like he did.
c. You **could HAVE heard** a pin drop.
d. I **must HAVE been** four or five years old. etc.

(3) What is the nature of this null element?

- a. is it the silent version of the morpheme *have*? (i.e., silent HAVE), or
b. is there a morpheme *have* in the underlying representation but it's just deleted via a phonological reduction rule? or
c. is there no such morpheme in the underlying representation at all? (radically missing)
d. AND FINALLY: should we be thinking of this morpheme (which is typically pronounced [\u0259v] or [\u0259]) not as a form of the auxiliary *have*, but rather, as the “complementizer” *of*, a la Kayne 1997?

To put aside for now. Though please note that wherever I say *have*, it's in reference to a form that is always pronounced [\u0259v] or [\u0259]

2.3 Summary: modal-perfects and *to*-perfects

- an **infinitival perfect** is form of the auxiliary verb *have* (typically pronounced [əv] or [ə]) plus a participle
 - UNDER MODAL: She would **have said**...
 - UNDER *TO*: She has to **have said**...
- The data in (1) and (2), exhibiting *to* + participle and modal + participle, might reasonably be taken to involve infinitival perfects with a silent HAVE (with caveats in (3))

3. The AAPCAppE: the behavior of silent infinitival perfects with modals versus *to*

- The data in (1) and (2) at first glance give the impression that silent HAVE is a uniform phenomenon with infinitival *to* and modals.
- Previous literature, Montgomery & Hall (2004): “[a]uxiliary *have* and *had* are sometimes elided in Smokies speech, especially between a modal verb and a past participle...; [this phenomenon is] more favored under a modal verb.”
- **Our study reveals the opposite.**

It is true that infinitival perfects with modals are overall more frequent than infinitival perfects with *to*:

- (4) a. modals + infinitival perfect: 797 tokens
 b. *to* + infinitival perfect: 69 tokens

However, despite their differences in absolute frequency, silent HAVE is relatively less frequent with modals, and relatively more frequent with infinitival *to*:

- (5) **TABLE 1:** Frequency of [modal + infinitival perfect] versus [*to* + infinitival perfect]

	overt <i>have</i> [hæv] or [əv] or [ə] or [v]	silent <i>HAVE</i>	total
modals + infinitival perfect	715 (90%)	82 (10%)	797
<i>to</i> + infinitival perfect	19 (28%)	50 (72%)	69

- there is a total of 797 infinitival perfects with modals
- there is a total of 69 infinitival perfects with *to*
- only 10% of the infinitival perfects with modals occur with silent HAVE
- in contrast, 72% of the infinitival perfects with *to* occur with silent HAVE

- (6) **TABLE 2:** [modal + infinitival perfect] broken down by modal

	can	could	may	might	must	need	ought	shall	should	will	would	zero	total
[hæv]	0	16	14	15	65	0	0	0	4	1	51	0	166
[əv] / [ə]	0	83	7	67	69	0	0	0	24	0	272	1	523
‘ve	0	9	0	2	2	0	0	0	0	0	12	1	26
silent <i>HAVE</i>	0	13	4	1	7	0	2	0	2	0	49	4	82
total	0	121	25	85	143	0	2	0	30	1	384	6	797

4. What's different about *to*-perfects?

4.1 The “pleonastic” Sequence-of-Tense (SoT) infinitival perfect vs. true anterior tense

There may be a previously unnoted semantic difference between **silent HAVE vs. overt *have* with *to*-perfects** indicating that the “silent HAVE construction” is a marker of a particular tense interpretation (or lack thereof!)

4.1.1 The “Sequence of Tense” (SoT) phenomenon

Zagona (2014):

- (12) Terry believed that Sue was pregnant.
- The time of Sue's pregnancy precedes the time of Terry's belief (precedence)
 - The time of Sue's pregnancy overlaps with the time of Terry's believe (simultaneity)

The **simultaneity** interpretation is the **pleonastic / SoT** interpretation (= not a true anterior)

Another example:

- (13) I thought you **meant** how big the entire corpus **was**, in terms of sentence tokens. (*was = is*)

Here, the form *was* is “pleonastic” (or put differently, has an “SoT interpretation”); it does not have a true anterior interpretation. This is simply a case of **formal agreement** in tense features (Zagona 2014).

4.1.2 Perfect SoT in written English

English has also historically exhibited a pleonastic (SoT) tense interpretation with the infinitival perfect.

PPCHE

- (14) a. He had intended **to have gone** to London with Mr. Oates. (= *He had intended to go to London*)
 b. If the prisoner had chosen **to have staid** in France... (= *If X had chosen to stay in France...*)

Molencski (2003): pleonastic perfect has essentially been rigorously proscribed out of the language, though one can still find the following variant in *would have loved to...* type contexts:

- (15) I would have liked **to have settled** down sooner.

However, in writing, editors (e.g. of newspapers) routinely correct forms like that in (15) before going to press, to one of two possible variant forms:

- (16) a. I would have liked **to settle** down sooner.
 b. I would like **to have settled** down sooner.

Do not lose sight of fact that *to*-perfects can also have a true anterior interpretation (cf. (12)):

- (17) a. Mary believes Sue **to have won** the prize. (true anterior)
 b. Mary believed Sue **to have won** the prize. (ambiguous between SoT and anterior, like (12))

4.1.3 Perfect SoT in vernacular speech (the AAPCAppE)

Some preliminary notes:

- **to-perfects are rare:** Recall from Table 1 in (5) that we have only 69 tokens of *to*-perfects in the corpus. The corpus is comprised of 127,375 sentence tokens. This means that *to*-perfects occur only .05% of the time.
- **the rarity is not just a function of vernacular speech:** This rarity is not due to the vernacular-speech nature of the corpus: the *Penn Parsed Corpora of Historical English* (PPCHE) exhibit an equally rare rate of occurrence of *to*-perfects:
 - Early Mod. English (years 1500-1710): .05% of all tokens (58 of 105,614 sentence tokens)
 - Mod. British English (years 1700-1899): .07% of all tokens (129 of 178,160 sentence tokens)

Hypothesis: silent HAVE in *to*-perfects is a marker pleonastic SoT.

Recall (1): the majority of our *to*-perfect examples with silent HAVE correlate with a SoT interpretation:

- (1) a. And them mules went right on up there, just where they was supposed **to went**.
cf. *...just where they was supposed to go.*
- b. But the men should've refused **to went** in. (> into the mines)
cf. *The men should've refused to go in.*
- c. Course, if it'd've lasted much longer, he would've had **to went**. (> off to war)
cf. *If it'd've lasted much longer, he would've had to go off to war.* etc.

Consider also the following example:

(18) Wouldn't you like **to worked** in the mines today?

(asked by an AppE speaker, to an AppE interviewee, who was talking about how hard it was to work in the mines back in the day; the AppE interviewer notes that things are a lot better today, and then asks (18))

The example in (18) indicates that the tense agreement is purely **formal agreement**, and not interpretive: the formal [+past] marking of *would* triggers the SoT.

TABLE 13 Frequency of *to*-perfects with **SoT** (= pleonastic) versus **anterior** interpretation:

	overt have	silent HAVE	total
anterior interpretation	4 (57%)	3 (43%)	7
SoT interpretation	14 (25%)	41 (75%)	55
unclear	1	6	7
total	19	50	69

- does this indicate that the silent HAVE structure is a form specializing for the SoT interpretation?
- if so, why has it specialized in this way? a response to the prescriptivist pressure to avoid pleonastic *to*-perfects?

4.1.4 Problems with corpora

- we have to rely on a very careful checking of the interpretation of the *to*-perfect in context; there are cases where we cannot tell (see example (19))
- the numbers are so low, any error could tip the scale in the direction of the simple fact that silent HAVE is more frequent than overt *have*, possibly not for any reasons to do with morpho-syntax / semantics (see 4.2)
- leads to the need for back up study, using experimentation

(19) Um and I thought that I would like(d) to [æv] seen him uh mm do something with that, that talent that he had there.

4.1.5 Conceptual problems (independent of whether data is experimental or from a corpus)

[A] **What do we count** when counting examples of silent HAVE with infinitival *to*? What to do with examples like (20)?

(20) a. He should've refused **to give** it to him.

- From the perspective of the “standard,” example (20a) unambiguously involves an embedded infinitival
- However, from the perspective of an AppE speaker, the bare verb form *give* (and many other bare forms!) is robustly used as a participle. So what parse do we assume that the speaker intended? That in (20b), or that in (20c)?

(20) b. He should've refused **to HAVE_{silent} give_{past-participle}** it to him.

(20) c. He should've refused **to give_{bare-inf}** it to him.

- Side note speculation: is it possible that...
 - (a) the existence of strings such as that in (20a), combined with
 - (b) the fact of variable use of the bare form as a past participle in other contexts, combined with
 - (c) historic use of the perfect form as a pleonastic tense (as in 14b)

...is what gave rise to the silent HAVE grammar (as in e.g. (1))?

[B] **What do we count** when counting examples of silent HAVE with “modals” ?

- Note that the corpus exhibits numerous variants of the following (leading to the question of whether *had* in such contexts should be counted as a modal):

(21) a. If only he had ___ left sooner...

b. If only he had [æv] done that sooner... (v. Kayne 1997)

- From the perspective of the “standard,” example (21a) is to be expected, and we wouldn't assume a silent HAVE
- However, from the perspective of AppE speakers: we have no reason to assume they would NOT posit a silent have in (21a) (given the existence of (21b)), as in (21c):

(21) c. If only he had HAVE_{silent} done that sooner.... (cf. (21a))

[C] How do we count examples with 'd ?

Consider (22a):

(22) a. If only he'd left sooner...

- From the perspective of the “standard,” example (22a) is to be expected, under a parse where 'd = *had*
- However, from the perspective of AppE speakers: there are actually three possible parses for the string in (22a)

How are there three possible parses for (22a)? Four independent (?) issues conspire to give rise to the possible parses for AppE speakers:

FOUR CONSPIRING PHENOMENA IN APPE:

- i. The form 'd is ambiguous between *had* and *would* (cf. *he'd put it on the table*)
- ii. The modal *would* variably allows for an embedded infinitival perfect with silent HAVE
- iii. The verb *had* (in this context) variably allows for an overt [əv] between it and the participle (v. 21b)
- iv. Speakers robustly exhibit (22b) as a variant of (22a); thus, *had* < > *would've*:

(22) b. If only he would've left sooner... (cf. (21a) and (22a))

THREE POSSIBLE PARSES FOR (22a):

- (22) a. If only he'd left sooner... >>
- (23) a. If only he had left sooner... (where 'd = *had*)
- b. If only he had HAVE_{silent} left sooner... (where 'd = *had*)
- c. If only he would HAVE_{silent} left sooner... (where 'd = *would*)

4.2 Or is it just phonology?

Does the phonological context of a preceding [tu] or [tə] make it difficult for transcribers to hear a following [əv] / [ə] (= *have*) ?

Work in progress. However, preliminary measurements indicate that transcribers were not failing to perceive something in the speech signal. E.g.

Emily Adams (SKCTC)

(7) *They wouldn't knowed what **to done** with the underclothes they've got.*
vowel between [t] and [d] = 0.04 seconds

(8) *When he got this cooked just right -- (which he =ud had enough L-- “Life Everlasting” in that **to =uv [tə ə]** **made** forty gallons, but he just had about a quart of it) -- he got him a teacup and dipped down in that, and he drunk that, and all at oncet he had a pain in his stomach.*
vowel(s) [ə ə] between [t] and [m] = 0.24 seconds

(9) *She was too old **to be** put in the orphanage.*
vowel between [t] and [b] = 0.11 seconds

(10) *Mamie Whitaker and Eli Whitaker had decided they wanted to keep all four of us.*
vowel between [t] and [k] = 0.079 seconds

Trained transcribers not failing to hear ‘have’ in the speech signal.

Of course, even if all the measurements show no evidence of two segments in a context like *to done* e.g. in (7) — one for the [ə] of [tə], and one for the [ə] of *have* [ə]...

...this does not entail that there is no phonological rule of deletion, for example as follows:

- (11) a. [ə] ‘have’ --> ∅ / [tə] ___ or
 b. [tə] --> [t] / ___ [ə] ‘have’

5. Final remarks

- the speech represented in the AAPCAppE exhibits the phenomenon of silent HAVE with infinitival perfects with modals and with infinitival *to*
- the relative frequency of silent HAVE is different for the two contexts; it occurs more frequently with infinitival *to* than it does with modals
- still not clear whether higher frequency of silent HAVE with *to*-perfects (compared to the modal context) is due to phonological considerations (delete [ə]?), or due to semantic considerations (use silent HAVE for the SoT interpretation)
- we will continue to investigate this to see if we can find ways to get at an answer
- we hope to have also shown what an audio-aligned and parsed corpus (like the AAPCAppE) is useful for:
 - the syntactic annotation makes it straightforward to search for particular constructions
 - the accompanying speech signal allows researchers to empirically verify the appropriateness of the transcription and test for indicators that might have otherwise been missed by previous researchers

APPENDIX: NON-PRESENT variants for five speakers from DOHP (from Tortora, Blanchette, O’Neill, & Arriaga 2015)

began, begin	keep, kept	saw, seen, seemed
bring, brought	knew, knowed	scald, scalded
brought, brung	laid, lay	start, started
burned, burnt	learned, learnt	send, sent
came, come	load, loaded	set, sit
catch, caught	lose, lost	start, started
cause, caused	lost, losted	swore, sworn
did, done	made, make	taken, takened, took
done, doned	open, opened	taken, took
drill, drilled	paid, pay	take, took
drop, dropped	push, pushed	taught, teached
get, got	ran, run	tell, told
give, given	ran, run, runned	turn, turned
gone, went	run, runned	walk, walked
go, gone, went	rent, rented	want, wanted
hand, handed	sang, sung	work, worked
heard, heard	saw, seen	
held, held		

Acknowledgments [please see AAPCAppE website for a complete list, for the AAPCAppE in general]:

We would like to thank Cristina Schmitt for inspiring discussion about the data in this paper. For the building of the AAPCAppE, the people we wish to thank are quite numerous; we list some folks here, but fuller acknowledgment is given on the AAPCAppE website: Alexia Ault, Greta Browning, Kathleen Currie Hall, Marcel den Dikken, Aaron Ecay, Robert Gipe, Fred Hay, Anton Karl Ingason, Tyler Kendall, Larry LaFollette, Michael Montgomery, Paul Reed, John Shean, Edward Snajdr, Laura Smith, Doug Whalen, Tiffany Williams, Walt Wolfram, Jiahong Yuan, Raffaella Zanuttini. Research has been supported by NSF Grants #BCS-0617197 (Tortora) and #BCS-0616573 (Den Dikken); by an NSF REG Award made to Blanchette on NSF Grant #BCS-0963950 (Snajdr); by a RISLUS Fellowship (Blanchette); by a Graduate Center Fellowship (Blanchette); by the College of Staten Island's *Provost Research Scholarship* (2010-11); by an NEH 2011-12 Fellowship (Tortora); and by an NEH Digital Humanities Start-Up Grant (2012-14) #HD-51543-12 (Tortora); and by NSF BCS Awards #BCS-1152148 (PI: Tortora) & #BCS-1151630 (PI: Santorini).

Selected Bibliography:

- Boersma, Paul, and David Weenink. 2011. *Praat: doing phonetics by computer*, version 5.2.32.
- Bowie, Jill & Bas Aarts. 2011. "Change in the English Infinitival Perfect Construction."
- Johnson, G. 2016. "The syntax of *liketa*: A case of restructuring in Appalachian English." Under review.
- Kayne, R. 1997. "The English Complementizer of," *Journal of Comparative Germanic Linguistics* 1: 43-54. also in Kayne 2002, *Parameters & Universals*, OUP.
- Kendall, Tyler (2007). "Enhancing Sociolinguistic Data Collections: The North Carolina Sociolinguistic Archive and Analysis Project," *Penn Working Papers in Linguistics* 13.2: 15-26.
- Kroch, A. & Ann Taylor. 2000. *The Penn-Helsinki Parsed Corpus of Middle English* (PPCME2). Department of Linguistics, University of Pennsylvania. CD-ROM, 2nd edition, (<http://www.ling.upenn.edu/hist-corpora/>).
- Kroch, A., Beatrice Santorini, and Lauren Delfs. 2004. *The Penn-Helsinki Parsed Corpus of Early Modern English* (PPCEME). Department of Linguistics, University of Pennsylvania. CD-ROM, first edition, (<http://www.ling.upenn.edu/hist-corpora/>).
- Kroch, A., Beatrice Santorini, and Ariel Diertani. 2010. *The Penn-Helsinki Parsed Corpus of Modern British English* (PPCMBE). Department of Linguistics, University of Pennsylvania. CD-ROM, first edition, (<http://www.ling.upenn.edu/hist-corpora/>).
- Marcus, M., B. Santorini, & M. Marcinkiewicz. 1993. "Building a large annotated corpus of English: The Penn Treebank," *Computational Linguistics* 19.2:313-330.
- Molencki, Rafal. 2003. "Proscriptive Prescriptivists: On the Loss of the 'Pleonastic' Perfect Infinitive in Counterfactual Constructions in Late Modern English," *Linguistic Insights: Studies in Language and Communication* 7: 175-196.
- Montgomery, M. & J.S. Hall. 2004. *A Dictionary of Smoky Mountain English*. Knoxville: UT Press.
- Randall, Beth. 2009. *CorpusSearch 2: A tool for linguistic research*. Includes CorpusDraw, a graphical interface for displaying and correcting parsed corpora. <http://corpussearch.sourceforge.net/>
- Stowell, Tim. 2007. "Sequence of Perfect," in Louis de Saussure, Jacques Moeschler and Genoveva Puskas (eds.) *Recent advances in the syntax and semantics of tense, mood and aspect* (Trends in Linguistics Vol. 185) Berlin: Mouton De Gruyter, pp. 123-146.
- Taylor, A. 1994. "Variation in Past Tense Formation in the History of English," in R. Izvorski, M. Meyerhoff, B. Reynolds, & V. Tredinnick (eds.), *UPenn Working Papers in Linguistics* 1, pp. 143-159.
- Tagliamonte, S. 2012. *Roots of English: Exploring the History of Dialects*. CUP.
- Tortora, C. 2014. "Evidence for the non-finiteness of English 'present' and 'past' verb forms," talk given at the NYU Syntax Brown Bag series, February 28, 2014.
- Tortora, C., F. Blanchette, T. O'Neill, & S. Arriaga. 2015. "Variation in Appalachian non-present forms," talk given at the FWAV2 Workshop, May 2015, Reykjavik, Iceland.
- Tortora, C., B. Santorini, F. Blanchette, & C.E.A. Diertani. to appear. *The Audio-Aligned and Parsed Corpus of Appalachian English* (AAPCAppE). <http://csivc.csi.cuny.edu/aapcapp/>
- Wolfram, W. 1984. "Is there an Appalachian English?" *Appalachian Journal* 11:215-226.
- Wolfram, W. & D. Christian. 1976. *Appalachian Speech*.
- Zagona, K. 2014. "Sequence-of-tense and the Features of Finite Tenses," *Nordlyd* 41.2: 261-272.