MAIN in the Tajik and Shughni languages of Tajikistan

Qurbonidin Alamshoev

NGO Kuhhoi Pomir

Aleksandra Trifonova

Universität Potsdam

After a brief overview of the linguistic situation in Tajikistan, this paper describes the adaptation and use of the Multilingual Assessment Instrument for Narratives (LITMUS-MAIN) in the Tajik and Shughni languages of Tajikistan.

1   Introduction

Nowadays, children growing up bilingually or even multilingually is a highly wide-spread phenomenon. Although linguistic and cultural contacts between nations have always been commonplace, research on bilingual populations is still relatively scarce. Furthermore, investigation of bilingual language development is challenging for both speech-language therapists and linguists due to the fact that, as a rule, standardized diagnostics tests are not suitable for bilingual language acquisition assessment, which, in turn, can lead to misdiagnoses (cf. Fleckstein et al., 2018; Grimm & Schulz, 2014). Things get even more complex when it comes to understudied minority languages. Therefore, it is vital to develop assessment tools that are sensitive enough to fully take into account the nature of bilingual language acquisition.

In the literature, narrative abilities have been reported to be one of the most ecologically valid measures of communicative competence in various speakers' populations (Botting, 2002, p. 1). One test of narrative ability is the *Multilingual Assessment Instrument for Narratives* (MAIN; Gagarina et al., 2012; 2015; 2019), which is part of the language test battery *Language Impairment Testing in Multilingual Settings* (LITMUS; Armon-Lotem et al., 2015). MAIN constitutes a reliable assessment tool for investigation of language attainment in children with different linguistic backgrounds (mono-, bi- and multilinguals of different ages). It can be used to evaluate children's narrative comprehension and production in three elicitation modes (telling, retelling, model story), and to assess measures of macro- and microstructure of a narrative (Gagarina et al., 2019). MAIN has received much public attention, has been

recognized by the global linguistic community and has already been adapted to a large number of languages from different parts of the world (German, Russian, Turkish, Mandarin, etc.)

In this paper, the linguistic situation in Tajikistan is outlined, including an overview of Tajik and Shughni, two Indo-Iranian languages, after which the initial steps of the process of adapting MAIN to Tajik and Shughni are described.

2   Languages in Tajikistan

In the Republic of Tajikistan, Tajik is the official language. Apart from Tajik, people in Northern Tajikistan also speak Yagnob. The Pamir languages Shughni, Rushani, Khufi, Bartangi, Roshorvi, Sariqoli, Yazghulami, Wakhi, and Ishkashimi are spoken in the Gorno-Badakhshan Autonomous Region of Tajikistan; Tajik is used for communication between different ethnic groups in this region (Dodikhudoeva, 2004). Therefore, the population in Gorno-Badakhshan is multilingual: 140,000 out of 211,000 inhabitants spoke different East-Iranian Pamir languages in 1999 (Dodikhudoeva, 2004). Additionally, Russian is also widely used in Tajikistan and is officially recognized by the constitution as a language of inter-ethnic communication, and a small part of the work of state organizations is done in Russian (The Constitution of the Republic of Tajikistan). Today, the language of instruction in schools of the Gorno-Badakhshan Autonomous Region is Tajik. However, before entering school, children usually do not speak Tajik or have very poor knowledge of the language (Dodikhudoeva, 2004).

2.1   *The Tajik and Shughni languages*

Tajik is an Indo-Iranian language of the Indo-European family and is spoken by more than eight million people (Ethnologue). It is mainly spoken in Tajikistan but is also used in Afghanistan, Pakistan, India, China. Tajik is an analytic language with the SOV basic word order (Ethnologue). Recently, a corpus consisting of more than 5 million words in Tajik has been introduced, which was reported to be the largest computer corpus of the Tajik language so far (Dovudov et al., 2011). After the establishment of the Tajik Soviet Socialist Republic in 1929, Tajik became the official language in Tajikistan, and the Arabic script, traditionally used by Tajiks for centuries, was first transformed into Latin script, and later into Cyrillic script (Avezova, 2017; see Dodikhudoeva, 2004 for a detailed discussion). During the Soviet period (1929-1991), the Tajik language was heavily influenced by Russian, which resulted in numerous borrowings from Russian in the domain of vocabulary (Avezova, 2017).

Shughni belongs to the East Iranian group of languages spoken by approximately 130,000 people in several districts in the Gorno-Badakhshan Autonomous Region of Tajikistan and the Shighnan district in Afghanistan (Mueller, 2015; Olson, 2017). Shugnhi has a number of different dialects: Badjuvi, Khufi, Rushani, Bartangi, and Roshorvi (Mueller, 2015; Olson, 2017). However, there is an ongoing debate about the status of the dialects: whether some are independent languages, subdialects, or all, including Shughni, are varieties of one single language (Mueller, 2015; Olson, 2017). Shughni, like the other Pamir languages spoken in the

Gorno-Badakhshan Autonomous Region, has no tradition of written language (Dodikhudoeva, 2004). However, several attempts for establishing a written form for Shughni have been made (Mueller, 2015).

In the 19th century, Robert Shaw was the first to describe the Shughni language. Later, the Russian researcher D.L. Ivanov visited the Pamirs and collected linguistic and folklore samples of the Shughni language, based on which the first Russian-Shughni dictionary was compiled by K.G. Zaleman in 1895. In the second half of the 20th century, the study of Shughni and other Pamir languages was revived by establishing the department of world studies in Dushanbe, where Pamir scholars started to explore these languages. Today, on the basis of the Institute of Humanities of the Academy of the Republic of Sciences of Tajikistan, a department for the study of Pamir languages is active in the Gorno-Badakhshan Autonomous Region.

Shughni has a number of characteristic linguistic features. It has an unstrict SOV word order and allows for scrambling (Parker, 2020). Additionally, it is mainly head-final and displays the following linear structure: DEM-ADJ-NOUN (Parker, 2020). Shughni is also a vestigial ergative language (Parker, 2020). Interestingly, Shughni determiners have three degrees of distance: proximal, medial, and distal (Mueller, 2015).

3   Adapting MAIN for use in Tajikistan

During the Fall School on Documentary Linguistics in the MENA region (the Middle East and North Africa) at the Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS) in Berlin in October 2018, the first author of this paper Qurbonidin Alamshoev, who is a staff member of the public organization *Kuhhoi Pomir* in Tajikistan, was introduced to MAIN. After receiving training on the use of MAIN, Qurbonidin Alamshoev started the adaptation of MAIN to Shughni and Tajik based on the revised Russian version of MAIN. A personal Skype-training for testers was organized by Alyona Sternharz, a research assistant at ZAS. During this training, the process of data collection and assessment guidelines were discussed.

In order to ensure correct adaptations of MAIN to Tajik and Shughni (see Bohnacker & Gagarina, 2019), assistants, who were to help with data collection in Tajikistan, were selected from scholars and postgraduate students of the Khorog Institute of Humanities. These assistants (Nurijahon Kurbonkhonova, Shiringul Azorabekova, Ibodat Karamova, Chilla Nazarshoeva, and Muslima Broimshoeva) were philologists, historians, and specialists in the languages and culture of the Pamirs, and also had the experience of working with children. Within a one-day workshop, the testers were introduced to the MAIN-materials and were trained in carrying out data collection with MAIN. After this training, pilot data from Shughni-Tajik bilinguals were collected following the MAIN procedure (Gagarina et al., 2019).

Each tester collected data from four typically-developing bilingual Shughni-Tajik children (mean age=10.75). Audio recordings were made of each testing session. Each child was tested with MAIN in both languages: first, in their native language Shughni and 10 days later in Tajik. Since children who have a Pamir language as their native language hardly speak Tajik before entering school, only children who attended the third and fourth grades were

chosen to participate in the pilot study, as at this age they were already conversant in Tajik. A total of 40 testing sessions was conducted: 20 in Shughni and 20 in Tajik. All four stories (Cat, Dog, Baby Birds, Baby Goats) in the three testing modes (telling, retelling, model story) were administered to the participants. In addition, a parent survey was carried out. These pilot data will contribute to the development of the Tajik and Shughni MAIN-versions, which are planned to be finished by the end of 2020.

4   Some preliminary results from the pilot study

In the *Shughni* narratives, the mean story structure score was 6.9 out of 17 (score range: 2–11). Only 2.3% of all episodes consisted of a GAO-sequence, i.e. were complete episodes. The percentage of children who produced at least one GAO-sequence in Shughni, was 12.5%. On average, the Shughni narratives contained 2.5 IST tokens. For *Tajik*, the numbers were similar: the mean story structure score was 6.8 out of 17 (score range: 3–9) and only 2.9% of the episodes were complete (GAO-sequences); 13.6% of the children produced at least one GAO-sequence in Tajik. The average number of IST tokens in Tajik was 2.7.

A more detailed analysis comparing narratives elicited in the telling and retelling modes revealed that the mean scores for story structure in Shughni were slightly higher in the retelling as compared with telling: 8 points vs. 6.7 points. However, the mean scores in the Tajik narratives were equal: 6.8 points. It seems that story structure in the reproduction of the story in Shughni, which the children heard beforehand, is more developed than in storytelling.

A comparison of the number of Goals across the three episodes yielded interesting results. In both languages, the number of Goals was much lower in the third episode: 0 for Shughni and 3 for Tajik. In contrast, the number of Goals in the first episode was 10 for Shughni and 11 for Tajik, and the second episode elicited a total of 17 Goals in the Tajik narratives and 16 Goals in the Shughni narratives.

5   Conclusion

This paper gave a brief overview of the language situation in the Republic of Tajikistan and presented a summary characterization of the Tajik and Shughni languages. The current example of the MAIN implementation contributes to the growing body of research on understudied languages and provides useful information on the two Indo-Iranian languages (Tajik and Shughni).

6   References

Armon-Lotem, S., de Jong, J. & Meir, N. (Eds.). (2015). *Assessing multilingual children: Disentangling bilingualism from language impairment*. Bristol, UK: Multilingual Matters.

Avezova, B. (2017). The influence of Russian on Tajik. In Polenova, et al. (eds.), *Language and Speech in Synchrony and Diachrony*: Papers from an International Linguistics Conference. (pp. 180–203). Cambridge Scholars Publishing.

Bohnacker, U. & Gagarina, N. (2019). Background on MAIN – Revised, how to use it and adapt it to other languages. *ZAS Papers in Linguistics*, *63*, iv-xii.

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy*, *18*(1), 1–21.

Dodikhudoeva, L. (2004). The Tajik Language and the Socio-linguistic situation in the Mountainous Badakhshan, *Iran and the Caucasus*, *8*(2), 281–288.

Dovudov, G., Pomikálek, J., Suchomel, V., Šmerk, P. (2011). In Building a 50M Corpus of Tajik Language. Aleš Horák, Pavel Rychlý (Eds.), *Proceedings of Recent Advances in Slavonic Natural Language Processing* (pp. 89–95). Tribun EU.

Ethnologue (2020). Languages of the world. www.ethnologue.com (accessed: 19.06.2020).

Fleckstein, A., Prévost, P., Tuller, L., Sizaret, E. & Zebib, R. (2018) How to identify SLI in bilingual children: A study on sentence repetition in French. *Language Acquisition, 25*(1), 85–101.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). MAIN: Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics*, *56*.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2015). Assessment of Narrative Abilities in Bilingual Children. In Armon-Lotem, et al. (Eds.), *Assessing multilingual children disentangling bilingualism from language impairment* (pp. 243–269). Bristol, UK: Multilingual Matters.

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., & Walters, J. (2019). MAIN: Multilingual Assessment Instrument for Narratives – Revised. *ZAS Papers in Linguistics*, *63*.

Grimm, A., & Schulz, P. (2014). Specific Language Impairment and Early Second Language Acquisition: The Risk of Over- and Underdiagnosis. *Child Indicators Research*, *7*(4), 821–841.

Mueller, K. S. (2015). Deixis in Shughni: Grammatical and Semantic Considerations. ND: University of North Dakota MA thesis.

Olson, K. (2017). Shughni phonology statement. *Silk Road Ascent, SIL International*. www.zabanha.af (accessed: 19.06.2020).

Parker, C. (2020). Vestigial ergativity in Shughni: At the intersection of alignment, clitic doubling, and feature-driven movement. *Glossa: A Journal of General Linguistics*, *5*(1), 52.

The constitution of the Republic of Tajikistan (2003). www.unesco.org (accessed: 19.06.2020).