Prosodic words across large domains in Choctaw[*]

Matthew Tyler
University of Cambridge

George Aaron Broadwell
University of Florida

Abstract: Choctaw (ISO:cho) has some large prosodic words which appear to include an overt or null copula. This article first discusses several phonological tests for the right edge of a prosodic word, which we use to support the claim that the overt copula may encliticize onto its complement. This sometimes results in prosodic words that span over domains that include a clause boundary. Our account of this phenomenon relies on the idea of prosodic pre-specification. Additionally, we discuss the prosodic behavior of the null copula, and implications for the syntax-prosody interface.

Keywords: prosody, Choctaw, cliticization, Match Theory, syntax-prosody interface

1 Introduction

Choctaw (ISO:cho) is a Western Muskogean language spoken in Mississippi and Oklahoma. The language generally shows head-final order, as in (1), where the verb is final in the clause, 'dog' is final in the noun phrase 'Henry's dog', and 'under' is final in the postpositional phrase 'under the table'.

(1) Henry im-ofi-yat   aa-ípa   nóta   ittóla-h.
     Henry III-dog-NOM LOC-eat.NMZ under lie:NG-TNS
     'Henry's dog is lying under the table'.

(1) also shows that the subject of the clause receives nominative case. Because there is a third person subject in this example, there is no overt agreement. Sentences with non-third person arguments show a complex agreement system of the 'active' type, discussed in detail in Davies (1986), Ulrich (1986), Broadwell (2006), and Tyler (2019a), among others.

One of the difficulties for both linguists and learners of the language is the frequent appearance of long multisyllabic words, as in (2a) and (2b).[1]

---

[1] Abbreviations used in the paper: 1 = first person, 2 = second person, 3 = 3rd-person, COMP = complement, COP = copula, DPST = distant past tense, DS = different-subject, FOC = focus, I = class I, II = class II, III = class III, LG = l-grade, MOD = modal, NEG = negative, NG = n-grade, NMZ = nominalization, NOM = nominative, OBL = oblique, PL = plural, POSS = possessor prefix, PST = past tense, Q = question particle, SG = singular, SS = same-subject TNS = default tense.

---

(2) a. Chi-tikba-y**aa**tokm-a̱ ...
         2sg.ii-front-topic-ds
         'In front of you, ...'

     b. Ano̱polii-li-h-**aa**tok-oosh ...
         speak-1sg.i-tns-because-ss
         'Because I was speaking ...'

The questions of interest for this article are: are the items in (2) single prosodic words (PWds)? Just how much syntactic structure do they contain? And what are the theoretical consequences of the answers to these questions?

Key to the discussion is the copula. The expressions in (2) appear to contain the overt copula *a(a)-* (the initial parts of the strings glossed as 'topic' and 'because'), and as we will argue, this copula often sits embedded inside a prosodic word. Choctaw also has a null copula, which, similarly, may be deeply embedded inside a prosodic word. In (3), for example, a pronoun (*chishn-ak*) is followed by a tense marker and a switch-reference marker. The presence of such verbal morphology on apparent nouns leads Broadwell (2006:84ff.) to argue that these examples are covert cleft constructions with a null copula.

(3) Chishn-ak-Ø-baano-h-o̱ chi-pı́sa-li-tok.
     you-foc-cop-only-tns-ds 2sg.ii-see:ng-1sg.i-pst
     'I saw only you' (Lit. 'It was only you that I saw.')

This article investigates the prosodic behavior of the copula, and argues that it can 'knit together' a PWd composed of the copula and its complement: this is possible even when the complement of the copula is a finite clause, and the resulting PWd thus spans a clause boundary. We focus mainly on the overt copula, but turn to the null copula towards the end of the article.

This article is organized as follows. Section 2 shows how the right edge of a PWd can be diagnosed in Choctaw. Section 3 then provides evidence that the overt copula *a(a)-* optionally encliticizes onto the material to its left, leaving no PWd boundary between the copula and its complement. Section 4 provides a formal account of this optional encliticization behavior, via variable *prosodic pre-specification*. Section 5 argues against an alternative 'syntax-first' analysis, whereby optional encliticization of the copula would be derived from optional syntactic incorporation of the copula's complement into the copula itself. Section 6 discusses an interesting theoretical consequence of the proposal: that PWds can span across domains that include a clause boundary within them. Section 7 extends the proposal to the null copula, and section 8 briefly concludes.

1.1 Data and methodology

This paper uses examples of both Mississippi and Oklahoma Choctaw. All of the Mississippi Choctaw examples presented here are from recordings with first-language speakers of Choctaw, made by the first author at the Mississippi Band of Choctaw Indians (MBCI) reservation, between 2016 and 2019. They were transcribed by the first author. The recordings, alongside transcriptions and English translations, are currently held by the Language Program at MBCI.

In addition, we employ some 19th-century textual examples from Oklahoma. These come from a 400,000-word corpus of 19th century Choctaw, developed by the Choctaw Language and History working group. The examples cited in this paper come from the legal records of the Choctaw Nation prior to Oklahoma statehood.

2 Prosodic words in Choctaw

We make some basic assumptions about prosody and the prosodic hierarchy. We assume that utterances, in addition to their hierarchical syntactic structure, also have a parallel hierarchical prosodic structure. The overt phonological material in an utterance is distributed across the terminals in this prosodic structure, such that the prosodic structure is isomorphic to the utterance's syntactic structure, up to a point (Selkirk 1981 et seq., Beckman and Pierrehumbert 1986; Nespor and Vogel 1986). The constituents in a prosodic structure bear prosodic category labels, which determine various properties of their segmental and suprasegmental realization. We follow recent work in assuming that the set of prosodic categories includes, at least, prosodic words (PWds, a.k.a. phonological words), phonological phrases (PPhs), and intonational phrases (IPs) (Ito and Mester 2013). In this article we are concerned mainly with one prosodic category within Choctaw—we assume this category is a PWd, since it tends to correspond to just one morphosyntactic word, though the exact label of this category is not important.[2]

In this section we provide three diagnostics for the presence or absence of the right-edge of a phonological word.

2.1 [ʔ] is pronounced only in PWd-final position

Some words end in phonetic glottal stops, such as *im-alla* 'his child' in (4a). This glottal stop becomes a phonetic glide (orthographic <y>; IPA [j]), when a vowel-initial suffix is added, as in (4b).

(4) a. Bill im-alla[ʔ] i-chókfih-at   abiika-h.
         Bill 3.III-child 3.III-rabbit-NOM sick-TNS
         'Bill's child's rabbit is sick.'

     b. Bill im-alla[j]-at   chókfih im-abiika-tok.
         Bill 3.III-child-NOM rabbit   3.III-sick-TNS
         'Bill's child rabbit was sick.'

This alternation also reflects a larger pattern: glottal stops are unattested in word-medial position, and *only* occur word-finally. Thus, the presence of a phonetic glottal stop evidences the right edge of a phonological word.

Note that there is some debate about the phonemic status of the final glottal stop. Some authors have argued that the glottal stop is not phonemic, and is inserted at the right edge of phrases whose underlying phonemic representation ends in a vowel—in turn, *phonetically* vowel-final phrases are analyzed as having an underlying final /h/ (Munro 1984, 1985). The non-phonemic view is implicit in most work on Choctaw, as most scholars have simply not marked glottal stops (e.g. Byington 1870, 1915; Nicklas 1974. Other authors (Broadwell 2006; Katenkamp 2021; Ulrich 1986) argue that [ʔ] must be phonemic, pointing to a (phonetically subtle) distinction between vowel-final words, [ʔ]-final words, and [h]-final words. The phonemic vs. non-phonemic status of [ʔ] also impacts the analysis of the [ʔ]/[j] alternation in (4): if the glottal stop is phonemic, then [j] could be analyzed as

---

[2] We abstract away from the possibility of PWd recursion in Choctaw, although there *is* evidence for sub-PWd domains in Choctaw (Broadwell 2006:21ff. shows that the domain of word-internal iambic lengthening does not always correspond to a full morphosyntactic word). One simple ancillary assumption that we could adopt is that our 'PWd' is in fact a *maximal* PWd.

a contextual allophone of it; if not, then [j] must be epenthetic instead. We set this debate aside here: what's important for present purposes is that [ʔ] is only ever pronounced at the right edge of a word.

2.2 /h/ is weak or absent in PWd-final position

As mentioned above, Choctaw syllables may end in phonemic /h/, as in *Chahta* 'Choctaw'. When an /h/-final syllable occurs in word-final position, the /h/ is either unrealized, or phonetically weak (see Katenkamp 2021 for detailed phonetic analysis).[3] We notate this weak/absent /h/ as [ʰ], as in *am-aafoh* 'grandfather' in (5a). It alternates with a 'full' [h], when a suffix is added, as in (5b).

(5) a. Am-aafo[ʰ]   ich-chǫkash achokma-kiyo-h.
         1SG.III-grandfather POSS-heart   good-not-TNS
         'My grandfather's heart isn't good.'

    b. Am-aafo[h]-at   oshíitik   im-alla   ilaawiili-h.
         1SG.III-grandfather-NOM daughter 3.III-child care-TNS
         'My grandfather takes care of his daughter's child.'

We analyze this alternation as a consequence of resyllabification: in (5b) the /h/ is re-syllabified into the onset of the nominative case suffix *-at*, and thus can be given its full (non-weak) realization. We assume that resyllabification is possible word-internally, but is not possible across a word boundary, explaining why the /h/ in (5a) cannot resyllabify into the onset of the first syllable in *ichchǫkash*. Thus, where a final /h/ has resyllabified, we can safely assume that there is *no* PWd boundary at this syllable juncture.

2.3 /k/ is optionally voiced and lenited in PWd-medial position

Choctaw has one phonemic velar stop: /k/. In word-final position, it is generally voiceless, as in (6a). But in word-medial positions, /k/ is often voiced to [g], as in (6b), and is sometimes lenited too, to [ɣ].

(6) a. Nitta[k] istáayoopiya-t mį́ti-h.
         day   final-NOM   come:NG-TNS
         'The final day is coming.'

    b. Ofíi-m-at   i[g]-ahchíib-o   masaali-tok.
         dog-that-NOM NEG-long.time:LG-NEG heal-PST
         'The dog healed in no time.'

Thus, where a /k/ has undergone voicing (and perhaps lenition), we can infer that the /k/ does not sit at a PWd boundary.

The three diagnostics presented in this section for the presence or absence of a PWd edge are listed in (7).

---

[3] The distinction between a phonetically weak vs. absent /h/ isn't terribly important for out purposes—what's important is that it is different from a full /h/.

(7) a. The presence of a glottal stop evidences the right edge of a PWd.

    b. The presence of a phonetically-full [h] immediately before a vowel-initial morpheme (which does not typically start with /h/) indicates that there is no PWd boundary at that syllable juncture.

    c. The presence of a voiced (and perhaps lenited) /k/ indicates that there is no PWd boundary at that syllable juncture.

Next, we use these diagnostics to show that, often, there is *no* PWd boundary before the overt copula *a(a)-*.

3   The overt copula *a(a)-* optionally encliticizes

As noted in section 1, Choctaw has an overt copula (*a(a)-*) and a null one. The form of the copula that arises in a given context is determined by morphosyntactic factors: broadly, in the presence of certain prefixes (8a) or suffixes (8b) that require support, then the overt copula shows up. Otherwise, it is null (8c).[4]

(8) a. Ohooyo-p-a̱   siy-**a**-h.
    woman-this-OBL 1SG.II-**COP**-TNS
    'I am this woman.'

    b. Mary **aa**-tok   am-ahwa-h.
    Mary **COP**-PST 1SG.III-think-TNS
    'I think it was Mary.'

    c. A̱-ki-t   alíkchi **Ø**.
    1SG.III-father-NOM doctor **COP**
    'My father is a doctor.'

In this section we investigate the phonological behavior of the overt copula *a(a)-*. In section 7 we will see that the null copula—or rather, the phonological material that immediately follows the null copula—behaves in a similar way.

We claim in this section that the overt copula *a(a)-* may optionally encliticize onto its complement ('COMP') (due to Choctaw's head-final syntax, the the copula always follows its complement). The evidence for this comes from applying the diagnostics for PWd edges that were discussed above. Sometimes it turns out to be possible to discern the presence of a PWd boundary immediately before the copula, and sometimes it is possible to discern the absence of one. When there is a PWd boundary, we assume that encliticization has *not* occurred, and that the prosodic structure of the copula and its complement is like (9a). When there is *no* PWd boundary, we assume that encliticization *has* occurred, and that the prosodic structure is like (9b) instead (no left edge is shown for the first Pwd, since this PWd may be coterminous with the left edge of the morphosyntactic complement, or the complement may extend out beyond the bounds of this PWd).

(9) a. ... COMP] [$_{PWd}$ *a(a)-*...]

    b. ... COMP   *a(a)-*...]

---

[4] There are suffixes that do *not* require support. For instance, the default tense suffix *-h* does not, by itself, force the overt copula to be inserted: **a(a)-h* is not attested.

The evidence that the overt copula optionally encliticizes comes from applying the three diagnostics for PWd boundaries that were discussed in the last section. Note finally that we are restricting attention to the prefix-less copula—the diagnostics are not applicable when it has a prefix, as in (8a).

3.1 Vowel hiatus at the COMP/*a(a)*- border is resolved as [ʔ] or [j]

In section 2.1, we saw that glottal stops are only ever found at the right edge of a PWd. We also saw that PWd-internal vowel hiatus may be resolved by inserting a [j]. We find that when the complement to the copula *a(a)*- ends in a vowel, then both options are attested: either a glottal stop is inserted as in (10a), diagnosing the presence of a PWd boundary immediately before the copula, or an epenthetic [j] is inserted as in (10b), diagnosing the *absence* of a PWd boundary.

(10) a. John-at p̲ashi lósa [ʔ] aa-tok.
      John-NOM hair black COP-PST
      'John had black hair.'

    b. Chokkaa-m-a̲ p̲ísa-li-tok hómma [j] aa-tok.
      house-that-OBL see:NG-1SG.I-PST red COP-PST
      'I saw the house that was red.'

We also find evidence for this alternation in late 19th-century Choctaw texts. Glottal stops are unwritten, but the [j] is written as <y>. When the copula takes a vowel-final complement, we find it written as both <a...> and <ya...>, illustrated by the examples in (11) (the copula is used here to introduce a topic, cf. (2a)).

(11) a. Holitopa ish binnili m̲a: ch̲i Petitioner yokat anumpa yakohmi h̲o **chitikba yatuk ma** ch̲i bohlili shki.
      'To the sitting judge: I submit the following petitioner **before you**.'

<div align="right">(1889, Red River Co., case #124)</div>

    b. Yohmima Court ʊt anumpa **ittola atukma** achukmalit silhhit pisamʊt anumpa **ittola atukmat** nan isht ikayono a̲sha hatuk osh ...
      'And when the court carefully tracked and looked at **the case**, **the case** was not enough, and so ...'   (1885, Jack's Fork Co. Court, March-April affairs)

3.2 COMP-final /h/ is optionally resyllabified as the onset of *aa*-

We saw in section 2.2 that /h/ in word-final position is weak or unpronounced, but /h/ can be syllabified into the onset of an onsetless syllable, provided that no PWd boundary intervenes between them. Where the complement of *a(a)*- ends in /h/, we find both possibilities (though syllabification into the onset is by far the more common option). Syllabification into the onset of *aa*- is illustrated in (12a), showing that there must be no PWd boundary between the complement and the copula. A *potential* example of /h/ failing to syllabify into the onset of *a(a)*- is given in (12b).[5]

---

[5] It is difficult to identify clear examples in which a complement-final /h/ has failed to syllabify into the onset of *a(a)*-, rather than simply being absent from the input. The example in (12b) is a good candidate for a non-resyllabified /h/, since the recent-past interpretation of the sentence would lead us to expect an underlying *-h* tense suffix.

(12) a. Am-aafoh   hicha sa-ppókni   [h] aa-tok-m-a̲   ...
          1SG.III-grandfather and   1SG.II-grandmother   COP-PST-that-DS
          'As for my grandfather and grandmother...'

      b. Maryh-at   nokoowa-h, baliili [ʰ/Ø] aa-tok-o̲
          Mary-NOM   angry-TNS   run   COP-PST-DS
          'Mary was angry because she ran.'

There are a couple of things to note about these examples. For one thing, both of the words featuring the copula *a(a)-* here are in some sense idiomatic. *Aatokma̲* is used to (re)introduce topics (see also (2a)), and *aatoko* means that the clause in its complement was the cause of the clause that embeds it (most scholars do not decompose it, and treat it as a single element translated as 'because'). Both of these forms contain the past tense suffix *-tok*, but neither have clearly past-tense interpretations.

There is historical textual evidence for this alternation too. When certain phrases appear before the copula, the copula is sometimes written as <*ha-*>, sometimes as <*a-*>. One such phrase is bolded in (13).

(13) a. Kiamichi County Court ʋt nan anoli hoyot im aponaklʋchit micha **nan ʋlhpisa hatuk**ma
          ai ittapisʋt pihinsa ma ...
          'Kiamichi County Court has examined witnesses and **in view of the law...**'
                                      (1886 Choctaw Nation vs. Joseph Tanitʋbbi Trial- Pt. 2)

      b. **Nan alhpisa atuk** pulla mako̲ Court i̲ shahli pit afoka mako̲ apisashki ...
          '**Due to the law**, the District Court should examine the case ...'
                                      (1896 Red River County Court Affairs November)

We assume that the initial <h> shows that /h/ has syllabified into the onset of the copula. Where /h/ is absent, then we *could* be dealing with a weak or absent word-final /h/, which is not written in the Traditional Orthography, *or* with a glottal stop, also unwritten. Fortunately the fact that we have multiple tokens of the same phrase, some written as <*ha-*> and some as <*a-*>, shows us that there must be a word-final /h/ in the input, which is sometimes not syllabified into the onset of *a(a)-*.

3.3 COMP-final /k/ is optionally voiced before *aa-*

We saw in section 2.3 that /k/, which is realized as [k] in word-initial and word-final positions, may be voiced, and sometimes lenited too, in word-medial position. Where the copula *a(a)-* follows a /k/-final complement, the /k/ may be voiced, as in (14a), though need not be, as shown by (14b). Where a final /k/ is voiced, we can infer that there is no PWd boundary between the complement and the copula.

(14) a. Kaniiya-li-ttoo[g] aa-tok.
          leave-1SG.I-DPST   COP-PST
          'I had already gone.'

      b. Anaa[k] aa-tok   tákla-na   famaa-tok.
          Me   COP-PST   be.with:LG-and.DS be.whipped-PST
          'It was me that was with him when he got whipped.'

7

To summarize the three parts of this section together, we have seen evidence, from three diagnostics, that there is *optionally* a PWd boundary between a copula and the complement to its left. We propose that the presence or absence of this PWd boundary is determined by whether or not the overt copula *a(a)-* undergoes an optional process of encliticization—see the schemas in (9). In section 4, we propose that optional encliticization is driven by the phonological properties of the overt copula: specifically, we propose that the overt copula is pre-specified, in the lexicon, with two distinct prosodic behaviors, and we model these behaviors with the technology of *prosodic subcategorization frames*.

In section 5, we contrast this account with an alternative 'syntax-first' account, in which (non)-encliticization is determined by the syntactic structure of the copula and its complement, and *not* by the lexically-idiosyncratic properties of the *a(a)-* morpheme. We argue that our 'phonology-first' account is more parsimonious.

4   Proposal: variable prosodic pre-specification

We have shown that the overt copula *a(a)-* sometimes encliticizes onto its complement, and sometimes does not. The absence of encliticization leads to the presence of a PWd boundary between the complement and the copula. Here, we provide a formal account of this optional encliticization.

The basic intuition is that individual morphemes may exhibit their own idiosyncratic prosodic behavior, and that this specialized behavior may 'overrule' more general prosodic tendencies of the language (Bennett, Harizanov, and Henderson 2018; Inkelas 1989; Inkelas and Zec 1990; Tyler 2019b). So in Choctaw (as in, probably, most languages), there is a tendency for verbs to form a separate prosodic word from their complement. But the overt copula *a(a)-* is different: in its lexical entry it is *pre-specified* with an instruction to encliticize onto the phonological material that sits to its right—although this instruction must be only variably active, since *a(a)-* does not always encliticize. This prosodic pre-specification overrules the more general pattern exhibited by verbs in the language, in which they form their own, separate prosodic words.[6]

We model prosodic pre-specification using the technology of *prosodic subcategorization*. The idea is that morphemes are associated with particular 'prosodic subcategorization frames', which restrict the licit prosodic structures in which the morpheme can be found. The part of the grammar responsible for prosodic structure-building must attend to the prosodic subcategorization requirements of individual items, as well as to general principles of prosodic structure-building (for instance, those introduced by Selkirk's 2011 Match constraints). In particular we propose that the overt copula *a(a)-* is optionally associated with the prosodic subcategorization frame in (15).

(15)   [[...] *a(a)*]$_{\text{PWd}}$

This frame specifies that *a(a)-* should form a prosodic constituent with some phonological material to its left, and that the resulting prosodic constituent should be of category 'PWd'.

Thus when this frame is active, and encliticization takes place, the resulting prosodic structure is as in (16a)—see how the frame in (15) is satisfied. When the frame is not active, and encliticization does not take place, the resulting structure is as in (16b). We assume that when the prosodic subcategorization frame is not active, general principles of prosodic structure-building, such as Selkirk's

---

[6] We assume that the Choctaw copula is a verb, albeit an unusual one.

Match constraints, cause the copula to form its own prosodic word, to the exclusion of the material to its left.

(16) a.   PWd   b.   (PPh)

                 ... *a(a)-*   PWd PWd

                                                                   ... *a(a)-*
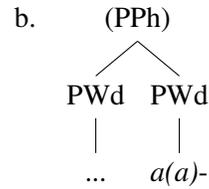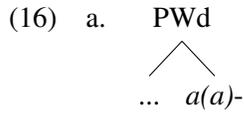
Note that when the copula carries suffixes (e.g. past tense *-tok*), this does not change the effects of the prosodic subcategorization frame. We assume that suffixes like *-tok* are characterized by their own (obligatory) prosodic subcategorization frames which make them form a PWd with the material to their left, but we do not pursue this line of investigation here.

5   Against a 'syntax-first' account

The account presented in the previous section, in which the (non-)encliticization of the overt copula is driven by the optional activation of an idiosyncratic prosodic pre-specification requirement, can be contrasted with an alternative analysis, which we term the 'syntax-first' analysis.

Essentially, in the prosodic pre-specification analysis, the underlying syntax of the copula is the same regardless of whether or not it encliticizes. The variability emerges in the phonological component. In the syntax-first analysis, however, the variability emerges in the syntax. The encliticizing output and the non-encliticizing output derive from *different* underlying syntactic structures, and the phonology straightforwardly interprets the different syntactic structures that are sent to it.

To create a plausible syntax-first analysis, we make use of long-known property of the phonology-syntax interface—that phonological words tend to correspond to syntactic heads (Selkirk 2011). Those syntactic heads may be atomic, or they may be complex, built by syntactic operations that combine heads together. This means that when a lexical item appears to form a prosodic word with its complement, one potential interpretation of this fact is that the complement has undergone some form of syntactic amalgamation with the head that selected it, and that this amalgamated syntactic head is interpreted by the phonology as a single prosodic word. The most widely-accepted syntactic 'amalgamation' operation is head-movement (Travis 1984), where a head X^0 that is selected by a c-commanding head Y^0 moves to form a complex head with Y^0.[7]

With head-movement at our disposal, we can posit two separate syntactic structures for the encliticized copula and the non-encliticized copula. When encliticization has occurred, we can infer that the X^0 heading XP, the complement to the copula, has head-moved, and formed a complex head with the copula. This is schematized in (17a). By contrast, when encliticization has *not* occurred, we can infer that the corresponding head-movement has not occurred either. This is schematized in (17b).

---

[7] Scholars have posited various operations to supplement or replace classical head-movement—see Harizanov and Gribanova (2019) for an overview of the issues involved, and a recent alternative proposal.

(17) a.

```
            VP
           /  \
        XP      V
       /  \    / \
     ...  t_X  X   V
            ↳     a(a)-
```

b.

```
            VP
           /  \
        XP      V
       /  \    a(a)-
     ...  X
```
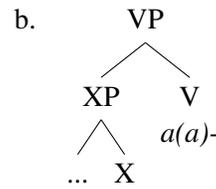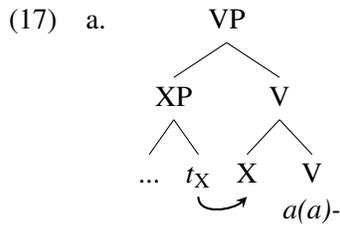
In a sense, this analysis captures the data just as well as the prosodic pre-specification analysis. However, there are two considerations that bias us towards the prosodic pre-specification analysis. Firstly, the syntactic amalgamation operation must be able cross over a clause boundary. Consider (18a), repeated from (14a), and (18b). In both of these examples, the complement to the copula is a *clause*, with its own syntactic subject and tense-mood-aspect marking.

(18) a. Kaniiya-li-ttoo[g] aa-tok.
　　　leave-1SG.I-DPST COP-PST
　　　'I had already gone.'

b. Anopolii-li-[h] aa-tok-oosh ...
　　speak-1SG.I-TNS COP-PST-SS
　　'Because I was speaking...'

In both of these examples, we know that the copula is able to encliticize, because of the telltale phonological processes that we see immediately to the left of the copula (lenition of /k/ and resyllabification of /h/, respectively). So under the syntax-first analysis, it must be possible for head-movement to cross over a clause boundary. However, it is a well-known property of head-movement that it does not cross clause boundaries (Chomsky 2000; Grimshaw 2000).[8] This issue disappears under the prosodic pre-specification analysis, since no unusual clause-crossing head-movement (or whatever syntactic operation would be responsible for amalgamating the copula and its complement) is necessary.

The second issue with the syntax-first analysis is simply that it puts the variability in the wrong place. As far as we are aware, the (non-)encliticization of the overt copula onto its complement is conditioned by phonological or phonetic factors. We observed, for instance, that encliticization seems very likely to happen after an /h/-final complement, but somewhat less likely after the other kinds of complement that we investigated (vowel-final, /k/-final). We also observed that it is more likely to happen at a higher speech rate than in more careful speech. By contrast, we don't have clear evidence that it is conditioned by morphosyntactic factors.[9]

Thus we submit that a prosodic pre-specification analysis has more to recommend it than a syntax-first analysis, at least based on what we know at the moment.

---

[8] So-called 'long head movement', in which apparent syntactic terminals move in a manner that is not constrained by Travis's Head Movement Constraint, is not clause-bound (Lema and Rivero 1990). We exclude it from the discussion here—indeed, in Harizanov and Gribanova's (2019) bifurcation of 'head-movement' into two distinct operations, 'amalgamation', the word-building operation which bundles together adjacent syntactic heads, *is* clause-bound.

[9] Encliticization does not seem to take place when the copula bears agreement prefixes, which could be construed as a 'morphosyntactic factor'. But (a) it could be that the the copula encliticizes onto its prefix, and (b) the diagnostics we have are not applicable when the right-hand element starts with a consonant, as all the copula's agreement prefixes do. This requires more careful investigation.

6 PWds that span large morphosyntactic domains

One theoretically-interesting consequence of the findings here is that the edge of the morphosyntactic clause does *not* constitute a 'hard' edge for prosodic word formation, at least in Choctaw. When the overt copula *a(a)-* takes a clause as its complement, we have seen that it is possible for *a(a)-* to encliticize across this clause boundary.

Various theories of the syntax-prosody interface predict that clause edges should line up with the edges of prosodic constituents, of various categories. For instance, in the version of Match Theory developed by Selkirk (2009, 2011), clause edges are matched to the edges of intonation phrases (IPs), although she proposes that languages differ in terms of what counts as an IP-matching clause. Some languages match only root clauses (which she characterizes as a ForceP in the sense of Rizzi 1997), others match all finite clauses (CPs). Other authors have proposed that in some languages, clauses/CPs are matched to phonological phrases instead (Weber 2020). And others still have proposed that in some languages, clauses/CPs are matched to the edges of prosodic words (Compton and Pittman 2010; Ershova 2020).[10] For all of these accounts, a prosodic word that happily spans across a clause edge presents a problem. That clause edge should have been matched to the edge of some prosodic constituent: a word, a phonological phrase or an intonation phrase.

Fortunately, Match Theory need not be the final say on how prosody and syntax correspond. Indeed, in *indirect reference* theories of the syntax-prosody interface, the pressure towards syntax-prosody isomorphism (which Match Theory describes) may be overruled by other competing pressures. For instance, one kind of competing pressure is that prosodic structures prefer to come in certain desirable shapes and sizes, regardless of how well they correspond to syntactic structure (so-called 'well-formedness' constraints). The final output prosodic structure is a negotiation between these competing pressures, and this negotiation process can be modelled using Optimality Theory (Prince and Smolensky 1993), with the various competing pressures supplying the constraints.

Prosodic subcategorization frames associated with individual lexical items, as described in section 4, represent one of these pressures. In Bennett et al. (2018) and Tyler (2019b), the pressure towards syntax-prosody correspondence, embodied for instance by Match constraints, is outweighed by the pressure to satisfy items' prosodic subcategorization frames, embodied in a constraint SUB-CAT.

This is essentially what we think is going on with these 'too big' prosodic words involving the overt copula *a(a)-*: when the overt copula's encliticizing subcategorization frame is active (recall that it is only optional), then the constraint that ensures this frame is satisfied is active too. This constraint, SUBCAT, outranks the constraint or constraints that insert a prosodic boundary at a clause edge, and the resulting prosodic structure contains an unusually large, domain-spanning prosodic word.[11]

---

[10] Some of the proposals mentioned here propose that prosodic categories are matched to syntactic *phases*, in the sense of Chomsky (2001). The finite clause or CP is canonically a phase, so for our purposes these proposals amount to a more theoretically-grounded version of CP-matching accounts.

[11] There is an alternative interpretation of the behavior of the overt copula, which we are unable to consider here: it could be that some Choctaw embedded clauses—specifically those that can appear in the complement of the copula—are structurally-reduced. The reason why encliticization is possible across these clause boundaries would be because these clauses are not capped with a functional syntactic layer that forces the insertion of a prosodic boundary.

7 More large PWds: the null copula

We have so far restricted attention to the overt copula *a(a)-*. But as noted in section 3, Choctaw also has a null copula: the overt copula appears in the presence of certain prefixes or suffixes which require its 'support' (e.g. agreement prefixes, or past-tense suffix *-tok*), but in all other cases, the copula is null.

Interestingly, not all verbal suffixes force the copula to be overt. Some suffixes can appear with either the overt copula *or* the null copula. The clearest case of a suffix that behaves this way is the polar question suffix *-o̱*. The example in (19a) shows that *-o̱* is found at the right edge of the clause, as the outermost suffix of the verb complex—the voicing of /k/ to [g] shows that *-o̱* forms part of the same PWd as the verb. (19b) shows that *-o̱* can attach to the overt copula.

(19) a. Ish-shalalli-to[g]-o̱?
          2SG.I-slip-PST-Q
          'Did you slip?'

  b. Hattak-m-ak   chiy-aa-h-o̱?
      man-that-FOC 2SG.II-COP-TNS-Q
      'Are you that guy?'

Most interestingly, however, *-o̱* can appear in clauses with a null copula. Here, *-o̱* encliticizes directly onto the copular object. Three examples of this are shown in (20).

(20) a. Aayíshko-ya̱ talaaya-m-at   áno Ø-[j]o̱?
          cup-OBL   stand-that-NOM mine COP-Q
          'Is that cup there mine?'

  b. Hattak-m-at   kánah   yokáachi Ø-[j]o̱?
      man-that-NOM someone catch.NMZ COP-Q
      'Is that man a police officer?' (lit: 'people catcher')

  c. Yamm-at hattak-m-a[g] Ø-o̱?
      that-NOM man-that-FOC COP-Q
      'Is he that guy?'

In examples (20a-b), a glide ([j]) is inserted between the copular object and the suffix *-o̱*. According to the diagnostic provided in section 2.1, the insertion of a glide before a morpheme, and the concomitant absence of a glottal stop, indicates that there is no PWd boundary between that morpheme and the preceding one. In (20c), the final /k/ of the copular object voices to [g]. According to the diagnostic in section 2.3, this indicates that the /k/ is word-medial, and accordingly that the *-o̱* morpheme is part of the same PWd as the copular object.[12]

We can understand the phonological behavior of the question particle *-o̱* as obligatory encliticization: it must form a prosodic word with the phonological material immediately to its left. Typically that material is the verb complex, but when the verb is null, as with the null copula, then it encliticizes onto the copular object. As with the overt copula, we can formalize this prosodic behavior with the technology of prosodic subcategorization frames: *-o̱*, in its lexical entry, is pre-specified with a frame that requires it to encliticize onto whatever's to its left.

The behavior of *-o̱* in examples like (20) thus supports the broad theoretical point made in the previous two sections: item-specific prosodic pre-specification can result in prosodic words that span

---

[12] We have not yet investigated what happens when *-o̱* attaches to copular objects that end in /h/. We predict, on the basis of the diagnostic in section 2.2, that /h/ is resyllabified into the onset of *-o̱*.

across 'typical' boundaries. In the case of the overt copula *a(a)*-, pre-specification gives us PWds that span clause boundaries; in the case of the question particle -*o̲*, we get PWds that span across the boundary between a noun and a (null) verb.

Null copulas may be more widespread in Choctaw, and Broadwell (2006:86ff.) argues that they occur in a wider range of environments than just in 'obvious' copular clauses like (20). He proposes that many long words in Choctaw should be analyzed as containing a null copula. Some examples are given in (21). Note that here, in keeping with Broadwell's practice, we write the null copula as suffixed to the material preceding it (with a '-'), though the prosodic and morphosyntactic status of this 'link' is not currently clear.

(21) a. Chishn-ak-Ø-baano-h-o̲   chi-písa-li-tok.
          you-FOC-COP-only-TNS-DS 2SG.II-see:NG-1SG.I-PST
          'I saw only you.'

      b. Hattak-m-ak-Ø-fíihna-h-oosh   ala-h.
          man-that-FOC-COP-very-TNS-SS come-TNS
          'That very man came.'

It may be the case that the material following the null copula forms a PWd with the material preceding it, though we have not yet developed diagnostics to determine whether these form a single PWd or separate ones—the diagnostics in section 2 are only useful where certain pairs of segments abut (V+V; /h/+V; /k/+V). *If* the first word in the examples in (21) does turn out to form a single PWd, then we would attribute it to the encliticizing prosodic subcategorization frame associated with the adverbial suffixes *baano* 'only' and *fíihna* 'very'/'indeed'.

One good candidate for a suffix which can attach to a null copula and which *does* encliticize onto the preceding phonological material is the modal suffix *aachi̲/aachi̲ni*. It can takes a clausal complement ending in the past-tense suffix -*tok*, and adds a degree of uncertainty to the assertion it embeds, as in (22).

(22) a. Kátos-at hali̲bis faláaya-to[g] Ø-aachi̲ni-h.
          cat-NOM tail   long:LG-PST   COP-MOD-TNS
          'It seems that the cat has a long tail'

      b. Icho̲li-li-tok-ak-oosh   aatábli-li-to[g]   Ø-aachi̲-h.
          write-1SG.I-PST-FOC-SS too.much:NG-1SG.I-PST COP-MOD-TNS
          'It seems I wrote too much.'

As shown, the final /k/ of the complement clause can be voiced—per the diagnostic in section 2.3, the /k/ is therefore PWd-medial, and so the material immediately following the /k/ must be part of the same PWd as the material before it. So *if* the morphemic decomposition provided in (22) is the correct one, then we have a case in which the material following a null copula encliticizes across that null copula, and across a clause boundary, and forms a PWd that spans across two clauses.

However, as an alternative decomposition, the initial /aa/ of *aachi̲* could be the overt copula. If this is the correct analysis, then the encliticization of *aachi̲* is less mysterious: it encliticizes because it begins with *a(a)*-, which is pre-specified to (optionally) encliticize. And in another alternative analysis, -*tokaachi̲ni/-tokaachi̲* is an atomic morpheme in itself, in which case the structure is not

13

biclausal (see Broadwell 2006:211ff.). However, this would leave unexplained why the alternation between -*aachi̱* and -*aachi̱ni* is preserved in the forms beginning *tok*-.

8 Conclusion

We have shown that Choctaw allows the formation of prosodic words that span across clause boundaries. The domain-spanning prosodic words we investigated all have a copula (null or overt) inside them. With the overt copula *a(a)*-, domain-spanning prosodic words are formed when the copula's optional encliticization frame is active. If the material directly to its left is a complement clause, then the copula will cliticize onto it, across the clause boundary. We argued that this account is preferable to an account which requires the complement of the copula to optionally undergo syntactic incorporation into the copula. We also argued that it challenges accounts of the syntax-prosody mapping that make clause boundaries 'hard' edges for prosodic word formation.

We further showed that within this model, we can account for the behavior of clause-level morphemes like the question particle -*o̱* as well. These typically attach to the verb, but, when the verb is a null copula, they can encliticize directly onto the copular object. We attribute this to -*o̱* having a prosodic pre-specification to encliticize, just like the overt copula *a(a)*-.

References

Beckman, Mary, and Janet Pierrehumbert. 1986. Intonational structure in English and Japanese. *Phonology* 3:255–309.

Bennett, Ryan, Boris Harizanov, and Robert Henderson. 2018. Prosodic smothering in Macedonian and Kaqchikel. *Linguistic Inquiry* 49:195–246.

Broadwell, George Aaron. 2006. *A Choctaw reference grammar*. Lincoln, NE: University of Nebraska Press.

Byington, Cyrus. 1870. Grammar of the Choctaw language (edited by Dr. Brinton). In *Proceedings of the American Philosophical Society*, volume 11, 317–67.

Byington, Cyrus. 1915. *A dictionary of the Choctaw language*. Number 46 in Bureau of American Ethnology Bulletin. Washington, D.C. (Reprinted: Oklahoma City Council of Choctaws, 1973. Reprinted: St. Clair Shores, MI: Scholarly Press, 1978).

Chomsky, Noam. 2000. Minimalist Inquiries: the Framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. Roger Martin, David Michaels, and Juan Uriagereka, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A Life in Language*, ed. Michael Kenstowicz, 1–52. Cambridge, MA: MIT Press.

Compton, Richard, and Christine Pittman. 2010. Word-formation by phase in Inuit. *Lingua* 120:2167–2192. Publisher: Elsevier.

Davies, William D. 1986. *Choctaw Verb Agreement and Universal Grammar*. Dordrecht: Reidel.

Ershova, Ksenia. 2020. Two paths to polysynthesis. *Natural Language & Linguistic Theory* 38:425–475. Publisher: Springer.

Grimshaw, Jane. 2000. Locality and extended projection. In *Lexical specification and insertion*, ed. Peter Coopmans, Martin Everaert, and Jane Grimshaw, 115–134. Amsterdam: John Benjamins.

Harizanov, Boris, and Vera Gribanova. 2019. Whither head movement? *Natural Language & Linguistic Theory* 37:461–522.

Inkelas, Sharon. 1989. *Prosodic constituency in the lexicon*. New York: Garland.

Inkelas, Sharon, and Draga Zec. 1990. *The phonology-syntax connection*. University of Chicago Press.

Ito, Junko, and Armin Mester. 2013. Prosodic subcategories in Japanese. *Lingua* 124:20–40.

Katenkamp, Seth. 2021. A description and analysis of pitch accent in the verbal morphology of Mississippi Choctaw: An Optimality Theory Account. Master's thesis, University of Florida.

Lema, José, and María-Luisa Rivero. 1990. Long head movement: ECP vs. HMC. In *Proceedings of NELS*, volume 20, 333–347.

Munro, Pamela. 1984. On the Western Muskogean source for Mobilian. *International Journal of American Linguistics* 50:438–450.

Munro, Pamela. 1985. Chickasaw accent and verb grades. In *Studia linguistica diachronica et synchronica: Werner Winter sexagenario*, 581–593. Berlin: Mouton de Gruyter.

Nespor, Marina, and Irene Vogel. 1986. *Prosodic phonology*. Berlin: Mouton de Gruyter.

Nicklas, Thurston Dale. 1974. The elements of Choctaw. Doctoral Dissertation, University of Michigan.

Prince, Alan, and Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Technical Report RuCCS TR-2, Center for Cognitive Science, Rutgers University, New Brunswick, NJ. Published, Malden, MA: Blackwell (2004).

Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of Grammar*, ed. Liliane Haegeman, 281–337. Dordrecht: Kluwer.

Selkirk, Elisabeth. 1981. On prosodic structure and its relation to syntactic structure. In *Nordic Prosody II*, ed. Thorstein Fretheim, 111–140. Trondheim: Tapir.

Selkirk, Elisabeth. 2009. On clause and intonational phrase in Japanese: The syntactic grounding of prosodic constituent structure. *Gengo Kenkyu* 136:35–73.

Selkirk, Elisabeth. 2011. The syntax-phonology interface. In *The Handbook of Phonological Theory*, ed. John Goldsmith, Jason Riggle, and Alan Yu, 435–483. Blackwell.

Travis, Lisa. 1984. Parameters and effects of word order variation. Doctoral Dissertation, MIT.

Tyler, Matthew. 2019a. Choctaw as a window into the clitic/agreement split. *Studia Linguistica* 73:299–338.

Tyler, Matthew. 2019b. Simplifying MATCH WORD: Evidence from English functional categories. *Glossa: a journal of general linguistics* 4.

Ulrich, Charles. 1986. Choctaw morphophonology. Doctoral Dissertation, UCLA.

Weber, Natalie. 2020. Syntax, prosody, and metrical structure in Blackfoot. PhD Thesis, University of British Columbia.