Phonology (2026), **Vol. 43**:e6 1–26
doi:10.1017/S0952675726100281

PHONOLOGY

ARTICLE

Stress assignment in Abkhaz verbs: a corpus-based approach

Samuel Andersson

Department of Linguistics, Yale University, USA
E-mail: rubyvendela@gmail.com

Received: 4 December 2022; **Revised:** 19 October 2023; **Accepted:** 13 November 2024

Keywords: corpus phonology; Abkhaz; lexical stress; prosody; Caucasian languages

Abstract

This article tests theories of verb stress in the Northwest Caucasian language Abkhaz using a new database corpus of 3,115 inflected forms of 445 verbs. I describe the creation of the corpus and show how it can be used to gain new insights into principles of Abkhaz stress assignment, which depend on complex interactions between phonology and polysynthetic verbal morphology. I implement a previous theory of Abkhaz stress assignment (Dybo 1977) in a computer program and use the corpus to assess empirically how well it accounts for stress patterns across the lexicon of eventive verbs in Abkhaz. I show how this empirical evaluation identifies both strengths and weaknesses of the theory, and use these to propose a revised theory of Abkhaz stress assignment. The revised theory ties with or outperforms the original in all verb categories, accounting for the stress alternations in 40 additional verbs, which comprise almost 10% of the corpus. This shows that corpora, combined with computational implementations of phonological theories, can be used to further our understanding of highly complex phonological data sets.

Contents

1. Introduction **2**
2. Language background **3**
2.1. Phonology . . . 3
2.2. Verbal morphology . . . 5
3. A database corpus of Abkhaz verbs **7**
3.1. The data . . . 7
3.2. The creation of the corpus . . . 9
4. Methodology: using corpora to evaluate phonological theories **12**
4.1. Dybo's Rule: a proposal for Abkhaz stress assignment . . . 12
4.2. Methodology: evaluating Dybo's Rule empirically . . . 13
5. Results **16**
6. Revising the theory **17**
6.1. Motivating the revisions . . . 17
6.2. Evaluating the revised theory . . . 22
7. Discussion **23**
8. Conclusion **24**
References **25**

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

1. Introduction

This article reports on a corpus study of stress assignment in verbs from the Northwest Caucasian language Abkhaz. This language is known for its phonological and morphological complexity, and features contrastive stress with frequent stress alternations within morphological paradigms, as well as polysynthetic verb morphology. Understanding how phonological and morphological factors interact to produce the attested patterns of stress alternation is a difficult challenge, and I argue that a large database corpus of inflected verb forms is necessary to make progress.

I describe the creation of a corpus containing 3,115 inflected forms of 445 Abkhaz verbs, based on dictionary data from Yanagisawa (2010). All forms are phonologically and morphologically analysed through a combination of automatic computer parsing and hand corrections. I use the corpus to provide an empirical evaluation of a previously proposed theory of Abkhaz stress assignment known as Dybo's Rule (Dybo 1977; Spruit 1986), which relies on a binary division into underlyingly accented and unaccented morphemes. I show how the corpus aids in the identification of the empirical strengths and weaknesses of this theory. Based on generalisations about where Dybo's Rule fails, I propose a revised theory of stress assignment in Abkhaz verbs, which is designed to keep the advantages of Dybo's Rule while remedying its flaws. I show that the revised theory I propose either ties with or outperforms the previous implementation of Dybo's Rule in every area of the verbal lexicon. In total, the revised theory correctly predicts the location of stress in 95% of the forms in the corpus, compared to 90% for the original theory. Eighty-six percent of verbs in the corpus have stress correctly predicted on all of their inflected forms, compared to 77% for the original theory.

Methodologically, I highlight the advantages that corpora and digital tools give phonologists. Much of the time-consuming work of glossing thousands of verb forms was done by computer, albeit with human oversight. Theories of Abkhaz stress assignment were also implemented in the programming language Python 3, allowing for an instantaneous, fully reproducible computer evaluation of the theories against the entirety of the verb corpus. Outsourcing these steps to a computer ensures that the linguist-programmer can devote more of their time to tasks not suitable for a computer, such as analysing patterns in the data and making theoretical revisions. I also argue that the methodology of implementing and re-implementing revisions of theories allows phonologists to make gradual, incremental improvements to their theories, even if the original assumptions are imperfect. The full code and final version of the corpus are made available through GitHub (Andersson 2024a), for researchers who wish to explore the methods and data in greater detail.

Empirically, I provide quantitatively grounded data on verb stress in Abkhaz from over 3,000 verb forms. This allows for the identification of novel generalisations about the data, as well as refinements of previous generalisations based on data collected through fieldwork. I show quantitative evidence that previous proposals about stress in causative verb forms are sound, and provide a new characterization of the stress behaviour of the negative marker, which is sensitive both to phonological factors such as root length, and morphological factors such as the presence of certain person prefixes. Identification of such intricate interactions between phonology and morphology would not be possible without a large data set of this kind.

Theoretically, I improve on previous work using Dybo's Rule, and introduce a new proposal for Abkhaz verb stress. While the stress assignment algorithm from Dybo (1977) remains central, I argue that verbs with several different argument structures make use of a completely different way of assigning stress. These verbs are pre-stressing, with stress always falling immediately before the root. Such stress behaviour is familiar from other languages with movable lexical stress (e.g., Revithiadou 1999 on Greek and Yates 2017 on Cupeño), and from some descriptive generalisations on Abkhaz. In this article, I implement in software an explicit theory which combines Dybo's Rule and pre-stressing behaviour, and show that it accounts for the data better than a theory which only uses Dybo's Rule. Of the verbs which have all of their inflected forms correctly predicted by the revised theory, just under one third crucially rely on one or more of the modifications I propose, and just over 8% rely on pre-stress rather than Dybo's Rule.

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

The remainder of this article is structured as follows: §2 provides relevant language background on Abkhaz, including overviews of the phonology and verbal morphology. §3 describes the creation of a database corpus of Abkhaz verbs. §4 introduces Dybo's Rule, and the methodology used to evaluate this theory of stress assignment against the corpus data. §5 reports on the results of this evaluation. In §6, I identify the strengths and weaknesses of Dybo's Rule, using these to propose a revised theory of Abkhaz stress assignment which is also evaluated against the empirical data in the verb corpus. §7 provides discussion, including of limitations and weaknesses of the present methodology, as well as questions in theoretical phonology that are raised by the Abkhaz data. §8 concludes.

2. Language background

In this section, I introduce relevant background information about Abkhaz, the language which is the empirical focus of this article. Abkhaz (ISO 639–3 [abk]; Eberhard *et al.* 2021) is a Northwest Caucasian language. The three main dialect groups are Abzhywa, Bzyp and Sadz (O'Herin 2021; see Chirikba 2003 for a more detailed classification). The standard Abzhywa dialect is the focus of this article, and is spoken primarily in Abkhazia, on the eastern coast of Black Sea (O'Herin 2021). Recent estimates suggest that Abkhaz has 124,000 native speakers (Dobrushina *et al.* 2021: 29), although Abkhazia uses Russian as a lingua franca, and Abkhaz is increasingly replaced by Russian especially in urban settings (Chirikba 2003: 8).

2.1. Phonology

Table 1 shows the consonant inventory of Abzhywa Abkhaz, which comprises 58 phonemes in total (Jakovlev 2006: 309; Hewitt 2010: 10; O'Herin 2021: 451). The transcription system in this article is broad, and does not reflect the exact phonetic realization of the consonants, for which see Hewitt (1979), Chirikba (2003) and Yanagisawa (2013) on Abzhywa.

The inventory of vowels, shown in (1), is considerably smaller. The native phonology of Abkhaz is typically analysed as having two vowel phonemes, /a ə/ (Hewitt 1979; Colarusso 1988; Chirikba 2003; Yanagisawa 2013; O'Herin 2021). The phonemic status of schwa has been questioned in Abkhaz and related languages (see Kuipers 1960 on Kabardian; Allen 1965 on Abaza, with discussion of Kabardian and Abkhaz; Anderson 1978 on Abaza and Kabardian; and Colarusso 1988 on Northwest Caucasian in general), but since it does not affect the present article, schwa is assumed to be phonemic here (*pace* Andersson 2024b). On the surface, additional vowel qualities beyond [a] and [ə] appear, often conditioned by adjacent glides (Vaux & Pysipa 1997; Arstaa & Č'ķadua 2002: 23–29; Jakovlev 2006: 306; O'Herin 2021: 453–454). In the numerous loanwords of Abkhaz, /i u e o/ also appear freely.

(1) *Vowel inventory, with non-native vowels in parentheses and orthography in angle brackets*

	Front	*Central*	*Back*
High	(i) ⟨и⟩		(u) ⟨y⟩
Mid	(e) ⟨e⟩	ə ⟨ы⟩	(o) ⟨o⟩
Low		a ⟨a⟩	

Abkhaz has contrastive word stress, as can be seen in (2). In Gulia's (1939: 97–106) list of homographs, Hewitt (1979: 263) counts 185 minimal pairs for stress. As (2) and (3) make clear, the

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

Table 1. *Abzhywa Abkhaz consonant inventory, with orthography in angle brackets.*

		Labial	Dental	Other coronal	Velar	Uvular	Pharyngeal
Stop	Voiced	b ⟨б⟩	d dʷ ⟨д дә⟩		g gʲ gʷ ⟨г гь гә⟩		
	Aspirated	pʰ ⟨ԥ⟩	tʰ tʷʰ ⟨т тә⟩		kʰ kʲʰ kʷʰ ⟨қ қь кә⟩		
	Ejective	p’ ⟨п⟩	t’ t’ʷ ⟨т тә⟩		k’ k’ʲ k’ʷ ⟨к кь кә⟩	q’ q’ʲ q’ʷ ⟨ҟ ҟь ҟә⟩	
Fricative	Voiced	v ⟨в⟩	z ⟨з⟩	ʐ ʒ ʒʷ ⟨ж жь жә⟩		ʁ ʁʲ ʁʷ ⟨ғ ғь ғә⟩	
	Voiceless	f ⟨ф⟩	s ⟨с⟩	ʂ ʃ ʃʷ ⟨ш шь шә⟩		χ χʲ χʷ ⟨х хь хә⟩	ħ ħʷ ⟨ҳ ҳә⟩
Affricate	Voiced		d͡z ⟨ҙ⟩	d͡ʐ d͡ʒ d͡ʒʷ ⟨џ џь ҙә⟩			
	Aspirated		t͡sʰ ⟨ц⟩	t͡ʂʰ t͡ʃʰ t͡ʃʷʰ ⟨ҿ ч цә⟩			
	Ejective		t͡s’ ⟨ҵ⟩	t͡ʂ’ t͡ʃ’ t͡ʃ’ʷ ⟨ҿ ч ҵә⟩			
Nasal		m ⟨м⟩	n ⟨н⟩				
Liquid			l ⟨л⟩	r ⟨р⟩			
Glide		(w) ⟨у⟩		j ɥ ⟨и ҩ⟩	w ⟨у⟩		

modern orthography does not indicate word stress in any way, although dictionaries typically provide this information.[1]

(2) *Contrastive word stress* (Gulia 1939: 97, 99)

	Orthography	*Surface form*	*Gloss*	*Translation*
a.	ала	[á-la]	DEF-eye	‘eye’
b.	ала	[a-lá]	DEF-dog	‘dog’
c.	Избылуада?	[jí-z-bəl-wa-da]	REL.ABS-1SG.ERG-burn-DYN-who	‘Who do I burn?’
d.	Избылуада?	[ji-z-bəl-wá-da]	3NH.ABS-REL.ERG-burn-DYN-who	‘Who burns it?’

Although stress is contrastive, there are ubiquitous stress alternations across different forms of the same root. It is such stress alternations that the present article attempts to shed light on.

(3) *Stress alternations within paradigms* (Dybo 1977: 41–42; Hewitt 1979: 272)

		Orthography	*Surface form*	*Gloss*	*Translation*
a.	i.	ала	[a-lá]	DEF-dog	‘dog’
	ii.	алақәа	[a-la-kʷʰá]	DEF-dog-PL	‘dogs’
	iii.	алақәагьы	[a-la-kʷʰa-gʲə́]	DEF-dog-PL-and	‘and dogs’

[1] Glosses in this article follow the Leipzig Glossing Rules, with the following additions: DYN ‘dynamic (eventive)’ (see §2.2), FIN ‘finite’, H ‘human’, NH ‘nonhuman’, PREV ‘preverb’ (see §2.2), PSTABS ‘past absolute (having VERBed)’ (see §2.2), TAME ‘tense–aspect–mood–evidentiality’.

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

b. i.	Ишәфа!	[jə-ʃʷ-fá]	3NH.ABS-2PL.ERG-eat	‘Eat it/them!’
ii.	иафеит	[j-á-fe-jt’]	3NH.ABS-3SG.NH.ERG-eat-DYN.FIN	‘it ate it/them’
iii.	илфеит	[ji-l-fé-jt’]	3NH.ABS-3SG.F.ERG-eat-DYN.FIN	‘she ate it/them’

Previous impressionistic work has highlighted intensity as a main correlate of stress (Aršba 1979: 7). Recent data based on a small-scale acoustic experiment suggest that F0 and duration are also relevant, with the role of F0 showing considerable between-speaker variation (Andersson 2024b).

2.2. Verbal morphology

The morphology of the Abkhaz verb is highly complex. Here, I give only a brief introduction to the small subset of verbal functional morphology that will be relevant for the rest of this article. A fundamental division in the Abkhaz verbal system is between dynamic verbs (denoting events) and stative verbs (denoting states). Relative to statives, dynamic verbs have different morphology, additional possibilities for person prefixes, and a larger tense–aspect system. Since the complex morphology of dynamic verbs poses additional challenges, it is on these verbs that the present study focusses, and the simpler stative verbs will not be analysed here.

Abkhaz has three sets of person markers, which appear as prefixes. These are absolutives (direct objects of transitive verbs and single arguments of intransitive verbs), obliques (such as datives and locational arguments) and ergatives (subjects of transitive verbs). All verbs require at least an absolutive prefix. Some verbs idiosyncratically take person markers that would not be expected from the semantics of the verb. (4) illustrates individual bivalent verbs which select different person prefixes.

(4) *Argument structure of bivalent verbs* (Yanagisawa 2010: 391, 429, 572)

a. исшәеит
ji-s-ʃʷé-jt’
3NH.ABS-ISG.ERG-measure-DYN.FIN
‘I (ergative) measured it (absolutive)’

b. сасит
s-á-si-jt’
ISG.ABS-3SG.NH.OBL-hit-DYN.FIN
‘I (absolutive) hit it (oblique)’

c. исоуит
ji-s-ówi-jt’
3NH.ABS-ISG.OBL-receive-DYN.FIN
‘I (oblique) received it (absolutive)’

Oblique person markers often mark the argument introduced by a preverb. In Abkhaz, preverbs are prefixes which typically introduce an additional argument conveying spatial information about an event. However, preverbs do not always add an argument, and their semantic contributions to a given root are not always transparent. They can be compared to particles in English, which may have transparent spatial meanings (like *out* in *bring out*) or may be semantically opaque (like *up* in *give up*).

(5) *Examples of preverbs*

a. *Semantically transparent: ‘sit down’ + ‘on’ = ‘sit down on’* (Yanagisawa 2010: 396, 237)

i. дтәоит
d-t’ʷó-jt’
3SG.H.ABS-sit.DYN-DYN.FIN
‘I sit down’

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

ii. сақәтәоит
s-á-k^{wh}-t'^{w}o-jt'
ISG.ABS-3SG.NH.OBL-PREV(on)-sit.DYN-DYN.FIN
'I sit down on it'

b. *Semantically opaque: 'take' + 'in' = 'start'*

i. ҳалгоит
ħa-l-gó-jt'
IPL.ABS-3SG.F.ERG-take.DYN-DYN.FIN
'she takes us'

ii. далагоит
d-á-la-go-jt'
3SG.H.ABS-3SG.NH.OBL-PREV(in)-take.DYN-DYN.FIN
'(s)he starts it'

The only other preverbal element discussed in this article is the causative prefix /r-/, which obligatorily occurs with an ergative prefix for the causer. The negative marker /m/ is a suffix in the present tense, but a prefix elsewhere. All tense–aspect–mood–evidentiality marking is suffixal. Present tense forms additionally have a dynamic suffix /-wa/, indicating that the verb is dynamic (eventive). This suffix /wa/ coalesces with root-final /a/: /a-wa/ → [o], as in [d-t͡s^{h}ó-m] '(s)he doesn't go' (from *[d-t͡s^{h}a-wá-m] 3SG.H.ABS-go-DYN-DYN.FIN; Yanagisawa 2010: 494).

All affixes in Abkhaz follow a strict templatic order, shown in simplified form in (6). Not all slots in the template are filled in all verb forms, but when a given morpheme is present, it appears in its templatic slot. (6) gives an example verb form in which most of the slots are filled.

(6) *Verb template* (simplified; see Chirikba 2003: 38–39)

ABS	OBL	PREV	ERG	NEG	CAUS	ROOT	DYN	NEG	TAME
иақәдыршәуеит									
j-	á-	k^{wh}-	də-		r-	ʃ^{w}	-we		-jt'
3NH.ABS-	3SG.NH.OBL-	on	3PL.ERG-		CAUS-	be.on	-DYN		-DYN.FIN

'They throw it/them over it, they cover it with it/them' (Yanagisawa 2010: 236)

One additional verbal category requires comment, since its name does not straightforwardly indicate its semantic function. This is the so-called 'past absolute', which is the name some scholars use for forms which translate English 'having VERBed'. (7) gives an example of how it is used in a sentence. The form даԥхьаны [dáp^{h}χ^{j}anə] 'him/her having read it' is a past absolute form of 'to read'.

(7) *The past absolute* (Yanagisawa 2013: 373)

Ашәҟәы	даԥхьаны	сара	сахь	дааит.
a-ʃ^{w}q'^{w}ə́	d-á-p^{h}χ^{j}a-nə	sará	s-áχ^{j}	d-aá-jt'
DEF-book	3SG.H.ABS-3SG.NH.OBL-read-PSTABS	ISG	ISG-towards	3SG.H.ABS-come-DYN.FIN

'Having read the book, (s)he came to me.'

There is much more that could be said about the verbal morphology of Abkhaz. However, the brief overview on the preceding pages is sufficient for the purposes of this article. Arguments are marked through a series of prefixes, while tense and related functions are marked through suffixes. There are preverbs and causative forms, which can change the number of arguments a verb takes. All affixes are arranged linearly according to a template. With this background, we are ready to turn to the corpus used to investigate Abkhaz verb stress in this article.

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

3. A database corpus of Abkhaz verbs

There are several challenges that face linguists who want to understand stress assignment in Abkhaz verbs. It is not possible to pick out a small set of representative forms to get an overview of the system. Each verb has hundreds or thousands of possible inflected forms, and verbs can be divided into a wide array of categories based on the presence or absence of certain person prefixes, of a preverb, or of a causative marker. These problems are compounded by phonological factors such as root length, and by the fact that different roots may belong to different accentual categories.

I believe that a large corpus of inflected Abkhaz verbs is necessary to overcome these challenges. It is not possible to study the interaction of all of these phonological and morphological factors without a large number of data. In this section, I describe the creation of a database corpus of Abkhaz verbs from a dictionary of the language (Yanagisawa 2010). In §3.1, I discuss the data, focussing on choices of which forms to include and exclude. In §3.2, I then describe how these data were extracted and analysed, paying special attention to the division of labour between automatic parsing by computer and hand correction of the data. The final result is a phonologically and morphologically parsed corpus of Abkhaz verbs, including seven forms each of 445 verbs, for a total of 3,115 word-forms. The corpus and the code used to create it are available in a GitHub repository (Andersson 2024a).

3.1. The data

The corpus is based entirely on data from Yanagisawa's (2010) dictionary of Abkhaz. The data come from previous published sources as well as Yanagisawa's original fieldwork, primarily with Anna Tsvinaria-Abramishvili, a native speaker of Abzhywa Abkhaz from Kutol (Кәтол, [k'^{w}t'ól]) in the Ochamchira region of Abkhazia. In particular, Tsvinaria-Abramishvili was the source of all stress judgements in the dictionary, in which primary stress is marked on orthographic forms of every word (Yanagisawa 2010: 5).

This dictionary includes a wide variety of inflected forms of each verb, all with stress marked, allowing linguists to study stress alternations across many different forms of the same lemma. These forms were elicited over the course of seven years of fieldwork (Yanagisawa 2010: 5), which is why the verbal morphology sections are more extensive than in other dictionaries. Verb forms are segmented into morphemes, and dictionary entries include useful grammatical information about verb morphology. Since a single speaker provided all the stress judgements, there is less of a concern about dialect variability and different forms representing the grammars of different speakers. Finally, Yanagisawa kindly provided me with a digital PDF copy of the dictionary, so that there was no need for scanning a printed dictionary in order to extract digital text for computer analysis. Here is a sample entry from the dictionary, showing the citation form, grammatical information, inflected forms in various tenses and translated example sentences (Yanagisawa 2010, s.v. *аԥпра*):

> **а-рт-ра́** [tr.] [C1-a-C3-S / C1-a-C3-Neg-S [C3 open C1] (**Fin.** [pres.] и-а-сы-рт-уе́-ит / и-а-сы-рт-уа́-м (-рт-зо́-м), [aor.] и-а-сы-рт-и́т / и-а-с-мы-рт-и́т (рт-зе́-ит), [imper.] и-а-рты́! / и-а-б-мы-рты́-н!, и-а-шәы-рты́! / и-а-шә-мы-рты́-н!; **Non-fin.** [pres.] (C1) и-а-лы-рт-уа́ / и-а-л-мы-рт-уа́, (C3) и-а-зы-рт-уа́ / и-а-з-мы-рт-уа́, [aor.] (C1) и-а-лы-рты́ / и-а-л-мы-рты́, (C3) и-а-зы-рты́ / и-а-з-мы-рты́, [impf.] (C1) и-а-лы-рт-уа́-з / и-а-л-мы-рт-уа́-з, (C3) и-а-зы-рт-уа́-з / и-а-з-мы-рт-уа́-з, [past indef.] (C1) и-а-лы-рты́-з / и-а-л-мы-рты́-з, (C3) и-а-зы-рты́-з / и-а-з-мы-рты́-з; **Abs.** и-а-рт-ны́ / и-а-мы-рты́-кәа) 1. to open: А-ҧе́нцьыр а-ды-рт-уе́-ит. *They are opening the window.* А-ҧе́нцьыр а-рт-ра́ ҟало́-у? *Is it possible to open the window? Можно открыть окно?* А-ҧе́нцьыр а-сы-рты́-р ҟало́-ма? *May I open the window? Мне можно открыть окно?* Апҳсуаа р-еиҳабы́ ды́-ҟа-м-кәан а́изара с-з-а́а-рт-уа-м, и-а́арласны у-ца-ны́ д-аа-у-га́-р-о-уп! (AF) *I cannot open the assembly without the leader of the Abkahzians being present.* [cf. **а-т-ра́** 'to open']

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

I decided to extract seven forms of each verb. Inspecting the dictionary showed that these seven forms were consistently reported for a large number of verbs, and that they often exhibited stress alternations. All forms come from so-called dynamic verbs, which denote events (as opposed to stative verbs). Dynamic verbs have greater morphological complexity than stative verbs, and therefore provide a greater challenge for phonological theories of stress assignment.

Below are glossed examples of the seven categories of verb forms I will study, labelled with the slots of the verbal template that they fill. Note the prefixal person markers and the suffixal tense–aspect–mood–evidentiality morphology throughout. The infinitive behaves rather more like a noun than like a verb, and does not follow the verb template; instead it is marked by the definite nominal prefix /a-/ and the verbal noun suffix /-ra/. The forms are from the verb 'to disappear', with the preverb /k'əla-/ 'through' and the root /d͡z/ 'disappear'.

(8) *Glossed examples* (Yanagisawa 2010: 213)

a. *Infinitive (verbal noun)*
аКылаʒра
a-k'ə́la-d͡z-ra
DEF-through-disappear-INF
'to disappear'

b. *Present (affirmative)*
дКылаʒуеит
ABS-PREV-ROOT-DYN-TAME
d-k'ə́la-d͡z-we-jt'
3SG.H.ABS-through-disappear-DYN-DYN.FIN
'(S)he disappears'

c. *Present (negative)*
дКылаʒуам
ABS-PREV-ROOT-DYN-NEG
d-k'ə́la-d͡z-wa-m
3SG.H.ABS-through-disappear-DYN-NEG
'(S)he does not disappear'

d. *Imperative (affirmative)*
бКылаʒ
ABS-PREV-ROOT
b-k'ə́la-d͡z
2SG.F.ABS-through-disappear
'Disappear!'

e. *Imperative (negative)*
бКыламʒын
ABS-PREV-NEG-ROOT-TAME
b-k'ə́la-m-d͡zə-n
2SG.F.ABS-through-NEG-disappear-NEG.IMP
'Don't disappear!'

f. *Past absolute (affirmative)*
дКылаʒны
ABS-PREV-ROOT-TAME
d-k'ə́la-d͡z-nə
3SG.H.ABS-through-disappear-PSTABS
'him/her having disappeared'

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

Table 2. *Morphological verb categories found in the corpus.*

Person markers	Causative	Preverb	Example	Translation	Verbs in corpus
Absolutive	—	Absent	[ats͡hará]	‘to go’	72
		Present	[ak’áɦara]	‘to fall’	24
Absolutive–oblique	—	Absent	[áχara]	‘to pull’	12
		Present	[apʰə́lara]	‘to meet’	72
Absolutive–ergative	Absent	Absent	[azɻ́rá]	‘to dig’	59
	Present		[argə́lara]	‘to build’	64
	Absent	Present	[agʷátʰara]	‘to notice’	54
	Present		[atʰart͡ʃʷʰrá]	‘to make empty’	14
Absolutive–oblique–ergative	Absent	Absent	—	—	0
	Present		[arbará]	‘to show’	7
	Absent	Present	[ákʷʰtʷʰara]	‘to pour on’	57
	Present		[aχart’ʷará]	‘to seat at’	10
Total					445

g. *Past absolute (negative)*
дкыламʒкәа
ABS-PREV-NEG-ROOT-TAME
d-k’ə́la-m-d͡z-k’ʷa
3SG.H.ABS-through-NEG-disappear-NEG.PSTABS
‘him/her not having disappeared’

The verbs included have all attested common combinations of person prefixes in Abkhaz. All verbs have at least an absolutive marker. Some verbs also have an oblique marker, an ergative marker or both. These verb types can appear with and without preverbs, and verbs with an ergative marker may also have a causative prefix. In Table 2 in §3.2, I summarise how many verbs belong to each of these categories in the final version of the corpus.

3.2. The creation of the corpus

In order to create this corpus, I copied the text of the PDF version of the dictionary into a plain text document, which I then modified using scripts in the Python 3 programming language. The purpose of this subsection is not to describe every line of code in this process, but to give a brief overview of how the corpus was created, including a discussion of the division of labour between human and computer analysis. The code and final corpus are publicly available in a GitHub repository (Andersson 2024a). I’m grateful to Tamio Yanagisawa for allowing me to make the final corpus derived from his work publicly accessible.

The extraction of the data was carried out by a script which searches for all entries whose headwords look like verbs morphologically and are followed by transitivity information, a property unique to verbs. The headword of a verb is the infinitive, marked by a prefix /a-/ and a suffix /-ra/. Irrelevant information such as page numbers and line breaks was removed by this script. The seven verb forms of interest were extracted with the help of the dictionary formatting: present tense forms are those which appear immediately after a ‘[pres.]’ tag, and so on. Also included was information given in the dictionary about how the verb is conjugated: whether it takes a preverb, and which person prefixes it requires. Subsequent scripts remove extraneous information such as punctuation which has incorrectly been identified as part

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

of an Abkhaz verb form, and convert the text encoding from the dictionary, which uses a proprietary font, into modern Abkhaz orthography in Unicode encoding.

In addition to the phonemic orthography of the language, basic phonological information about each entry was automatically added to the corpus. All consonants were replaced by C, and all glides and high vowels were replaced by G, while other vowels were replaced as follows: А Е О Ы /a e o ə/ → A E O Y. Although Abkhaz orthography is generally phonemic, the letter ⟨И⟩ can represent both /j/ and /i/, and ⟨У⟩ both /w/ and /u/. Their distribution is very often predictable, but no attempt was made to encode the phonological generalisations that separate glides from high vowels (see discussion in §3.1). Vowels are uppercase if stressed, and lowercase otherwise.

After this, an attempt was made to supply a simplified automatic gloss for each form. This was only possible because the forms in the dictionary include morpheme boundaries, and because Abkhaz verbal morphemes are ordered strictly according to a template (as described in §2.2). For example, if one knows that a particular verb takes only an absolutive person marker and lacks a preverb, then its imperative form will consist of an absolutive marker followed by the verb root. Based on knowledge about the verbal template of Abkhaz, I supplied the scripts with information about how each verb type is conjugated; that is, which morphemes occur in which order in each of the verb's seven forms. This allowed for an automatic alignment of the morpheme-separated orthographic form in the dictionary with the gloss. Note that these glosses are simplified, in the sense that they only specify which templatic slots are filled; for example, the form [b-t͡sʰá] 'go!' was glossed as 'ABSOLUTIVE-ROOT' rather than '2SG.F.ABS-go'.

Errors in the automatic glossing procedure can arise in several ways. Some forms have missing morpheme boundaries in the dictionary; others are problematic because the hyphens separating forms broken across multiple lines are identical to the hyphens that separate morphemes within a form; and still others contain segmentally null morphemes which still affect stress. This latter problem is common with certain preverbs, where a singular third-person oblique argument is not expressed segmentally, but stress falls as if it were. For example, in (9a) the oblique person prefix 'them' causes stress to fall immediately before the root 'cross'. In (9b), the corresponding singular oblique prefix 'it' causes stress to fall in the same place, even though this prefix is segmentally null. These forms all have incorrect or misleading glosses, since there is a discrepancy between the expected number of morphemes and the number of morphemes actually found in the surface form.

(9) *Null oblique arguments* (Spruit 1986: 62)

a. Дрыруама?
d-rə́-r-wa-ma
3SG.H.ABS-3PL.OBL-CROSS-DYN-Q
'Does (s)he cross them?'

b. Дыруама?
də́-∅-r-wa-ma
3SG.H.ABS-3SG.NH.OBL-CROSS-DYN-Q
'Does (s)he cross it?'

Some attempts were made to fix morpheme boundary errors automatically, but many verbs would require a hand-supplied gloss for all seven forms to ensure that there are no mistakes. In the future, it may be worth taking the time to gloss such cases by hand, but for the present version of the corpus, they were instead excluded.

Manual inspection of the data was necessary, however, to identify cases where morphemes appeared in an unexpected position. According to the verbal template in §2.2, the negative prefix is expected to appear between a preverb and its verb root: PREV-NEG-ROOT. In some forms, however, we see NEG-PREV-ROOT instead. Such forms may have been reanalysed by speakers as not having a preverb at all (PREV-ROOT > ROOT), which would explain why the 'preverb' is never separated from its root.

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

These forms were excluded, since it is not clear that the morphological analysis given in the dictionary corresponds to the structure speakers assign it synchronically.

I decided not to analyse verbs containing non-native vowels, such as /e/ or /o/. I excluded roots and preverbs containing the grapheme ⟨И⟩, which can represent either native /j/ or non-native /i/, and ⟨У⟩, which can represent either native /w/ or non-native /u/. This is a coarse-grained way of removing loanwords, since some words with native glides are excluded. However, glides and high vowels are not orthographically distinct, so separating /j/ from /i/ and /w/ from /u/ would require manual analysis of each verb. Including such forms without hand-correction would also be problematic for stress assignment, where vowels and consonants may behave very differently. Only roots and preverbs with /j i w u/ were excluded. The affixes with the relevant graphemes in the corpus are all native, and contain the underlying glides /j w/ only. For this reason, they were left in the corpus.

In forms with a long or geminate [aa], stress can fall on either [a], as is typically reflected in dictionaries (Genko 1998; Kaslandzija 2005). Since his consultant lacked this [áa]–[aá] contrast, Yanagisawa (2010: 14) assigned stress to either the first or second [a] based on etymological factors and other dictionaries. I decided to exclude any forms with stress on /aa/, to ensure that all stress data in the corpus came from the same source.

Also excluded are forms which, presumably because of typographical errors, did not have stress marked in the dictionary at all, as well as forms with multiple stresses. The dictionary uses multiple stress marks as a notational convention to indicate that stress is judged to be possible on either marked syllable (Yanagisawa 2010: 14). When considering forms like [jirə́bʒarts͡'ojt']~[jirəbʒárts͡'ojt'] 'they conclude it' as well as the related past absolute [jirəbʒats͡'anə́] 'them having concluded it' without stress variation, it is not clear which variants belong together. When evaluating theories of Abkhaz stress assignment, should a theory be penalised for predicting only one of the present tense forms? Does it matter which present tense form is predicted by the theory? Instead of making arbitrary decisions about data points like these, I decided to exclude them. In cases of optionality concerning factors other than stress, such as vowels which can optionally be present or absent, I arbitrarily included the first variant given.

Before any processing of the verbal forms, there were 951 verbs extracted from the dictionary. Approximately, 100 were excluded for having stress on /aa/. The process of segmenting verbs morphologically removed 200, while another 200 were excluded for containing non-native vowels, segmentally null morphemes, and the like. This led to the 445 verbs present in the final version of the corpus.

Two additional data points for each verb were entered by hand. First, a translation into English was included. Second, the dictionary does not include a morpheme boundary between the causative prefix and the following root, with the rationale that they are always immediately adjacent to each other. This meant that automatic identification of causative verb forms was impossible, and this information was instead entered manually. Causative and non-causative verb forms have previously been described as having different stress patterns (Dybo 1977: 43; Spruit 1986: 71), so this information cannot be ignored in a study about stress. This is an example of a situation where the benefits of hand-annotating forms outweighs the time cost, and where excluding data would not be suitable.

After this combination of automatic and manual processing, the corpus contains seven forms each of 445 Abkhaz verbs, for a total of 3,115 forms. All have orthographic and phonological information in addition to a gloss. The lemmas are translated, and include morphological information about preverbs, causatives, and person markers. The corpus is stored in tab-separated plain text format (UTF-8 encoding), to ensure that it can be read on any device and can be easily manipulated by a wide variety of programs. A partial entry for the verb 'to feel' is shown in (10). The first three columns are the infinitive form [ak'rá] 'to feel', written a-ĸ-pA (column 1) with morpheme boundaries and a capital A to indicate stress on the final vowel); the phonological transcription a-C-CA (column 2); and the simplified gloss DEF-ROOT-INF (column 3) indicating that the first morpheme is the definite prefix, the second the root, the third the infinitive suffix. The full entry includes six other such triplets of forms for the remaining six verb forms analysed (not shown). At the end of the entry is grammatical information about the verb: it has absolutive and oblique person markers (column 4; there is no preverb); it means 'to feel' (column 5); and it lacks a causative (column 6).

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

(10) *Partial verb entry from the corpus*

a-к-pA a-C-CA DEF-ROOT-INF … Absolutive, oblique 'to feel' No causative

The breakdown of verbs into morphological categories is shown in Table 2.

How can we begin to understand the patterns of stress assignment across over 3,000 forms? This question is addressed in the next section, where I discuss a methodology for using this corpus to evaluate and improve on previous theories of Abkhaz verb stress.

4. Methodology: using corpora to evaluate phonological theories

As mentioned at the start of §3, it is difficult to get an overview of Abkhaz verb stress by considering a small, carefully selected set of representative forms. The morphological and phonological complexity of the data make such an approach less useful for Abkhaz than it would be for phonological processes sensitive to fewer morphophonological factors. In this article, I instead begin with a hypothesis about Abkhaz stress assignment, and evaluate it empirically against the data in the corpus. This initial hypothesis may be incorrect; in fact, it is expected to be incomplete at the very least. However, by evaluating even an imperfect hypothesis, we can identify where its empirical problems lie, looking for patterns in where the theory makes incorrect predictions. Equipped with this knowledge, we can then modify the initial hypothesis to attempt to remedy its flaws. By evaluating the new theory against the corpus again, we receive quantitative feedback on how much we have improved. In this way not only do we use corpora to evaluate hypotheses from previous work, but we also have a mechanism for how to improve on those hypotheses and come up with new theories with increased empirical coverage.

In §4.1, I discuss the previous proposal for Abkhaz stress assignment which I take as a starting point here. This is an algorithm for stress assignment known as Dybo's Rule, originally proposed by Dybo (1977), and later developed by others, especially Spruit (1986). In §4.2, I then discuss the implementation of the methodology outlined above in greater detail. I used Python to implement Dybo's Rule computationally, allowing a computer to evaluate the empirical merits of this stress assignment algorithm against all 3,115 forms in the corpus. However, the examination of patterns and generalisations in the data, and the proposed amendments to Dybo's Rule, are left to linguists. In this way, the methodology I employ takes advantage both of what computers can do best – extremely quick implementation of stress assignment algorithms – and of what human linguists can do best – identifying patterns in phonological data, and suggesting new theories to account for these.

4.1. Dybo's Rule: a proposal for Abkhaz stress assignment

In previous literature on Abkhaz, stress has been assumed to work according to Dybo's Rule (Dybo 1977). In Dybo's original proposal, morphemes are divided into one of two classes: accented and unaccented (originally called dominant and recessive, respectively). Dybo's Rule states that primary stress falls on the leftmost accented morpheme that is not immediately followed by an accented morpheme. For illustrative purposes, in this section, I import wholesale (Dybo's (1977) assumptions about which morphemes are accented. Details about the exact assumptions used when evaluating Dybo's Rule against the corpus of Abkhaz verbs are discussed in §4.2.

Dybo's Rule is quoted in its original formulation in (11a), and paraphrased in the equivalent formulation (11b).

(11) *Dybo's Rule*

a. '[T]he stress is always placed at the end of the first sequence of morphemes of the highest valence' [i.e., accented morphemes] (Dybo 1977: 43).[2]

b. Primary stress falls on the leftmost accented morpheme not immediately followed by an accented morpheme.

[2]The translation is mine; the original Russian reads 'ударение всегда ставится в конце первой последовательности морфем высшей валентности.'

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

By Dybo's Rule, if a word has a single accented morpheme, that morpheme receives stress. Underlying accents are shown with a plus sign (+), while the lack of an accent is shown with a minus sign (−). In (12)–(14), morphemes are separated by a hyphen as usual. Between a prefix and a root, the hyphen is placed with the prefix, and between a root and a suffix, the hyphen is placed with the suffix.

(12) *One underlying accent* (Dybo 1977: 42)

сыла	
/sə-	la/
+	−
[sə́-	la]
ISG.POSS-	eye

'my eye'

If there are multiple immediately adjacent accents, stress falls on the rightmost such accent, as in (13). Dybo's Rule ensures that stress does not fall further to the left, since accents which are immediately followed by other accents do not receive stress.

(13) *Multiple adjacent accents* (Dybo 1977: 43)

аҵара		
/a-	t͡sʰa	-ra/
+	+	+
[a-	t͡sʰa	-rá]
DEF-	go	-INF

'to go'

If there are multiple non-adjacent accents, stress is on the leftmost of them, as in (13). This is the case even if there is a longer sequence of accents later in the word.

(14) *Multiple non-adjacent accents*

a. аҵара

/a-	pʰa	-ra/
+	−	+
[á-	pʰa	-ra]
DEF-	jump	-INF

'to jump'

b. сысасцәагьы

/sə-	sas	-t͡ʃʷʰa	-gʲə/
+	−	+	+
[sə́-	sas	-t͡ʃʷʰa	-gʲə]
ISG.POSS-	guest	-H.PL	-and

'and my guests'

Subsequent phonological work on Abkhaz stress has expanded on Dybo's (1977) original proposal in several ways (see especially Spruit 1986). The formulation in (11), for example, says nothing about what happens when there are no underlying accents. Others have also found it necessary to assign longer morphemes more than one underlying accent, perhaps one per syllable (Trigo 1992), one per mora (Kathman 1992; Vaux & Samuels 2018), or one per segment (Andersson 2024b). In the next subsection, I discuss the precise implementation of Dybo's Rule that this article uses, which is based heavily on Spruit (1986).

4.2. Methodology: evaluating Dybo's Rule empirically

Recall from the start of this section that I aim to implement a version of Dybo's Rule computationally, so that a computer program can automatically evaluate this stress assignment algorithm against all 3,115 forms in the corpus of Abkhaz verbs. In order to do this, it is necessary to state Dybo's Rule and any

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

surrounding assumptions in precise enough detail that a computer can implement it unambiguously. Below I describe all relevant assumptions, drawing primarily on Spruit's (1986) version of the theory, and discuss how the corpus evaluation was carried out. I repeat the point from the beginning of §4 that these assumptions are not necessarily correct. It is necessary to start with a theory, which may be more or less promising. Once it is evaluated empirically against the corpus data, it will be possible to make improvements based on where the original assumptions proved inadequate.

The first issue to address is that the formulation of Dybo's Rule from §4.1 does not state where stress falls if there is no underlying accent. In such cases, the location of stress appears to depend on morphological factors (Spruit 1986: 38), but in both Dybo's (1977) and Spruit's (1986) data, stress generally falls at the end of the root. For this reason, the precise implementation of Dybo's Rule in this article is as follows:

(15) *Dybo's Rule (implementation in this article)*

a. If there is at least one underlying accent:
   Primary stress falls on the leftmost accent not immediately followed by another accent;

b. Else:
   Primary stress is root-final.

Trigo (1992: 229) discusses cases in nouns and adjectives where stress alternates between word-final and root-final position, such as [lá-k']~[la-k'ə́] ('eye-INDF') 'an eye'. However, verbs do not show similar variation, and stress is always root-final rather than word-final: [d-sasí-jt'], *[d-sasi-jt'ə́] ('3SG.H.ABS-guest-DYN.FIN') '(s)he was a guest' (Yanagisawa 2010: 494). Since this article is concerned only with accentuation in verbs, I retain the root-final elsewhere clause in (15b).

I wrote a program in Python 3 to implement this stress assignment algorithm. Several other implementational possibilities exist, and Meurer (2011) presents a finite-state transducer for Abkhaz stress assignment. The present article relies instead on the Python programming language, although it would be interesting in future work to explore finite-state implementations of the theories I discuss.

Of course, in order to apply a stress assignment algorithm, it is necessary to specify the underlying accent status (accented or unaccented) of the morphemes that appear in the corpus. Here, I rely mainly on Spruit (1986), who gives the presumed accent status of many functional morphemes in Abkhaz based on the stress behaviour of these morphemes in his fieldwork data. The complete set of assumptions that the program uses is in (16).

(16) *Accent status of functional morphemes in verbs*

a. *Accented*

i. All oblique person markers:
/s-/ 1SG, /w-/ 2SG.M, /b-/ 2SG.F,
/j-/ 3SG.M, /l-/ 3SG.F, /a-/ 3SG.NH,
/ħ-/ 1PL, /ʃʷ-/ 2PL, /r-/ 3PL

ii. One ergative person marker:
/(n)a-/ 3SG.NH

iii. The dynamic (eventive) suffix: /-wa/

iv. The past absolute suffix: /-nə/

v. The definite prefix: /a-/

vi. The infinitive suffix: /-ra/

b. *Unaccented*

i. All absolutive person markers:
/s-/ 1SG, /w-/ 2SG.M, /b-/ 2SG.F,
/d-/ 3SG.H, /j-/ 3SG.NH, /ħ-/ 1PL,
/ʃʷ-/ 2PL, /j-/ 3PL

ii. Most ergative person markers:
/s-/ 1SG, /w-/ 2SG.M, /b-/ 2SG.F,
/j-/ 3SG.M, /l-/ 3SG.F, /ħ-/ 1PL,
/ʃʷ-/ 2PL, /r-/ 3PL

iii. The negative marker (prefix and suffix): /m/

iv. The dynamic finite suffix: /-jt'/

v. The negative imperative suffix: /-n/[3]

vi. The negative past absolute suffix:
/-k'ʷa/

[3]As far as I can tell, Spruit (1986) does not discuss the accent status of this morpheme, but since it never bears stress in any surface form, I have treated it as unaccented here.

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

Spruit's (1986) analysis also involves a number of affixes which have segmentally identical but accentually different allomorphs in certain environments. Implementing the exact proposal in Spruit (1986) was not possible for a number of reasons. In some cases, the proposals are too narrow, discussing the effects that individual verb roots have on other morphemes (Spruit 1986: 72). In other cases, Spruit notices and exemplifies an exception to Dybo's Rule, but does not identify the context precisely enough to allow a computer to interpret it unambiguously. In §6, I return to the question of which of the above assumptions need to be modified to better account for Abkhaz verb stress, including some of the patterns of allomorphy proposed by Spruit (1986).

The morphemes in (16) above are all relatively short, but Spruit proposes that longer morphemes have more than one accent specification. Spruit (1986: 37) calls the units which carry accent 'elements', where a /CV(V)/ sequence constitutes an element, as does a single /C/ if there is no following vowel, and a single /V/ if there is no preceding consonant. There is general agreement that something like these elements are relevant for stress assignment, although others have attempted to identify them with syllables (Trigo 1992), moras (Kathman 1992; Vaux & Samuels 2018) or segments (Andersson 2024b). Single consonants often alternate between having and not having a following schwa, [C]~[Cə], which may perhaps explain why they seem to pattern as elements together with non-alternating /CV/ sequences such as [Ca]. Thus, a root like /ʁəz(ə)/ 'groan' has the two elements /ʁə/ and /z(ə)/, each of which can be accented or unaccented. This system is intended to explain the fact that longer roots appear to have additional stress patterns relative to shorter roots (Spruit 1986: ch. 2). Regardless of whether they should be identified with syllables, moras or something else, I implement Spruit's (1986) elements in the evaluation of Dybo's Rule, letting longer roots and preverbs contain more than one accent.

There is, understandably, no source which attempts to list the accent status of all roots and preverbs in Abkhaz. But in order for a computer to evaluate Dybo's Rule against the corpus, it needs to have precisely this information. Here, the program evaluates all possible options by trial and error, and uses whichever set of accent specifications works best empirically on all seven forms of each verb. In other words, the program answers the question *Is there any possible underlying accent specification for this verb such that Dybo's Rule yields the pattern of stress alternation observed in the corpus?* In the verb aʯkʲapa [atʰqʼʲará] 'to jump out' (Yanagisawa 2010: 414), there is a preverb /tʰ(ə)/ which can be accented or unaccented, and a root /qʼʲa/ which can be accented or unaccented. This gives $2 \times 2 = 4$ total options for the accent specifications of this verb. The program evaluates all four possibilities against the seven forms of this verb in the corpus, and uses the accent specifications that lead to the highest number of forms being correctly predicted. The causative prefix /r-/, conspicuously absent in (16), is segmented as part of the root in the dictionary, and the current implementation of the Python program treats it as such, potentially allowing its accent specification to vary across different roots. In §6, I return to the stress behaviour of the causative prefix, and modify this assumption.

With all of these assumptions specified in detail, we can run the Python 3 program on the corpus data to evaluate the particular implementation of Dybo's Rule I have outlined above. For each of the 445 verbs in the corpus, it applies Dybo's Rule to each of its seven forms (choosing the accent specification of the root and any preverb by trial and error), and asks whether the stress predicted by Dybo's Rule matches the stress recorded in the corpus. It outputs each verb's citation form and translation, accompanied by a list of seven numbers, one for each of the verb's seven forms. A 1 indicates that the predicted form matches the data in the corpus, while a 0 marks an incorrect prediction. A verb where all seven forms are correctly predicted therefore comes back as [1, 1, 1, 1, 1, 1, 1], while a verb where only the second and third forms (the affirmative and negative present tense, respectively) are correctly predicted displays as [0, 1, 1, 0, 0, 0, 0]. The program also outputs the accent specifications for the root and any preverb that it used to derive the reported results. As a summary, the program also outputs the total number of verb forms correctly predicted across the corpus (out of 3,115), as well as the total number of verbs for which all seven forms were correctly predicted (out of 445). In §5, I discuss the results of running this program on the full corpus of Abkhaz verbs, and attempt to identify patterns and generalisations in where Dybo's Rule makes incorrect predictions. These data are then used to inform a new proposal for stress assignment in Abkhaz verbs in §6.

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

5. Results

In this section, I report the results of the computer evaluation of Dybo's Rule against the corpus of Abkhaz verbs. The numbers are promising, with the theory accounting for 90% of total verb forms. Seventy-seven percent of verbs have the location of stress on all seven of their inflected forms correctly predicted. Below, I discuss these results in more detail, by breaking up how well the theory performs on certain subsets of the data: on particular argument structures, on verbs with preverbs *vs.* those without, on causative *vs.* non-causative verbs, and so on. Fortunately for linguists who want to understand Abkhaz verb stress, its errors are not randomly distributed. Instead they form patterns, suggesting that a revised theory that accounts for these patterns may be more successful. In §6, I examine the incorrect predictions more closely, in an attempt to use the patterns in the data to arrive at a theory of Abkhaz stress assignment with even greater empirical coverage.

(17) summarises the results of applying Dybo's Rule to all forms in the verb corpus, showing the overall proportion of forms correctly predicted in (17a) and the proportion of verbs whose stress was correctly predicted on all seven of their forms in (17b):

(17) *Results for Dybo's Rule*

a. Correct / total (forms): 2,816 / 3,115 = 90.40%

b. Correct / total (verbs): 343 / 445 = 77.08%

These numbers and percentages are difficult to interpret without any reference point or baseline. Considering the fact that approximately two-thirds of English words have initial stress (Cutler 2005: 271), perhaps a theory with a system of abstract underlying accents and a stress assignment rule with an elsewhere clause, is a dramatic overcomplication if it only accounts for 77% of the verbal lexicon. This is not the case. Simpler alternative stress assignment algorithms such as 'assign root-initial stress' account for only 12.42% of the total forms, and 5.84% of the verbs. Predicting 90% of the total forms, and 77% of the verbs, is therefore an impressive result. Table 3 breaks down the results for Dybo's Rule in (17) by morphological factors. The table illustrates the fact that some verbal categories are predicted nearly perfectly, while others are close to 0%.

Table 3. *Breakdown of results by verb category for the initial implementation of Dybo's Rule.*

			Forms			Verbs		
Person markers	Causative	Preverb	Correct	Total	% correct	Correct	Total	% correct
Absolutive	—	Absent	489	504	97.02	68	72	94.44
		Present	147	168	87.50	18	24	75.00
Absolutive–oblique	—	Absent	61	84	72.62	1	12	8.33
		Present	495	504	98.21	65	72	90.28
Absolutive–ergative	Absent	Absent	407	413	98.55	56	59	94.92
	Present		321	448	71.65	38	64	59.38
	Absent	Present	340	378	89.95	36	54	66.67
	Present		82	98	83.67	3	14	21.43
Absolutive–oblique–ergative	Absent	Absent	—	—	—	—	—	—
	Present		20	49	40.82	0	7	0.00
	Absent	Present	392	399	98.25	50	57	87.72
	Present		62	70	88.57	8	10	80.00
Total			2816	3115	90.40	343	445	77.08

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

By examining the statistics in a table like Table 3, it is possible to identify which types of verbs should be investigated further. For example, verbs with an absolutive and oblique person marker have a high proportion of verbs correctly accounted for when there is a preverb (65/72 = 90.28%), but a much lower proportion when there is no preverb (1/12 = 8.33%). In its current implementation, Dybo's Rule also appears to perform poorly on many causative verbs, and the proportion of causative verbs correctly predicted is always lower than the proportion of non-causative verbs with the same person prefixes and preverb status. In the next section, I examine the verb forms behind the statistics in Table 3, and attempt to use the patterns that emerge from the data to create a revised theory of stress assignment in Abkhaz verbs.

6. Revising the theory

In §5, I explained that identifying patterns in where the current implementation of Dybo's Rule fails may be helpful in finding improvements and revisions to the theory. In this section, I take up this task. The current implementation has several moving parts, notably the algorithm for stress assignment (Dybo's Rule as defined in (15)) and the assumptions about which morphemes are accented and unaccented (see §4.2). I propose revisions to both, with the aim of improving the empirical coverage of the theory. The goal of modifying Dybo's Rule in light of exceptions is not novel, and the work in this section builds in part on previous analyses (Kathman 1992; Trigo 1992; Bruening 1995; Vaux & Samuels 2018; Andersson 2024b; among others). I compare the revisions I propose to previous theoretical work on Abkhaz where relevant.

In §6.1, I discuss the revisions that I propose, which are summarised in (21). I modify the assumptions about accent status by arguing that the unaccented negative prefix has an accented allomorph for certain absolutive–oblique verbs. I modify the stress assignment algorithm by proposing that there is a class of pre-stressing verbs, in which stress falls immediately before the root instead of being assigned by Dybo's Rule. This is a more radical departure from previous literature on Abkhaz stress, but pre-stressing verbs are familiar from other languages with contrastive but movable stress (Revithiadou 1999 on Greek and Yates 2017 on Cupeño). I also adopt several changes to the accent status of morphemes in causative verb forms which are familiar from earlier work on Abkhaz.

Finally, in §6.2, I implement all of the proposed revisions in the Python 3 programming language, and evaluate the revised theory empirically against the 3,115 forms in the verb corpus. I show that the revisions improve empirical coverage across the verbal lexicon. The revised theory correctly predicts stress for 95% of the verb forms in the corpus, and 86% of verbs have stress correctly predicted on all seven forms. The performance is either the same as or better than that of the implementation of Dybo's Rule in §4 for all categories of verbs. These results suggest that even when dealing with complex and multidimensional data such as those found in Abkhaz verbs, the use of corpora and a combination of computerised and human-guided analysis can be useful methodological tools in phonological research.

6.1. Motivating the revisions

In this section, I propose several revisions to the implementation of Dybo's Rule presented in §4. The reasoning below follows the methodology outlined in §4, in which the patterns of correct and incorrect predictions in the corpus evaluation from §5 are used to guide the analytical work and steer it in the right direction. I begin by discussing the behaviour of the negative prefix in absolutive–oblique verbs, which I cover in some detail. For other changes to the theory, including the existence of pre-stressing verbs, I give a shorter summary of some relevant data.

Recall from §5 that when there is no preverb, almost none of the verbs with absolutive and oblique person markers have stress correctly predicted on all seven forms. As shown in Table 4, investigating these verbs in detail reveals a pattern consistent across verbs: all forms are correctly predicted (✓) by the implementation of Dybo's Rule from §4, except for the negative imperative and negative past

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

Table 4. *Results for absolutive–oblique verbs (AFF. = affirmative, NEG. = negative).*

		Infinitive	Present tense		Imperative		Past absolute	
Verb	Translation		AFF.	NEG.	AFF.	NEG.	AFF.	NEG.
[anə́rra]	'to influence'	✓	✓	✓	✓	✗	✓	✗
[áp^{h}χ^{j}ara]	'to read'	✓	✓	✓	✓	✗	✓	✗
[afəqrá]	'to smell'	✓	✓	✓	✓	✗	✓	✗

absolute (✗). The fact that all verbs of this type pattern together suggests there is a missing generalisation here.

What separates the cells with incorrect predictions from those with correct predictions? Both sets of incorrect predictions are for negative forms, but the negative present tense is correctly predicted. Recall from §2.2, however, that the negative marker /m(ə)/ is a suffix in the present tense, but a prefix elsewhere. A possible generalisation, then, is that Dybo's Rule fails for these verbs when there is a negative prefix. Below are the three negative forms of 'to read', with the negative marker bolded; it is a suffix in (18a), but a prefix in (18b) and (18c). The remainder of the forms in this section come from the corpus – and thus ultimately from Yanagisawa (2010) – unless otherwise noted.

(18) *Negative forms of 'to read'*

a. сапҳьом
s-á-p^{h}χ^{j}o-**m**
1SG.ABS-3SG.NH.OBL-read.DYN-**NEG**
'I don't read it'

b. Бамыпҳьан!
b-a-**mə́**-p^{h}χ^{j}a-n
2SG.ABS-3SG.NH.OBL-**NEG**-read-NEG.IMP
'Don't read it!'

c. дамыпҳьакәа
d-a-**mə́**-p^{h}χ^{j}a-k'^{w}a
3SG.H.ABS-3SG.NH.OBL-**NEG**-read-NEG.PSTABS
'him/her not having read it'

The forms in (18) also suggest how the theory can be revised to account for them. In (18b) and (18c) the negative prefix, which is unaccented in the current implementation of Dybo's Rule (§4.2), carries primary stress. Since the morpheme carries stress here, it seems to behave as accented for these verbs. Below, I show two derivations of the negative imperative: first in (19a) with an unaccented negative prefix, which leads to an ungrammatical surface form, and then in (19b) with an accented negative prefix, which correctly predicts the location of stress.[4] Recall from (15) that Dybo's Rule assigns stress to the leftmost accent not immediately followed by an accent.

(19) *Derivations of negative imperative forms*

a. Бамыпҳьан!

/b-	a-	mə-	p^{h} χ^{j}a	-n/
–	+	–	– –	–
*[b-	á-	mə-	p^{h} χ^{j}a	-n]
2SG.ABS-	3SG.NH.OBL-	NEG-	read	-NEG.IMP

Intended: 'Don't read it!'

[4]The root is shown with two accent specifications here in accordance with Spruit's (1986) theory of elements: there is one accent for the /p^{h}(ə)/ and one for the /χ^{j}a/ (see §4.2).

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

b. Бамыԥхьан!

/b-	a-	mə-	pʰ χʲa	-n/
−	+	+	− −	−
[b-	a-	mə́-	pʰ χʲa	-n]
2SG.ABS-	3SG.NH.OBL-	NEG-	read	-NEG.IMP

'Don't read it!'

Although I do not show derivations for all the relevant forms here for reasons of space, treating the negative prefix as accented in fact turns all of the incorrect predictions in Table 4 into correct ones. The computer evaluation in §6.2 shows a dramatic improvement on verbs in this category. For this reason, I propose the following pattern of accent allomorphy:

(20) *Negative allomorphy*

a. In absolutive–oblique verbs without a preverb, the negative prefix /m(ə)-/ is accented;

b. Elsewhere, it is unaccented.

This illustrates how to use the results of the corpus evaluations to perform the type of analytical work familiar from theoretical phonology, namely revising a theory in response to new data. We began by noticing an area of the verbal lexicon where Dybo's Rule performed poorly. We examined those verbs and found a pattern, isolating the negative prefix as the morpheme causing the problems. Since this prefix can, exceptionally, carry stress in the relevant verbs, this suggests that the prefix has an accented allomorph. By introducing this allomorph all of the incorrect predictions are fixed.

In (21), I give a summary of the revised theory I propose for Abkhaz stress assignment, followed by discussion of data points that justify some of the complexity in the conditions for particular stress assignment rules.

(21) *Revised theory of stress assignment in Abkhaz verbs*

a. The negative prefix /m(ə)-/ is accented in absolutive–oblique verbs without preverbs (a restatement of (20)).

b. The causative prefix /r-/ is invisible to stress assignment.

c. The negative prefix /m(ə)/ and all ergative markers are accented in causatives.

d. Stress falls immediately before the root in short (one accent specification) unaccented absolutive-only verbs with no preverbs.

e. Stress falls immediately before the root in non-accent-initial absolutive-only verbs with a causative and no preverbs.

f. In all other cases, stress is assigned by Dybo's Rule with the accent specifications in §4.2.

The changes regarding causative verbs are mostly familiar from earlier work. Dybo's (1977:43) original proposal has ergative markers as accented in causatives, while Spruit (1986: 72) argues that the negative prefix shows the same allomorphy (see also Trigo 1992: 225). Trigo (1992) additionally proposes that the negative prefix /m(ə)/ and the ergative person markers are infixed. This infixation triggers a new cycle of phonological rule application, which is intended to explain why they exceptionally carry stress despite being underlyingly unaccented. Although these sometimes pattern together (see (21c)), I believe that this grouping of negative and ergative markers as infixes is not justified in general. Both negatives and ergatives can appear in the same morphological environment: between an absolutive on the left, and a short unaccented stem combined with suffixes on the right. In these cases, the negative is stressed, but ergatives are not:

(22) *Lack of parallelism between the negative and ergatives* (Yanagisawa 2010: 331, 433)

a. ҳмыԥеит
ħ-**mə́**-pʰe-jt'
IPL.ABS-**NEG**-jump-DYN.FIN
'we didn't jump'

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

b. исфеит
ji-**s**-fé-jt'
*ji-**sə́**-fe-jt'
3NH.ABS-**1SG.ERG**-eat-DYN.FIN
'I ate it'

In the revised theory in (21), the stress pattern in (22a) arises not from infixation but from pre-stressing, which is not applicable in forms with ergatives like (22b). For discussion of pre-stressing, including predictive differences between the revised theory and the analysis in Trigo (1992), see later in this section.

The invisibility of the causative /r-/ to stress assignment results in causative and non-causative versions of verbs having the same stress patterns in the infinitive, as if the causative marker were not there. In Kathman's (1992: 216–218) analysis of Abkhaz using metrical grids, the invisibility of the causative is accounted for by the assumption that this affix exceptionally does not have a grid mark. Trigo (1992: 226) analyses the causative prefix as extrametrical, also in a grid-based framework.

(23) *Causative /r-/ is invisible in stress assignment*

a. i. адыдра
á-dəd-ra
DEF-thunder-INF
'to thunder'

ii. ардыдра
á-r-dəd-ra
DEF-CAUS-thunder-INF
'to make (it) thunder'

b. i. абара
a-ba-rá
DEF-see-INF
'to see'

ii. арбара
a-r-ba-rá
DEF-CAUS-see-INF
'to show' (lit. 'cause to see')

The conditions for assigning stress immediately before the root in (21), which I call pre-stressing behaviour, are complex. Non-causative verb roots must contain at most a single (unaccented) accent specification, and there must not be a preverb or any person marker other than an absolutive. The data in (24) show that these conditions are necessary. Verbs which meet these criteria are pre-stressing, as in (24a). Polysyllabic verb roots, which have multiple accent specifications (see §4.2), are not, as in (24b), nor are verbs with preverbs, even if both the verb root and preverb are short, as in (24c). In (24b) and (24c), stress is root-final by Dybo's Rule. Roots are shown in bold for clarity.

(24) *Conditions on pre-stressing behaviour*

a. Быҧ!
bə́-**ҧ**
2SG.F.ABS-**run**
'Run!'

b. Бкəаша!
b-**k'ʷaşá**
2SG.F.ABS-**dance**
'Dance!'

c. Бнаʒа!
b-na-**d͡zá**
2SG.F.ABS-there-**reach**
'Reach thither!'

That person markers other than the absolutive block pre-stressing behaviour is easiest to see for verb roots which sometimes have an ergative and sometimes do not. Although such verbs still have an agent semantically, ergative markers are not present on imperatives to a singular addressee. If addressing a group, an ergative marker is present even in imperatives. This results in a difference in stress. In the singular form in (25a), there is no ergative marker, so the verb /ʒʷ(ə)/ 'drink' is pre-stressing. In the corresponding plural in (25b), there is an ergative marker, so the verb is instead assigned root-final stress by Dybo's Rule.

(25) *Ergative markers and pre-stressing behaviour* (Yanagisawa 2010: 123)

a. Ижə!
jí-ʒʷ
3NH.ABS-drink
'Drink it! (singular addressee)'

b. Ижəжəы!
ji-ʒʷ-ʒʷə́
3NH.ABS-2PL.ERG-drink
'Drink it! (plural addressee)'

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

Causative verbs can also display pre-stressing behaviour. Unlike non-causative verbs, pre-stress in causatives does not depend on the length of the verb root. This predicts that polysyllabic verb roots will follow Dybo's Rule when non-causative, while being pre-stressing as causatives. This prediction is correct, as shown in (26). The root 'shine' follows Dybo's Rule in a non-causative form, leading to root-final stress in (26a), but is pre-stressing when causative, as in (26b). Recall that the causative prefix itself is invisible to stress, which explains why pre-stress in (26b) does not result in the ungrammatical *[ji-rə́-laʂa].

(26) *Causative and non-causative forms of the same verb*

a. Блаша!
b-laʂá
2SG.F.ABS-shine
'Shine!'

b. Ирлаша!
jí-r-laʂa
3NH.ABS-CAUS-shine
'Illuminate it/them!'

Verb roots with underlying accents can also be pre-stressing in causative forms, which is another difference between pre-stress in causative and non-causative verbs. Pre-stress is found as long as the root begins with an unaccented element, even if accents are found later in the root. The root /ʂkʷák'ʷa/ 'white' has three elements in Spruit's (1986) analysis: /ʂ(ə)/, /k'ʷá/ and /k'ʷa/. The second of these bears an underlying accent, which explains why stress falls in this location on many inflected forms of this stem, as seen in (27a). Despite the presence of an underlying accent, the causative 'whiten' is pre-stressing, since the first element /ʂ(ə)/ is unaccented, as shown in (27b).

(27) *Causative pre-stressing for verbs with underlying accents* (Yanagisawa 2010: 556, 383)

a. ишкəакəахеит
/jə-ʂkʷák'ʷa-χa-jt'/
[ji-ʂkʷák'ʷa-χe-jt']
3NH.ABS-white-become-DYN.FIN
'it became white'

b. Иршкəакəа!
/jə-r-ʂkʷák'ʷa/
[jí-r-ʂkʷak'ʷa]
3NH.ABS-CAUS-white
'Whiten it!'

Previous analyses of Abkhaz do not include pre-stress, although the possibility has been alluded to (see, in particular, Kathman 1992: 220, fn. 10; Yanagisawa 2013: 136). This means that previous research has used other analytical tools to explain the stress patterns of the relevant verbs. I highlight here the analysis in Trigo (1992), which is one of the more detailed revisions to the theory of Abkhaz stress from Dybo (1977). Trigo analyses short unaccented roots in Abkhaz as extrametrical, explaining why stress falls before the root rather than on it. The pre-stressing analysis above accounts for more data than that of Trigo (1992) because of the causative generalisations illustrated in (26). Trigo (1992) does not predict that a polysyllabic stem like /laʂa/ 'shine' should ever be pre-stressing, since it is too long to be extrametrical. As shown in (26), this is correct only in non-causative forms.

Pre-stressing behaviour is also found on verb roots with underlying accents, whereas Trigo's (1992) extrametricality analysis requires a fully unaccented root. This predictive difference is illustrated in (27), where [jí-r-ʂk'ʷak'ʷa] 'Whiten it!' is pre-stressing despite the presence of an underlying accent in the root /ʂk'ʷák'ʷa/ 'white'.

There are many more sets of data that could be considered to show why particular conditions in (21) were or were not included. For reasons of space, I turn instead to the evaluation of the revised theory. Without access to a large corpus of inflected Abkhaz verb forms, the only way to evaluate this theory would be to apply it to minimal or near-minimal pairs of hopefully representative verb forms, in much the same way as I have done above. However, equipped with a corpus, we can implement the revised theory and evaluate it empirically against the verbal lexicon of Abkhaz. Below, I report the results of such an evaluation.

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

Table 5. *Breakdown of results for the revised theory.*

			Forms			Verbs		
Person markers	Causative	Preverb	Correct	Total	% correct	Correct	Total	% correct
Absolutive	—	Absent	501	504	99.40	70	72	97.22
		Present	147	168	87.50	18	24	75.00
Absolutive–oblique	—	Absent	79	84	94.05	10	12	83.33
		Present	495	504	98.21	65	72	90.28
Absolutive–ergative	Absent	Absent	410	413	99.27	57	59	96.61
	Present		414	448	92.41	50	64	78.13
	Absent	Present	340	378	89.95	36	54	66.67
	Present		90	98	91.84	12	14	85.71
Absolutive–oblique–ergative	Absent	Absent	—	—	—	—	—	—
	Present		47	49	95.92	6	7	85.71
	Absent	Present	392	399	98.25	50	57	87.72
	Present		66	70	94.23	9	10	90.00
Total			2981	3115	95.70	383	445	86.07

6.2. Evaluating the revised theory

The evaluations reported in this section follow the same methodology as in §4, but with the accent allomorphy and pre-stressing patterns described in (21) instead of the assumptions in §4.2. The results are summarised in (28), alongside those of the original implementation of Dybo's Rule from §4.

(28) *Results for two theories of Abkhaz stress assignment*

a. *Original implementation* (results repeated from (17))

i. Correct / total (forms): 2,816 / 3,115 = 90.40%

ii. Correct / total (verbs): 343 / 445 = 77.08%

b. *Revised theory*

i. Correct / total (forms): 2,981 / 3,115 = 95.70%

ii. Correct / total (verbs): 383 / 445 = 86.07%

The difference between the original implementation and the revised theory is that 40 additional verbs, almost 10% of the corpus, now have stress correctly predicted on all seven forms. Over 150 additional verb forms, almost 5% of the total number of forms, are now assigned stress correctly. Of the verbs which have all seven forms correctly predicted, 30.55% are assigned stress using one of the theoretical modifications in the revised theory presented in (21), including accentual allomorphies and pre-stress. Among the same verbs, 8.38% have pre-stressing forms. These numbers highlight the fact that even in the revised theory Dybo's Rule forms the core of Abkhaz verbal stress assignment. However, for three verbs in ten, the revised theory is relevant in assigning stress, showing that a considerable portion of the verbal lexicon obeys slightly different principles than previously assumed.

Table 5 shows the results of the revised theory for different subsets of the corpus. For every row in Table 5, the revised theory either ties with or outperforms the original implementation in §5 (cf. Table 3). All verb categories with extremely low scores are now gone, with no row having less than 87.5% of its forms predicted, and no row having fewer than two-thirds of its verbs fully accounted for. Several classes of verbs are approaching the maximum of 100%.

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

There are 46 verbs which are fully accounted for in the revised theory but not in the original implementation, and only six verbs which were fully accounted for in the original implementation but not in the revised theory. This results in the net difference of 40 verbs in (28). Five of the six verbs that the original implementation fared better on are causative verb forms, suggesting that there may be nuances in causative stress behaviour not captured by the revised theory. However, the revised theory outperforms the original implementation on causatives overall, suggesting that despite occasional exceptions, it still represents a step in the right direction. The next section provides closing discussion for the article, including discussion of some of the challenges in accounting for the remaining verbs.

7. Discussion

In this article I have described the creation of a corpus of 3,115 verb forms from 445 Abkhaz verbs in order to study the patterns of stress assignment in this language. The corpus was used to empirically evaluate a theory of stress assignment against a large database, via a Python 3 program. I also implemented a revised theory, designed to improve on the weakness of the original, and evaluated it against the same corpus. This project can be seen as a case study for how to use corpora in phonological research, and I hope that other linguists are able to adapt this methodology to improve on previous phonological theories from other languages.

A recurring theme in this type of digital phonological work is the division of labour between computer software and human linguists. There are several advantages to computer-assisted phonological analysis. Testing whether Dybo's Rule accounts for 3,115 different verb forms would take a human weeks or months, but because the individual steps of the algorithm are so simple, it takes a modern computer less than a second to accomplish the same task. Unlike humans, computers also do not make errors due to fatigue or distractions, ensuring that the analysis is fully reproducible. By implementing a theory on a computer, the linguist is also forced to state it precisely and unambiguously, something that may be difficult to accomplish in natural language. Even if an analysis is presented in a formal theoretical framework, specifying which forms are subject to which rules in which contexts is sometimes done using natural language generalisations in linguistics papers.

Of course there are also many disadvantages. In order to be parsed by a computer, all data must conform strictly to some very precise format. Any forms that do not meet the specifications must either be tediously hand-corrected, a process which may take longer than simply analysing the forms by hand, or else discarded, risking biases in the form of artificially regular data sets. Computers also make errors, but of a different kind than those human linguists tend to make. If I analyse Abkhaz forms by hand, I can notice morphemes appearing in unexpected positions, or words which appear to be loanwords, or contain typos. By contrast, unless it is told in advance about every single possible exception and edge case, a computer will not care if it is fed Abkhaz verb forms or song lyrics or randomly generated strings. A computer program will only notice what it is told to notice, meaning that much of the burden of the intellectually challenging work involved in phonological analysis still falls on the shoulders of the linguist.

Although the revised theory in §6 markedly improves on the original implementation of Dybo's Rule from §4, it still has many flaws. There are 60 verbs in the corpus that it fails to fully account for, and it is not clear how to account for their stress patterns. It is possible that some include typos, or judgement errors because Yanagisawa's (2010) consultant was distracted, or confused a verb with a similar-sounding one. Some verbs may have two possible patterns of stress alternation, with the dictionary reporting a mix of the two. It is also doubtless the case that Abkhaz has lexical exceptions – forms which, for whatever diachronic reason, do not follow the general synchronic rules. Given the complexity of the data, it is also surely true that there are generalisations which I have overlooked, because I never thought to put the right two subsets of the lexicon next to each other to see that they in fact behave the same. In §6, I have presented one possible theory of Abkhaz verb stress, which appears to be relatively empirically successful. But countless imaginable theories have not been implemented or evaluated, and I do not doubt that there are many patterns in the data which remain to be discovered.

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

Several theoretical questions have also not been addressed in this article. Phonologists typically implement proposals in theoretical frameworks rather than computer code. I have not attempted this here. Previous work on Abkhaz stress does offer implementations of Dybo's Rule (see Kathman 1992; Trigo 1992; Vaux & Samuels 2018; Andersson 2024b for analyses in metrical grid theories of stress). All of these authors also attempt to bring Spruit's (1986) 'elements' in line with more familiar phonological units such as the syllable, the mora or the segment.

There are several previous proposals for how to analyse pre-stressing behaviour in languages with lexical stress, whether in terms of feet (as in Özçelik 2014), through a theory of weak *vs.* strong lexical accents (as in Revithiadou 1999 and Yates 2017), Turbidity Theory and Coloured Containment (as in Revithiadou 2007), or morpheme-specific parenthesis projections (as in Halle & Idsardi 1995). Andersson (2024b) proposes a metrical grid analysis following Idsardi (1992) which accounts for the data in a single grammar, but it would be interesting to know which other phonological theories are capable of implementing both Dybo's Rule and pre-stressing behaviour in the same language. The revised theory in §6 also includes morphemes like preverbs and ergative markers which appear to block exceptional pre-stressing behaviour, and cause verbs to behave regularly in terms of stress. Such overriding of exceptional stress assignment is familiar from languages such as Armenian (Dolatian & Özçelik 2021), and is also modeled in Abkhaz in Andersson (2024b). However, future work remains to determine which theories can and cannot capture such behaviour in general.

8. Conclusion

This article has reported on a corpus study of stress assignment in Abkhaz verbs. Because of the complex interactions between the prosodic phonology and the polysynthetic verb morphology of the language, I created a database corpus which allows for the study of these interactions across the verbal lexicon. The corpus, which was compiled from dictionary data (Yanagisawa 2010), contains seven inflected forms each from 445 verbs, for a total of 3,115 forms, each annotated with phonological and morphological information. The code and final corpus are publicly available (Andersson 2024a).

In order to make use of the corpus, I implemented Dybo's Rule, a previous proposal for Abkhaz stress assignment (Dybo 1977; Spruit 1986), in the programming language Python, which allowed me to test the theory empirically against all of the forms in the verb corpus. Dybo's Rule relies on a binary system of accent status – accented *vs.* unaccented – and states that stress falls on the leftmost accent not immediately followed by an accent. The version of Dybo's Rule implemented was relatively successful, correctly predicting stress on 90% of the data in the corpus, and correctly predicting stress on all seven forms of a large majority of the lexicon (77%). Despite this, there were several verb categories where the theory performed poorly. Based on the generalisations about where it performed well and where it did not, I proposed both minor and major changes to Abkhaz stress assignment. This included both novel and previously proposed patterns of allomorphy, environments where certain affixes which are normally unaccented become accented instead. I have argued that some verbs in Abkhaz do not use Dybo's Rule at all, and that stress instead falls immediately before the verb, a pre-stressing pattern familiar from other languages with similar stress systems. The conditions for pre-stressing behaviour are complex, relying on both phonological (root length, root accent status) and morphological factors (presence or absence of person markers, preverbs, and causativisation). Without access to a large corpus, identifying these conditions would likely not have been possible.

I implemented the revised proposal in software and tested it empirically on the verb corpus, showing that its performance was the same as or better than the previous implementation in all verb categories. The revised theory accounts for 95% of verb forms in the corpus, and correctly predicts all forms for 86% of the lexicon. For approximately three verbs in ten whose forms were correctly predicted, at least one of the modifications from the revised theory was used to assign stress. Methodologically, I have relied heavily on digital technology in general, and computer programming in particular. Python scripts were used in every step of the corpus creation and in the evaluation of phonological theories. This

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

ensures a fast and reproducible analysis, which was indispensable when dealing with such a large and complex data set. The use of computational evaluation of theories also has the advantage of forcing us to be precise and unambiguous about what claims our phonological theories make.

I have left several important tasks for future work. Many verbs in Yanagisawa (2010) could not be parsed and glossed automatically, and hand-correcting these verbs may increase the size of the corpus considerably. Empirically, there is likely much more work to be done in identifying the stress patterns of the 14% of the verbal lexicon where even the revised theory fails to correctly predict all forms. On the theoretical side, I have not attempted here to translate the complex principles of Abkhaz stress assignment from the language of code into the language of mainstream theoretical frameworks. See Andersson (2024b) for such work, including detailed discussion of the theoretical consequences of Abkhaz stress assignment. Although much work remains to be done, I hope that the data and generalisations I have identified here are of some use to the field of Abkhaz studies and Caucasian linguistics, and that the methodologies I have used may be of interest to other phonologists who seek quantitative answers to complex phonological questions.

Data availability statement. The corpus and Python code used in this study are available in a GitHub repository at https://github.com/SJAndersson/abkhaz-verbs (cited in the text as Andersson 2024a).

Acknowledgements. First and foremost, I would like to thank Tamio Yanagisawa for allowing me access to his work and allowing me to publish the results. I have had many useful discussions about Abkhaz phonology over the years with Claire Bowern, George Hewitt, Björn Köhnlein, Jason Shaw, Bert Vaux, and Natalie Weber among others. I thank the reviewers and editors for their comments, which have made this article better. Remaining errors are my own.

Competing interests. The author declares no competing interests.

References

Allen, William Sidney (1965). On one-vowel systems. *Lingua* **13**, 111–114.

Anderson, Stephen R. (1978). Syllables, segments, and the Northwest Caucasian languages. In Alan Bell & Joan B. Hooper (eds.) *Syllables and segments*. Amsterdam: North-Holland Publishing Company, 47–58.

Andersson, Samuel (2024a). abkhaz-verbs. GitHub repository. https://github.com/SJAndersson/abkhaz-verbs.

Andersson, Samuel (2024b). *The phonetics and phonology of Abkhaz word stress*. PhD dissertation, Yale University.

Arstaa, Šoṭa Q. & Lidia P. Č'ķadua (2002). *Apsua liṭeraṭuraṭ^{o} byzš^{o}a agrammaṭiķa: afoneṭiķa, afonologia, až^{o}axyrčiaara, amorfologia [Grammar of the literary Abkhaz language: phonetics, phonology, word formation, morphology]*. Sukhum: Sukhum Branch Soros Foundation.

Aršba, Nelli V. (1979). *Dinamičeskoe udarenie i redukcija glasnyx v abxazskom jazyke [Dynamic stress and vowel reduction in Abkhaz]*. Tbilisi: Mecniereba.

Bruening, Benjamin (1995). Lexical stress interactions in Abkhaz. Ms, Massachusetts Institute of Technology.

Chirikba, Viacheslav A. (2003). *Abkhaz*. Munich: Lincom Europa.

Colarusso, John (1988). *The Northwest Caucasian languages: a phonological survey*. New York: Garland.

Cutler, Anne (2005). Lexical stress. In David B. Pisoni & Robert E. Remez (eds.) *The handbook of speech perception*. Oxford: Blackwell, 264–289.

Dobrushina, Nina, Michael Daniel & Yuri Koryakov (2021). Languages and sociolinguistics of the Caucasus. In Maria Polinsky (ed.) *The Oxford handbook of languages of the Caucasus*. Oxford: Oxford University Press, 26–66.

Dolatian, Hossep & Öner Özçelik (2021). Cyclic overriding of irregular stress in Armenian. Paper presented at the 28th Manchester Phonology Meeting, online, May 2021.

Dybo, Vladimir Antonovič (1977). Zapadnokavkazskaja akcentnaja sistema i problema eë proisxoždenija [The West Caucasian accentual system and the problem of its origin]. In Vyacheslav V. Ivanov (ed.) *Nostratičeskie jazyki i nostratičeskoe jazykoznanie [The Nostratic language and Nostratic linguistics]*. Moscow: Institute of Slavic and Balkan Studies of the USSR Academy of Sciences, 41–45.

Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.) (2021). *Ethnologue: languages of the world*. 24th edition. Dallas, TX: SIL International.

Genko, Anatolij Nestorovič (1998). *Apsua–auryṣ^{o} ž^{o}ar [Abkhaz–Russian dictionary]*. Sukhum: Alašara.

Gulia, Dmitrij I. (1939). *Sbornik abxazskix poslovic, zagadok, skorogovorok, omonimov i omografov, narodnyx primet o pogode, zagovorov i nagovorov [Collection of Abkhaz proverbs, riddles, tongue-twisters, homonyms, and homographs, folk observations on the weather, charms, and spells]*. Sukhum: Abnii AN SSSR.

Halle, Morris & William J. Idsardi (1995). General properties of stress and metrical structure. In John Goldsmith (ed.) *The handbook of phonological theory*. Oxford: Blackwell, 403–443.

Hewitt, Brian G. (1979). *Abkhaz: a descriptive grammar*. Amsterdam: North-Holland Publishing Company.

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press

Hewitt, Brian G. (2010). *Abkhaz: a comprehensive self-tutor*. Munich: Lincom Europa.
Idsardi, William J. (1992). *The computation of prosody*. PhD dissertation, Massachusetts Institute of Technology.
Jakovlev, Nikolaj F. (2006). *Grammatika abxazskogo literaturnogo jazyka [Grammar of the literary Abkhaz language]*. Sukhum: Abkhazian Academy of Sciences.
Kaslandzija, Vladimir Arušanovič (2005). *Abxazsko–russkij slovar' [Abkhaz–Russian dictionary]*. Sukhum: Abkhazian State Academy.
Kathman, Dave (1992). Stress and accent in Abkhaz. In Germán F. Westphal, Benjamin Ao & Hee-Rahk Chae (eds.) *ESCOL '91: proceedings of the eighth Eastern States Conference on Linguistics.* Columbus, OH: Department of Linguistics, Ohio State University, 210–221.
Kuipers, Aert (1960). *Phoneme and morpheme in Kabardian*. The Hague: Mouton.
Meurer, Paul (2011). A finite state approach to Abkhaz morphology and stress. In Nick Bezhanishvili, Sebastian Löbner, Kerstin Schwabe & Luca Spada (eds.) *Logic, language, and computation*. Heidelberg: Springer, 271–282.
O'Herin, Brian (2021). Abaza and Abkhaz. In Maria Polinsky (ed.) *The Oxford handbook of languages of the Caucasus*. Oxford: Oxford University Press, 447–488.
Özçelik, Öner (2014). Prosodic faithfulness to foot edges: the case of Turkish stress. *Phonology* **31**, 229–269.
Revithiadou, Anthi (1999). *Headmost accent wins: head dominance and ideal prosodic form in lexical accent systems*. PhD dissertation, Leiden University.
Revithiadou, Anthi (2007). Colored turbid accents and containment: a case study from lexical stress. In Sylvia Blaho, Patrik Bye & Martin Krämer (eds.) *Freedom of analysis?* Berlin: Mouton de Gruyter, 149–174.
Spruit, Arie (1986). *Abkhaz studies*. PhD dissertation, Leiden University.
Trigo, Loren (1992). Abkhaz stress shift. In Brian G. Hewitt (ed.) *Caucasian perspectives*. Munich: Lincom Europa, 191–235.
Vaux, Bert & Zihni Pysipa (1997). The Cʷəž'ə dialect of Abkhaz. *Harvard Working Papers in Linguistics* **6**, 193–208.
Vaux, Bert & Bridget Samuels (2018). Abstract underlying representations and prosodic structure. In Diane Brentari & Jackson L. Lee (eds.) *Shaping phonology*. Chicago: University of Chicago Press, 146–181.
Yanagisawa, Tamio (2010). *Analytic dictionary of Abkhaz*. Tokyo: Hitsuji Shobo.
Yanagisawa, Tamio (2013). *A grammar of Abkhaz*. Tokyo: Hitsuji Shobo.
Yates, Anthony D. (2017). Against root faithfulness in Cupeño stress. In Karen Jesney, Charlie O'Hara, Caitlin Smith & Rachel Walker (eds.) *Proceedings of the 2016 Annual Meeting on Phonology.* Washington, DC: Linguistic Society of America, 11 pp.

https://doi.org/10.1017/S0952675726100281 Published online by Cambridge University Press