Using corpus data for testing a working hypothesis about Haitian Creole prosody

Alexander M. Teixeira Kalkhoff

Abstract: Le but de cette étude à base de corpus est de recueillir des informations sur la prosodie du Créole Haïtien, à savoir la place et la nature phonétique des événements prosodiques, des mouvements tonals, des structures et la complexité des syllabes, la constituance prosodique et l'allongement de la pénultième à partir d'un échantillon du Corpus of Northern Haitian Creole. Outre les questions techniques et méthodologiques du traitement des données audio et textuelles comme l'alignement temporel du signal avec le transcrit et de l'annotation prosodique, l'hypothèse d'un système prosodique hybride sera évaluée dans la lumière des observations.

Key words: Haitian Creole/Créole Haïtien; prosody/prosodie; hybrid prosodic system hypothesis/l'hypothèse de système prosodique hybride; prosodic annotation/annotation prosodique; corpus-based/à base de corpus

1. Introduction

Haitian Creole (henceforth HC) is a sociohistorical and linguistically well-described creole language (Fattier 2013; Valdman 2014, see there for related references). However, to date, linguists have focused by the majority on its grammar and lexicon, and there are only a few studies about HC prosody and intonation (Cadely 1997; Brousseau 2003; Fattier 2005; Bhatt & Nikiema 2006; Brousseau & Nikiema 2006).

For heuristic reasons I distinguish between prosody and intonation – notions, which are often used synonymously (cf., for example, *The Cambridge Handbook of Phonology* 2007). Here, prosody only extends to accentuation patterns, tones, rhythm, syllable structure, duration of segments, pauses, and prosodic phrasing. In contrast, intonation deals with speech acts (illocution), the epistemic status of the utterance, information structuring, syntactic and semantic phrasing, and turn taking. Intonation is generally related to the final pitch contour (nuclear accent) of the utterance or intonational phrase. This article studies only prosodic features as defined above. Research on intonation needs another methodological access (e.g. Prieto 2001; Interactive Atlas of Romance Intonation 2010-2014; Jun & Fletcher 2014; Frota & Prieto 2015).

In the last decades, creolists have focused on two major prosodic research questions. First, they have wondered about the role of the (West) African substrate and the European superstrate

languages by shaping the prosodic systems of today's creole languages. And, second, how we can understand the tonal character of many creole languages, knowing that the vast majority of creole languages do not have lexical tones (among many others, McWhorter 1998, 798; Bhatt & Plag 2006; Rivera Castillo & Faraclas 2006; Gooden et al. 2009; Clements & Gooden 2011).

The goal of the corpus-based study in hand is to gather information about HC prosody and to assess the working hypothesis of a hybrid prosodic system exploring the linguistic resources of the Corpus of Northern Haitian Creole (henceforth CNHC) freely accessible online. The article accounts as well as for technical and methodological aspects, such as automatic audio–text alignment of the corpus data and prosodic annotation, as for HC prosodic features.

The article is organized as follows. In section 2, I will expose a working hypothesis according to which HC is characterized by a hybrid prosodic system shaped by influences from West African languages and French. Section 3 describes how to establish the workflow for prosodic annotation and analysis for a short test piece taken from the CNHC. Based on that corpus fragment, section 4 discusses prosodic features of HC and reconsiders the working hypothesis. Finally, section 5 discusses the pros and cons of my corpus-based approach to HC prosody and intonation.

2. Working hypothesis: Haitian Creole prosody as a hybrid prosodic system

HC genesis took place in Saint-Domingue, the western part of the Caribbean Hispaniola Island, within the sociohistorical context of slave trade and plantation society (Chaudenson 2002; McWhorter 2005; Eltis & Richardson 2010). In the late 17th century, French colonialists started to buy West African slaves for their growing sugar plantations. Thus, linguistic sources for HC have been French as the major lexifier language (superstrate), languages from the Kwa group of West Africa, especially Gbe, Bantu languages of Central Africa (both groups as substrate), and some influences of Spanish, English, and Amerindian languages (e.g. Arawak-Taíno, Tupi) (Lefebvre 2006, 52-57; Fattier 2013, 195s.). Researchers acknowledge a large predominance of the West African Gbe languages.

While the origin of the HC lexicon is traditionally linked to French and the origin of the internal structuration of the HC grammar to the involved African substrate languages, there is no clear statement about the origin of HC prosody yet. According to Shelome Gooden and her colleagues' model of a hybrid prosodic system for Jamaican Creole (Gooden et al. 2009, see there for related references), I argue also for a hybrid prosodic system approach for HC. In the case of HC that

means that its prosodic system constitutes a new third that has been shaped by both French and African substrate languages (for a similar approach to HC see Brousseau 2003). Moreover, I am fully aware that we talk about linguistic events in the past, which started around 1700, and that prosodic systems may evolve like the grammar and the lexicon of a given language. But for lack of diachronic information, I am using synchronic accounts of French and West African languages for heuristic reasons.

Typologically, the French prosodic system is considered as a *stress* accent system (Hyman 2006). Consequently, pitch does not play a dominant or exclusive role to mark prominences and phrase boundaries in French. In connected speech, French does not have neither word accent, i.e. a lexically distinctive fixed syllable-related pitch movement within words, nor word stress, but a fixed phrase accent on the right edge of prosodic phrases (Hulst et al. 2010: 460). The prosodic phases or constituents of French are intonational phrase (IP) (in the traditional French terminology called "mot phonétique"), intermediate phrase (ip), and accent phrase (AP) (for the most recent account of the prosodic constituency and intonational phonology of French see Delais-Roussarie et al. 2015, see there for related references).[1] The most influential prosodic constituent of French is the AP.[2] Its importance results from the fact that it is the phonological domain of the accent assignment, therefore, the French AP governs even the primary stress accent of the superordinate IP (Jun & Fougeron 2002; Delais-Roussarie et al. 2011; Delais-Roussarie et al. 2015, 76). Furthermore, French is classified as a syllable language (once, syllable-timed language) (Auer 2001) with a very strong tendency to realize CV (consonant-vowel) and other open syllable structures (Abry & Veldeman-Abry 2007: 17).

The most salient prosodic event in French occurs within the domain of the stressed last full syllable on the right edge of the IP (right-oriented language). Speakers of French realize phonetically this primary stress accent by cumulative means of lengthening of the vocalic nucleus, articulatory strengthening (Fougeron & Keating 1997), and pitch movement, although the vocalic lengthening is

[1] Intonational phrases (IP) represent intonationally self-contained and in most cases semantically coherent prosodic units, which are boundary marked by an audible discontinuity. Intermediate phrases (ip) are assumed as coherent subunits of morpho-syntactically complex structures. Regardless of its rather syntactic definition, ip boundaries are often prosodically marked by prosodic events, such as changes within the pitch track. Accent phrases (AP) contain one or more content words accompanied by its or their depending grammatical function words; prosodically the form one single unit. Phonological words (ω) are single accentuable words (Nespor & Vogel 2007).

[2] An AP contains one or more content words accompanied by its/their functional words; an IP may contain one or more APs; and an ip maps onto coherent sub-constituents of complex syntactic units. The most important prosodic domain in French is the AP with its facultative, positionally free left initial phrase accent and its obligatory, right pitch accent, which is aligned with the last full syllable of the AP (Jun & Fougeron 2000 and 2002).

the most salient one. Hence, the specific linguistic rhythm of French results from the alternation of lengthened and non-lengthened syllables (Dufter 2004), in which the prominent lengthened syllables appear on the right edge of the IPs, and from the AP related tonal pattern of the pitch track whose phonological default shape is /LHiLH*/ (Jun & Fougeron 2002: 149-153).[3]

West African prosodic systems are considered typologically as *tone* systems (Creissels et al. 1997; Hyman 2001 and 2006; Hulst et al. 2010: 381-427). So, the West African prosodic systems deal primarily with tonal or *pitch* movements. There are lexical, grammatical, and pragmatical tones. Lexical tones are lexically distinctive syllable/mora-specific pitch movements within words (Yip 2002). Grammatical tones mark function word classes or morphemes and pragmatical tones mark focus (Welmers 1959; Stahlke 1971: 133-202; Fiedler & Jannedy 2013; Kügler 2015). Like in French, prosodic events, such as lexical tones and prosodic phrasing, occur on the right edge of the prosodic word (right-oriented languages) (Brousseau 2003).

In summary, we can say that the French stress accent system has encountered West African Gbe tonal systems during HC genesis. According to my working hypothesis of a hybrid HC prosodic system, I will pay special attention to HC prosodic phrasing, lengthening patterns, and tonal movements on the right edge of prosodic constituents, expecting traces from the contributing original prosodic systems.

3. Establishing the workflow for prosodic analysis

Recorded in 2007 and subsequently published by the Indiana University Creole Institute under the auspices of Albert Valdman, the Corpus of Northern Haitian Creole nowadays constitutes a freely available and downloadable online resource for linguistic research on HC. It documents speakers of Capois, i.e. the language variety of HC spoken in the north of Haiti near Cap Haïtien. The mentioned webpage gives access to the audio files of the recordings, to the text files of the transcripts, and to the transcription conventions of ten interviews of altogether approximately ten hours of speech data. Audio files are saved as MP3 and the transcripts as Microsoft® Word files with an overall size of 509 MB. Audio and text files are nor aligned. The transcripts are not translated in any other language.

[3] /LHiLH*/ means a tonal pattern as a sequence of a left i(nitial) pitch movement of a L(ow) und a H(igh) tone and a right tonal pattern of a L(ow) and a H(igh) tone, where the H(igh) stared tone is associated with the metrical strong syllable, i.e. the stressed syllable.

To establish the workflow for a test run, I chose the longer turn of speaker R from the first interview of the CNHC that can be seen in (1). Selection criteria were, first, the selected test piece should not be too close to the beginning, because a more natural and relaxed social and linguistic interaction between the speakers was expected after a certain time of contact and the fading of the awareness of being recorded, and, second, the test piece should represent a longer turn of one single speaker. The whole conversation was recorded in Thibeau near Cap Haitian and involves three participants, i.e. one interviewer (K), a native school teacher, and two consultants (S, R), two middle-aged farmers, both illiterate monolingual speakers of Capois, the Northern HC variety:[4]

(1) oke pou mayi ya ‖ m te ka bay yon ti fòmasyon nan mayi ya ‖ mayi ya yon epòk ‖ nan tan lontan lè papanm ‖ tè yo te pi kiltivab ‖ sezon an te pi mache pi byen ‖ ou sè ou konn rive nan tè a ‖ mèt gen rajo ‖ ou fouye tou mayi ou plante mayi ‖ apre lè w fini w sèrkle mayi ya ou fè mayi ‖ men nou rive nan on moman konnya nou vin pa gen sezon menm jan ankò ‖ nou nou sèrkle tè a ‖ lè w fin sèrkle tè a ‖ ou mete latè ‖ tè a vin rive nan lè w rive nan ‖ ou vin rive nan mwa desam ‖ janvye w ap sèrkle tè nou konn plante mayi janvye ‖ nou plante mayi fevriye ‖ nou plante mayi mas ‖ nou plante mayi avril se kat mwa sa n te konn gen n te konn kiltive mayi ya ‖ pou sezon ‖ lè w plante ‖ mayi ya ‖ mayi ya vin i leve ‖ ou sèrkle men nan tan lontan granmoun yo te konn plante i yon lòt jan ‖ tan yo te pi bon ‖ yo konn plante mayi ya senk grenn nan tou ‖ yo fouye tou mayi ya yo mete senk grenn mayi ‖ men nou menm konnya nou vin pa gen sezon menm jan ‖ nou plante mayi ya twa grenn ‖ e nou vin plante a yon distans tou ‖[5]

As a first step of the editing process, the downloaded audio signal was normalized, denoised, and saved as WAV audio file by using Audacity® (version 2.2.2), freely available software for recording and sound editing. It is a big technical deficit of the CNHC audio data for further automatic phonetic analysis that the recordings were saved originally in the lossy compressed MP3

[4] The test piece corresponds to the time interval from 00 h 19 min 26.540 sec to 00 h 20 min 27.165 sec of the audio signal and to the lines 211 to 222 of the transcript. I removed all metalinguistic information, punctuation, and capitalization, which the transcribers originally used for sentence marking of the orthographic transcription, from the transcript. Audible discontinuities were marked by the IPA symbol for intonational grouping ‖.

[5] ‘okay for the corn / I can give you a short instruction for the corn / the corn, in a time far in the past, at the age of our fathers / their soil was more cultivable / the season worked better / you know, you could arrive at a ground / you weeded / you dug all the corn, you planted the corn / after you finished to weed, you made corn / but, we arrived at a time where the season does not work longer in the same way / we weed the ground / when we finished to weed the ground / you put into the soil / you come along in the month of December / in January you will weed the ground we can plant corn in January / we plant corn in February / we plant corn in March / we plant corn in April, these four months we could have, we could cultivate the corn / for the season / when you plant / the corn / the corn grows / you weed, but, at the time far in the past, the elderly people they could plant in another way / their time was better / they used to plant the corn with five seeds in a hole / they dig a hole for the corn, they put five corn seeds in it / but we ourselves use, (because) we do not have the seasons in the same way / we plant corn with three seeds / and we plant at a totally different distance’ [My translation].

(MPEG-1 Audio Layer 3) audio format (128 kBit per second, 44.1 kHz) instead of WAV (Waveform Audio File Format) (1411 kBit per second, 44.1 kHz). But, both the Praat (Boersma & Weenink 2017) algorithm to calculate automatically the pitch and the MAUS (Munich AUtomatic Segmentation) algorithms (Schiel 1999; Kisler et al. 2012) to align the audio signal to the text and to segment the sound signal turned out to be robust enough to deal with the lossy compressed audio signal. Nevertheless, regarding the pitch track visualized in Praat, there are some artefacts, such as cracks and discontinuities, in our test piece.

The second step was the automatic sound–text alignment and word-and-segment-sized segmentation. This step had rather an experimental character because, of course, for my short sample, I could align the audio with the transcript file and segment the words and sound segments manually. I did this step rather in the perspective of taking soundings for a further use of bigger amounts of corpus data. To accomplish this step, I took advantage of the tools and services provided openly and free of charge by the Bavarian Archive for Speech Signals (BAS), which is part of the German CLARIN-D (Common Language Resources and Technology Infrastructure) research infrastructure for humanities and social sciences. The BAS team used the well-trained German acoustic model to perform the HC alignment and segmentation procedure because the German and HC phonemic spelling systems share many features (for HC spelling see Valdman 1981, x-xiii). The output format of the MAUS procedure is a Praat text grid containing the interval tiers ORT, i.e. word-tokenized orthographic transcript, KAN, i.e. canonical pronunciation, and MAU, i.e. SAM-PA (Speech Assessment Methods Phonetic Alphabet) encoded phonemes. As a result, the formerly unconnected audio signal and the text of the transcript are now temporally aligned and decomposed into words and sound segments, ready to be re-used for further phonetic analysis in Praat.

The third step was to create an annotation hierarchy for prosodic analysis in Praat taking into account the prosodic features mentioned above, i.e. prominences, syllable structure, local tonal movements, duration of segments, and prosodic phrasing. For that purpose, I adapted the Intonational Variation Transcription System (IVTS). IVTS is a language-non-specific annotation system to annotate intonational variation of still unknown phonological systems developed by Brechtje Post and Elisabeth Delais-Roussarie (Post & Delais-Roussarie 2006) on the basis of the language-specific Intonation Variation in English (IViE) annotation system (Grabe & Post 2004).

The IVTS originally encodes orthographic, prosodic, and intonational information on six annotation levels, i.e. (i) a comment, (ii) a phonological (or tonal), (iii) a global phonetic, (iv) a local phonetic, (v) a rhythmic (or prominence), and (vi) an orthographic tier. Starting point for the annotation

process is the identification of prominent syllables and a narrow phonetic annotation of local intonational events. Intonation at the discourse level is annotated phonetically at the global phonetic tier. First phonological assumptions can be also annotated at the phonological tier. The IViE/IVTS set of labels is transparent, easy to manage and re-uses well-established ToBI symbols, such as L, H, and % (Grabe 2001).

According to my strong focus on the listed prosodic features of HC, I adapted the IVTS and established the following annotation hierarchy for Praat: (i) a segmental tier (SEG) to enable durational measurements, (ii) a syllable tier (SYL) to evaluate syllable patterns and complexity, (iii) a rhythmic tier (RHY) to annotate prominent syllables and intonational phrasing, (iv) a local phonetic or micro-prosodic tier (MPR) to annotate phonetically the realized pitch levels or glides within the tonal implementation domain[6] of prominent syllables, (v) a break indices tier to annotate prosodic constituency (PCO), (vi) an orthographic tier (ORT) for an orthographic transcription of the speech signal, and (vii) a glossed tier (GLO) to enable to evaluate pitch movements upon function word classes.

Sound segments were labeled automatically by the MAUS algorithm with SAM-PA symbols. Labels for syllable constituents are basically C for consonants and V for vowels. On the rhythmic tier, phonetic salience of a syllable relative to adjacent syllables is labeled with P for prominence and rhythmic boundaries are marked by the boundary symbol %. On the micro-prosodic tier, the shape of the local pitch movement relative to the prominent syllables is annotated with six target labels. Capital labels H(igh), M(id), and L(ow) mark the pitch level of the accented syllables; minuscule labels h(igh), m(id), and l(ow) mark the pitch level of unaccentuated syllables preceding or following the prominent syllable (Grabe 2001).[7] Break indices indicate boundaries and junctures between prosodically sensitive constituents. The break index 4 indicates intonational phrase boundaries, 3 indicates intermediate phrase boundaries, 2 indicates clitic group boundaries, BI 1 indicates phonological word boundaries, and 0 indicates junctions between content and function words (Frota 2012). On the orthographic tier, the words from the transcript appear time-aligned with the audio signal as a result of the MAUS alignment process. Each word is glossed according to the *Leipzig Glossing Rules* (The Leipzig Glossing Rules 2015).

[6] The tonal implementation domain contains the pre-accentual syllable, the accented syllable, and all following syllables up to the next accented syllable or up to the end of the IP (Grabe 2001).

[7] According to the speaker's pitch range ranging from 130 to 240 Hz, H(igh) tones were assigned to pitch values ranging from 200 to 240 Hz, M(id) tones were assigned to pitch values ranging from 165 to 200 Hz, and L(ow) tones were assigned to pitch values ranging from 130 to 165 Hz.

4. Some prosodic features of Haitian Creole

With the aid of an extensive prosodic Praat annotation, I analyzed (i) prosodic phrasing, (ii) the location and phonetic nature of the HC accentuation, (iii) prominences upon function word classes, (iv) syllable patterns and complexity, and (v) penultimate lengthening. The goal was to identify the prosodic domain that governs the HC stress accent and to specify factors of the HC speech rhythm.

Based on the auditory analysis of discontinuities, changes in pitch, and stress patterns of the whole sample, I advocate the intonational phrase (IP), the intermediate phrase (ip), the clitic group (C)[8], and the phonological word (ω) as the prosodic domains of HC. Intonational phrase boundaries are marked by clear-cut breaks within the audio signal; intermediate phrase boundaries are marked by a change within the pitch track; and clitic groups/accent phrases and phonological words are the domain of the realization of the HC stress accent. Prosodic phrasing of a section of the sample is given in (2).

(2) [[ou vin rive]$_C$ [nan [mwa]$_ω$]$_C$ [desam]$_ω$]$_{IP}$ [[[janvye]$_ω$ [w ap [sèrkle]$_ω$]$_C$ [tè]$_ω$]$_{ip}$ [[nou konn [plante]$_ω$]$_C$ [mayi]$_ω$ [janvye]$_ω$]$_{ip}$]$_{IP}$ [[nou [plante]$_ω$]$_C$ [mayi]$_ω$ [fevriye]$_ω$]$_{IP}$ [[nou [plante]$_ω$]$_C$ [mayi]$_ω$ [mas]$_ω$]$_{IP}$ [[[nou [plante]$_ω$]$_C$ [mayi]$_ω$ [avril]$_ω$]$_{ip}$ [[[se [kat]$_ω$]$_C$ [[mwa]$_ω$ sa]$_C$]$_{ip}$ [n te konn [gen]$_ω$]$_C$ [n te konn [kiltive]$_ω$]$_C$ [[mayi]$_ω$ ya]$_C$]$_{ip}$]$_{IP}$ [[pou [sezon]$_ω$]$_C$]$_{IP}$
'you come along in the month of December / in January you will weed the ground we can plant corn in January / we plant corn in February / we plant corn in March / we plant corn in April, these four months we could have, we could cultivate the corn / for the season'

Clearly, we can state that HC is a word accent language and not a phrase accent language like French (see Hulst et al. 2010: 3-53 for a discussion on word accent terminology). As we can see in figure 1, every phonological word and clitic group are accentuable in HC by the phonetic means of a stress accent. Salient phonetic means of the HC stress accent is a rise in intensity upon vocalic segments (the dashed line indicates the intensity track). In the IP [yo konn plante mayi ya senk grenn nan tou]$_{IP}$ ('they used to plant the corn with five seeds in a hole'), the informational focus is on <senk> (five) at the center of the IP. Here, the speaker emphasizes the fact that once they used FIVE seeds of corn instead of today's three seeds to plant corn. In addition to a local intensity peak, the word <senk> is articulatorily strengthened, i.e. pronounced very clearly, and its segments are

[8] A clitic group is a combination of a word and clitic(s) that forms a coherent phonological domain (Nespor & Vogel 2007: 145-163).

considerably lengthened compared to adjacent segments. This kind of intonational focusing, i.e. the intonational highlighting of an element within connected speech, is only possible in word accent languages.

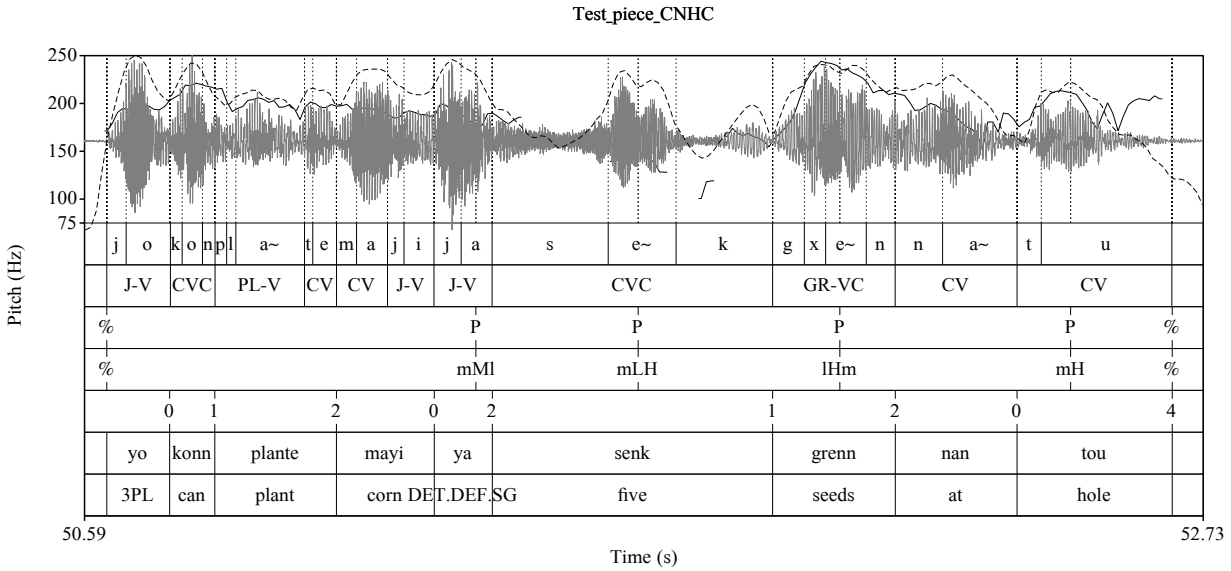

Figure 1: Indication of HC stress accent patterns by displaying the intensity track (dashed line) and the pitch track (solid line) of the IP [yo konn plante mayi ya senk grenn nan tou]$_{IP}$ (‘they used to plant the corn with five seeds in a hole’).

Stress patterns of polysyllabic words and clitic groups indicate that the HC stress accent is assigned to the right edge of the respective prosodic domain (in accordance with Brousseau 2003: 123 and Fattier 2013: 196). An example is given in figure 2 with the polysyllabic word <kultivab> (cultivable) where the rightmost syllable receives the stress accent. Interestingly, clitics are fully accentuable and extent the accentuable domain of the connected content word. By the way, external sandhi is the expression of the firm phonological connection between the content word and the connected clitics in HC (Dejean 1980: 140-149; Cadely 1997), e.g. the postpositive HC singular definiteness marker <la> alternates phonetically as <a, an, nan, lan, ya> according to the preceding last sound of the connected content word. In figure 1, we can see the stress pattern of the two clitic groups [yo konn plante]$_{C}$ (3PL can plant) and [mayi ya]$_{C}$ (corn-DET.DEF.SG). Both are stressed and <ya> is even one of the most prominent syllables of the whole IP. In figure 2, the HC past marker <te> is the most prominent syllable of the IP [tè yo te pi kiltivab]$_{IP}$ (‘their soil was more cultivable’). Here, the speaker emphasizes the fact that the soil WAS more cultivable in the past. The same can be confirmed for all function word classes of the sample, as well as for the

postpositive nominal markers DET.DEF.SG/PL <a, ya, yo>, DET.INDF.SG/PL <yon>, POSS.PL <yo>, and DEM <sa> as for the prepositive verbal markers PST <te> and FUT <ap>.[9]

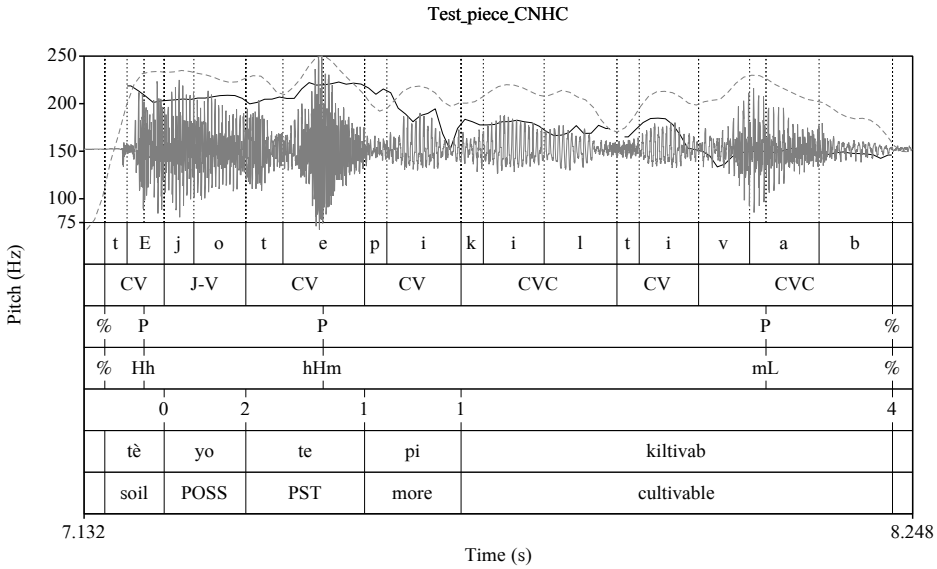

Figure 2: Stress pattern of the past marker <te> and the polysyllabic word <kultivab> within the IP [tè yo te pi kiltivab]$_{IP}$ ('their soil was more cultivable'). The dashed line indicates the intensity track and the solid line the pitch track.

Our sample contains 287 syllables overall. Table 1 shows the ranked distribution of a detailed account of used syllable patterns and syllabic complexity. As well as the glides, symbolized by J and W, as the consonant clusters known as *muta cum liquida*, i.e. the articulatory connection between a plosive or fricative with a liquid, symbolized by PL, KL, GR, and VR, and their combinations with following sounds are indicated. For each syllable pattern, an example is given. Table 2 displays a condensed version of the syllable pattern ranking. Here, structures, such as CV and J-V or CCV and *muta cum liquida* structures, are merged to CV respective to CCV.

Rank	Syllable structure	Number	Percentage
1.	CV <te> (PST)	148	51,57%
2.	J-V <ya> (DET.DEF.SG)	40	13,94%
3.	CVC <kat> (four)	39	13,59%
4.	V <e> (and)	24	8,36%
5.	PL-V <plante> (plant-V)	10	3,48%
6.	KL-V <sèrkle> (weed-V)	5	1,74%
7.	CJ-V <janvye> (January)	4	1,39%

[9] DET determiner; DEF definite; INDF indefinite; SG singular; PL plural; POSS possessive; DEM demonstrative; PST past; FUT future.

7.	W <w> (2SG)	4	1,39%
8.	GR-VC <grenn> (seed)	3	1,05%
8.	CW-V <mwa> (month)	3	1,05%
9.	C <n> (1PL)	2	0,7%
10.	VC <ap> (FUT)	1	0,35%
10.	GR-V <granmoun> (adult)	1	0,35%
10.	PR-V <apre> (after)	1	0,35%
10.	VR-V <fevriye> (February)	1	0,35%
10.	VR-VC <avril> (April)	1	0,35%
		Σ 287	Σ 100%

Table 1: A detailed account of syllable patterns and syllabic complexity of the HC sample.

Rank	Syllable structure	Number	Percentage
1.	CV	188	65,51%
2.	CVC	39	13,59%
3.	CCV	25	8,71%
4.	V	24	8,36%
5.	C	6	2,09%
6.	CCVC	4	1,39%
7.	VC	1	0,35%
		Σ 287	Σ 100%

Table 2: A condensed account of syllable patterns and syllabic complexity of the HC sample.

Undeniably, there is a very strong tendency for CV syllable structures (65,51% in table 2) and open syllable structures (82,64%), i.e. syllables ending with a vocalic sound (CV, CVC, CCV, V), in our HC sample. Within creole language genesis, the CV syllable pattern is considered as the prototypical and uniform outcome of a typologically non-marked language pattern (Schramm 2015: 4-13). Nevertheless, our sample shows also typologically marked complex onsets (CCV and CCVC); complex codas are absent in our sample (in accordance with Cadely 2003 and Fattier 2013).

In the sample, we can find a very specific rhythmic pattern that involves the lengthening of the vocalic nucleus of the penultimate syllable followed by the stressed ultimate syllable. In figure 3, such pattern appears within the realization of the word <avril> (April). Here, the vowel /a/ is lengthened compared to the vocalic nucleus of the stressed last syllable. Such rhythmic pattern is known and described as penultimate lengthening for Bantu languages (Hyman 2009) and for Palenquero, a Spanish based creole language (Hualde & Schwegler 2008; Correa 2012).

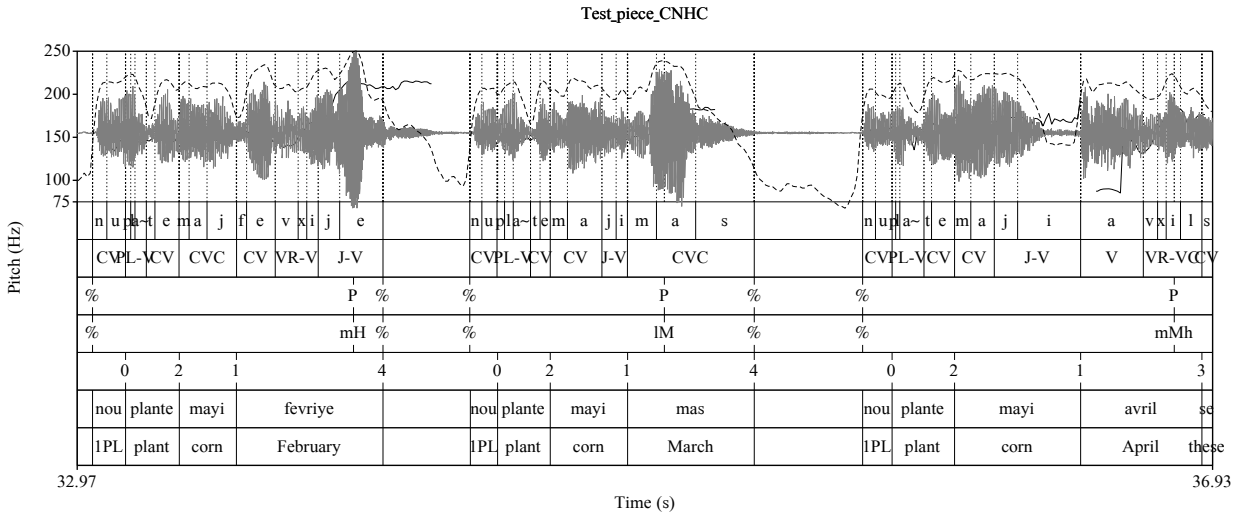

Figure 3: Penultimate lengthening rhythmic pattern within the word <avril> (April) of the ip <nou plante mayi avril> (‘we plant corn in April’) and the dynamic rising enumeration pitch pattern for the enumeration sequence of <fevriye> (February), <mas> (March), and <avril> (April). The dashed line indicates the intensity track and the solid line the pitch track.

To conclude this section, HC prosodic features are reflected in the light of French and West African prosodic systems. First, unlike French, HC does not have a phrasal accent but word accentuation patterns like West African languages. But, unlike the West African tonal marking of lexical differences, HC uses stress accents phonetically realized as rises in intensity and articulatory strengthening. However, like both French and West African languages, HC is a right-oriented language, i.e. the salient prosodic events take place at the right edges of the relevant prosodic domains. So, we can state the same location for prominences but realized by different phonetic means.

Second, although HC does not have lexical tones, function word classes, such as DET and PST, and focused elements of the utterance can be set prominent by a rise in intensity or/and by pitch movements. HC definiteness markers represent a case of convergence because their origins can be attributed to both French and West African languages for their formal identity (Bollée 1982: 394s.; Stein 2017: 119-122). But, there is also an important parallelism between the prosody of the HC and Ewe definiteness markers <la, a>. In addition to their formal identity, in Ewe those grammemes are realized prosodically by grammatical tones (Stahlke 1971: 133-202). So, the HC stress/intonational pattern upon function word classes and focused elements can be interpreted as a residuum of West African grammatical and pragmatic tones and be understood as a “re-interpretation” (Devonish 2002) of them by means of stress accent. Prominent elements are louder and pitch marked.

Third, penultimate lengthening, which is attested for West African languages and other creole languages, occurs also in HC. Fourth, both the HC and French lexicon have a very strong tendency for open CV syllable structures, which represents more than half of all possible syllable patterns.

The hybrid system hypothesis is an illuminative approach to creole grammar genesis (Aboh 2015) as well as to prosody and intonation (Brousseau 2003; Gooden et al. 2009), which goes hand in hand with the perception-based idea of the filter function of the African substrate languages while interpreting lexifier language structures and features during uncontrolled language acquisition scenarios (Hazaël-Massieux 1993). The observed HC stress pattern upon function words and focused elements might be an illustrative example for prosodic hybridization, i.e. the re-interpretation of West African grammatical and pragmatical tones by means of French stress accent.

5. Pros and cons of a corpus-based approach to HC prosody and intonation

Currently, spoken and written, synchronic and diachronic creole language corpora are a prolific and quite quickly evolving field (see amongst others Hagemeijer et al. 2014 for the building of a Gulf of Guinea Creole corpus and Kriegel 2015 for an overview of French-based creole corpora). Besides the CNHC, online resources for HC of varying quality and functionality are the short APiCS online (2013) HC sample, several recordings of spoken HC from the 1980s provided by the French meta-platform for oral corpora CoCoON, and the Haitian Creole language data corpus (2010) built by the Language Technologies Institute of Carnegie Mellon University's School of Computer Science. But, none of them complies with the full range of desirable corpus functionalities and technical standards, i.e. audio–text alignment, glossing, translation, metadata, annotations, license for free data use, download and publishing (e.g. creative commons), clear indication for citation, high fidelity audio recordings, and TXT or XML text format for transcripts, metadata and annotations (see amongst others Gries & Newman 2013 and the CLARIN-D standards information system).

Corpus-based approaches to language are economic and sustainable, especially if we consider language communities that live far from our own home country. Instead of prior time, money, and infrastructure-consuming fieldwork, we can start with linguistic analysis nearly immediately by re-using already existing data. Of course, besides that undeniable merit, many technical and methodological problems arise while using corpus data. On the one hand, these are related to the general properties of linguistic corpora, which linguists usually create for specific research questions. In the case of the used CNHC, linguists have explored the corpus data for lexical,

morphophonological, syntactic, and sociolinguistic concerns (for recent research see, for instance, Valdman et al. 2015). Thus, the design of the interviews and the quality of the recordings did not pay special attention to the usability of the speech data for later phonetic, prosodic, or even intonational analyses. On the other hand, many problems of today's usability have to do with technical standards and limits of the date of creation of the corpus. For instance, in 2007, the audio files of the CNHC have been saved in the disk space-saving but lossy MP3 audio format. Today the memory capacity of electronic recording devices and of computer hard drives as well as the online data transmission are not a major problem anymore. Large WAV files neither limit field recordings nor website functionalities of online-corpora.

In face of all undeniable challenges and inconveniences of field research, creolists in the field should keep in mind well-established technical and methodological standards for language resources and technology (see Maddieson 1999 and Podesva & Zsiga 2013 for phonetic field work and the CLARIN-D standards information system for data formats). Audio files should be recorded and saved as WAV and if any audio compression is needed, it should be executed by lossless compressed digital audio formats such as FLAC (Free Lossless Audio Codec). For the subsequent transcription and annotation process, computer programs such as freely downloadable ELAN or EXMARaLDA should be used instead of creating separate Word or Excel documents to ensure suitable data formats.

A major challenge for creole language prosody and intonation research are the background noises due to the natural environment of the interview settings such as crickets, barking dogs, human voices, and human activities such as playing children or the use of tools. In the case of the CNHC, interviews 5, 6, 7, 8, and 10, i.e. half of the corpus data, cannot be used at all for phonetic analysis because of strong background noises. Therefore, researchers in the field creating new audio or multimedia data should seek noise-reduced recording facilities.

Last but not least, a major methodological problem for corpus-based approaches to *intonation* is that the researcher has to content him/herself with the speech data from the corpus. For instance, the ten interviews of the CNHC are designed in a very similar way: In the presence of Albert Valdman, a local teacher chairs the conversation between two speakers about different topics, such as agriculture, traditions, festivities, working activities, and so on. There is no controlled elicitation of particular speech acts or intonational patterns. An outlook of how corpus-based analysis of intonational patterns can look like might give the enumeration pattern of [nou plante mayi fevriye]$_{IP}$ [nou plante mayi mas]$_{IP}$ [nou plante mayi avril]$_{ip}$. ('we planted corn in February / we planted corn in

March / we planted corn in April') displayed in figure 3. Here the speaker enumerates months using a dynamic, domain-final rising pitch pattern. However, this inductive, yet knowledge-driven, approach to illocution needs deeper methodological consideration, e.g. by comparing elicitated intonational patterns with contextually embedded ones of the same speech act. For the moment, corpus-based and data-driven approaches to intonation still remain a desideratum. Nevertheless, some automatic parametric but yet highly idiosyncratic data-driven approaches to intonation (Möhler 1998; Taylor 2000; Reichel 2010 and 2014, see there for related references) are promising for the future for exploring bigger amounts of corpus data in a user-friendly way.

Acknowledgements

I am very grateful to Uwe Reichel from the CLARIN-D infrastructure of the Ludwig Maximilian University of Munich who enabled the use of the BAS tools for HC alignment and segmentation.

6. References

Aboh, Enoch O. (2015) *The emergence of hybrid grammars: Language contact and language change* (Cambridge approaches to language contact), Cambridge: Cambridge University Press.
APiCS online (2013) = *Atlas of Pidgin and Creole Language Structures online*, accessed 2018-06-10. URL: http://apics-online.info/
Abry, Dominique & Veldeman-Abry, Julie (2007) *La Phonétique – audition, prononciation, correction*. Paris: CLE International.
Audacity® (version 2.2.2), accessed 2018-06-10. URL: http://audacity.sourceforge.net/
Auer, Peter (2001) "Silben- und akzentzählende Sprachen", in *Language Typology and Language Universals / Sprachtypologie und sprachliche Universalien / La typologie des langues et les universaux linguistiques* (Handbücher zur Sprach- und Kommunikationswissenschaft 20.2), Berlin: De Gruyter, p. 1391-1399.
Bhatt, Parth & Nikiema, Emmanuel (2006) "Empty positions in Haitian Creole syllable structure", in Parth Bhatt & Ingo Plag (eds.) *The structure of Creole words: Segmental, syllabic and morphological aspects* (Linguistische Arbeiten 505), Tübingen: Niemeyer, p. 85-105.
Bhatt, Parth & Plag, Ingo (eds.) (2006) "Stress, tone and intonation in Creoles and contact languages", *Language Typology and Universals* 59(2), p. 131-218.
Boersma, Paul & Weenink, David. 2017. Praat: doing phonetics by computer [Computer program]. Version 6.0.30, retrieved 22 July 2017 from http://www.praat.org/

Bollée, Annegret (1982) “Die Rolle der Konvergenz bei der Kreolisierung”, in Ureland, P. Sture (ed.) *Die Leistung der Strataforschung und der Kreolistik: Typologische Aspekte der Sprachkontaktforschung* (Linguistische Arbeiten 125), Tübingen: Niemeyer, p. 391-405.
Brousseau, Anne-Marie (2003) “The accentual system of Haitian Creole: The role of transfer and markedness values”, in Plag, Ingo (ed.) *Phonology and morphology of creole languages* (Linguistische Arbeiten 478), Tübingen: Niemeyer, p. 123-145.
Brousseau, Anne-Marie & Nikiema, Emmanuel (2006) “From Gbe to Haitian: The multi-stage evolution of syllable structure”, in Lefebvre, Claire & White, Lydia & Jourdan, Christine (eds.) *L2 acquisition and creole genesis*, Amsterdam: Benjamins, p. 295-330.
Cadely, Jean-Robert (1997) “Prosodie et cliticisation en créole haïtien”, *Études Créoles* 20, p. 77-88.
Cadely, Jean-Robert (2003) “Les sons du Créole Haïtien”, *Journal of Haitian Studies* 9(2). p. 4-41.
Robert Chaudenson (2002) “Une théorie de la créolisation : le cas des créoles français”, *Études Créoles* 25(1), p. 25-44.
CLARIN-D (Common Language Resources and Technology Infrastructure), accessed 2018-06-10. URL: http://www.clarin-d.de
CLARIN-D standards information system, accessed 2018-06-01. URL: https://clarin.ids-mannheim.de/standards/index.xq;jsessionid=5vjm9hrn6sjz1nc0ypi9mest3
Clements, J. Clancy & Gooden, Shelome (2011) *Language change in contact languages. Grammatical and prosodic considerations*, Amsterdam: Benjamins.
CoCoON (Collections de Corpus Oraux Numériques), accessed 2018-06-10. URL: http://cocoon.huma-num.fr/exist/crdo
Corpus of Northern Haitian Creole (2007), accessed 2018-06-10. URL: http://www.indiana.edu/~creole/index.shtml
Correa, José Alejandro (2012) “La entonación del criollo palenquero y del kateyano de San Basilio de Palenque (Colombia)”, in Maglia, Graciela & Schwegler, Armin (eds.) *Palenque, Colombia. Oralidad, identidad y resistencia*, Bogotá: Pontificia Universidad Javeriana, p. 31-56.
Creissels, Denis & Chebanne, A.M. & Nkhwa, H.W. (1997) *Tonal morphology of the Setswana verb*, München: LINCOM Europa.
Dejean, Yves (1980) *Comment écrire le créole d’Haïti* (Collectif Paroles), Québec: Outremont.
Delais-Roussarie, Elisabeth & Post, Brechtje & Avanzi, Mathieu & Buthke, Carolin & Di Cristo, Albert & Feldhausen, Ingo & Jun, Sun-Ah & Martin, Philippe & Meisenburg, Trudel & Rialland, Annie & Sichel-Bazin, Rafèu & Yoo, Hi-Yon (2015) “Intonational phonology of French: Developing a ToBI system for French”, in Frota, Sónia & Prieto, Pilar (eds.) *Intonation in Romance*, Oxford: Oxford University Press, p. 63-100.

Delais-Roussarie, Elisabeth & Yoo, Hiyon & Post, Brechtje (2011) “Quand frontières prosodiques et frontières syntaxiques se rencontrent”, *Langue Française* 170(2), p. 29-44.
Devonish, Hubert (2002) *Talking Rhythm Stressing Tone: The role of prominence in Anglo-West African languages*, Kingston (Jamaica): Arawak Publications.
Downing, Laura J. & Rialland, Annie (eds.) (2016) *Intonation in African Tone Languages*, Berlin: De Gruyter.
Dufter, Andreas (2004) “Ist das Französische eine silbenzählende Sprache?”, in Meisenburg, Trudel & Selig, Maria (eds.) *Nouveaux départs en phonologie: Les conceptions sub- et suprasegmentales* (Tübinger Beiträge zur Linguistik 470), Tübingen: Narr, p. 139-159.
ELAN, accessed 2018-06-10. URL: https://tla.mpi.nl/tools/tla-tools/elan/
Eltis, David & Richardson, David (2010) *Atlas of the transatlantic slave trade*, New Haven, CT: Yale University Press.
EXMARaLDA, accessed 2018-06-10. URL: http://exmaralda.org/de/
Fattier, Dominique (2005) “Remarques sur l’interrogation en créole haïtien”, *La Linguistique* 41(1), p. 41-55.
Fattier, Dominique (2013) “Haitian Creole”, in Michaelis, Susanne Maria & Maurer, Philippe & Haspelmath, Martin & Huber, Magnus (eds.) *The survey of pidgin and creole languages*, vol. 2, Oxford: Oxford University Press, p. 195-204.
Fiedler, Ines & Jannedy, Stefanie (2013) “Prosody of focus marking in Ewe”, *Journal of African Languages and Linguistics* 34(1), p. 1-46.
Fougeron, Cécile & Keating, Patricia A. (1997) “Articulatory strengthening at edges of prosodic domains”, *Journal of the Acoustical Society of America* 101(6), p. 3728-3740.
Frota, Sónia (2012) “Prosodic structure, constituents and their implementation”, in Cohn, Abigail C. & Fougeron, Cécile & Huffman, Marie K. (eds.) *Handbook in Laboratory Phonology*, Oxford: Oxford University Press, p. 255-265.
Frota, Sónia & Prieto, Pilar (eds.) (2015) *Intonation in Romance*, Oxford: Oxford University Press.
Gooden, Shelome (2014) “Aspects of the intonational phonology of Jamaican Creole”, in Jun, Sun-Ah (ed.) *Prosodic typology II: The phonology of intonation and phrasing*, Oxford: Oxford University Press, p. 273-301.
Gooden, Shelome & Drayton, Kathy-Ann & Beckman, Mary (2009) “Tone inventories and tune-text alignments. Prosodic variation in ‘hybrid’ prosodic systems”, *Studies in Language* 33(2), p. 396-436.
Grabe, Esther (2001) “The IViE Labelling Guide, version 3”, accessed 2018-06-10. URL: http://www.phon.ox.ac.uk/files/apps/IViE/guide.html

Grabe, Esther & Post, Brechtje (2004) “Intonational variation in the British Isles”, in Sampson, Geoffrey & McCarthy, Diana (eds.) *Corpus Linguistics: Readings in a widening discipline*, London: Continuum International, p. 474-481.
Gries, Stefan Th. & Newman, John (2013) “Creating and using corpora”, in Podesva, Robert J. & Sharma, Devyani (eds.) *Research Methods in Linguistics*. Cambridge: Cambridge University Press, p. 257-287.
Hagemeijer, Tjerk & Généreux, Michel & Hendrickx, Iris & Mendes, Amália & Tiny, Abigail & Zamora, Armando (2014) The Gulf of Guinea creole corpus, *Proceedings LREC 2014*, Reykjavik (Iceland), p. 523-529.
Haitian Creole language data (2010), accessed 2018-06-10. URL: http://www.speech.cs.cmu.edu/haitian/
Hazaël-Massieux, Guy (1993) “The African filter in the genesis of Guadeloupean Creole: at the confluence of genesis and typology”, in Mufwene, Salikoko S. (ed.) *Africanisms in Afro-American language varieties*, Athens: University Press of Georgia, p. 81-110.
Hualde, José Ignacio & Schwegler, Arnim (2008) “Intonation in Palenquero”, *Journal of Pidgin and Creole Languages* 23(1), p. 1-31.
Hulst, Harry van der & Goedemans, Rob & Zanten, Ellen van (eds.) (2010) *A Survey of World Accentual Patterns in the Languages of the World*, Berlin: De Gruyter Mouton.
Hyman, Larry M. (2001) “Tone systems”, in *Language Typology and Language Universals / Sprachtypologie und sprachliche Universalien / La typologie des langues et les universaux linguistiques* (Handbücher zur Sprach- und Kommunikationswissenschaft 20.2), Berlin: De Gruyter, p. 1367-1380.
Hyman, Larry M. (2006) “Word-prosodic typology”, *Phonology* 23(2), p. 225.-257.
Hyman, Larry M. (2009) “Penultimate lengthening in Bantu: Analysis and spread”, *UC Berkley Phonology Lab Annual Report* 1, p. 195-209.
Interactive Atlas of Romance Intonation (2010-2014), coordinated by Pilar Prieto & Joan Borràs-Comes & Paolo Roseano, accessed 2018-06-10, URL: http://prosodia.upf.edu/iari/
Jun, Sun-Ah & Fletcher, Janet (2014) “Methodology of studying intonation: From data collection to data analysis”, in Jun, Sun-Ah (ed.) *Prosodic typology II: The phonology of intonation and phrasing*, Oxford: Oxford University Press, p. 493-519.
Jun, Sun-Ah & Fougeron, Cécile (2000) “A phonological model of French intonation”, in Botinis, Antonis (ed.) *Intonation: Analysis, modelling and technology*, Dordrecht: Kluwer, p. 209-242.
Jun, Sun-Ah & Fougeron, Cécile (2002) “Realizations of accentual phrase in French intonation”, *Probus* 14, p. 147-172.

Kisler, Thomas & Schiel, Florian & Sloetjes, Han (2012) Signal processing via web services: The use case WebMAUS. in *Proceedings Digital Humanities 2012*, Hamburg (Germany), p. 30-34.
Kriegel, Sibylle (2015) "La documentation linguistique des franco-créoles", in Iliescu, Maria & Roegiest, Eugeen (eds.) *Manuel des anthologies, corpus et textes romans* (Manuals of Romance Linguistics 7), Berlin: De Gruyter, p. 647-658.
Kügler, Frank (2016) "Tone and intonation in Akan", in Downing, Laura J. & Rialland, Annie (eds.) *Intonation in African Tone Languages*, Berlin: De Gruyter, p. 89-129.
Lefebvre, Claire (2006) *Creole genesis and the acquisition of grammar: The case of Haitian Creole* (Cambridge Studies in Linguistics 88), Cambridge: Cambridge University Press.
Maddieson, Ian (2001) "Phonetic fieldwork", in: Newman, Paul & Ratliff, Martha (eds.) *Linguistic Fieldwork*. Cambridge: Cambridge University Press, p. 211-229.
McWorther, John H. (1998) "Identifying the creole prototype: Vindicating a typological class", *Language* 74(4), p. 788-818.
McWhorter, John H. (2005) *Defining creole*, Oxford: Oxford University Press.
Möhler, Gregor (1998) Describing intonation with a parametric model, in *Proceedings ICSLP 1998*, Sydney (Australia), p. 2851-2854.
Nespor, Marina & Vogel, Irene (2007) *Prosodic phonology* (Studies in Generative Grammar 28), Berlin: Mouton de Gruyter.
Podesva, Robert J. & Zsiga, Elizabeth (2013) "Sound recordings: acoustic and articulatory data", in Podesva, Robert J. & Sharma, Devyani (eds.) *Research Methods in Linguistics*. Cambridge: Cambridge University Press, p. 169-194.
Post, Brechtje & Delais-Roussarie, Elisabeth (2006) "Transcribing intonational variation at different levels of analysis", in *ISCAM Archives, Speech Prosody 2006*, Dresden (Germany).
Prieto, Pilar (2001) "L'entonació dialectal del català: El cas de les frases interrogatives absolutes", in Bover, A. & Lloret, M.-R. & Vidal-Tibbits, M. (eds.) *Actes del novè colloqui d'estudis catalans a Nord-Amèrica*, Barcelona: Publicacions de l'Abadia de Monserrat, p. 347-377.
Reichel, Uwe D. (2010) *Datenbasierte und linguistisch interpretierbare Intonationsmodellierung*, München: Ludwig-Maximilians-Universität. (Doctoral dissertation.)
Reichel, Uwe D. (2014) "Linking bottom-up intonation stylization to discourse structure", *Computer Speech and Language* 28, p. 1340-1365.
Rivera Castillo, Yolanda & Faraclas, Nicholas (2006) "The emergence of systems of lexical and grammatical tone and stress in Caribbean and West African Creoles", *STUF - Sprachtypologie und Universalienforschung* 59(2), p. 148-169.
Schiel, Florian (1999) Automatic Phonetic Transcription of Non-Prompted Speech, in *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, p. 607-610.

Schramm, Mareile (2015) *The emergence of Creole syllable structure. A cross-linguistic study*, Berlin: De Gruyter.
Stahlke, Herbert Frederic Walter (1971) *Topics in Ewe phonology*, Los Angeles: University of California. (Doctoral dissertation.) (URL: http://linguistics.ucla.edu/images/stories/Stahlke.1971.pdf) (Accessed 2018-06-10.)
Stein, Peter (2017) *Kreolisch und Französisch* (Romanistische Arbeitshefte 25), Berlin: De Gruyter.
Sylvain, Suzanne ([1936] 1979) *Le créole haïtien. Morphologie et syntaxe*, Genève: Slatkine Reprints.
The Atlas of Pidgin and Creole Language Structures (2013), edited by Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath, and Magnus Huber. Oxford: Oxford University Press.
The Leipzig Glossing Rules (2015), accessed 2018-06-10. URL: https://www.eva.mpg.de/lingua/resources/glossing-rules.php
Taylor, Paul (2000) "Analysis and synthesis of intonation using the tilt model". *Journal of the Acoustical Society of America* 96, p. 2064-2075.
The Cambridge Handbook of Phonology (2007), edited by Paul de Lacy, Cambridge, Cambridge University Press.
Valdman, Albert (ed.) (1981) *Haitian Creole-English-French dictionary*, Bloomington: Indiana University, Creole Institute.
Valdman, Albert (2014) *Haitian creole: Structure, variation, status, origin*, Sheffield: Equinox.
Valdman, Albert & Villeneuve, Anne-José & Siegel, Jason F. (2015) "On the influence of the standard norm of Haitian Creole on the Cap Haïtien dialect: Evidence from sociolinguistic variation in the third person singular pronoun". *Journal of Pidgin and Creole Languages* 30(3), p. 1-43.
Welmers, William E. (1959) "Tonemics, morphotonemics, and tonal morphemes". *General Linguistics* 4, p. 1-9.
Yip, Moira (2002) *Tone*, Cambridge: Cambridge University Press.