

Notes on two markers of contrast in Terek Kumyk

Elizaveta Dorofeeva*

Abstract. This paper introduces two discourse particles of Terek Kumyk (<West Kipchak<Kipchak<Turkic): =*msA* and =*č*. The first one is found commonly within Turkic languages as a contrastive topic marker (Iatridou 2013), while the latter one is only observed in this dialect, and its functions remain quite vague. In our notes we will distinguish these two particles from homonymic markers, observe some contexts where they occur, take a look at their impact on sentential prosody and will also provide a brief sketch of semantic analysis of these two markers.

Keywords. information structure; discourse markers; (contrastive) topic; (contrastive) focus

1. Background information.

1.1. TEREK KUMYK. Terek Kumyk is a dialect of Kumyk language, which belongs to the Kipchak-Cuman subfamily of the Kipchak family of the Turkic languages. It is spoken mostly in Mozdok district (the Republic North Ossetia — Alania)¹. Despite the fact that Standard Kumyk, which is spoken in Dagestan, is fully described in (Gadjiaxmedov et al. 2014), we cannot rely on this description in our research, because the Terek dialect differs from Standard Kumyk in several aspects. One of this differences is that Terek Kumyk has a more free unmarked word order (1) than the standard variant. Terek Kumyk word order can be SOV or SVO. Notably, the latter one is preferred by the native speakers, all of which are Russian-Kumyk bilinguals.

- (1) a. Kerim hudur-nu aša-j.
 Kerim porridge-ACC eat-IPFV
 b. Kerim aša-j hudur-nu.
 Kerim eat-IPFV porridge-ACC
 ‘Kerim eats porridge.’

Subject tends to be the leftmost word in the sentence and to precede direct (1) and indirect (2) objects. The only class of constituents that can precede the subject, when the interpretation of the sentence is unmarked, is high adverbs, but we shall not observe them in this sketchy study.

- (2) Kerim jurt-da.
 Kerim village-LOC
 ‘Kerim is in the village.’

The pitch contour of an isolated sentence with basic word order is relatively smooth (Fig. 1). The pitch falls on every (prosodic) word, but the most significant pitch lowering is located at the end of the phrase.

* I would like to thank Sergey Tatevosov, Pëtr Rossyaykin, Daniar Kasenov, the audiences at the TU+9 conference, and two anonymous reviewers for valuable discussion and comments. All remaining errors are my own. Authors: Elizaveta Dorofeeva, Lomonosov Moscow State University (anrie.llnfd@gmail.com).

¹ All the data for this study was collected during two field trips to Predgornoye village (Mozdok district, the Republic North Ossetia — Alania) in August 2023 and July 2024.

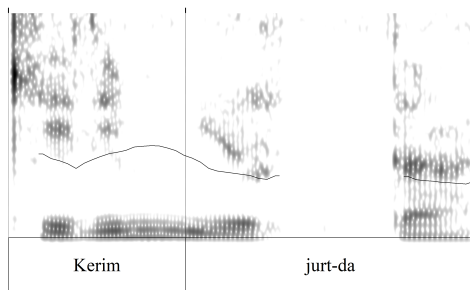


Figure 1. Pitch track for sentence *Kerim jurt-da* (Kerim village-LOC) ‘Kerim is in the village.’

All the violations of SOV or SVO order will lead to marked interpretations. For example, in a sentence with OSV order like (3a) the direct object *hudur* ‘porridge’ would possibly be interpreted as topic. For (3b), where indirect object *jurt* ‘village’ occurs at the beginning of the sentence, the situation is quite similar. The sentences in (3) cannot serve as answers to the question “What is happening?”, albeit they are among possible answers to the questions about the porridge and about the village.

- (3) a. *Hudur-nu Kerim aša-j.*
 porridge-ACC Kerim eat-IPFV
 ‘As for the porridge, Kerim eats it.’
 b. *Jurt-da Alim.*
 village-LOC Alim
 ‘As for the village, Alim is there.’

As for prosody, pitch (Fig. 2) goes up and reaches its peak at the right border of the topicalized constituent. After this, it acts like the pitch contour of a sentence with unmarked word order. As mentioned in (Vallduvé & Engdahl 1996) (*via* (İşsever 2003)), the same pattern of pitch movement on sentence-initial topics is observed in Turkish. It is also worth noting that there is (almost) no pause between topicalized and non-topicalized part of the sentence.

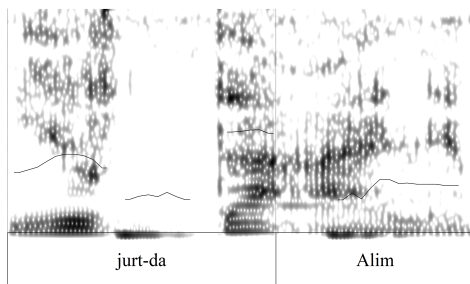


Figure 2. Pitch track for sentence *Jurt-da Alim* (village-LOC Alim) ‘As for the village, Alim is there.’

1.2. SOME (PRE)THEORETICAL NOTIONS. In the previous section we have mentioned topicalization and its impact on the sentential prosody. At this point, let us define what we mean by topic, as well as by focus and contrast, the information structure-related notions that will also appear in the following text.

As mentioned in (Reinhart 1981), (sentence) topic is the equivalent of the term *theme*, i. e. a discourse entity which the sentence is about. Topic is distinguished from the background, which stands for all the information that is not focused. Following (Stevens 2017), we define focus as “the portion of an utterance which is especially informative or important within the context”. If topic occurs in the question, focus occurs in its answer, so, as noted in (Van der Wal 2016), the question-answer test is “the most widespread and most widely accepted test for focus”. The results of application of this test are presented below (e. g. (12) and (13)).

Contrast is perhaps the most disputable of the aforementioned notions. One can treat it as an independent feature which is not a part of topic and focus (as in Neeleman & Vermeulen 2012). However it is also possible to analyse contrast as a feature dependent on topic and focus (see (Dyakonova 2009; Constant 2014)). According to the second point of view, the reason why we interpret contrastive focus more contrastive than discourse-new focus is because it has a closed set of alternatives, while discourse-new focus has an open set of alternatives. In our study we adopt this approach, following previous researches about information structure in Turkic languages (e. g. (Gürer et al. 2015)).

2. Note #0: Introducing =č and =msA, two markers of contrastive topic. Let us first look at contrastive topics. For example, when answering a question about the location of Alim and Kerim, one can say [*Alim*]_{TOPIC} *is in the town*, [*Kerim*]_{CONTR. TOPIC} *is in the village*. In Terek Kumyk there are three possible ways to express this meaning (4): without any discourse markers, with clitic =msA attached to the contrastive topic *Kerim* or with clitic =č attached to the same constituent. All these ways to express contrastive topic are evaluated as equally acceptable by native speakers of Terek Kumyk, and in (4) or in similar contexts =č and =msA are interchangeable.

- (4) Alim Mazlek-de, Kerim / Kerim=mse / Kerim=č jurtda.
 A. Mozdok-LOC K. / K.=MSA / K.=č village-LOC

Alim is in the town (i.e. Mozdok), and Kerim is in the village.’ (Answering the question *Where is Alim, and where is Kerim?*)

The pitch tracks for all three variants are provided below, and we shall now compare them to the ones of sentences with non-contrastive topics and to each other in order to detect possible impact of the two aforementioned contrastive topic markers on sentence prosody. We also aim to compare pitch contour of a sentence with =č to the one of a sentence with =msA — despite the fact that they are totally fungible in the current context, they may trigger different prosodic behavior.

As can be seen in Fig. 3, prosody contour of both parts of the sentence without any of contrastive topic markers is as smooth as one of the sentence with basic word order (Fig. 1): we observe a minor pitch rise on the topicalized constituent [*Alim*]. The main difference is that at the right border of the first clause [*Alim Mazlek-de*] Alim town-LOC ‘Alim is in the town’ the pitch goes up, which does not resemble its behavior at the end of the second clause.

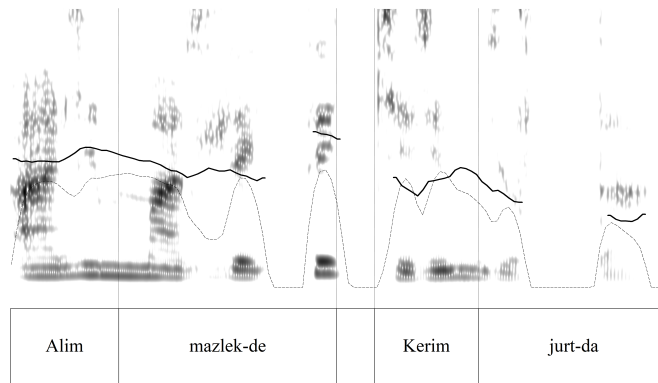


Figure 3. Pitch (and intensity) track for sentence *Alim Mazlek-de, Kerim jurt-da* (Alim town-LOC, Kerim village-LOC) ‘Alim is in the town, and Kerim is in the village.’

The same pitch contour is observed in Fig. 4: pitch goes up at the end of the first clause and then rises up on a topicalized constituent at the left edge of the second clause. The picture resembles Fig. 3.

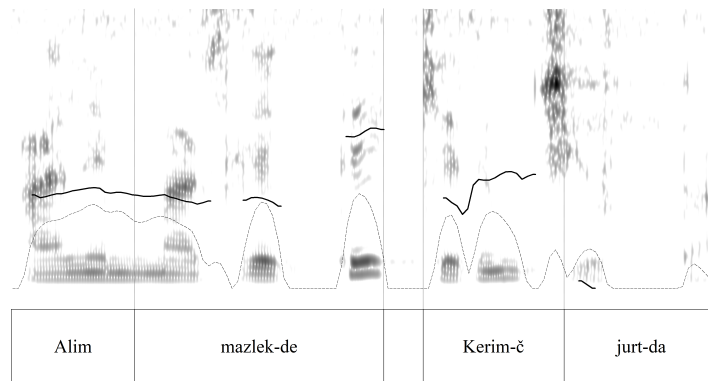


Figure 4. Pitch (and intensity) track for sentence *Alim Mazlek-de, Kerim=č jurt-da* (Alim town-LOC, Kerim=č village-LOC) ‘Alim is in the town, and Kerim is in the village.’

While the pitch contour of the first clause of the sentence with *=msA* expectedly goes up at the right border, the prosody of the second one is slightly different. We observe pitch rise at the right border of the *=msA*-marked constituent accompanied by a pause after it — in two aforementioned cases the pause occurred only between two clauses, but definitely not after the topicalized constituent. This might as well indicate that a constituent with *=msA* prosodically acts like an autonomous clause, similarly to the first clause in (4), and this cannot be said about topicalized constituents with *=č* or without any markers.

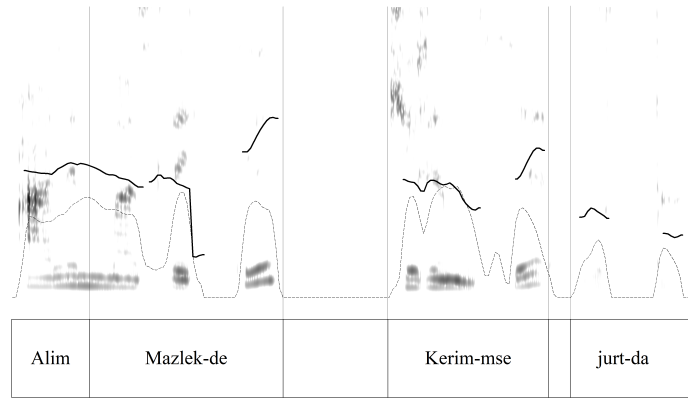


Figure 5. Pitch (and intensity) track for sentence *Alim Mazlek-de, Kerim-mse jurt-da* (Alim town-LOC, Kerim=*msA* village-LOC) ‘Alim is in the town, and Kerim is in the village.’

To sum up our introductory note, we have observed that notwithstanding the interchangeability of these two markers in the context of contrastive topic, one can catch the difference between them on the level of prosody. This note offers the prospect of looking for other differences between *=č* and *=msA* as CT-markers: for example, if a *=msA*-marked constituent is a clause and a *=č*-marked one is not, then there might be different restrictions in extraction from these two kinds of topicalized constituents.

To figure out why does *=msA* makes a constituent followed by it act like a clause, while *=č* does not, we feel compelled to say something about the origins of these two markers. Our Note #1 is dedicated to the comparison of *=č* and *=msA* to the homonymous units.

3. Note #1: What can homonyms tell about *=msA* and *=č*. Let us start with separating our discourse markers from their homonyms in order to discover some peculiarities of their meaning and possibly some restrictions on their attachment.

3.1. *=msA*: NOT ONLY A CONDITIONAL MARKER. Terek Kumyk particle *=msA* originates from *bu-sa* (< *bol-sa* in Literal Kumyk) ‘be-COND’. *=msA* can also be a conditional marker that is placed after an (im)perfective (5).

- (5) a. Tünegün gel-gen=**mse**-ng, šorpa aša-žadi-ng.
yesterday come-PFCT=**be**+COND-2SG soup eat-FUT1+PST-2SG
‘If you came yesterday, you would eat some soup.’
- b. O uhla-j=**msa**, men o-nu čaqir-ma-žadi-m bizin=**mnen**.
3SG sleep-IPFV-**be**+COND 1SG 3SG-ACC call-NEG-FUT1+PST-1SG 1PL.GEN=COM
‘If she was asleep, I wouldn’t have called her to come with us.’

In Terek Kumyk, if not attached to a phrasal argument, *=msA* may only derive the conditional reading, but not a contrastive one (6). However, it is possible to have *=msA* after conditional verb form (7), if contrastive topic is located inside the antecedent of a conditional clause.

- (6) a. Mišik uhla-j, it=mse rapla-j
 cat sleep-IPFV, dog=MSA bark-IPFV
 ‘The cat is sleeping, the dog is barking.’
- b. Mišik uhla-j, it rapla-j=msa
 cat sleep-IPFV, dog bark-IPFV=MSA
 ‘The cat sleeps, if the dog barks.’ (int.: ‘The cat is sleeping, the dog is barking.’)
- (7) a. Biz-ge Alim gel-se, it-ibiz uhla-p bol-a, Kerim
 1PL-DAT Alim come-COND dog-1PL.POSS sleep-CVB be-IPFV Kerim
 gel-se-mse, hapla-p tur-a.
 come-COND-MSA bark-CVB be-IPFV
 ‘When Alim comes to us, our dog sleeps, but when Kerim comes to us, it barks.’

In paper about analogous (Turkish contrastive topical *y-sa* is also conditional form of the verb ‘to be’) marker in Turkish, (Iatridou 2013, there was proposed a reduction of both Conditional *-sA* (12) and free-relative *-sA* to *-sA* as a marker of correlativity, but “the question of whether contrastive topic *-sA* (13) can be brought into the fold still remains open”.

- (8) a. Turkish (from Iatridou 2013:(12-13))
 (Eğer) yağmur yağ-ar-sa sinema-ya gid-eceğ-iz.
 (in.case) rain rain-AOR-COND movies-DAT go-FUT-1PL
 ‘(In case) it rains, we will go to the movies.’
- b. Ali çok hoş bir insan-dır, Oya-ysa çok sevimsiz-dir.
 Ali very nice a person-is Oya-sA very unlikeable-is
 ‘Ali is a very nice person; Oya, on the other hand, is very unlikeable.’

For Terek Kumyk we propose to combine the contrastive topicalization marker with the conditional inflectional marker, since the former, together with the topicalized component, forms a (conditional) clause that is on the left periphery; we assume that it is a cleft (i.e. a construction that means something like *Speaking about X, ...*). The evidence for treating a constituent with CT- indicating *=msA* as a clause just like an antecedent of conditional from prosody is provided in Note #0.

3.2. =č: NOT ONLY AN ANTIHONORIFIC IMPERATIVE MARKER. The grammar of literary Kumyk (Gadjiaxmedov et al. 2014) informs us that =č is a “dictive-objective particle” — and it says nothing more about this marker, not a single word about its functioning as a marker of contrast. In Terek Kumyk =č serves to form a rude order (Kasenov 2023), but a homonymous particle marking a contrastive topic can be added to it (9).

- (9) a. Get!
 go.away.IMP
 ‘Go away!’

- b. Get=*č*!
 go.away.IMP=ANTIION
 ‘Go away!’ (*rude*)
- c. Get=*č=či*!
 go.away.IMP=ANTIION=*č*
 ‘Go away finally!’ (*rude*)

Nor is one entity a marker of contrastive topicality with the indicator that makes pronouns specific unknown (10) and that is also only met in Terek Kumyk.

- (10) a. Birewde=*č* meni mahta-ma-di.
 someone=*č*_{SP} 1SG.ACC praise-NEG-PST
 ‘Nobody has praised me.’
- b. ..., birewde=*č=či* meni mahta-ma-di.
 ... someone=*č*_{SP}=*č* 1SG.ACC praise-NEG-PST
 ‘(Everyone loved my work,) but nobody has praised me.’

The homonymy of these two markers allows us to draw a parallel with the Russian clitic =*to* (11), that can be attached to an interrogative pronoun to make an indefinite (regarded as specific unknown in (Haspelmath 1997), but also allowed in non-specific contexts) one or can serve as a marker of emphasis or contrast (not only contrastive topic), like it is mentioned parenthetically in (Rosseyaykin et al. 2022).

- (11) a. Kto-to menja poxvali-l-Ø.
 who-TO_{SP} 1SG.ACC praise.PFV-PST-3SG
 ‘Somebody has praised me.’
- b. ..., kto-to=*to* menja poxvali-l-Ø.
 ... who-TO_{SP}=TO_{CT} 1SG.ACC praise.PFV-PST-3SG
 ‘(Despite the fact that nobody loved my work,) somebody still has praised me.’

Bearing in mind that Russian affects the lexicon (one can observe a lot of loanwords from Russian in Kumyk dictionary and especially in fluent speech of native speakers of Terek Kumyk) and the grammar of Terek Kumyk (e.g. the aforementioned basic/neutral word order), we may suppose that polysemy of Russian =*to* has triggered the development of additional meanings of Kumyk =*č*: it started to mark some kind of emphasis and then the choice of one single item from the set of possible variants — this mechanism is common for focus (as seen in (Rooth 1992)) and plausible for specific unknown pronouns.

Summing up this note, we may say that (diachronically) the origins and (synchronically) the homonyms of the two markers of contrastive topic introduced in the previous note can tell a lot about their possible meanings: cleft-like =*msA* can only be a topical marker, while =*č*, which has evolved from an emphatic marker that chooses one single item from a set, may serve not only as a contrastive topic marker, but also as a focal particle.

4. Note #2: On alternative sets. After setting some restrictions on possible meanings of =č and =msA in the previous section, in current note we shall observe interaction of =č and =msA with sets of alternatives.

Given an open, contextually unrestricted, set of alternatives (e.g. when answering an open question), none of the markers suits to indicate focus.

- (12) Speaker A. *Kim gel-di?*
 who come-PST
 ‘Who came?’
- Speaker B. *Alim / *Alim=č / *Alim=mse gel-di*
 A. / A.=č / A.=MSA come-PST
 ‘[Alim]_F came.’

As for the closed set (e. g. when answering an alternative question), the particle =č is suitable for pointing out the most important alternative of the proposed in the question set (13) — at least it is more suitable than its counterpart, =msA.

- (13) Speaker A. *Kim gel-di jaš-lar-dan, Alim gel-di-m, Kerim?*
 who come-PST boy-PL-ABL A. come-PST-Q K.
 ‘Who of the boys came, Alim or/and Kerim?’
- Speaker B. *Alim / ?Alim=č / *Alim=mse gel-di*
 A. / A.=č / A.=MSA come-PST
 ‘[Alim]_F came.’

As we have already seen in (4), both =č and =msA are acceptable as contrastive topic markers, when alternative set is definitely contextually restricted, i.e. closed.

How can we model the contrast demonstrated in (12-13)? We are not absolutely sure about accurate semantic representations for =č and =msA, so we shall not provide them here and save this for future research. However, we suggest that semantics of these two discourse markers should consider these facts:

- For both =č and =msA there should be contextual restriction on the alternative set: there might be a subset of C and $\{X\}_F$ for any constituent X followed by any of the two discourse markers;
- Because =msA can only indicate contrastive topic, and =č can mark focus in a restricted set of alternatives (i. e. contrastive focus) and contrastive topic, following (Wagner 2012), we propose to analyze =č as a FOCUS operator, choosing one salient variant from a set of items, and =msA — as two nested FOCUS operators, which introduces a set of alternatives for every (not only one) salient constituent.

5. Conclusion. In this paper, we presented data on two contrast markers in the Terek dialect of Kumyk. We have shown how they behave in terms of prosody, and which units they can be associated with in order to better understand their meaning, and more clearly defined the function of each of them. Further work on this topic includes an accurate formal description of the markers under investigation, as well as an investigation of the syntactic constraints on their use.

References

- Constant, Noah. 2014. Contrastive topic: Meanings and realizations
<https://doi.org/https://doi.org/10.7275/5694973.0>.
- Dyakonova, Marina. 2009. *A phase-based approach to russian free word order*, vol. 230. LOT Utrecht.
- Gadjiaxmedov, Nurmagomed, Aziza Abdullaeva, Kalsyn Kadyradžiev, Ibragim Kerimov, Nuramat Olmesov & Džangiši Hangišiev. 2014. *Sovremennyy kumyjskij yazyk [modern kumyk language]*. Makhachkala: IYALI DNC RAN.
- Gürer, Aslı et al. 2015. Semantic, prosodic, and syntactic marking of information structural units in Turkish .
- Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford University Press.
- Iatridou, Sabine. 2013. Looking for free relatives in turkish. In *Proceedings of the 8th workshop on altaic formal linguistics*, 129–152.
- İşsever, Selçuk. 2003. Information structure in Turkish: the word order–prosody interface. *Lingua* 113(11). 1025–1053. [https://doi.org/https://doi.org/10.1016/S0024-3841\(03\)00012-3](https://doi.org/https://doi.org/10.1016/S0024-3841(03)00012-3).
- Kasenov, Danyar. 2023. Morphosyntax of politeness and indexical shift: evidence from terek kumyk. 20th Conference on Typology and Grammar for Young Scholars.
- Neeleman, Ad & Reiko Vermeulen. 2012. The syntactic expression of information structure. *The syntax of topic, focus, and contrast: An interface-based approach* 1–38.
<https://doi.org/http://dx.doi.org/10.1515/9781614511458.1>.
- Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27. <https://doi.org/https://doi.org/10.21825/philosophica.82606>.
- Rooth, Mats. 1992. A theory of focus interpretation. *Natural language semantics* 1(1). 75–116.
<https://doi.org/https://doi.org/10.1007/BF02342617>.
- Rossyaykin, Pëtr, Regina Nasyrova & Elizaveta Dorofeeva. 2022. Indefinite pronouns in Terek Kumyk. Field trip report.
- Stevens, Jon Scott. 2017. Pragmatics of focus. In *Oxford research encyclopedia of linguistics*,
<https://doi.org/https://doi.org/10.1093/acrefore/9780199384655.013.207>.
- Vallduvé, E. & Elisabet Engdahl. 1996. The linguistic realisation of information packaging. *Linguistics* 34. <https://doi.org/10.1515/ling.1996.34.3.459>.
- Wagner, Michael. 2012. Contrastive topics decomposed. *Semantics and Pragmatics* 5. 1–54.
<https://doi.org/http://dx.doi.org/10.3765/sp.5.8>.
- Van der Wal, Jenneke. 2016. Diagnosing focus. *Studies in Language* 40(2). 259–301.
<https://doi.org/http://dx.doi.org/10.1075/sl.40.2.01van>.