

## Chapter 1

# There's a(nother) quotative on the rise in Macedonian

Izabela Jordanoska

CNRS LLACAN

Leonardo Contreras Roa

CNRS LLACAN

This paper shows that the speech verb *reče* 'says' in Macedonian is in the initial stages of grammaticalization towards a quotative marker. Morphologically, syntactically and phonetically *reče* shows properties of a quotative clitic. First, because unlike other speech verbs, *reče* tends to follow the first phonological word within a discourse report. Secondly, it can co-occur with other speech verbs. Furthermore, it is not always inflected in contexts where a verb would be. Finally, *reče* is sometimes reduced to *eče* or just *če*. However, there is a high degree of variation and the quotative use of *reče* can only be found in spontaneous, colloquial spoken language. There are also speakers who seem to use another speech verb, *veli*, as a quotative marker, rather than *reče*.

## 1 Introduction

This paper examines the verbs used to introduce reported speech in Macedonian (Eastern South Slavic, Indo-European) and argues that the verb *reče* 'say' has properties of a quotative marker. The Macedonian verbs corresponding to 'say' are *reče*, the perfective stem, and the suppletive imperfective stem: *veli*, or *vika*, the colloquial form of it.<sup>1</sup> Mushin (1997: 298) has shown that the verb *veli* is grammaticalizing into a 'hearsay marker'. The speech data used in the current paper, however, comes from speakers who do not use the form *veli* at all, neither as a

---

<sup>1</sup>The original meaning of *vika* is 'shouts'.

speech verb nor as a hearsay marker. Instead, they only use *vika* and *reče*. While *vika* is imperfective, and therefore on par with *veli* on an aspectual level, it is the verb *reče* which for these speakers is grammaticalizing into a quotative marker. In this paper we compare the use of *vika* and *reče* and show that *reče* is grammaticalizing into a quotative marker, whereas *vika* is not. Thus, there is usage difference between *vika* and *reče* that goes beyond the aspect they encode.

The methodology is detailed in §2. In §3 we present known grammaticalization sources for quotative markers, and we discuss the role of evidential marking, which Macedonian also employs. In §4 we compare the use of *reče* and *vika* and show that *reče* has taken on some grammatical properties associated with quotatives. In §5 we show phonetic properties of *reče* which indicate that it is becoming reduced. We discuss the implications of our findings in §6. §7 concludes.

## 2 Methodology

Macedonian is spoken in North Macedonia, northern Greece and western Bulgaria by approximately 2 million speakers in total. Notable diaspora communities exist in North America and Australia. It has three main dialects, although the exact classification varies slightly across sources. [Vidoeski et al. \(2003\)](#); [Markovič \(2007\)](#) divide it into the Western, Northern and Southeastern groups, whereas [Friedman \(2001\)](#) divides it into the Western, Eastern and the Kostur-Korča group. Friedman's Kostur-Korča group is a sub-part of the Southeastern dialects in [Vidoeski et al. \(2003\)](#); [Markovič \(2007\)](#). The data used in this paper all come from Western dialects, which are all spoken in North Macedonia, and which are classified in the same way in [Friedman \(2001\)](#) and [Vidoeski et al. \(2003\)](#); [Markovič \(2007\)](#). The main data source is a 19-minute corpus consisting of three folktales told by an elderly female speaker living in Skopje, but originally from Sekirci, a village near Prilep. The speaker was recorded with her permission on a smartphone and the recordings were annotated in ELAN CorpA ([Chanard 2015](#)), using the Speechreporting template ([Nikitina et al. 2019](#)).

Additional data sources are the corpus collected by Georges Drettas 1976 in the Pangloss Collection ([Michailovsky et al. 2014](#)), and the recordings in the Research Center for Areal Linguistics of the Macedonian Academy of Arts and Sciences (MANU, [Vidoeski et al. \(s.a\)](#)). From the MANU collection, data was taken from the recordings made of speakers from the following locations: Bučište and Brodec (both classified as Northern dialects), Trpejca and Lukovo (both classified as peripheral Western dialects), and Bitola and Šiševo (both classified as central Western dialects). These recordings included instances of *reče*. Other recordings were

excluded because either the speech didn't appear to be spontaneous enough, the audio quality was not good enough, the dialect was too different from the Western dialect, they didn't contain discourse reports, or they did contain discourse reports but they were introduced by other speech verbs than *reče*.

Finally, there are also secondary data sources used and indicated as such. In these cases the glosses and orthography are adjusted to be consistent with the glosses and orthography used in this paper and occasionally the translations are slightly adjusted to better match the original utterance.

The methodology used for the phonetic study is detailed in section §5.

### 3 From speech verbs to quotatives and beyond

#### 3.1 Grammaticalization of quotatives cross-linguistically

Quotatives are particles that are used to mark reported speech. A definition from is given in (1).

- (1) QUOTATIVE MARKER: A grammaticalized function word which is normally adjacent to the quote. (Güldemann 2008: 14)

Cross-linguistically, sources for quotative markers include speech verbs, prepositions and similatives (Kouteva et al. 2019). A list including both general properties of grammaticalized markers, as well as specific properties of quotatives that have been grammaticalized from speech verbs, is given in (2). This list is based on criteria found in Deutscher (2011); Nikitina et al. (2021); Güldemann (2008); Hopper & Traugott (2003):

- (2) a. they can get repeated multiple times throughout a discourse report  
b. in the case of clitics: they often occur following the first phonological word  
c. they can co-occur with speech verbs  
d. their form is invariant  
e. they are mono- or disyllabic

Languages with fully grammaticalized quotatives from the verb 'say' include Hausa (*cēwā* from *cē* 'say', see Lord (1990)), colloquial Russian (*grit* from *govorit* 'tells, says', see Makartsev (2012: 31)) and *mol* from *molvit* 'speaks, says', see Vykypěl (2010: 140)), Tatar (*dip*, from *dip*, 'saying', see Greed (2014: 86)), West African Pidgin English (*sey*, from *sey* 'say', see Lord (1990)) and Buru (*fen* from

*fen(e)* ‘think’, ‘say’, ‘affirm’, see Klamer (2000)). An example of the use of a quotative from Buru (Central Malayo-Polynesian, Austronesian) is given in (3a). In Buru, *fen* is grammaticalized even further and can also function as a complementizer, i.e., it also introduces complements of non-speech verbs, as shown in 3b.

- (3) a. Da prepa **fen**, “Sira rua kaduk.”  
 3:SG say FEN 3:PL two arrive  
 ‘She said, “The two of them came”.’  
 Buru (Klamer 2000: 76)
- b. Sira kita **fen** de iko.  
 3PL saw FEN 3SG go  
 ‘They saw that he left.’  
 Buru (Klamer 2000: 78)

In other languages, there are quotative markers which are not yet fully grammaticalized into particles or complementizers. The Turkic languages Chuvash and Bashkir, for example, each have two semi-grammaticalized quotatives from speech verb, in which one of them is further grammaticalized than the other (Nikitina et al. 2021; Knyazev 2019), see also Aplonova (this volume) and Nikitina (this volume). An example from Chuvash is shown in (4) and from Bashkir in (5).

- (4) a. es **t-et** man-[a] **t-et** pitar-s[a] lar-t  
 2SG QY-PRS.3SG 1SG.OBL-ACC/DAT QY-PRS.3SG bury-CV.COORD sit-CAUS  
**t-et**.  
 QY-PRS.3SG  
 ‘You, he says, hide, he says, me, he says.’  
 Chuvash (Nikitina et al. 2021: 7)
- b. kaçç-ax **te-s[e]** şotl-at<sup>j</sup>  
 suitor-EMPH say-CVB think-PRS.3SG  
 ‘She thinks he is a suitor.’  
 Chuvash (Nikitina, this volume)

In (4a) *tet* ‘say’ – referred to as a quotative verb in Nikitina et al. (2021) and Nikitina (this volume) – appears three times throughout the discourse report. This repetition is a property of a quotative, rather than a full speech verb. Nonetheless, the stem can still get inflected and is used as a speech verb, thus the verb is not fully grammaticalized into a quotative marker (Nikitina, this volume). The converb *tese* is further grammaticalized, as unlike *tet*, it also appears with non-speech verbs, such as ‘think’ in (4b), where it has a complementizer function.

Bashkir follows a similar pattern: it has a more verb-like quotative, *tigän*, and a more particle-like quotative, *tip*. The Bashkir example in (5a) shows that the quotative *tip* can co-occur with another speech verb, such as *äjtermen* ‘tell’. *Tigän* in (5b), on the other hand, is more verb-like and can also occur on its own.

- (5) a. [min heð-ðeŋ olata-əyǝð-ðəŋ olata-hə-nəŋ duθə  
 1SG 2SG-GEN grand.father-P.2PL-GEN grand.father-P.3-GEN friend  
 ine-m] **ti-p** äjt-er-men  
 be.PST say-CVB tell-POT-1SG  
 ‘I will say: “I am a friend of your great-grandfather”’
- b. [äjðä mineŋ menän] **ti-gän**  
 come.on 1SG.GEN with say-PTCP.PST  
 “Come on, come with me” – (she) said.’  
 Bahskir (Aplonova 2021)

We claim that *reče*, like Chuvash *tet* and Bashkir *tigän*, is still in the early stages of the process of grammaticalization.<sup>2</sup> That is to say, there are instances of *reče* being used both as a quotative marker and as a full verb.

Thus, this study adds to the amount of diversity in the spectrum between regular speech verbs and fully-grammaticalized complementizers, and helps us to grasp this phenomenon for other languages in which this process is still ongoing. The grammaticalization process is schematically illustrated with the continuum in Figure 1.

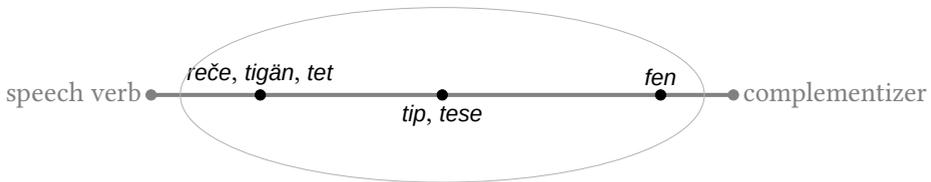


Figure 1: the quotative space

The leftmost node represents speech verbs that fully function as verbs and are not undergoing any form of grammaticalization. Macedonian *reče*, Bashkir *tigän*

<sup>2</sup>There are other particles in Macedonian that originate from speech verbs. One of them is *demek* ‘as if’, from Turkish *demek* ‘say’. Turkish itself has the particle *diye* ‘in order to’, ‘as’, from *demek* (Özyıldız et al. 2018). According to Makartsev (2012) *demek* is quite often used as a reportive with an additional epistemic meaning of disbelief. Another such particle is *goa* ‘allegedly’, from Persian *gū* ‘say’, through Turkish *güya/güya* (Makartsev 2012). There are also other particles that have grammaticalized from *reče*, but which have taken on an approximative instead of a quotative meaning: *rečisi* and *tukureče*, both meaning ‘approximately, almost’.

and Chuvash *tet* are already moving towards the complementizer node on the right side. They are not as far, however, as Bashkir *tip* or Chuvash *tese*, which in their turn are as not as far as Buru *fen*. Note that the scale is just for visualization and there is no formal correspondence to the distance between the nodes. The gray circle around the scale represents the ‘quotative space’, i.e. the area in which any given element on the continuum can have a quotative function.

### 3.2 Evidentiality and quotative markers

The grammaticalization process described in the previous subsection also takes place in a language more closely related to Macedonian than Buru, Chuvash or Bashkir. Namely in the Banar dialect of Bulgarian (South Slavic), which has the quotative particle *kaj*, from *kâe* ‘says’ (Makartsev 2012: 31), as exemplified in (6).

- (6) Do-sega da e stana-l veke doktor, ama inat  
until-now SBJ be.PRS.3SG become-AOR.LPTCP.M already doctor but spite  
hora tukašni-te. Rektora, kaj, ne go ostava. Ne  
people from.here.PL-DEF.PL director QUOT NEG 3SG.M.ACC leave NEG  
može, kaj, kaza-l mu, za tri mesec-a da staneš doktor,  
can QUOT tell 3SG.M.DAT for three month-PL SBJ become-2SG doctor  
kaj. A be kak da ne može, kogato momče-to znae?  
QUOT but INTJ how SBJ NEG can when boy-DEF.N know  
‘He should have been a doctor already by now, but the people here are  
spiteful. The rector, he says, won’t let him. “It is impossible” – he says he  
said to him – “for you to become a doctor in three months” – he says.  
Well, but how can it be impossible when the fellow knows [everything].’<sup>3</sup>  
Bulgarian (Friedman 1998: 4)

Furthermore, the use of quotative markers is also found in (at least) two other languages that are considered part of the Balkan Sprachbund: Albanian (*gjoja*) and Turkish (*diye*) (Friedman 1998). Another feature that these languages share is evidential marking on the verb. In Macedonian, evidentiality is marked with the so-called *l*-participle form of the verb (Fielder 1999; Friedman 2000; 2003; Gvozdanovic 1996; Mushin 2000), as in (7).

- (7) Тој беше во Скопје – односно бил, не го видов.

---

<sup>3</sup>Glosses added by first author.

Toj beše vo Skopje – odnosno bil, ne  
 3SG.NOM be.IMPF.2SG in Skopje actually be.AOR.LPTCP.M NEG  
 go vidov.  
 3SG.ACC.M see.AOR.1SG  
 'He was in Skopje, or actually he was (supposedly), I didn't see him.'  
 (Lunt 1952: 93)

In (7) the verb 'be' appears in two forms: the participle form *bil*, marked for reported evidentiality, and the imperfect finite form *beše*, which would (in this case wrongfully) convey that the speaker can vouch for the truth of the utterance. The English translation can only be approximated by adding the adverb 'supposedly'.

According to Friedman (2003) the use of the *l*-form in Macedonian can be used to mark any of the following: reported speech, dubitativity and mirativity. Tomić (2006) additionally shows they are used in narratives. Mushin (2000), in an empirical study on narratives with speakers from Bitola (Western dialect), has found that the *l*-participles are indeed frequently used in narratives, except in direct speech reports, since this involves deictic shift towards the character's perspective. Direct speech was not the only instance in which non-evidential forms were used, however. An unexpected use of the non-evidential form that Mushin (2000) observes is shown in (8).

- (8) Си дошла свекрвата од пазар и прашала “Како е лебот?” Таа му рекла невестата “Арен е,” рече, “Ама внатре како жила постана. Не допечен.’ Свекорот му рекол “Арен е. Арен е. Ќе го изедеме.” Рече “Не е лошо. Убаво печен.”
- a. Si= doš-l-a svekrva-ta od pazar i  
 REFL= come-AOR.LPTCP-F mother.of.husband-DEF.F from market and  
 prava-l-a “Kako e leb-ot.”  
 ask.PFV-AOR.LPTCP-F how be.PRS.3SG bread-DEF.M  
 ‘The mother-in-law returned from the market and asked, “How is the bread?”’
- b. Taa mu= rek-l-a nevesta-ta “Aren e”  
 3SG.NOM 3SG.DAT= say.PFV-LPTCP-F bride-DEF.F good.M be.PRS.3SG  
 reče, “Ама внатре како жила по-стана. Ne=  
 say.PFV.3SG but inside how tendon PFV-become.AOR.3SG NEG=

do-peče-n.”

PFV-bake-PTCP.M

She said, the bride, “It is good.” She said, “But it became like a tendon inside. Not fully baked.”

- c. Svekor-ot mu= rek-ol “Aren  
 father.of.husband-DEF.M 3SG.DAT= say-AOR.LPTCP.M good.M  
 e. Aren e. Āe= go= iz-edi-me.”  
 be.PRS.3SG good.M be.PRS.3SG FUT= 3SG.M.ACC= PFV-eat-PRS.1PL  
**Reče** “Ne e loš-o. Ubav-o peče-n.”  
 say.PFV.3SG NEG be.PRS.3SG bad-N nice-N bake-PTCP.M  
 ‘The father-in-law said to her, “It’s good. It’s good. We will eat it.” He  
 said. “It’s not bad. Nicely baked.”’ (Mushin 2000: 943-4)

The unexpected forms in (8) are the two instances of *reče*. Mushin (2000: 944) calls these forms problematic because they are preceded by the *l*-participle form, but they themselves are in the perfective present form, creating a mismatch in tense. However, under the analysis proposed in this paper these data can easily be explained as examples of the quotative use of *reče*.

Furthermore, Mushin (2000: 954) observes that “despite the existence of a grammaticalized system in Macedonian for coding information as reportive, speakers used alternative strategies as well to achieve a reportive effect.” These alternative strategies include the use of a so-called ‘framing verb’, such as *veli* in (9).

- (9) И вели “Татко ми беше после на крајот” вели. “А лицитацијата  
 почна од пет илјади. И нагоре” вели “кренавме до кај седум вели и  
 пол илјади” вели.  
 I **veli** “Tatko =mi beše после na kraj-ot”  
 but say.IPFV-PRS.3SG father =1SG.POSS be.IMPF.3SG after on end-DEF.M  
**veli** “A listatsija-ta počna od pet ilijadi.  
 say.IPFV-PRS.3SG and auction-DEF.F start.PFV.AOR.3SG from five thousand  
 I **veli** “krena-vme do kaj sedum”  
 and upward say.IPFV-PRS.3SG raise.PFV-AOR.1PL until at seven  
**veli** “i pol ilijadi” **veli**.  
 say.IPFV-PRS.3SG and half thousand say.IPFV-PRS.3SG  
 ‘And she says, “My father was last in the end.” She says, “and the auction  
 started from \$5000 and up” she says. “We raised (it) to seven,” she says,  
 “and a half thousand.” she says. (Mushin 2000: 950)

In (9) the speech verb *veli* is repeated throughout the discourse report, much

like a quotative (Mushin 1997).

We speculate about an explanation for the multiple strategies, i.e., the use of evidential verb forms as well as emerging quotatives. It is possible that due to the use of the evidential verb forms being part of the grammar of Macedonian, speakers are already used to marking reported speech. However, since in the reported speech in narratives almost always involves deictic shift towards the reported speaker's perspective, the *l*-form cannot be used within the speech report. (Using an *l*-form with a first person subject would give a reading of self-doubt, i.e., not the desired reading.) Indeed, in our corpus, out of 154 reported speech constructions, only one did not involve deictic shift towards the reported speaker's perspective.

Thus, in narratives, there are two levels of reportedness. First of all, the whole narrative is marked as reported speech with the *l*-forms, since it is a retelling and the speaker did not witness any of the events. Secondly, there is reported speech within the narrative that is attributed to the characters, which can be marked with a quotative. Since Macedonian speakers are used to marking reported speech, the quotative marker emerges a way for the storyteller to mark reported speech when the *l*-form is unavailable. This proposal remains to be fully explored in future work.

In the next section we present the morphosyntactic evidence that *reče* is indeed grammaticalizing into a quotative clitic.

## 4 Grammatical properties of *vika* and *reče*

The following section contains counts that only refer to the corpus of three narratives that we recorded. However, some of the examples for illustration, come from the MANU corpus.

In this section we show that *reče* 'say.PFV' and *vika* 'say.IPFV' behave in different ways, that are unexpected if the difference between them were purely aspectual. First, it is shown that *reče* is more frequent than *vika*. Second, *reče*, but not *vika* is relatively the most frequent after the first phonological word of a reported clause. Third, there are examples of co-occurrence of *reče* with other speech verbs. *Vika*, on the other hand, only co-occurs with *reče*, but not with other speech verbs. Fourth, there are cases of *reče* not inflecting like a verb.

#### 4.1 Frequency and position

In our own corpus, we counted the occurrences of *vika* compared to *reče*.<sup>4</sup> With 55 occurrences, *reče* is more frequent than *vika* (25 occurrences). Table 1 shows the number of tokens for *vika* and *reče* respectively in the corpus.<sup>5</sup> In Table 1 the label ‘before DR’ refers to all cases of the verb occurring before the discourse report (DR) starts. ‘Second in DR’ means that the verb appears after the first phonological word within the discourse report. All other cases of the verb appearing within a discourse report, but not after the first phonological word, are captured with the label ‘medial in DR’. This could be after the second or after the third or fourth phonological word, these cases are conflated.

Table 1: Position of *vika* and *reče*.

	<i>vika</i>	<i>reče</i>
before DR	13	5
second in DR	6	31
medial in DR	4	9
after DR	2	10
<b>total</b>	<b>25</b>	<b>55</b>

As was illustrated with the Chuvash example (4a) in §3, quotatives and quotative verbs can occur multiple times within one discourse report, whereas speech verbs usually only occur once. Furthermore, Table 1 also illustrates in which position *reče* and *vika* most often occur. We see that *vika* mostly occur before the discourse report, as illustrated in (10).

- (10) На сестра му вика: “Ѓерданчето кај ти је?”  
 na sestra =mu vika ġerdan-če-to kaj  
 to sister =POSS.3SG say.IPFV.PRS.3SG necklace-DIM-DEF.N where  
 ti= je  
 2SG.DAT= be.PRS.3SG  
 ‘To his sister, he says: “Where is your necklace?”’ siljan\_the\_stork.136-7

Canonically, speech verbs in Macedonian precede the discourse report. Other verbs that are used with discourse reports in our corpus are *znae* ‘know’, *prašuva*

<sup>4</sup>*Velj* did not occur at all. There were a few instances of some other speech verbs, such as *kaže* ‘tell’.

<sup>5</sup>This includes inflected forms such as *rekle* ‘say.AOR.LPTCP.PL’ and *vikaat* ‘say.IPFV.PRS.3PL’.

'ask', *se raduva* 'rejoice', *dojde* 'come', *veruva* 'believe' and *kaže* 'tell'. These always occur before the discourse report.

*Reče*, on the other hand, most frequently occurs in the second position of the discourse report, i.e., after the first phonological word. This is illustrated in (11), where *reče* is placed after the first phonological word, *dajte* 'give'.<sup>6</sup>

- (11) “Дайте,” рече, “ваму една канта.”  
 daj-te      reče              vamu edn-a kanta  
 give.IMP.PL say.PFV.3SG here one-F bin  
 “Give me,” he said, “a bin.” simpletons.39

The tendency of *reče* to appear after the first phonological word suggests that it is grammaticalizing into a second-position clitic. An example of a second-position clitic in Macedonian is *li*, a focus clitic used mostly in polar questions. Consider (12). Clitics also appear in second position when hosted by an imperative, adjective or predicative nominal (Tomić 1996).

- (12) a. Мусли ли сакаш?  
 Musli      =li              saka-š?  
 muesli=LI want-PRS.2SG  
 ‘Do you want MUESLI?’  
 b. Сакаш ли мусли?  
 Saka-š              =li      musli?  
 want-PRS.2SG=LI muesli  
 ‘DO you want muesli?’  
 c. \*Сакаш мусли ли?  
 \*saka-š              musli =li?  
 want-PRS.2SG muesli =LI  
 Intended: ‘Do you want muesli?’  
 (based on the examples in Rudin et al. (1999: 579))

Compare (11) also to the Chuvash example in (13), in which the quotative verb *tet* appears after the first phonological word, *esir* ‘you’.

<sup>6</sup>*Reče* is not glossed for tense in our data, because the form *reče* is syncretic between present and past tense. While most of the verbs in the narratives are in past tense, other speech verbs, such as *vika*, are in the present tense. Thus, it is not clear whether the present or past form of *reče* is used in the narratives.

- (13) esir **t-et** jəvan-ən aʃʃə-p[e] aməʃ=i?  
 2PL QV-PRS.3 Ivan-GEN father+POSS.3-INSTR mother+POSS.3=Q  
 ‘You are, he says, Ivan’s father and mother?’ (Nikitina et al. 2021: 9)

According to Nikitina et al. (2021) *tet* in Chuvash has a tendency to appear after the first phonological word, because it is grammaticalizing into a second-position clitic.

This section has shown that *reče* is more frequent than *vika* and that it more often occurs in second position than *vika*, making it similar to the quotative verb *tet* in Chuvash (Nikitina et al. 2021). In the next subsection we look at instances of *vika* and *reče* appearing together in the same discourse report.

## 4.2 Co-occurrence

There are some instances of *vika* and *reče* co-occurring. This is exemplified in (14).

- (14) Вика на жената “сеа,” рече, “чекај, ја ќе му кажам нив,” рече, “друг пат!”  
**vika** na žena-ta sea **reče** čekaj ja  
 say.IPFV.PRS.3SG to wife-DET.F now say.PFV.3SG wait.IMP.SG 1SG.NOM  
 ќе= mu= kaža-m niv **reče** drug pat.  
 FUT= 3SG.M.DAT= tell.PFV-PRS.1SG 3PL.DAT say.PFV.3SG other time  
 ‘He told his wife: “Now, wait,” he said “I will show them!”’  
 Role\_and\_Rolejca.023-24

In (14) we see that *vika* occurs before the discourse report, followed by two instances of *reče* within the discourse report. The first *reče* appears in the second position, after *sea* ‘now’. The second *reče* appears after *niv*, the first phonological word in the embedded clause.

The co-occurrence of two speech verbs within one discourse report often means that one of them has taken up a different function, such as a quotative function. Recall from §3 that quotatives can appear with speech verbs, this was illustrated with the Bashkir example in (5). Another example of this phenomenon is the West African Pidgin English example in (15).

- (15) Mása **tók sey**, kom-ow.  
 master talk (say) come-?  
 ‘The master said, “Come here”.’  
 West African Pidgin English (Lord 1990: 332)

In (15) the speech verb *tók* 'say' co-occurs with the quotative marker *sey*, which itself is grammaticalized from the speech verb *sey* 'say'.

The total number of examples of co-occurring speech verbs in our corpus is shown in Table 2.

Table 2: Co-occurrence of *vika* and *reče*.

verbs	frequency
<i>vika+vika</i>	1
<i>reče+reče</i>	1
<i>reče+vika</i>	4
<i>2reče+vika</i>	2
<b>total</b>	<b>8</b>

(14) is an example of what is labeled as *2reče+vika* in Table 2: *vika* co-occurring with two instances of *reče*. An example of *vika* with only one instance of *reče*, labeled in Table 2 as *reče+vika*, is given in (16).

- (16) Сеа ѝ вика на жената, “сеа,” рече, “жено, коа ќе дојда ја сега тебе ќе те колам.”
- sea ì=            **vika**                    na zena-ta            sea reče  
 now 3SG.F.DAT= say.IPFV.PRS.3SG to woman-DEF.F now say.PFV.3SG  
 žen-o            koa ќе= dojd-a(t)            ja            sega тебе ќе=  
 woman-VOC when FUT= come.PFV.PRS.3PL 1SG.NOM now 2SG.ACC FUT=  
 te=            kola-m  
 2SG.ACC= slaughter.IPFV-PRS.1SG  
 ‘Now he tells his wife: “Now,” he said, “wife, when they come I will slaughter you.”’
- Role\_and\_Rolejca.176

An example of two occurrences of *reče* with no other speech verb, labeled as *reče+reče*, is shown in (17).<sup>7</sup>

<sup>7</sup>There is also the example labeled as *vika+vika*, this is example (i).

- (i) На жената му, “кај ти е,” вика, “тебе чорапот,” вика, “со иглите.”
- na žena-ta =mu            kaj ti=            e            **vika**                    tebe  
 to wife-DEF.F =POSS.M.3SG at 2SG.DAT= be.PRS.3SG say.IPFV.PRS.3SG 2SG.DAT  
 čorap-ot **vika**                    so igli-te  
 sock-DEF.M say.IPFV.PRS.3SG with needle.PL-DEF.PL  
 ‘To his wife, “where,” he says, “is your sock with the needles?”’
- siljan\_the\_stork.130

- (17) “Како,” рече, “ние, ние сме,” рече, “штркоите шо ги има на племната твоја.”  
 kako reče        nie        nie        sme        reče        štrk-oi-te  
 how say.PFV.3SG 1PL.NOM 1PL.NOM be.PRS.PL say.PFV.3SG stork-PL-DEF.PL  
 šo gi=        ima                      na plemna-ta tvoja  
 REL 3PL.ACC= have.IPFV.PRS.3SG on barn-DEF.F POSS.2SG  
 “‘What,’ [the stork] said, ‘we, we, are the storks that live on your barn.’”  
 siljan\_the\_stork.031

Finally, examples of *reče* co-occurring with other speech verbs are also found in the MANU recordings (Vidoeski et al. s.a). An example similar to the ones in our corpus, namely with *vika* and *reče*, is given in (18).

- (18) И змијата му вика од смреката на овчарот: “извади ме,” рече, “што сакаш бакшиш ќе ти дадам.”  
 I zmiја-ta        му=        vika                      од смрека-ta        на  
 and snake-DEF.F 3SG.M.DAT= say.IPFV.PRS.3SG from juniper-DEF.F to  
 овчар-от        из-вади                      =me        reče        што  
 shepherd-DEF.M PFV-take.out.IMP.SG =1SG.ACC say.PFV.3SG what  
 saka-š                      bakšiš        ќе=        ti=        dada-m.  
 want.IPFV-PRES.2SG present FUT= 2SG.DAT= give.PFV-PRS.1SG  
 ‘And the snake calls out to him from the juniper, to the shepherd, “take me out of here,” it says, “and I will give you a gift.”’  
 Šiševo<sup>8</sup>

While *vika* can co-occur with *reče*, we have not seen it co-occurring with other

---

In (i) we see that *vika* occurs twice in the discourse report, once between *kaj ti e tebe* ‘where is your’ and once between *čorapot so iglite* ‘the sock with the needles’. The overt indirect object of *vika*, *na ženata mu* ‘to his wife’ is also not adjacent to the verb. This illustrates that the constituency of discourse reports and discourse reporting events is generally more free in Macedonian than in English. Another example that shows this is (ii).

- (ii) “Абе, бе, од кај ме знаете,” рече, “вие, мене,” Силјан.  
 abe be od kaj me=        znai-te                      reče        vie        mene  
 INTRJ INTRJ from where 1SG.ACC= know.IPFV.PRS-2PL say.PFV.3SG 2PL.NOM 2SG.ACC  
 Siljan  
 Siljan  
 “‘Hey, but, where do you know me from,’ said Siljan.”  
 siljan\_the\_stork.030

In (ii) the subject of *reče*, *Siljan*, is not adjacent to the verb. Further details on this phenomenon are outside of the scope of this paper.

<sup>8</sup><http://damj.manu.edu.mk/audio/mp3/02-AudioTrack-02.mp3>

speech verbs, such as *kaže* 'tell'. We have only observed *reče* with other speech verbs, such as *odgovara* 'answer' in (19) and *kažuva* 'tell' in (20).

- (19) Петелот одговара: “па нека умре,” рече, “да му ја ебам мајката.”  
 Petel-ot            **odgovara**                    pa    neka    umre                    **reče**  
 rooster-DEF.M answer.IPFV.PRS.3SG well OPT die.PFV.PRS.3SG say.PFV.3SG  
 da=    mu=                    ja=                    eba-m                    majka-ta.  
 SBJ= 3SG.M.DAT= 3SG.F.ACC= fuck.IPFV-PRS.1SG mother-DEF.F  
 ‘The rooster answers: “Well, let him die,” he said, “fuck it.”’                    Šiševo

In (19) *reče* is combined with *odgovara* 'answer'. (18) and (19) are from the same speaker, (20) is from a different speaker.

- (20) Сретнува го и му кажува рече човек ја ќе ти ја земам душата.  
 Sretnuva                    =go                    i    mu=                    **kažuva**  
 meet.IPFV.PRS.3SG =3SG.M.ACC and 3SG.M.DAT= tell.IPFV.PRS.3SG  
**reče**                    čovek    ja    ќе=    ti=                    ja=                    zema-m  
 say.PFV.3SG person 1SG FUT= 2SG.DAT= 3SG.F.ACC= take.PFV-PRS.1SG  
 duša-ta  
 soul-DEF.F  
 ‘He meets him and tells him, he said: “Human, I will take your soul.”’  
 Bučište<sup>9</sup>

While in (18), (19) and (20) *reče* co-occurs with another speech verb, the position of *reče* differs across the examples. In (18) follows the first phonological word of the discourse report, like a clitic. In (19) it is not after the first phonological word, but still within the discourse report, whereas in (20) it precedes the discourse report (and immediately follows the other speech verb). The same speaker who produced (20) produced an example with *reče* coming immediately after *moli* 'beg'. Thus, there is still substantial variability in the placement of *reče*, as it is not fully grammaticalized. In the next subsection we look at a final piece of grammatical evidence before we move on to phonetic arguments for the grammaticalization of *reče*: the loss of inflection.

### 4.3 Inflection

In this subsection we show that *reče*, when used as a quotative, loses its inflectional morphology. This loss is a characteristic of grammaticalization (Hopper & Traugott 2003).

<sup>9</sup><http://damj.manu.edu.mk/audio/mp3/24-AudioTrack-24.mp3>

The form *reče* is the third person singular. Since Macedonian doesn't have an infinitive, the third singular form, which is usually the bare stem, is also the citation form. When *reče* remains invariant, it can create a mismatch – on the level of number, tense or aspect – with the previous verb that would not arise if *reče* were used as a full verb.

Every use of *reče* in combination with *vika* is an example of a mismatch in aspect: *vika* is imperfective, whereas *reče* is perfective. We have seen in the previous subsection that both verbs are used in combination to describe the same discourse reporting event. Similar examples are (19) and (20), also shown in the previous section, in which *odgovara* and *kažuva* respectively are both imperfective stems.

We have also seen an example of this mismatch in (8) from Mushin (2000), shown in §3, in which *reče* was used together *rekla*, the feminine *l*-participle, creating a mismatch in evidentiality marking. We claim that in (8) only *rekla* is actually used as a verb; *reče* is used as a quotative.

In this section we show additional examples which illustrate that the mismatch can also be on the level of number. Compare (21), where *reče* occurs in a third plural *l*-participle form, *rekle*, with (22), where *reče* doesn't match in number or TAM-marking with the preceding verb, *otišle* 'they went'.

- (21) Го стретнале на патот и му рекле: “Каж си тргнал, Роле?”  
 go=            sretna-l-e            na pat-ot            i    mu=  
 3SG.M.ACC= meet-AOR.LPTCP-PL on road-DEF.M and 3SG.M.DAT=  
 rek-l-e            kaj    si=    trgna-l            Role?  
 say-AOR.LPTCP-PL where REFL= set.off-AOR.LPTCP.M Role  
 ‘They met him on the road and told him: “Where are you headed to, Role?”’  
 Role\_and\_Rolejca.010

In (21) there is no mismatch between the verbs *sretnale* 'they met' and *rekle* 'they said': both verbs are in the third person plural aorist form.

- (22) Му отишле, “сеа,” рече, “мамата негова!”  
 Mu=            otiš-l-e            sea    reče            mama-ta            negov-a!  
 3SG.M.DAT= go-AOR.LPTCP-PL now say.PFV.3SG mother-DEF.F POSS.3SG-F  
 ‘They went to him, “Now,” he said, “He’s done!”’    Role\_and\_Rolejca.102

In (22), however, there is a mismatch: while *otišle* is in the third plural aorist, *reče* remains its invariant third singular. Thus, we see that there can even be a mismatch in number, *reče* remains singular, even though *otišle* is plural. Were *reče*

to match the tense, aspect and number of *otišle*, it would become *rekle*, as in (21). Thus, *rekle* in (21) is a real speech verb, whereas *reče* in (22) is a quotative. In total, there were 16 cases such as (22) – i.e., cases in which the singular form *reče* occurs with a plural subject – in our corpus. This is out of a total of 49 occurrences of *reče* in the third singular citation form. Table 3 shows how often *vika* and *reče* respectively occur in the citation form.

Table 3: Inflection of *vika* vs *reče*.

	inflected	3sg/citation form	total
<i>vika</i> before DR	2	11	13
<i>vika</i> other positions	2	10	12
<i>reče</i> before DR	3	2	5
<i>reče</i> other positions	3	47	50

Both verbs most frequently occur in the citation form, because the speaker is usually a third person. Nonetheless, we see that when *reče* occurs before the discourse report, three out of five times it is inflected. While *reče* in other positions is much more frequent, the amount of times it is inflected in those cases is not higher. This shows that in the canonical speech verb position, before the discourse report, *reče* is used as a verb and gets inflected, whereas in other positions there are some cases in which it remains invariant. For example, in (23) *rečiš* ‘you say’, which is both inflected and occurs before the DR, refers to a hypothetical speech report, which is normally introduced by overt verbs of speaking.<sup>10</sup>

- (23) “И сеа,” рече, “коа ќе дојдат, жено [...] ќе речеш “Роле е на орање.””  
i sea reče koa ke dojd-at ženo [...] ke  
and now say.PFV.3SG when FUT come.PFV-PRS.3SG woman.VOC FUT  
reči-š Role e na ora-nje  
say.PFV-PRS.2SG Role be.PRS.3SG on plow-NMLZ  
“‘And now,’ he said, ‘when they come, wife, you will say ‘Role is out  
plowing.’”” Role\_and\_Rolejca.112-4

For *vika*, on the other hand, Table 3 shows that the distribution of inflected vs. citation form is equal, regardless of whether *vika* occurs before the discourse report or not.

<sup>10</sup>Thank you Tatiana Nikitina for pointing this out to us.

#### 4.4 Intermediate summary

In this section we have presented four arguments in favour of the hypothesis that *reče* is grammaticalizing into a quotative clitic: i) *reče* can occur multiple times throughout a discourse report. ii) it frequently occurs within the discourse report following the first phonological word, iii) it co-occurs with other speech verbs and iv) it remains invariant. In the next section we show phonological and phonetic arguments: in production, *reče* is often reduced to *eče*, *rče*, *reč* or *če*.

### 5 Phonetic properties of *reče*: An exploratory corpus study

#### 5.1 Introduction

In this section we explore a small corpus of recordings to study the behavior of stressed and unstressed vowels in Macedonian to determine if the first syllable of *reče* is undergoing a phonetic loss of stress. To do so, we first need to understand the behavior of stress in Macedonian.

#### 5.2 Stress in Macedonian

The Western dialects of Macedonian have fixed stress placement in the antepenultimate syllable (Lunt 1952). Stress is characterized by variation in lengthening, changes in intonation and volume, and sometimes phonetic changes in spectral features of vowels.

Clitics are unstressed and attach to phonological words usually modifying stress patterns to conform to antepenultimate accent.<sup>11</sup> Sometimes, this phenomenon responds to free variation, as with the enclitic pronouns *ti* ‘you’ and *go* ‘it’ in (24):

- (24) /e'vetigo/ <sup>12</sup>  
/'eve 'tigo/  
Eve ti go  
here 2SG.DAT 3SG.ACC.M  
‘Here he/it is for you’. (Friedman 2001: 13)

---

<sup>11</sup>These prosodic words are also referred to as accentual units in the literature (Rudin et al. 1999; Alexander 1994)

<sup>12</sup>For the sake of simplicity, we decided to represent the mid vowels of Macedonian as /e/ and /o/ instead of /ɛ/ and /ɔ/ respectively, as in most of the literature (Lunt 1952; Friedman 2001). There is no phonological contrast between /e/ and /ɛ/, or /o/ and /ɔ/.

The negation *ne* and the conjunction *i* can be stressed or unstressed, depending on the context or the dialect (Rudin et al. 1999). An example of (dialectal) variation with *ne* ‘no’ from Alexander (1994) is shown in (25):

- (25) a. *ne GO gledam*  
 NEG 3SG.M.ACC see-PRS.1SG  
 ‘I don’t see him.’ Western pronunciation
- b. *NE go GLEdam*  
 NEG 3SG.M.ACC see-PRS.1SG  
 ‘I don’t see him.’ Central pronunciation

In (25a), *ne* is unstressed and part of the prosodic word together with *go* and *gledam*, with the antepenultimate stress falling on *go*. In (25b), *go* enclitizes to *ne* and forms a prosodic word, with *gledam* forming a separate prosodic word. Since both prosodic words are disyllabic, the stress falls on the first syllable. In our data, we found that the speaker used both patterns, as in (25a) and (25b), in seemingly free variation. Thus, we manually inspected all of the instances of *reče* to determine which ones were stressed and which not.

### 5.3 Variation in the production of *reče*

We have noticed that in some of the occurrences of quotative *reče* the first phone is sometimes omitted, becoming *eče*, and in other cases the complete first syllable is dropped, becoming *če*.

Nevertheless, complete elisions of the first phone or the first syllable of a full verb, or of a non-clitic word, are not attested in any grammar. However, we have seen cases of an inverse process. Final syllable synaeresis is common in connected speech, i.e. the last unstressed syllable of a word can be resyllabified with the first stressed syllable of the next word in CV#V contexts, becoming #'CVV (Savicka et al. 2021: 195). For example, a sequence like [ʒo] *срамеле и си отишол* ‘shamed [him] and he left’, pronounced as /'sramele 'i 'si 'otifol/ in careful speech, can be pronounced as [ˈsrame ˈlej̥ ˈsi ˈotifol] in fast speech. In some cases, the medial vowel can even be elided: [ˈsrame ˈli ˈsi ˈotifol], as it was pronounced in our recording. Another illustration of this phenomenon can be seen in Figure 2 below, in which *пече оту* ‘said because’ becomes [ˈretʃˈoti].

To rule out the possibility of the reduction of *reče* being idiolectal and specific to the speaker in our data, we extended our analysis to productions of other speakers in data obtained from the MANU corpus (Vidoeski et al. s.a) and found further evidence of elision of the initial phones of *reče*. This elision appears to

happen in different degrees, but at this stage it is unclear whether this happens because of pragmatic, idiolectal or phonological reasons.

In Figure 2, initial /r/ is elided in the same way as attested in our data. In Figure 3, it is ambiguous whether the elision concerns only /r/ or the whole first syllable, but acoustic data – the presence of two intensity peaks and the long duration of the vowel – seems to point towards there being two consecutive /e/, an unstressed one followed by a stressed one. The elision in this case might also be attributed to connected speech hapology. In Figure 4, however, the whole first syllable is elided and in Figure 5 the first stressed /e/ is elided while /r/ is still uttered, but attached to the coda of the previous syllable.<sup>13</sup>

We hypothesize that these occurrences are evidence that *reče* is being grammaticalized and adopting the morphophonological features of clitics in Macedonian. Figure 2 and Figure 4 are both from the same speaker from Trpejca<sup>14</sup>, Figure 3 is from the speaker from Šiševo and Figure 5 is from the speaker of our original recordings, from Sekirci.

---

<sup>13</sup>Waveform and spectrogram representations were made thanks to the *create\_pictures-with-tiers* Praat script made by Wendy Elvira at the Laboratori de Fonètica (Universitat de Barcelona) (2017)

<sup>14</sup><http://damj.manu.edu.mk/audio/mp3/06-AudioTrack-06.mp3>

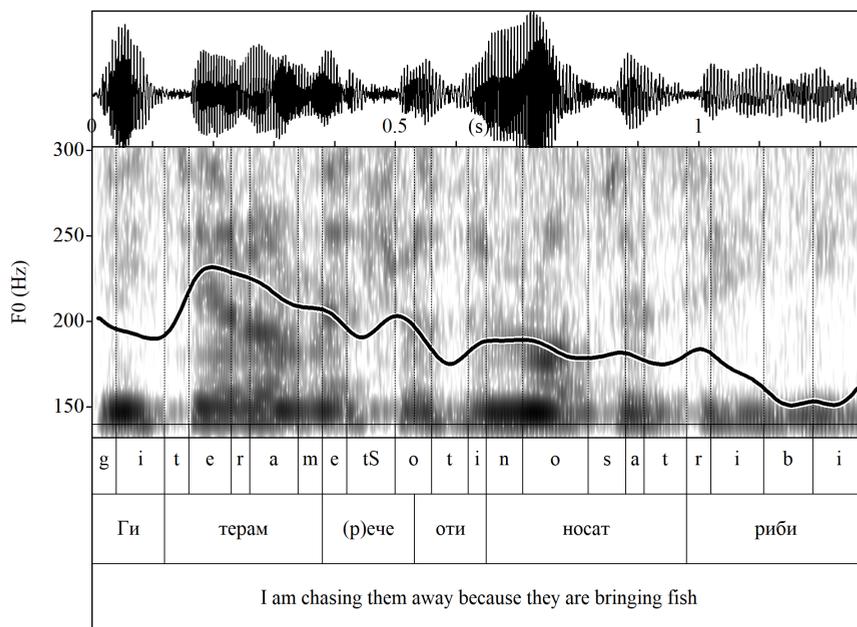


Figure 2: /r/ deletion in reče (Trpejca)

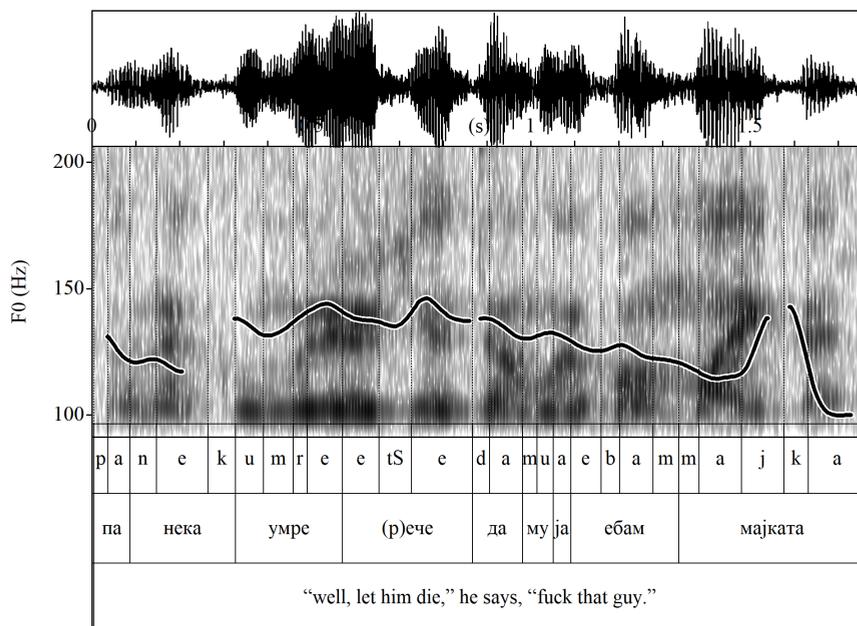


Figure 3: /r/ deletion in reče (Šiševo)

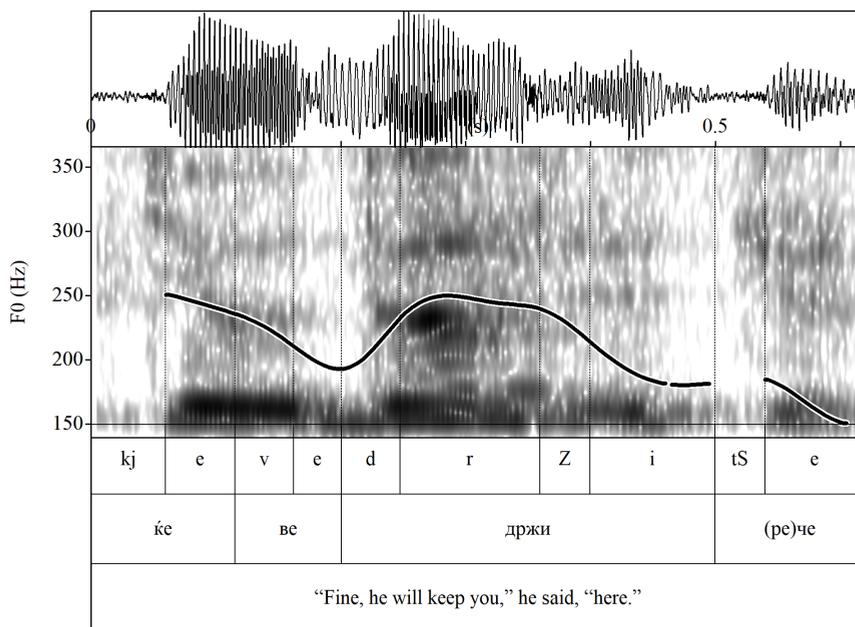


Figure 4: /re/ deletion in reče (Trpejca)

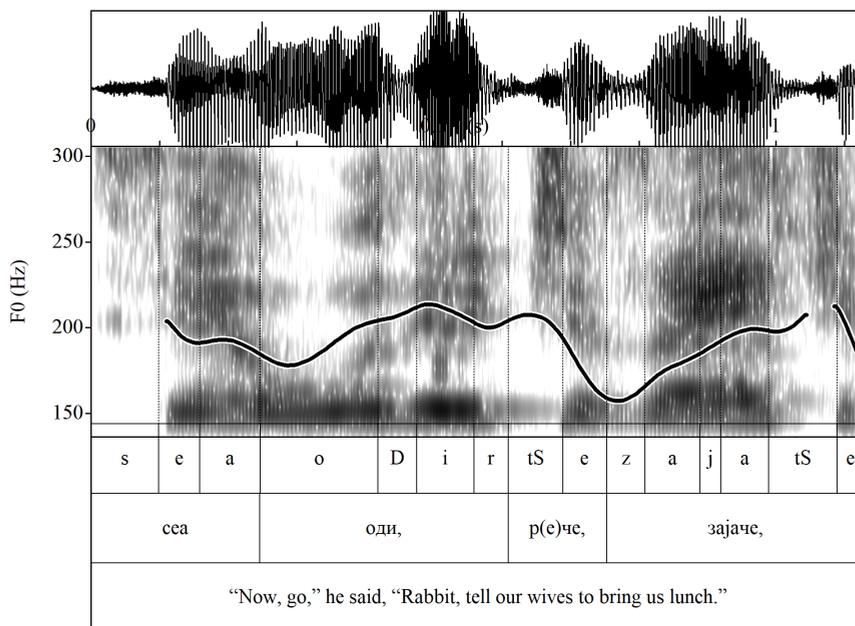


Figure 5: Deletion of first /e/ in reče (Sekirci)

Morphophonological reduction through elision of a stressed syllable like the one seen in Figure 4 is not unheard of in grammaticalization processes. We have seen in §3 that quotative particles are often reduced to monosyllabic forms from polysyllabic verbs, such as Russian *grit* from *govorit* ‘says’. This type of reduction is not restricted to quotatives, but is a common feature of grammaticalization. In Macedonian, for example, a similar process can be seen with the development of the future clitic *ќе* from the Proto-Slavic verb *xotěti* ‘to want’ (Tomić 2004). An example of a future construction from the 14th century, with the modal verb *xoščet*, is shown in (26a), whereas a parallel construction from synchronic standard Macedonian, with the clitic *ќе*, is shown in (26b).

- (26) a. **Xoščet** počiti moj brat.  
will.3SG die.INF my brother  
‘My brother will die.’ 14th Century,  
as cited in Koneski (1965: 204)
- b. **Ќе** umre moj-ot brat.  
FUT die.PFV-PRS.3SG POSS.1SG-DEF.M brother  
‘My brother will die.’ Modern Macedonian

While in Macedonian the word for ‘want’ is replaced by a different stem, Bosno-Serbo-Croatian has a cognate with both a full verb, *hoće* ‘wants’, and an auxiliary, *će* ‘will’, whereby the stress in *hoće* falls on the first syllable (Tomić 2006). Thus, while the deletion of a stressed syllable in *reče*, is unusual, it is not unprecedented.

It is interesting that reduction manifests itself either through the elision of the whole first syllable of *reče*. We would expect the second, i.e. unstressed syllable would become elided, since it already happens as a synaeresis phenomenon in other contexts in connected speech. A hypothesis could be that the first syllable is losing its stress.

Unlike, for example, Russian, English, or Catalan, Macedonian does not present phonological vowel reduction in unstressed syllables (Savicka 2014: 48). However, a non-categorical phonetic reduction can happen which can be observed acoustically at a duration and F1-F2 level. To explore this hypothesis, we analyzed the spectral and durational features of stressed and unstressed vowels and compared them with those of the first syllable of quotative and non-quotative *reče* to determine if phonetic reduction patterns are present in the quotative uses of the word, which would support our claims of an ongoing grammaticalization process.

## 5.4 Methodology

We analyzed a corpus of seven recordings of tales produced by five different speakers from five different locations: Sekirci, Brodec, Lukovo, Bitola and Šiševo, with Sekirci being the speaker of our own recordings and the rest from the MANU collection (Vidoeski et al. s.a).<sup>15</sup> Recorded data amounted to a total of 31 minutes and 04 seconds. Recordings were transcribed orthographically and force-aligned to the signal with WebMAUS (Kisler et al. 2017).<sup>16</sup> Stressed, unstressed and pre-pausal vowels were then labelled and their duration (in seconds) and midpoint F1 and F2 values were extracted (in Hertz, using frequency ceilings of 5000 Hz for low-pitched voices and 5500 Hz for high-pitched voices). A filter was used to discard outlier formant values. For the sake of comparison, the remaining values were normalized using the Lobanov method (Lobanov 1971) to discard inter-speaker, physiologically caused differences. We decided to use segmental cues like formants and one suprasegmental cue for stress (duration, and not intensity or fundamental frequency) due to the low audio quality of some of the recordings.

In order to assess the degree of phonetic vowel reduction, F1 and F2 values of stressed and unstressed vowels (/a/, /e/, /i/, /o/, and /u/) were plotted for visual analysis. Then, Pillai scores were calculated in order to evaluate the overlap of the distribution of F1 and F2 values of stressed and unstressed versions of all vowels (Nycz & Hall-Lew 2014; Hall-Lew 2010). Pillai scores have been used in sociophonetic research to assess the emergence of vowel splits or mergers. It is a score that varies from 0 to 1, 0 being complete a overlap between the distribution of two groups of vowels and 1 being no overlap at all. Pillai scores are obtained as output of MANOVA tests and can be calculated with two dependent variables (F1+F2)~vowel, three dependent variables (F1+F2+duration)~vowel, or more. For this study, we first inspected the distribution of stressed and unstressed vowels in a two-dimensional space (F1 and F2), then in a three-dimensional one, to separately assess the impact of duration.

---

<sup>15</sup> All audio material from MANU can be found here: <http://damj.manu.edu.mk/audio/>

<sup>16</sup> WebMaus does not provide a language-specific phonetic alignment system for Macedonian. However, based on similarities with the vowel and consonant inventory of Macedonian, the Spanish language transcription was used. This technique is recommended for phonologically similar languages (e.g. using Polish for Slovak), since WebMAUS uses a machine-learned probabilistic pronunciation model which can predict the transcription of sounds not present in the vocabulary database of the selected language. In our experience, the performance of Spanish WebMAUS for Macedonian was considerably good and transcription divergences, when present, were manually corrected.

## 5.5 Results

Table 4: Different realizations of *reče*.

form	frequency
<i>reče</i>	44
<i>(r)eče</i>	11
<i>(re)če</i>	5
<i>r(e)če</i>	2
<i>reč(e)</i>	1
<i>rečiš</i>	1
<i>rekle</i>	4
<i>rekla</i>	5
<i>rekol</i>	10
<b>total</b>	<b>78</b>
<b>of which reduced</b>	<b>19</b>

Full *reče* is the most common occurrence, as seen in Table 4, but cases of elisions were also common across some speakers, especially in the first syllable, as illustrated in Figure 2, Figure 3, Figure 4, and Figure 5.

If we observe stressed and unstressed vowels, we see that they show slightly different means – as illustrated in Figure 6 – and tend to become mid-centralized when unstressed. This difference is most marked in /a/. Pre-pausal vowels also present different formant value means although not always in the same direction: mid and close vowels are lowered, whereas the open vowel /a/ is raised. Although always unstressed as per the antepenultimate fixed stress of Macedonian, pre-pausal vowels are characterized by having a significantly longer duration due to final lengthening at the end of prosodic units, as shown in Figure 7.

When pronounced, stressed and unstressed /e/ in *reče* follow similar trends as all other vowels, both for formant mean values and duration. Furthermore, durational differences between stressed and unstressed /e/ proved to be statistically significant for all occurrences of this vowel as well as for its specific occurrences in *reče* (cf. Figure 8).

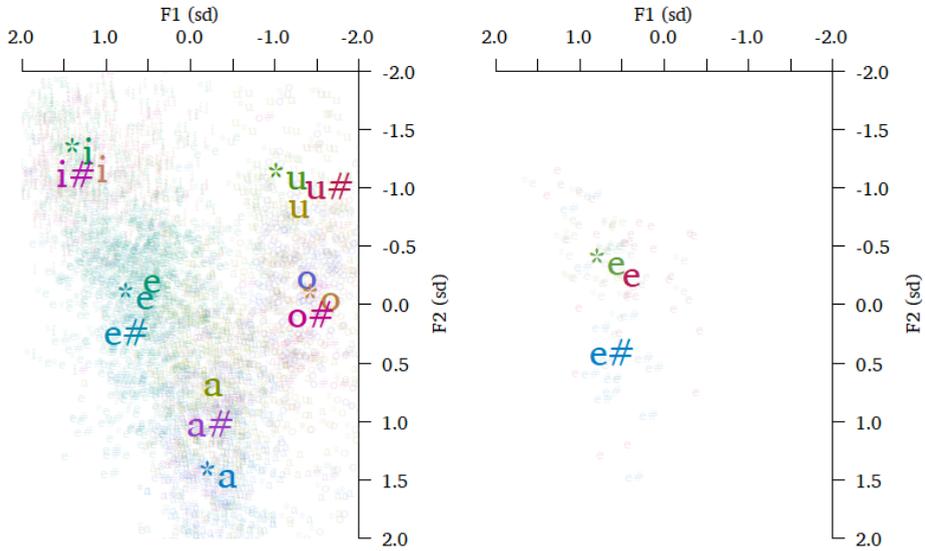


Figure 6: Mean values of F1 and F2 in stressed (\*V), unstressed (V), and pre-pausal (V#) vowels for: all vowels (left), vowels in *reče* (right).

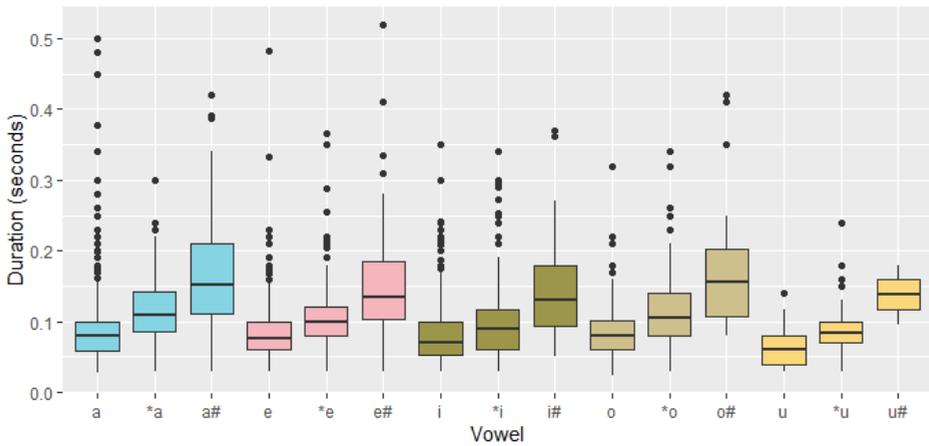


Figure 7: Durations of unstressed (V), stressed (\*V) and pre-pausal (V#) vowels.

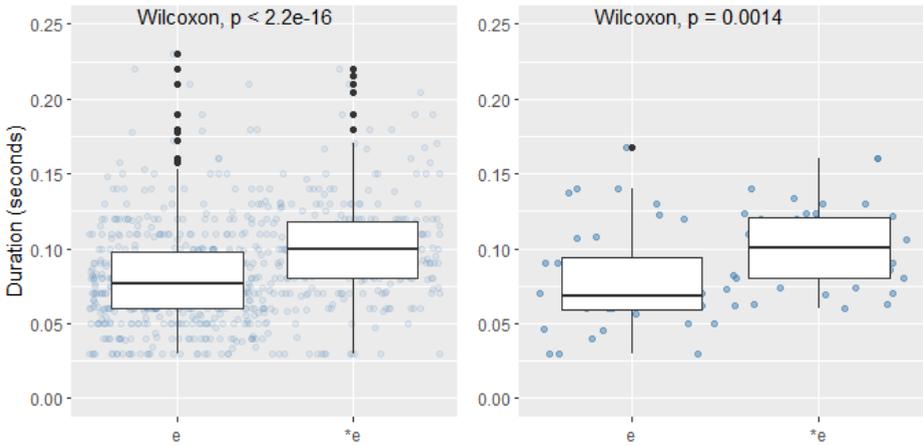


Figure 8: Overall durations of unstressed and stressed /e/ (left) vs. durations of unstressed and stressed /e/ in *reče* (right).

Table 5: Pillai score means for stressed and unstressed vowel pairs produced by 5 speakers, computed within a F1+F2+duration~vowel MANOVA.

	F1+F2~vowel <i>Pillai score</i>	F1+F2+duration~vowel <i>Pillai score</i>
/a/ vs /'a/	0.23	0.25
/e/ vs /'e/ <sup>17</sup>	0.05	0.09
/i/ vs /'i/	0.05	0.06
/o/ vs /'o/	0.05	0.13
/u/ vs /'u/	0.04	0.18
/e/ vs /'e/ in <i>reče</i>	0.08	0.19

The very low averaged Pillai scores statistically confirm what is descriptively observable in the vowel charts above, namely that i) vowels tend to be distributed rather homogeneously regardless of their stressed or unstressed quality, and ii) that, when uttered, stressed and unstressed /e/ in *reče* follow similar patterns as all other vowels. The effect of duration contributes to augmenting the split between stressed and unstressed vowel but at a very low degree, especially in the most frequent vowels in Macedonian, /e/ and /a/.

<sup>17</sup>Excluding cases of /e/ in *reče*.

An interesting indirect finding that came to our attention was that /a/ presented the highest Pillai scores degree of reduction across all speakers, reaching a score as high as 0.5 for one speaker. However, since they are a measure of numerical distance rather than perceptually significant one, there are no established Pillai score thresholds for determining whether these two groups of vowels are likely to become two distinct phonological categories. These values are to be considered as measures of “abstracted *difference*” (sic) (Nycz & Hall-Lew 2014: 5).

These findings confirm that, when not elided, the initial stressed /e/ in *reče* behaves like any other stressed /e/, i.e., it does not become phonetically reduced nor phonologically unstressed.

## 6 Discussion

### 6.1 Interpretation of results

We see no evidence of there being an intermediate stage in which *reče* loses stress while remaining disyllabic. We interpret these findings as showing that, during its process of grammaticalization and in order to assimilate with other clitics, *reče* is more likely to lose an initial consonant or a whole initial stressed syllable rather than only losing stress while remaining disyllabic. This might hint towards there being a hierarchy of the morphophonological features prone to diachronic reduction through grammaticalization.

Thus, *reče* might be experiencing a process analogous to the one undergone by *hoće* to *će* in Bosno-Serbo-Croatian – i.e., in which a stressed syllable is elided – but different from the one undergone by *govorit* to *grit* in Russian – in which the stressed syllable is the only one that remains. We speculate that this is related to the morphophonological features of clitics in Macedonian, which tend to be unstressed syllables.

Furthermore, the use of the quotative is very informal. It can be found in narratives, but not all recorded narratives in the MANU corpus (Vidoeski et al. s.a) contained it. Quotative *reče* is also restricted to spoken language and can, to our knowledge, not be found in written sources. Even folktales that have been written down and in which colloquial variants of the language are used – such as the ones found in the MANU corpus<sup>18</sup> and in the grammar by (Lunt 1952: 105-11) – the quotative use of *reče* cannot be found.

All in all, our findings point towards the conclusion that *reče* is in early stages of grammaticalization: there is still a high degree of variation in how it is pro-

---

<sup>18</sup>Texts from MANU can be found here: <http://ical.manu.edu.mk/Map/Map.html>

nounced: *reče*, *eče*, *če* or *rče*. We have shown that also syntactically and morphologically, *reče* only sometimes behaves as a quotative, while sometimes it still behaves as a prototypical speech verb.

## 6.2 Two quotatives?

Finally, it should be noted that not all speakers have the quotative *reče*, there seems to be sizable inter-speaker and dialectal variation. For comparison, we examined the data in the Pangloss Collection (Michailovsky et al. 2014), recorded in 1976 by Georges Drettas in Voden (Edessa), Greece.<sup>19</sup> The dialect spoken there belongs to the Kostur-Korča group in Friedman's (2001) classification, which in Vidoeski et al. (2003); Marković (2007) is a subset of the Southeastern dialects. In these data there was no instance of *reče* being used as a quotative. It was much less frequent than the imperfective stem *veli* 'says'. There were only 19 occurrences of *reče* and 106 of *veli*. Out of those 19 tokens of *reče*, it occurred before the discourse report 13 times and was mostly inflected (for example, *ričeja* '(they) said'). Thus, *reče* seems to behave as a real speech verb in that dialect.

*veli*, on the other hand, did occur often throughout reported speech, as shown in (27) (Drettas uses a phonological transcription rather than the standard orthography, hence 'velə for *veli*).

- (27) i 'čorbičkata mu 'velə se a isə'rkax i  
and soup-DIM-DEF.F 3SG.DAT say.3SG all 3SG.F.ACC PFV-SCOOP-AOR.1SG and  
pri'kazničkata mu 'velə 'sakam da se a  
story-DIM-DEF.F 3SG.DAT say.3SG want.IPFV-PRS.1SG SBJ all 3SG.F.ACC  
'čvam mu 'velə  
hear-PRS.1SG 3SG.DAT say.3SG  
'The soup, I've eaten it all,' he said, 'and I also want to hear the tale to the  
end.' trois\_pommes\_S117

This is similar to the example from Mushin (2000), (9). Recall from §1 that Mushin (1997) has found that *veli* behaves like a hearsay marker. The MANU corpus also contained recordings that contained multiple instances of *veli* throughout discourse reports, in a similar way to how *reče* is used in the examples shown in §4. Thus, there is variation as to whether *veli* or *reče* takes on a quotative role. Whether this difference is idiolectal or dialectal remains to be seen. We found no instance, however, of a speaker using both *veli* and *reče* as a quotative. As has

<sup>19</sup>The corpus can be found here: <https://pangloss.cnrs.fr/corpus/Bulgaro-Mac%C3%A9donien?lang=fr&mode=normal>

been mentioned in §1, the speakers in our sample who use *reče* as a quotative, do not use the verb *veli* at all, but use *vika* as the imperfective form of *reče*. The speakers in the Pangloss and MANU collections that use *veli*, on the other hand, do not use the verb *vika*.

We can compare this to the Turkic languages mentioned in §3, Chuvash (see Nikitina this volume) and Bashkir (see Aponova this volume), which each also have two semi-grammaticalized quotatives: *Tigän* and *tip* in Bashkir and *tet* and *tese* in Chuvash. The difference between Macedonian and the Turkic languages is that in the Turkic languages the two quotatives come from the same stem and both are used by the same speakers, i.e., they have taken up a different function. In Macedonian, on the other hand, the two quotatives are never used by the same speaker, and they seem to more or less have the same function.

## 7 Conclusion

We have presented the following evidence that the Macedonian speech verb *reče* is grammaticalizing into a quotative clitic: i) it is more frequent than the imperfective speech verb *vika*, ii) it most often occurs after the first phonological word, iii) it co-occurs with other speech verbs, such as *vika* ‘says’ or *kažuva* ‘tells’, iv) it loses its inflection and v) it is adopting the morphophonological stress features of clitics by losing (a part or the entirety of) its stressed syllable and orally becoming *eče*, *če*, or *rče*. This grammaticalization is not a conventionalized phenomenon, such as the future clitic *ke*, but still an ongoing process in its early stages. This can be seen from the fact that it was only attested in narrations by certain speakers in informal settings and that there is a high degree of variation of the realization of *reče* on the morphosyntactic, phonological and phonetic levels.

## Abbreviations

1, 2, 3	first, second, third person	M	masculine
ACC	accusative	N	neuter
AOR	aorist	NEG	negation
CAUS	causative	NMLZ	nominalizer
CVB	converb	NOM	nominative
CV.COORD	coordinative converb marker	OBL	oblique
DAT	dative	PFV	perfective
DEF	definite	PL	plural
DIM	diminutive	POSS	possessive
DR	discourse report	POT	potential
F	feminine	PROG	progressive
GEN	genitive	PRS	present
IMPF	imperfect	PTCP	participle
INF	infinitive	Q	question particle
INSTR	instrumental	QUOT	quotative marker
INTJ	interjection	QV	quotative verb
IPFV	imperfective	REFL	reflexive
LPTCP	<i>l</i> -participle	SBJ	subjunctive
		SG	singular

## Acknowledgements

We want to thank Marija Nikolova for the data. Thank you Tatiana Nikitina and everyone at the Helsinki workshop for the feedback.

This work was carried out in the scope of the “Discourse reporting in African storytelling” project. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 758232).

## References

- Alexander, Ronelle. 1994. The balkanization of Wackernagel’s law. *Indiana Slavic Studies* 7. 1–8.
- Aplonova, Ekaterina. 2021. Two quotative markers in Bashkir: a corpus-based study of reported speech constructions. In *Reported speech: new data studies*.

- Chanard, Christian. 2015. ELAN-CorpA: lexicon-aided annotation in ELAN. In *Corpus-based studies of lesser-described languages*, 311–332. John Benjamins.
- Deutscher, Guy. 2011. The grammaticalization of quotatives. In *The Oxford handbook of grammaticalization*.
- Evira García, Wendy. 2017. *Create pictures with tiers v.4.4. praat script*. <http://stel.ub.edu/labfon/en/praat-scripts>.
- Fielder, Grace E. 1999. The origin of evidentiality in the Balkans: linguistic convergence or conceptual convergence? *Mediterranean language review* 11. 59–89.
- Friedman, Victor A. 1998. The grammatical expression of presumption and related concepts in Balkan Slavic and Balkan Romance. In Michael Flier & Alan Timberlake (eds.), *American contributions to the 12th international congress of slavists*, 390–405. Slavica.
- Friedman, Victor A. 2000. Confirmative/nonconfirmative in Balkan Slavic, Balkan Romance, and Albanian with additional observations on Turkish, Romani, Georgian, and Lak. *Evidentials: Turkic, Iranian and neighbouring languages* 24. 329.
- Friedman, Victor A. 2003. Evidentiality in the Balkans with special. *Studies in evidentiality* 54. 189.
- Friedman, Victor A. 2001. *Macedonian* (SEELRC Reference Grammars). Durham: Duke University: The Slavic & Eurasian Language Resource Center. <http://www.seelrc.org:8080/grammar/mainframe.jsp?nLanguageID=3>.
- Greed, Teija. 2014. The expression of knowledge in Tatar. *The grammar of knowledge: A cross-linguistic typology* 69. 88.
- Güldemann, Tom. 2008. *Quotative indexes in African languages*. De Gruyter Mouton.
- Gvozdanovic, Jadranka. 1996. Reported speech in South Slavic. *Pragmatics and Beyond New Series*. 57–74.
- Hall-Lew, Lauren. 2010. Improved representation of variance in measures of vowel merger. *Proceedings of Meetings on Acoustics* 9(2010). DOI: [10.1121/1.3460625](https://doi.org/10.1121/1.3460625).
- Hopper, Paul J & Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge University Press.
- Kisler, Thomas, Uwe Reichel & Florian Schiel. 2017. Multilingual processing of speech via web services. *Computer Speech & Language* 45. 326–347. DOI: <http://dx.doi.org/10.1016/j.csl.2017.01.005>.
- Klamer, Marian. 2000. How report verbs become quote markers and complementizers. *Lingua* 110(2). 69–98.
- Knyazev, Mikhail. 2019. Two ‘say’-complementizers in Poshkart Chuvash: Subject-orientation, logophoricity and indexical shift under verbs of hearing. In *Pro-*

- ceedings of the 14th workshop on Altaic formal linguistics: MIT working papers in linguistics# 90*, 137–144.
- Koneski, Blaže. 1965. *Istorija na makedonskiot jazik*. Skopje: Kočo Racin.
- Kouteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog & Seongha Rhee. 2019. *World lexicon of grammaticalization*. Cambridge University Press.
- Lobanov, B. M. 1971. Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America* 49(2B). 606–608. DOI: [10.1121/1.1912396](https://doi.org/10.1121/1.1912396). <https://doi.org/10.1121/1.1912396>.
- Lord, Carol Diane. 1990. *Syntactic reanalysis in the historical development of serial verb constructions in languages of West Africa*. Los Angeles: University of California. (Doctoral dissertation).
- Lunt, Horace Gray. 1952. *A grammar of the Macedonian literary language*. Državno knigoizdatelstvo na NR Makedonija Skopje.
- Makartsev, Maxim. 2012. Towards a systemization of common Balkan lexical evidential markers. *Slověne. International Journal of Slavic Studies* 1(1).
- Markovik, Marjan. 2007. *Dijalektologija na makedonskiot jazik II [Dialectology of the Macedonian language II]*.
- Michailovsky, Boyd, Martine Mazaudon, Alexis Michaud, Séverine Guillaume, Alexandre François & Evangelia Adamou. 2014. Documenting and researching endangered languages: the Pangloss collection. *Language Documentation & Conservation* 8. 119–135.
- Mushin, Ilana. 1997. Maintaining epistemological stance: direct speech and evidentiality in Macedonian. In *Chicago linguistic society*, vol. 33, 287–300.
- Mushin, Ilana. 2000. Evidentiality and deixis in narrative retelling. *Journal of Pragmatics* 32(7). 927–957.
- Nikitina, Tatiana, Ekaterina Aplonova & Leonardo Contreras Roa. 2021. *The use of interjections as a discourse phenomenon: a contrastive study of Chuvash (Turkic) and Wan (Mande)*.
- Nikitina, Tatiana, Abbie Hantgan & Christian Chanard. 2019. *Reported speech annotation template for ELAN. discourse reporting in African Storytelling repository*.
- Nycz, Jennifer & Lauren Hall-Lew. 2014. Best practices in measuring vowel merger. *Proceedings of Meetings on Acoustics* 20(1). DOI: [10.1121/1.4894063](https://doi.org/10.1121/1.4894063).
- Özyıldız, Deniz, Travis Major & Emar Maier. 2018. Communicative reception reports as hear-say: Evidence from indexical shift in Turkish. In *Proceedings of wccfl*, vol. 36, 296–305.

- Rudin, Catherine, Christina Kramer, Loren Billings & Matthew Baerman. 1999. Macedonian and Bulgarian *li* questions: Beyond syntax. *Natural Language & Linguistic Theory* 17(3). 541–586. DOI: [10.1023/A:1006223423256](https://doi.org/10.1023/A:1006223423256).
- Savicka, Irena. 2014. *Fonetskite problemi vo arealnata lingvistika (vrz osnova na jazičnata situacija vo Južnoistočna Evropa) [Phonetic problems in areal linguistics (on the basis of the language situation in South-East Europe)]*. Skopje: MANU.
- Savicka, Irena, Branislav Gerazov, Veselinka Labroska, Anna Cychnerska & Agata Travińska. 2021. *Suprasegmentalna fonetika i fonologija: fonetika i fonologija na makedonskiot standarden jazik [Suprasegmental phonetics and phonology: phonetics and phonology of the macedonian standard language]*. Skopje: Makedonska Akademija na Naukite i Umetnostite (MANU).
- Tomić, Olga Mišeska. 1996. The Balkan Slavic clausal clitics. *Natural Language & Linguistic Theory* 14(4). 811–872.
- Tomić, Olga Mišeska. 2004. The syntax of the Balkan Slavic future tenses. *Lingua* 114(4). 517–542.
- Tomić, Olga Mišeska. 2006. *Balkan Sprachbund morpho-syntactic features*. Vol. 67. Springer Science & Business Media.
- Vidoeski, Božidar, Marjan Markovič, Angelina Pančevska, Sonja Milenkovska & Borko Jovanovski. s.a. *Audio materials of the Macedonian language*. Online collection of the Macedonian Academy of Arts and Sciences, last accessed 08 April 2021. <http://damj.manu.edu.mk/audio/>.
- Vidoeski, Božidar, Irena Savicka & Zuzanna Topolińska. 2003. *Polski-makedonski: gramatička konfrontacija. 2. prozodija [Polish-Macedonian: grammatical confrontation. 2. prosody.]* Skopje: Makedonska Akademija na Naukite i Umetnostite (MANU).
- Vykypěl, Bohumil. 2010. Slavonic-Baltic addenda to the World lexicon of grammaticalization. *Journal of Linguistics/Jazykovedný časopis* 61(2). 131–144.