Code-Switching Production Patterns Predicting Comprehension in Yakut-Russian Bilinguals

Research proposal

Anna Petukhova

Research Master *Linguistics and Communication Sciences*, Radboud University Nijmegen

Abstract

Switching between languages is a distinctive feature of being bilingual. Investigating code-switching in psycholinguistics can reveal the mechanisms behind language interaction in a bilingual mind since two language systems are activated. In this study, I will be testing a hypothesis introduced in Tamargo et al. (2016) that code-switching comprehension reflects production patterns of code-switching by using the data from eye movements of young adult Yakut-Russian bilinguals during a reading task. The prediction is that switches at more frequently witnessed syntactic sites (verb) in speech corpora would result in shorter reading times than switches that occur at less often observed boundaries (e.g., with a modal particle). In addition, participants will be divided into two groups based on their code-switching behavior as code-switching processing may differ depending on the frequency of being involved in such practice. The findings will contribute to the research of contact-induced phenomena in typologically distinct languages and the investigation of the relationship between production and comprehension.

Keywords: bilingualism, code-switching, language comprehension, eye-tracking

1. Introduction

Code-switching, 'the use of two language varieties in the same conversation' (Myers-Scotton, 2005, p. 239), is a widespread language phenomenon among bilinguals. There are two types of code-switching being distinguished: intersentential, code switches that happen between clauses, and intrasentential, code switches that occur within the clause (Poplack, 2013).

From a psycholinguistic perspective, code-switching draws attention as a window to the cognitive mechanisms behind the competition between languages as both languages remain activated (Traxler, 2011, p. 416). One prominent topic in the research of code-switching is switching costs which is a difference in comprehension time when encountering a switch or a non-switch. It is usually asymmetric, especially in unbalanced bilinguals: switching to a dominant language takes longer time than switching to a non-dominant one (Costa & Santesteban, 2004).

Experimental paradigms in the research on code-switching include corpus (e.g., Fricke & Kootstra, 2016), elicitation (e.g., Kootstra et al., 2020), self-paced reading (Adler et al., 2020), rapid serial visual presentation, and eye-tracking (e.g., Lipski, 2020) studies. The advantages of the latter method in code-switching processing research, according to Kroff et al. (2018), overcome such methodological challenges as the artificialness of the task and stimuli, allowing higher ecological validity in the studies of reading comprehension. In addition, using corpora data based on spontaneous speech as experimental material complements ecological validity even further.

Code-switching research based on corpus data has a number of advantages. First, it allows for high generalizability of the findings as they can be based on rich data. Second, corpora data, which is usually based on hundreds of hours of natural speech, allows for high ecological validity. Computer linguistics tools enable relatively fast and easy data extraction of any language phenomena of interest if it is annotated in the corpus.

One study that implemented both corpus and eye-tracking methods is a paper by Tamargo and colleagues (2016). It tested a hypothesis that distributional patterns of code-switching influence code-switching comprehension by implementing statistical knowledge to predict cues that will be followed by a code-switch. First, the authors made an exploratory spontaneous speech corpus analysis to determine frequent constructions where code-switch happens. They found out that there were more instances of code-switching at present participle or its auxiliary than at perfect participle and its auxiliary when switching from Spanish to English, which allowed them to build their hypothesis further, stating that they would expect higher processing costs for the code switches at perfect participle. Next, Tamargo et al. (2016) performed an eye-tracking experiment with Spanish-English bilinguals with early (*N*=42) and late exposure to code-switching (*N*=27). Participants' first-pass reading time and total time were measured while they were reading sentences in four conditions: 1) a code switch at the auxiliary (present), 2) a code switch at the participle (present), 3) a code switch at the auxiliary (perfect) and 4) a code switch at the participle (perfect). Linear mixed effects models analysis showed that both groups' comprehension facilitated from the production patterns found in corpora: code switches at the perfect auxiliary appeared to be more costly than code switches at present auxiliary, confirming the hypothesis.

The current study aims to provide more evidence for the patterns of code-switch comprehension for the pair of understudied and unrelated languages, Yakut (>Turkic) and Russian (>Indo-European). Yakut language is predominantly spoken on the territory of Sakha Republic, Russia, has a governmental status, and is compulsorily taught in schools (as L2 in Russian-speaking schools). The majority of the Yakut population speaks Russian which is the national language in all subjects of the Russian Federation. The use of Russian, however, is more predominant in towns, while Yakut is more frequently spoken in rural areas. For most families, Yakut is mainly spoken at home, as Russian remains to be a primary language of education. Hence, it is quite common to be a balanced Yakut-Russian bilingual in terms of the frequency of use, although the domains might not overlap.

The study follows the research line by Tamargo and colleagues (2016) and will investigate the influence of code-switching patterns in spoken production on reading comprehension. Even though production and comprehension modalities differ, Kroff and colleagues (2018) argue the patterns of code-switching are equivalent in written and spoken language. Narrowing the research question down, I will look at the presence of switching costs, reflected in the length of the reading time in an eye-tracking experiment, at the verb at the ending position when switching from Russian to Yakut. It has been attested in a Yakut-Russian spontaneous speech corpora (Petukhova & Sokur, 2021) that intrasentential switches occur predominantly when switching from Yakut to Russian, and the most frequent syntactic site that a code switch happens at is a verb/verb phrase. Thus, I hypothesize that code switches at the boundary with a verb will be read faster than switches with a boundary on other syntactic phrases. Furthermore, participants who code-switch regularly in their daily life are expected to read code switches quicker than the participant who rarely participate in such a language behaviour. In addition, the study can also pave the way for other research in Russian and a minority language pair. There are more than a hundred languages spoken in different regions of Russia that belong to various language families, especially with a lot of variety in the East Caucasus area.

2. Methods

2.1 Participants

Balanced/early young adult Yakut-Russian bilinguals will participate in the study, as the corpus I base our hypothesis on represents speech of that sample of a population. Participants will fill in a language background questionnaire (Anderson et al. 2018), including questions about code-switching habits (Rodriguez-Fornells et al., 2012), and complete language proficiency tasks in both languages. Participants whose scores in both tasks would differ in more than two standard deviations will be excluded from the analysis. Next, participants will be divided into the groups of frequent and non-frequent code switchers based on the questionnaire scores: participants with a mean score of up to 2 will be considered non-frequent switchers, participants with a mean score of 4 to 5 will be considered frequent switchers, the data of the rest will not be included into the analysis (5-point Likert scale used for the evaluation). Previous research showed that production can influence comprehension of code-switches (e.g., Beatty-Martínez & Dussias, 2017), so the differentiation based on the frequency of code switching might reveal differences in code-switch processing as well. Participants will be matched on age, socio-economic status, and place of residence to exclude the influence of dialect differences to avoid possible confounding variables and identify the specific linguistic factors that influence code-switching. Participants will be recruited via advertisement in social media, messengers, and word of mouth and will be compensated for their time.

The number of participants will be calculated using the sjstats package (Lüdecke, 2021) in R (R Core Team, 2022) with a large effect size determined by Cohen’s d (0.8).

2.2 Materials

Following Keating's (2014) suggestion, 12 experimental stimuli will be used per condition, resulting in 24 items: condition 1 is a switch at the verb at the end of the sentence (Example 1), and condition 2 is a switch at the last word of the sentence which is a modal particle (Example 2). The latter was chosen to be a control condition, as it has been attested to be less common than a code switch at the verb in the corpus. In addition, 48 fillers will be embedded into the final list of stimuli, consisting of sentences with one-word long switches at different positions in a sentence, except for the final one. All stimuli will represent switches from Russian to Yakut according to the switching frequency asymmetry reported in the corpus. To prevent spill-over effects, each stimulus will consist of two sentences: the first sentence will be a sentence starting in Russian with a code-switch to Yakut at the end of a sentence, and the second sentence will be a monolingual sentence in Russian (Russian and Yakut share the writing system). Regions of interest (ROIs) will be set at the first two words of the first sentence to analyze the factors influencing comprehension before encountering a code -switch; a critical region of a code switch, and two words at the beginning of the second sentence to analyze a spill-over effect of a code switch processing.

(1a)	*Представляешь,*	*сегодня*	*я*	*успел*	*сделать*
	imagine.IPF.PRS.2.SG	today	1.SG.NOM	manage.PRF.PST.M.SG	do.PRF.INF

все	*дела,*	*полезный*	*день*	*ааспыт.*
all.PL.ACC	deed.PL.ACC	efficient.M.SG.NOM	day.SG.ACC	pass.PRF.PST.M/N.SG

'Imagine, today I have managed to complete all my tasks, it was an efficient day.'

(1b)	*Надеюсь,*	*завтрая*		*наконец-то*	*доберусь*
	hope.IPF.PRS.1.SG	tomorrow	1.SG.NOM	finally	get.PRF.FUT.1.SG

до	*бассейна,*	*давно*	*там*	*не*	*был.*
to	pool.SG.GEN	long.ago	there	not	be.IPF.PST.M.SG

'Hope that tomorrow I will finally get to the pool, I haven't been there for a long time.'

(2a)

Ты	*не*	*купила*	*стиральный*	*порошок,*
2.SG.NOM	not	buy.PRF.PST.F.SG	washing.M.SG.ACC	powder.SG.ACC

как	*я*	*тебя*	*просил,*	*да?*
how	1.SG.NOM	2.SG.ACC	ask.IPF.PST.M.SG	MOD

'You haven't bought the washing powder as I asked you?'

(2b)

Тогда	*мы*	*перенесем*	*стирку*	*на*	*завтра,*
then	1.PL.NOM	postpone.PRF.FUT.1.PL	wash.SG.ACC	to	tomorrow

когда	*я*	*схожу*	*в*	*магазин.*
when	1.SG.NOM	go.PRF.FUT.1.SG	to	store.SG.ACC

'Then we'll postpone the laundry for tomorrow, when I go to the store.'

All sentences will be constructed to match in length (8-10 words, following corpus data). The words at the code-switch will be matched in length and frequency to meet the criteria of equal predictability. Moreover, as Yakut script has additional characters not present in Russian script, critical words will be chosen in such a way that no characters absent in Russian will be in a word to keep spelling uniform.

Stimuli will be followed by comprehension questions in Russian to ensure that participants pay attention and read sentences carefully for better reflection of natural reading patterns. Experimental items will comprise two lists, as there are two experimental conditions, and participants will be assigned the lists in a random order. Within each list stimuli will be presented in a pseudorandomized order, so no stimulus type (code switch at a verb, code switch at a modal particle, filler) will be encountered by a participant more than three times in a row.

2.3 Procedure

Participants will be sat in a noise-isolated booth 70 cm away from a computer monitor using a chin rest to prevent head movement. Eye movements will be recorded using a desktop Eyelink 1000 system (SR Research, Mississauga, Ontario, Canada) with the camera located below the screen. Monocular eye movements of a dominant eye will be recorded at a sampling rate of 1000 Hz given the size of the stimuli (text) and ROIs. Text will be presented in a Courier New font with 14 pt size and double line spacing. At the beginning of the experiment and after each break (two breaks in total, after 24 stimuli) eye tracker will be calibrated with the accuracy rate set to 0.3°. Participants will perform a nine-point calibration and validation before each trial. It will start with a calibration point appearing on the left side of the screen at the position of the first word in the sentence.

Participants will be asked to read sentences as they would usually read them, followed by a comprehension question regarding the content of the sentences to which they would need to answer 'yes' or 'no' by pressing Q and P buttons on the keyboard respectively. Prior to the experiment, participants will fill out the language background questionnaire and complete language proficiency tasks. At the beginning of the experiment, participants will perform a practice trial consisting of six stimuli to get acquainted with the procedure.

3. Data analysis

In the analysis, I will use first-pass reading time (a sum of fixation durations between first entering an ROI and first leaving it), and total time (a sum of all fixation duration in an ROI), following the experimental setting in Tamargo et al. (2016). As suggested by Conklin et al. (2017), inaccurate trials (trials where there was a wrong answer to a comprehension question), first fixation, and fixations with a duration of less than 100 ms. or two standard deviations above and below the participant's mean will be excluded from the analysis.

Additionally, linear mixed-effects models will be used via the lme4 package (Bates et al., 2015) in R (R Core Team, 2022), with a switch type (verb or modal particle) and participant group (frequent or non-frequent switchers) and their interaction as fixed factors and stimuli and participants as random factors.

References

Anderson, J. A. E., Mak, L., Keyvani Chahi, A., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods, 50*(1), 250–263. https://doi.org/10.3758/s13428-017-0867-9.

Adler, R. M., Valdés Kroff, J. R., & Novick, J. M. (2020). Does integrating a code-switch during comprehension engage cognitive control? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(4), 741. https://doi.org/10.1037/xlm0000755.

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Beatty-Martínez, A. L., & Dussias, P. E. (2017). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of Memory and Language, 95*, 173-189. https://doi.org/10.1016/j.jml.2017.04.002.

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research.* Cambridge University Press.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language, 50*(4), 491-511. https://doi.org/10.1016/j.jml.2004.02.002.

Fricke, M., & Kootstra, G. J. (2016). Primed codeswitching in spontaneous bilingual dialogue. *Journal of Memory and Language, 91*, 181-201. https://doi.org/10.1016/j.jml.2016.04.003.

Keating, G. D. (2014). Eye-tracking with text. *Research Methods in Second Language Psycholinguistics*, 69-92.

Kootstra, G. J., Dijkstra, T., & Van Hell, J. G. (2020). Interactive alignment and lexical triggering of code-switching in bilingual dialogue. *Frontiers in Psychology*, 1747. https://doi.org/10.3389/fpsyg.2020.01747.

Kroff, J. R. V., Tamargo, R. E. G., & Dussias, P. E. (2018). Experimental contributions of eye-tracking to the understanding of comprehension processes while hearing and reading code-switches. *Linguistic Approaches to Bilingualism, 8*(1), 98-133. https://doi.org/10.3389/fpsyg.2020.01747.

Lipski, J. M. (2020). Pronouns, Interrogatives, and (Quichua-Media Lengua) Code-Switching: The Eyes Have It. *Languages, 5*(2), 11. https://doi.org/10.3390/languages5020011.

Lüdecke, D. (2021). *sjstats: Statistical Functions for Regression Models (Version 0.18.1).*

Myers-Scotton, C. (2005). *Multiple voices: An introduction to bilingualism.* Wiley-Blackwell

Petukhova, A. A., & Sokur, E. O. (2021). *Yakut-Russian Corpus of Code-Switching*, Moscow: International Laboratory of Language Convergence, Higher School of Economics. Available online at: http://lingconlab.ru/cs_yakut, accessed on 06.04.2022.

Poplack, S. (2013). "Sometimes i'll start a sentence in spanish y termino en español": toward a typology of code-switching. *Linguistics, 51*(S1), 11–14. https://doi.org/10.1515/ling-2013-0039.

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from https://www.R-project.org/.

Rodriguez-Fornells, A., Kramer, U., Lorenzo-Seva, U., Festman, J., & Münte, T. F. (2012). Self-assessment of individual differences in language switching. *Frontiers in Psychology, 2*, 388. https://doi.org/10.3389/fpsyg.2011.00388.

Tamargo, R. E. G., Kroff, J. R. V., & Dussias, P. E. (2016). Examining the relationship between comprehension and production processes in code-switched language. *Journal of Memory and Language, 89*, 138-161. https://doi.org/10.1016/j.jml.2015.12.002.

Traxler, M. J. (2011). *Introduction to psycholinguistics: Understanding language science.* Wiley-Blackwell.zz

Appendix A: List of Used Abbreviations

1	first person	N	neuter
2	second person	NOM	nominative
3	third person	MOD	modal particle
ACC	accusative	PASS	passive
F	feminine	PL	plural
FUT	future	PRF	perfect
GEN	genitive	PRS	present
INF	infinitive	PST	past
IPF	imperfective	SG	singular
M	masculine		