



Heritage Speakers as Part of the Native Language Continuum

Heike Wiese^{1*}, Artemis Alexiadou^{2,4†}, Shanley Allen^{5†}, Oliver Bunk^{1†}, Natalia Gagarina^{1,4†}, Kateryna Iefremenko^{6†}, Maria Martynova^{3†}, Tatiana Pashkova^{5†}, Vicky Rizou^{2†}, Christoph Schroeder^{6†}, Anna Shadrova^{1†}, Luka Szucsich^{3†}, Rosemarie Tracy^{7†}, Wintai Tsehaye^{7†}, Sabine Zerbian^{8†} and Yulia Zuban^{8†}

¹ Department of German Language and Linguistics, Humboldt-Universität zu Berlin, Berlin, Germany, ² Department of English and American Studies, Humboldt-Universität zu Berlin, Berlin, Germany, ³ Department of Slavic and Hungarian Studies, Humboldt-Universität zu Berlin, Berlin, Germany, ⁴ Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin, Germany, ⁵ Center for Cognitive Science, Technische Universität Kaiserslautern, Kaiserslautern, Germany, ⁶ Department of German, Universität Potsdam, Potsdam, Germany, ⁷ Department of English, Universität Mannheim, Mannheim, Germany, ⁸ Department of Linguistics, Universität Stuttgart, Stuttgart, Germany

OPEN ACCESS

Edited by:

Antonella Sorace,
University of Edinburgh,
United Kingdom

Reviewed by:

Sarah Schimke,
Technical University Dortmund,
Germany
Archana Bhatia,
Florida Institute for Human
and Machine Cognition, United States

*Correspondence:

Heike Wiese
heike.wiese@hu-berlin.de

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 31 May 2021

Accepted: 25 November 2021

Published: 09 February 2022

Citation:

Wiese H, Alexiadou A, Allen S,
Bunk O, Gagarina N, Iefremenko K,
Martynova M, Pashkova T, Rizou V,
Schroeder C, Shadrova A,
Szucsich L, Tracy R, Tsehaye W,
Zerbian S and Zuban Y (2022)
Heritage Speakers as Part of the
Native Language Continuum.
Front. Psychol. 12:717973.
doi: 10.3389/fpsyg.2021.717973

We argue for a perspective on bilingual heritage speakers as native speakers of both their languages and present results from a large-scale, cross-linguistic study that took such a perspective and approached bilinguals and monolinguals on equal grounds. We targeted comparable language use in bilingual and monolingual speakers, crucially covering broader repertoires than just formal language. A main database was the open-access RUEG corpus, which covers comparable informal vs. formal and spoken vs. written productions by adolescent and adult bilinguals with heritage-Greek, -Russian, and -Turkish in Germany and the United States and with heritage-German in the United States, and matching data from monolinguals in Germany, the United States, Greece, Russia, and Turkey. Our main results lie in three areas. (1) We found non-canonical patterns not only in bilingual, but also in monolingual speakers, including patterns that have so far been considered absent from native grammars, in domains of morphology, syntax, intonation, and pragmatics. (2) We found a degree of lexical and morphosyntactic inter-speaker variability in monolinguals that was sometimes higher than that of bilinguals, further challenging the model of the streamlined native speaker. (3) In majority language use, non-canonical patterns were dominant in spoken and/or informal registers, and this was true for monolinguals and bilinguals. In some cases, bilingual speakers were leading quantitatively. In heritage settings where the language was not part of formal schooling, we found tendencies of register leveling, presumably due to the fact that speakers had limited access to formal registers of the heritage language. Our findings thus indicate possible quantitative differences and different register distributions rather than distinct grammatical patterns in bilingual and monolingual speakers. This supports the integration of heritage speakers into the native-speaker continuum. Approaching heritage speakers from this perspective helps us to better understand the empirical data and can shed light on language variation and change in native grammars. Furthermore, our findings for monolinguals lead us to

reconsider the state-of-the art on majority languages, given recurring evidence for non-canonical patterns that deviate from what has been assumed in the literature so far, and might have been attributed to bilingualism had we not included informal and spoken registers in monolinguals and bilinguals alike.

Keywords: heritage speakers, registers, participles, word order, bare NPs, boundary tone, referent introduction, relative clause formation

INTRODUCTION: HERITAGE SPEAKERS AS NATIVE SPEAKERS

Heritage speakers are speakers who grow up in a bi- or multilingual home with a minority language in addition to the majority language(s) dominant in the larger society (see, e.g., Montrul and Polinsky, 2011). Accordingly, these are speakers who acquire their heritage language early and naturally in a home environment as a first language, but who also acquire another language early on, which is more dominant in the larger society and will often be the only language supported in the formal context of schooling. This constitutes an interesting population that poses a challenge for the question of what counts as a “native speaker,” which has traditionally been conceptualized mostly from a monolingual standpoint [cf. criticism in Ortega (2009) and Cook (2016)]. Investigating heritage speakers can help us unravel underlying assumptions of “nativeness,” and contribute to our understanding of native grammars, language variation and change (e.g., Polinsky, 2015, 2018; Lowe, 2020).

The concept of “native speaker” as used in linguistic research can involve at least three different types of assumptions. First, a basic assumption is the requirement that a native speaker is “native” in the sense of being born into the language, that is, acquiring it from birth. In this sense, heritage speakers are uncontroversially native speakers of their heritage language, since they acquire it as a first language in a home environment. However, this is not the only requirement for native speakers used in the literature.

A second – implicit or explicit – requirement, often found in heritage language research, is that in order to be fully recognized as such, a native speaker has to master a repertoire that also involves standard or formal registers of their heritage language. For instance, Montrul and Polinsky (2019), while explicitly acknowledging that heritage speakers “are native speakers of their heritage language” (p. 420), require a range of registers, including formal writing, for a “proficient native speaker” (p. 426; cf. also Montrul, 2008:109). Since such registers are learned primarily in the context of formal education, heritage speakers often only acquire them for the majority language, which would then exclude them from the group of proficient native speakers of their heritage language.

It is not clear, though, why specific registers should be a necessary part of native-language proficiency, since the development of register distinctions is linked to social and communicative requirements that vary across social groups. This is independent of bi- or monolingualism, and in the case of monolinguals, it usually does not affect our view of speakers as native. For instance, we would not assume that a language

without formal writing does not have any proficient native speakers, for instance in historic stages before the invention of writing, or for minority languages without a written code, and we would not claim that preschool children and people who are illiterate are not native speakers of their language.

Furthermore, register knowledge may differ substantially across (monolingual) speaker groups and registers develop throughout the lifespan, so it is not clear why some registers, but not others, would be required for native-speaker proficiency. For instance, university students will typically acquire new spoken and written registers characteristic for academia, but this will not be regarded as a requirement for being a proficient native speaker, and it is uncontroversial that monolinguals without university experience are proficient native speakers of their languages.

This suggests that we ought to keep nativeness separate from register distinctions, so that speakers who use, for instance, only informal and/or spoken registers of their language, will also be considered proficient native speakers. This underlines the importance to tap into linguistic competences supporting language use in different real-life settings, including those outside standard language (cf. also Bayram et al., 2019). Avoiding a standard language bias in our research is also important given that formal standard varieties are exceptional: they are subject to codified norms that hamper normal patterns of language variation and change, and thus should be considered peripheral rather than central instances of native language grammars.

A third kind of assumption is related to attitudinal and language-ideological patterns constructing a native speaker as someone who has grown up monolingually. Heritage speakers are primarily investigated in societies with a strong monolingual habitus (Gogolin, 1994, 2002), in particular countries based on European nation state building (including those that developed out of former European settler colonies, e.g., the United States). Against this background, a “native speaker” is often taken to be monolingual [cf. criticism in Brutt-Griffler and Samimy (2001), Bonfiglio (2010), Cook (2016), and Ortega (2016)]. This conceptualizes monolinguals as the primary owners of a language, and as the gold standard for linguistic competence and attainment. Such a conceptualization was already implicit in earlier structuralist idealizations, such as the Chomskian “ideal speaker-listener” (Chomsky, 1965) or Saussure’s focus on one-to-one correlations of language and place as the “forme idéale” (de Saussure, 1916: Part 4, Ch. 2, §1). In current studies, it is implied in the use of monolinguals as a control group to test “native-like” behavior or “native competence” in heritage speakers, or to test whether some areas “develop at native levels” in heritage grammars.

However, such an idealization of the monolingual speaker as the primary bearer of a native language is not reasonable, and it is not even feasible. It is not reasonable given that multilingualism is the normal condition for human language, and, as has been amply stated (Grosjean, 1982, 2010; Romaine, 1989; Myers-Scotton, 2006, and others), most speakers in today's world are multilingual. A focus on monolinguals as native speakers in linguistics makes as little sense as the focus on males that has lately been criticized in medical research (Criado-Perez, 2019; McGregor, 2020). Medical research has, for a long time, focused only on males because studies did not want their data to be affected by hormonal changes thought to be characteristic for female bodies – but hormonal changes are part of the human condition, and if we want to know something about humans, we have to include females. In the same vein, if we want to know something about language, we have to include multilinguals, because being linguistically multi-competent is part of the human condition when it comes to language (e.g., Cook, 2016).

A restriction to monolinguals is not even feasible for empirical research, because it is not clear who would qualify as a “true monolingual.” Language is always variable, speakers' repertoires always involve a range of options which could be captured as different grammars (e.g., Tracy, 2002; Roeper, 2003), and the interaction of linguistic resources within repertoires is not categorically different for languages versus dialects, registers, or styles (cf. Li Wei, 2016). This suggests that there is no clear-cut distinction between bilingual speakers who use different languages and “monolinguals” whose repertoire will always include at least different registers. Furthermore, cross-linguistic effects on the L1 have even been attested for monolingual speakers who learn a second language in an instructed, non-immersion setting (Schmid and de Leeuw, 2019). Hence, if we restrict native-speaker status to monolinguals in a strict sense, then most of the world's population would not count as native speakers – including most linguists today, given that most of us are fluent L2 English speakers, and thus could no longer count as native speakers of our L1.

This calls for a perspective that integrates heritage speakers into the native language continuum. Heritage language research as a field has shown that language is flexible and open to change over the lifespan. We believe that the time is ripe to take a further step and to take seriously the fact that heritage speakers are native speakers of both their languages, as emphasized in recent discussions of heritage speakers and bilingualism¹. In what follows, we show what this means in terms of a research programme that does not take monolingual standard norms as a yardstick to identify what is missing or incorrect in heritage speakers' language use. On the basis of findings from cross-linguistic research, we show what is to be gained by overcoming a deficit-oriented view of heritage speakers. To this end, we explore the dynamics, rather than the vulnerability, of different linguistic domains and investigate development, variation, and innovation, rather than incomplete acquisition, attrition and loss.

Crucially, this means (1) that we do not concentrate on standard language and formal registers alone, but capture speakers' broader repertoires, including informal and spoken language, and (2) that we target heritage speakers and monolinguals alike – not as test group vs. control group, but as two groups of native speakers that we expect to both show interesting patterns of language variation.

In what follows, we present results from a large-scale, cross-linguistic investigation that realized such a research programme within the context of the Research Unit “Emerging Grammars in Language-Contact-Situations: A Comparative View” (short “RUEG”)². In our investigation, we approached bilinguals and monolinguals as two speaker groups to be investigated, rather than experimental vs. control group. Accordingly, we cast our net wide and targeted non-canonical patterns in general, that is, all patterns that would not be expected in standard grammar, and we did this for monolingual and bilingual groups alike.

This yielded a range of novel findings across languages, not only for heritage speakers, but also for the monolingual groups. In the following sections, we discuss evidence showing that a range of non-canonical phenomena in heritage speakers are also at work in monolingual speakers, pointing to language-internal tendencies of variation and change. These findings place multilinguals at the forefront of linguistic dynamics, and further support the integration of heritage speakers into the native language continuum. We have argued above that recognizing heritage speakers as native speakers is justified on conceptual and theoretical grounds. In what follows, we show that this perspective is also a better fit for the empirical data. We found a range of patterns that would be surprising if we saw monolinguals as a measure for “nativeness” and bilinguals as the deviant group. In contrast, these findings make a lot of sense if we see both groups as part of the native speaker continuum.

MATERIALS AND METHODS

The methods we used to elicit data meet the two demands formulated above: we need to include informal and spoken language, and we need to target non-canonical patterns and variation in bilinguals and monolinguals alike. In order to achieve this, we used the “Language Situations” (“LangSit”) set-up, which avoids a restriction to formal language and taps into broader repertoires across speaker groups (cf. Wiese, 2020). In this set-up, participants are familiarized with a fictional event (e.g., a car accident) and are asked to imagine themselves as a witness to this event, and then act out telling different interlocutors about it in different communicative situations. This yields naturalistic productions that are comparable across speaker groups, languages, and settings.

²<https://www.linguistik.hu-berlin.de/en/rueg>

³Funded by Deutsche Forschungsgemeinschaft, DFG (FOR 2537). Speaker: Heike Wiese; further PIs: Artemis Alexiadou, Shanley Allen, Oliver Bunk, Natalia Gagarina, Mareike Keller, Anke Lüdeling, Judith Purkharthofer, Christoph Schroeder, Anna Shadrova, Luka Szucsich, Rosemarie Tracy, Sabine Zerbian; postdoc: Kalliopi Katsika; Ph.Ds: Katerina Iefremenko, Esther Jahns, Martin Klotz, Thomas Krause, Annika Labrenz, Maria Martynova, Katrin Neuhaus, Tatiana Pashkova, Vicky Rizou, Wintai Tsehaye, and Yulia Zuban.

¹E.g., Rothman and Treffers-Daller (2014); Guijarro-Fuentes and Schmitz (2015), Schroeder (2016); Kupisch and Rothman (2018), Aalberse et al. (2019); Lohndal et al. (2019), Embick et al. (2020), and Flores and Rinke (2020).

All materials developed for RUEG's investigation, including stimuli, elicitor instructions, and a training video for elicitors, have been stored with the Open Science Foundation for open access at <https://osf.io/cm96g/>.

Stimuli

For our investigation, we developed a video showing a (minor) car accident that involved a young woman with a dog, a couple with a baby in a pram, and two cars. In this video, one sees the couple approaching a car park, with the man bouncing a ball. Across a lane, the woman with the dog is unloading groceries from her car. The two cars are seen approaching the lane, when suddenly the man loses control of his ball, which bounces in front of the first car. On the other side, the dog gets excited and runs into the lane toward the ball, and the woman drops her groceries. The first car comes to an abrupt halt, causing the second one to bump into it. The man with the ball helps the woman pick up her groceries, the two drivers get out of their cars, and one of them calls the police, which can be seen through a close-up of the emergency number on his phone.

We developed five versions for five countries (see below): Germany, Greece, Russia, Turkey, and the United States. In order to support cross-linguistic comparisons, these versions only differed with respect to the emergency number, but were otherwise identical.

Procedure

Participants or their parents, in the case of adolescents, gave informed consent. For the elicitation, they saw the video, were asked to imagine themselves as a witness to the accident, and then had to play-act telling different interlocutors about it. We constructed four different communicative situations by manipulating formality and mode: participants were asked to:

- (1) Leave a voice message for a friend, via instant messenger (informal-spoken).
- (2) Write a message to a friend, via instant messenger (informal-written).
- (3) Leave a voice message on a police "witness line" (formal-spoken).
- (4) Write a witness report for the police (formal-written).

For the informal language productions (1 and 2), participants used the WhatsApp® messenger on a mobile phone provided by the elicitor, where auto correction, swiping, and suggestions had been switched off. The formal-spoken message (3) was produced on the same phone, as a voice mail to the mail box of a (fictional) contact "Police Department – eyewitness line." The formal report (4) was typed in using a simple text editor on a laptop, with spelling correction switched off.

For the different language productions, the video was shown several times. Informal versus formal productions were elicited in two different rooms that were suitably decorated according to the (in-)formality, and with two different interlocutors who acted and were dressed informally vs. formally. Short breaks filled with (in-)formal conversations divided informal and formal parts of an elicitation session.

At the end of data elicitation, participants were asked to fill in a sociolinguistic questionnaire on biographical data including language use and personality traits.

Bilingual speakers were recorded twice, in their heritage language and in the majority language, with the two sessions at least three days apart. Monolingual speakers were recorded once, in the majority language. Order of elicitation was counterbalanced for the four communicative situations, and, in the case of bilingual speakers, for the two languages.

Participants

Participants were heritage speakers and monolingual speakers. Heritage speakers were defined as speakers who had grown up with a family language in addition to the country's majority language. In order to participate, they had to use the heritage language regularly with at least some members of their nuclear family, and to be able to speak and write in it (although not necessarily in the standard alphabet). Further conditions were that they were born in the country of the respective majority language or had arrived there at an early age⁴ and that they had lived in that country since, although not necessarily without interruptions. Monolingual speakers were speakers who used only one language regularly at home, namely the respective country's majority language, although they might have acquired additional languages, for instance through formal education.

The bilingual group covered heritage speakers of Greek, Russian, and Turkish in Germany and in the United States, and of German as a heritage language in the United States. The monolingual group consisted of speakers of English, German, Greek, Russian, and Turkish in the United States, Germany, Greece, Russia, and Turkey, respectively. In all categories, we covered two age groups: adolescents (14–18 years), and adults (22–35 years).

Participants had no reported speech disorders and normal or corrected-to-normal hearing and vision.

Data Processing and Corpus Generation

Elicitations yielded matched elicited, semi-spontaneous data across registers, contact-linguistic settings, and bilingual and monolingual speaker groups, in five languages:

- German as a majority language in Germany spoken by monolingual speakers, and by bilingual speakers with Greek, Russian, or Turkish as heritage languages, and as a heritage language in the United States spoken by bilingual speakers with English as a majority language;
- English as a majority language in the United States spoken by monolingual speakers, and by bilingual speakers with German, Greek, Russian, or Turkish as heritage languages;
- Greek, Russian, and Turkish as majority languages spoken by monolingual speakers in Greece, Russia, or Turkey, respectively, and as heritage languages spoken by bilingual speakers with English or German as majority languages in the United States or Germany, respectively.

⁴In general before the age of 23 months, although in some cases this was extended (up to 4 years) where otherwise it would not have been possible to recruit enough speakers.

In what follows, we refer to languages spoken as majority languages, e.g., German in Germany, or Greek in Greece, as “majority German/Greek” or short “maj-German/- Greek,” etc., and to languages spoken as minority languages in a heritage context, e.g., German or Greek in the United States, as “heritage German/Greek” or short “h-German/-Greek,” etc. We will use “HS” as an abbreviation for “heritage speaker.” When we give examples, we provide the transcriptions for spoken data, and keep to the original spelling (including possible typos) in the case of written data.

Codes identifying data from the RUEG corpus provide the following information, in this order:

- Country: DE – Germany; GR – Greece; RU – Russia; TU – Turkey, and US – USA.
- Bi-/monolingual speaker: bi vs. mo.
- Speaker number incl. age group: 1–50 – adults; from 51 onward – adolescents.
- Gender: M vs. F (there were no speakers who identified as non-binary).
- Heritage language for bilingual speakers or only family language for monolinguals: D – German; E – English; G – Greek; R – Russian; and T – Turkish.
- Communicative situation: f – formal/i – informal and s – spoken/w – written.
- Language of production: D, E, G, R, and T.

For instance, “DEbi51MT_isD” identifies data in Germany (DE) from a bilingual (bi) adolescent (51) male (M) speaker with Turkish (T) as a h-language in an informal (i) spoken (s) setting, communicating in German (D).

All corpus data has been anonymized and integrated into a unified corpus, the RUEG corpus (Wiese et al., 2019). The RUEG corpus is a multimodal and multi-layer corpus, which in its current version (0.4.0) contains approximately 520,100 tokens⁵ (appr. 146,000 for English, 157,000 for German, 66,000 for Greek, 88,000 for Russian, and 63,000 for Turkish), based on data from 716 speakers, of whom 393 are bilingual and 323 monolingual. **Table 1** gives the details for the different data sets⁶.

At the time of writing, the corpus continues to grow, with more data sets and improved annotations added. Corpus data includes language productions in all four communicative situations, with additional transcriptions for spoken data (conditions 1 and 3), and the biographical data from the speaker questionnaires. Language productions are annotated for syntactic spans, lemmata, language, and parts of speech in a universal and a language-specific set of categories. The corpus can be used via the ANNIS corpus search and visualization tool (Krause, 2019). The complete corpus, including its source data and all preliminary versions, is freely available in an open repository (doi: 10.5281/zenodo.3236068).

⁵We define “token” as the minimal annotated unit in our corpus: a string of characters between two spaces in written text, or in the written transcription of spoken text. Thus, a token is typically a word (e.g., “dog”) or an emoji or emoticon (e.g., :-/). Punctuation marks and filled pauses (e.g., “um”) are not included as tokens.

⁶Note that the “bilingual” rows represent the same speakers on two lines – e.g., maj-German (44) and h-Greek (47) are the same speakers.

TABLE 1 | RUEG corpus data.

Country	Bi-/monolingual	Languages	# speakers	# tokens
DE	Bilingual	maj-German	44	21,339
		h-Greek	47	19,783
	Bilingual	maj-German	56	34,503
		h-Russian	58	32,882
	Bilingual	maj-German	65	35,881
		h-Turkish	65	23,722
	Monolingual	maj-German	64	50,706
	Bilingual	maj-English	34	16,765
United States	Bilingual	h-German	34	14,888
		maj-English	64	30,913
	Bilingual	h-Greek	64	18,032
		maj-English	65	36,021
	Bilingual	h-Russian	66	29,214
		maj-English	59	32,905
	Bilingual	h-Turkish	56	18,502
		maj-English	64	29,238
GR	Monolingual	maj-Greek	64	27,931
RU	Monolingual	maj-Russian	67	25,930
TU	Monolingual	maj-Turkish	64	20,947

RESULTS

Our analyses yield three main findings: (1) cross-linguistically, we find non-canonical patterns not only in heritage speakers, but also in monolinguals, including patterns that, according to the literature, would not be expected for monolinguals; (2) we find extensive variation not only in heritage speakers, but also in monolinguals; (3) non-canonical patterns interact with register, underlining the importance of taking into account both formal and informal settings and, crucially, doing so for multilinguals and monolinguals alike.

Non-canonical Patterns: Not Just in Heritage Speakers, but Also in Monolinguals

In order to demonstrate what can be gained by approaching heritage speakers as native speakers of both their languages, we present non-canonical patterns that we observed in both heritage speakers and monolinguals. These are patterns that have so far been considered absent from native grammars and which might have been attributed to bilingualism had we taken a less inclusive approach. We cover domains of morphology and syntax, intonation, and pragmatics. In what follows, we present results from different languages, combining, in each case, qualitative and quantitative analyses. Qualitative analyses capture the relevant patterns and their distribution across bilingual and monolingual speaker groups. Quantitative analyses compare frequencies between different groups. In domains where corpus frequencies are high enough, this is supported by statistical tests⁷. For lower-frequency phenomena, we provide comparative

⁷In cases where we were interested in the impact of multiple mixed effects and/or inter-individual variation, linear mixed models were applied. In other cases, where

figures for the different groups through relative (rather than absolute) numbers for non-canonical cases as a proportion of all relevant cases.

Morphology and Syntax

In the domain of morphology and syntax, examples come from the formation and use of participles in Russian, word order in German, and bare NPs in German.

Non-canonical Participles in Russian

Participles in Russian are challenging in their morphology and syntax, and they are acquired later by monolinguals (Cejtlin, 2009; Tribushinina et al., 2013), which makes them an ideal domain to look for non-canonical forms. Results of our corpus study show that morphologically non-canonical participles can be found across all speaker groups, including monolinguals:

- | | | |
|-------------------------|---------------|-------------------|
| (1) <u>v''ezžaščaja</u> | belaja mašina | beloe avto (-) a: |
| driving.in | white car | white car |
| pritormozil | | |
| slowed.down | | |

[DEbi39MR_fsR]

(canonical: v''ezžajuščaja)

“A white car is driving in, the white car (-) a: slowed down.”

- | |
|--|
| (2) voditeli avtomobilej, kak pervogo, tak i <u>sledovšego</u> |
| drivers of.cars like first so also following |

we simply wanted to look at frequency differences across groups, we used the non-parametric Wilcoxon rank sum test for non-normally distributed data and parametric ANOVAs for normally distributed data.

za nim, rezko pritormozili
after him abruptly slowed.down

[RUmo06MR_fwR]

(canonical: sledovavšego)

“The drivers of the cars, both the first one and the one that followed it, abruptly slowed down”

A qualitative analysis reveals interesting dynamics in the morphological formation of participles. With suffixes, there is a widespread truncation of material, which can be seen both in (1), produced by a HS, and (2), produced by a monolingual. In (1), the expected suffix *-jušč-* for the formation of active present participles of open stems ending with a *j*-addition (Bogdanov et al., 2009) is truncated to *-šč-*. In (2), the base for the formation of the active past participle with the suffix *-vš-*, consisting of the stem and a thematic suffix *sled-ova-* “follow,” is truncated to *sled-o-*. Such a pattern can be interpreted as a reduction of morphological complexity pointing at a tendency for stem unification across paradigms (Gagarina, 2002:160).

In order to check whether the frequencies of participles differ significantly across groups, we ran a one-tailed unpaired Wilcoxon rank sum test. Results show that HSs produced more non-canonical forms than monolinguals ($M = 0.96$, $SE = 0.48$): $W = 36,480$, $p = 0.023$ for HSs in the United States ($M = 4.64$, $SE = 1.29$), and $W = 31,997$, $p = 0.032$ for HSs in Germany ($M = 4.11$, $SE = 1.28$). Overall, HSs produced fewer participles, both canonical and non-canonical ones, than monolinguals ($M = 0.69$, $SE = 0.09$): $W = 28,455$, $p < 0.001$ for HSs in the United States ($M = 0.15$, $SE = 0.04$), and $W = 26,118$, $p < 0.001$ for HSs in Germany ($M = 0.22$, $SE = 0.04$), see **Figure 1** for

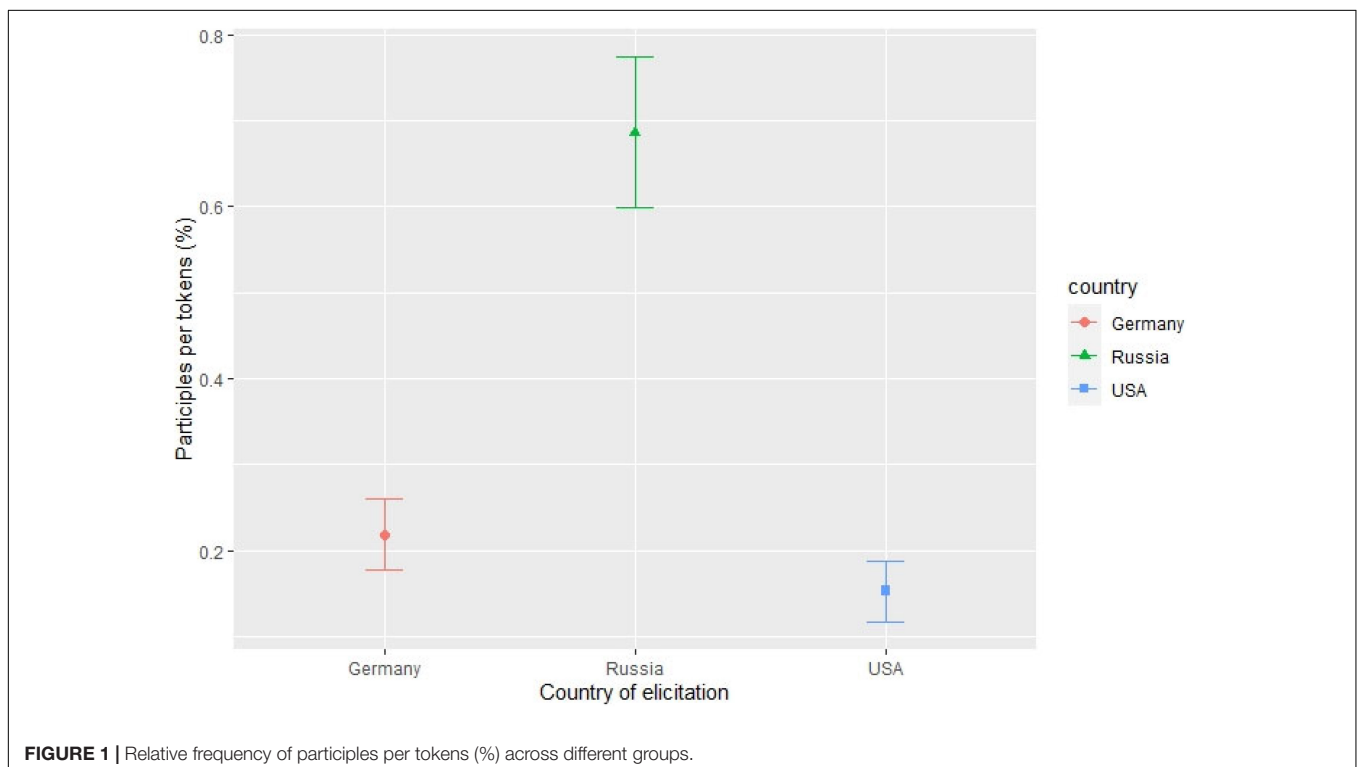


TABLE 2 | Frequency of tokens and participles across different groups.

Country	Tokens: overall	Tokens: canonical participles (% of all tokens)	Tokens: non-canonical participles (% of all tokens)
DE	32,882	57 (0.17%)	14 (0.04%)
US	29,214	40 (0.13%)	14 (0.05%)
RU	25,930	151 (0.58%)	6 (0.02%)

the relative frequencies and **Table 2** for absolute frequencies of tokens and participles across the groups. Both findings might be explained by the abovementioned status of participles, and by the fact that they are generally rare in HSs' oral input, since they are associated with formal registers (Zemskaja, 1973; Golub, 2001).

Non-canonical participle formation is also well documented for monolingual child acquisition (Cejtlin, 2009). Interestingly, such forms often follow the same patterns as those we find in our data:

- (3) Pečen'e uže *s'eto*. (Cejtlin, 2009:169)
cookies already eaten
(canonical: *s'edeno*)

"The cookies are already eaten"

Similar to (1) and (2), example (3) is a case of morphological truncation of a stem *s'ed-* "eat" - to *s'e-*, which then forms the base for participle formation with the suffix *-t-* rather than with canonical *-en(n)-*. The fact that we observed such patterns also in monolingual and bilingual (HS) adult speakers, suggests ongoing internal dynamics in this linguistic domain in Russian. That HSs use such a pattern with a higher frequency hence means that they can shed a spotlight on ongoing tendencies in native grammars.

Non-canonical Word Order in German

For German, we report relevant findings from two domains: word order and bare NPs. German has traditionally been described as an SOV language with verb-second (V2) word order requiring the finite verb in main declaratives to appear in second position, after exactly one constituent in the domain in front of it, the "forefield." This position of the finite verb constitutes one of two "sentence brackets" characteristic for the lay-out of German sentences. The other position is located at the right clausal periphery. In main declaratives, it contains non-finite verbs and separable verb particles. Together, the left and the right bracket delimit the "middle field," the canonical domain for complements and adjuncts. Typically, embedded clauses are extraposed, i.e., occur beyond the right sentential bracket in the "post-field."

The V2 requirement is usually regarded as a prime example of a rigid constraint in the grammar of German native speakers. Therefore, deviations from V2 in the maj-German of Turkish HSs, where an adverbial occurs in front of the subject at the left periphery, were taken to fall outside native German. For Auer (2013:37f), for instance, they indicated the reorganization of German V2 to SVO, which would "intervene deeply in the

structures of autochthonous German in its standard and non-standard forms" and, together with other non-canonical patterns, such as bare NPs, "would have the potential to constitute a new variety that would differ substantially from autochthonous German" (German originals, our translation).

This view has been challenged by accounts integrating this non-canonical pattern into the syntactic lay-out of German sentences (Wiese, 2013; te Velde, 2017; Walkden, 2017; Wiese and Müller, 2018). Findings suggest a systematic verb-third (V3) option that, unlike SVO, preserves the characteristic German sentence brackets. From the point of view of information structure, V3 has the advantage over V2 of allowing both a framesetter or discourse linker [e.g., *dann* "then," see (4) and (5) below] and a topic in the left periphery (Wiese, 2012, 2013; Walkden, 2017; Wiese et al., 2017; Bunk, 2020).

While V3 in German has mostly been associated with language-contact situations (e.g., Walkden, 2017), we have shown that it is also available in monolingual speakers (Wiese and Rehbein, 2015; Bunk, 2020). The present study confirms this for maj-German across populations: in the RUEG corpus, we find V3 not only in bilingual speakers with h-Greek, h-Russian, and h-Turkish, but also in monolingual speakers, cf. (4) and (5).

- (4) und *dann* er *lässt* sein ball einfach fallen
and then he lets his ball simply fall

[DEbi51MT_isD]

"And then, he just lets his ball fall down."

- (5) dann die polizei is auch (-) richtig schnell
then the police has also really quickly
gekommen
come

[DEmo68FD_isD]

"And then, the police arrived really quickly."

Findings point to the same V3 options in monolinguals and bilinguals, with an adverbial and a subject preceding the finite verb. As in previous studies, V3 is infrequent, though, with only 48 cases in the bilingual and 11 cases in the monolingual group, with the bilingual group in the lead quantitatively (cf. also Wiese and Rehbein, 2015). In order to compare the difference, we computed normalized frequencies per 100 CUs (communicative units). In the RUEG corpus, CUs were used as a means to segment utterances and were defined as an "independent clause with its modifiers" (following Loban, 1976:9). Accordingly, normalizing for 100 CUs gives us the numbers for V3 as a percentage of all independent clauses, and hence a good approximation for the proportion of non-canonical cases, since V3 is a pattern located at the clausal level. The quantitative difference between speaker groups we observed for absolute numbers is confirmed for such normalized frequencies: we find 0.96 V3 occurrences per 100 CUs in the bilingual group, compared to 0.41 occurrences per 100 CUs in the monolingual group.

Interestingly, the higher frequency is primarily due to speakers with h-Turkish: data from this group makes up 38 of the 48 V3 findings, or 1.88 occurrences per 100 CUs, compared to only 0.42 and 0.27 occurrences per 100 CUs (five cases each) coming from

the h-Greek and h-Russian group, respectively. These findings speak against contact-linguistic transfer, since Turkish, an SOV language, would support the basic SOV word order of German, while Russian and Greek both have a tendency to SVO, which shares a strong surface similarity with V3 as soon as additional left-peripheral constituents such as adverbials are involved. If cross-linguistic transfer were a relevant trigger for V3, we would expect bilingual speakers with h-Greek and h-Russian to be in the lead, rather than those with h-Turkish. This further supports a view of German grammar as the locus of this phenomenon.

For h-German speakers in the United States, we observe an increase in the non-canonical V3 pattern, similar to what we found for non-canonical participle formation in h-Russian: h-German speakers produce 3.47 V3 main clauses per 100 CUs (55 cases in total), with framesetters/linkers like *dann*, *nun* in addition to those probably adopted from English, such as *so* ([zo]), cf. (6) and (7) (with the latter exhibiting all three linkers).

- (6) *dann* um die ecke *sind* zwei autos gefahren
then around the corner are two cars driven
[USbi58FD_iwD]

“then two cars drove around the corner”

- (7) und *dann* der auto hinter ihm *war* nicht bereit *so*
and then the car behind him was not ready so
der *hat* ihn (-) dreingeschubst
he has him there-in-pushed
und *nun* es *is* alles überall *hingerollt*
and now it is all everywhere rolled
[USbi62FD_isD]

“and then the car behind it wasn’t ready so it pushed into it and now everything rolled everywhere”

In this case, the influence of English (X)SVO might further support V3 production, given that English is the majority and main contact language for h-German here. While we cannot rule out that some patterns are enhanced by parallels in English, this only holds partially, though, as shown in (6) and (7), where the clausal brackets remain canonical (“sind . . . gefahren;” “hat . . . dreingeschubst”), in contrast to English SVO. Qualitatively, the h-German data matches, to a large extent, what we find in monolingual and bilingual speakers of maj-German. The difference appears to be quantitative and due to an increased range of constituents involved in V3 clauses, such as non-subjects in the forefield [e.g., a PP as in (6)] and – trivially – borrowed linkers [such as *so* in (7)]. All other cases are attested in our monolingual data as well, even though with a lower frequency.

Had we only investigated h-German in contact with maj-English, we might have claimed that V3 is due to cross-linguistic transfer. Instead, we can now conclude these patterns are also available within the monolingual German repertoire but may be selectively strengthened in HSs by language contact.

Non-canonical Bare NPs in German

For the investigation of non-canonical bare NPs in German, we used additional data from the DNAM corpus of German in

Namibia (Wiese et al., 2017; Zimmer et al., 2020)⁸, in order to compare two groups of h-German speakers. Namibian German represents a rare case of h-German that is still grounded in a vital speech community that systematically uses German not only in informal, but also in formal settings. We focus on the LangSit subcorpus which contains register-differentiated data similar to that in the RUEG corpus (103 speakers; 51,509 tokens). It covers informal-spoken and formal-spoken productions (formal: 23,606 tokens, informal: 25,629 tokens), elicited with visual stimuli in the form of a photo story about a car accident.

Our findings indicate that monolinguals as well as bilinguals produced non-canonical bare NPs, i.e., those that we would not expect in standard German. Figures are overall low: non-canonical cases make up 0.34% of all NPs across maj-German data by both monolingual and bilingual speakers in Germany (22 and 50 occurrences, respectively), 0.98% (22 cases) in h-German in the United States, and 1.06% (44 occurrences) in Namibia.

Interestingly, the non-canonical cases differ qualitatively between the h-German group in the United States and the others. Non-canonical NPs in Germany (maj-German by mono- and bilingual speakers) and Namibia (h-German) can be subsumed under current trends of article decline in German triggered by hyperdetermination, as described by Leiss (2010): (a) generally in generic and unique reference and light verb constructions and in local and directional contexts, and (b) a decline of the definite article in initial, thematic position, and of the indefinite article in rhematic position, because these are already inherently definite or indefinite, respectively. (8) illustrates this for a non-canonical bare NP with generic reference (and in rhematic position) from Namibian German, and (9) for one in rhematic position, produced by a monolingual speaker in Germany:

- (8) *dann* hab ich auch *krankenwagen* angerufen
then have I also ambulance called
[DNAM_S_00066]

“Then, I also called (an) ambulance.”

- (9) ich hab eben *verkehrsunfall* beobachtet
I have just traffic.accident observed
[DEmo30FD_isD]

“I just observed (an) accident.”

The h-German data in the United States differs from this in that we find a distinctive pattern that accounts for almost half of the cases (11 occurrences) and does not occur in the other data. In this pattern, non-canonical bare NPs form the second element in a coordination, cf. (10):

- (10) Die *blaue* auto hat gehalten *wiel* die ball und
the blue car has stopped because the ball and
hund war vorne
dog were in.front
[USbi08MD_fwD]

“The blue car stopped because there was the ball and (the) dog in front of it.”

⁸http://agd.ids-mannheim.de/DNAM_extern.shtml

Unlike known patterns of determiner sharing (McCawley, 1993; Ackema and Szendrői, 2002), this is not restricted to coordination at the VP level, suggesting that an existing pattern can be further extended in heritage language contexts. As the contrast to Namibian German shows, this variation is not related to heritage German *per se*, but might differ across speech communities. Hence, heritage languages can participate in ongoing tendencies as well as further extend native-grammar options.

Intonation

For the domain of intonation, our data provides an example from Russian yes-no questions (YNQs). In the literature, Russian YNQs are reported to be realized with a bitonal rising nuclear pitch accent on the verb ($L^* + H$ or $L + H^*$; Rathcke, 2006b; Meyer and Mleinek, 2006) followed by a low final boundary tone (FBT; L%) (Igarashi, 2006; Rathcke, 2009), except if the nuclear pitch accent falls on the final syllable, in which case a high FBT (H%) is realized (Makarova, 2003; Rathcke, 2006a). The FBT can thus be considered truncated if no material follows the pitch accent (Rathcke, 2009, 2013). The intonation patterns of YNQ differ in Russian and English and are hence interesting to investigate in bilingual speakers.

In order to study the prosodic realization of YNQs of mono- and h-Russian speakers, we elicited experimental data in addition to the corpus data during data collection. This consisted of 10 read-aloud YNQs about details of the car accident. We recorded 20 speakers per group, i.e., (1) bilingual speakers of h-Russian in the United States, (2) bilingual speakers of h-Russian in Germany, and (3) monolingual speakers in Russia [see Zuban et al. (2020) and in prep. for details].

The elicited YNQs differed in the number of syllables following the nuclear pitch accent (or an additional pitch accent on the object for SVO questions). Each YNQ was

annotated for the location of the nuclear pitch accent and FBT, following a combined phonetic and auditory approach: presence of a pitch accent was detected auditorily, and the FBT was examined with respect to local F0 trajectories and changes using Praat. Labeling followed the autosegmental-metrical framework (Makarova, 2003; Igarashi, 2006; Rathcke, 2006b).

Results of the descriptive analysis showed that h-Russian speakers in the United States predominantly produced L% (82% of all cases) while h-Russian speakers in Germany and monolinguals produced L% less frequently (58% for both groups). In order to check for a possible impact of multiple fixed effects on the distribution of high and low boundary tones, we ran a binomial generalized linear mixed-effects model with FBT as the dependent variable, and the three speaker groups, number of syllables following the last pitch accent to the FBT (0–5), transitivity, nuclear contour as independent variables, and with speaker and item as random effects (Figure 2; see “Supplementary Appendix 1” for full model specifications and summaries). It was found (among other things) that all speaker groups produced an H% when the last pitch accent fell on the final syllable, in line with the literature on Russian. Along the same lines, according to what was reported on truncation of FBTs in Standard Russian, h-Russian speakers in the United States chose the L% FBT as soon as there was at least one syllable following the last pitch accent. However, h-Russian speakers in Germany and mono-Russian speakers preferred the L% only when there were more syllables following the last pitch accent.

Hence it is only h-Russian speakers in the United States who behave according to what was described in the literature, while mono- and h-Russian speakers in Germany do not. Had we investigated only h-Russian in Germany and found the significantly increased use of H% (i.e., the absence of categorical truncation), we might have been led to think

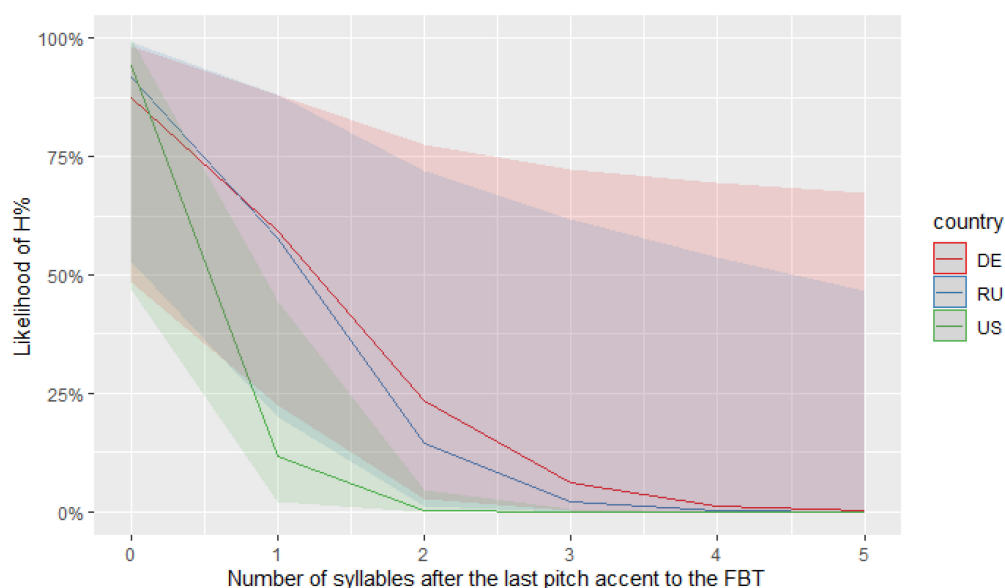


FIGURE 2 | Estimated likelihood of H% being chosen over L% based on the productions of the three speaker groups.

that this pattern is a specific feature of HSs' intonation grammar, possibly due to influence from maj-German, which has H% in YNQs (Grice et al., 2005). However, comparable realizations by mono-Russian speakers show that what we see here is a more general pattern: if we approach both HSs and monolinguals as native speakers and analyze language use across speaker groups, we may also find non-canonical patterns in monolinguals that might otherwise have been attributed to bilingualism.

Pragmatics

In the domain of pragmatics, examples come from data on the position of new referents in Turkish and on referent introduction in English.

Non-canonical Placement of New Referents in Turkish

As briefly mentioned in "Morphology and Syntax," the basic and pragmatically neutral word order in Turkish is SOV. However, for specific pragmatic purposes, elements can also be placed in the post-verbal position. In particular, this is possible for marking backgrounded information and afterthoughts (Erguvanli, 1984; Schroeder, 1995; Kornfilt, 1997), for constituents such as NPs, adverbs, discourse markers, forms of address, and finite subordinate clauses. What is crucial is that placement of new information is believed to be impossible here: according to the literature, the information placed in the post-verbal position has to be discourse-predictable or recoverable from previous discourse (Erguvanli, 1984:56).

Approaching HSs as native speakers of their languages, we investigated these information-structural restrictions for h-Turkish and mono-Turkish alike. We proceeded by selecting the 21 most frequently used nominal referents that played a role in the elicited narratives. We annotated every occurrence of these, adapting Riester and Baumann's (2017) referent annotation scheme, and identified information status through three categories (cf. Schroeder et al., to appear): (1) "new" (first mention), (2) "given" (referent that was introduced before) (3) "bridging" [referent that has not explicitly been introduced but belongs to the "pragmatic set" in the sense of Hawkins (1984) of a given referent (anchor)]. Afterthoughts, repairs and finite subordinate clauses were excluded from analysis.

Results of the analysis show that even though most of the referents placed post-verbally were indeed "given" and "bridging," there is a substantial number of new referents used in the post-verbal position. What is interesting in the context of the present manuscript, is that new referents in the post-verbal position occurred not only in the data of HSs [as seen in (11), but also in monolinguals in Turkey (12)]. In HSs in Germany, the occurrences of new referents in post-verbal position constitute 32.43% of the overall number of referents in the post-verbal position (24 new referents out of 74 referents in the post-verbal position), in monolinguals in Turkey the occurrences of new referents constitute 22.45% (11 new ones out of 49 referents), and in HSs in the United States, the new referents in the post-verbal position make up 21.43% (24 occurrences out of 112). This outcome contradicts what the literature says about canonical

Turkish, namely that new referents are not possible in the post-verbal position.

- (11) bi: çift vardı çocuk arabasıyla
one couple was child car.with

[DEbi18MT_fsT]

"There was a couple with a stroller."

- (12) az kalsın bi araba çarpyıodu çocuklu bi kadına
almost one car hit child.with one woman

[TUmo26MT_isT]

"A car almost hit a woman with a child."

Most of the new referents in post-verbal position stand in a close semantic relationship to the subject of the clause. This relationship is wider than the "bridging" relation, and it is often indicated by means of the possessive suffix *-(s)I* on the post-verbal constituent [like in (11)], or by free adjuncts that carry adverbial case. Less often, a new referent in post-verbal position is a lexical subject or object [as in (12)].

However, it would be misleading to call those new referents that are in a close semantic relationship with the subject "backgrounded information," since in about half of the cases in our data, the referent introduced in the post-verbal position is mentioned again in the subsequent discourse, and is treated as "given" when mentioned again. We propose to call such referents "secondary" new referents, in the sense that they are secondary (and related) to another new referent with a higher relevance to the discourse at that point.

Thus, we conclude that it is indeed possible to place new information in post-verbal position in Turkish, and this is not a feature that is typical only for HSs, since the pattern is also found in the monolingual data from Turkey. As we will discuss in more detail in section "Register Leveling in Heritage Languages," in monolinguals this pattern seems to be associated with informal registers. Hence, if we compare like with like and systematically include data from such registers from monolinguals as well, we can avoid misattributing some non-canonical patterns to bilingualism that form a more general part of native grammars.

Non-canonical Referent Introduction in English

Unlike Turkish, English marks newness and givenness of referents through indefinite and definite articles (Hickmann and Hendriks, 1999). The indefinite article *a* presupposes that the referent of the NP is new and the addressee is not familiar with it. The definite article *the* implies that the addressee can uniquely identify the given referent of the NP based on previous discourse, shared physical environment and/or general world knowledge (Payne and Huddleston, 2002:368–371).

Previous research has shown that bilingual speakers often differ in their production of articles from monolingual speakers of English. For example, child HSs of other languages have been reported to oversupply *the* in indefinite contexts and *a* in definite contexts in maj-English, regardless of their heritage language (Zdorenko and Paradis, 2008, 2012). Further, adult L2 English speakers with article-less L1s tend to overuse *a* and *the* in contexts with mismatching parameters of definiteness and specificity (i.e.,

subjective noteworthiness of the referent to the speaker) (Ionin, 2006; Ionin et al., 2008).

Most importantly, many studies on article use compare only two or more bilingual groups to each other (Lardiere et al., 2004; Zdorenko and Paradis, 2008, 2012; Ionin and Díez-Bedmar, 2021). If a monolingual comparison group is added, monolinguals usually supply articles in strict accordance with the expectations based on the literature (Hawkins et al., 2006; Ionin, 2006; Ionin et al., 2008; Sarko, 2009; Snape et al., 2013). Overall, these studies seem to suggest that variability in article production is a result of bilingualism.

We tested this assumption by examining article choice in new and given referents among bilingual and monolingual English speakers of maj-English – 214 HSs with various heritage languages and 64 English monolinguals.⁹ Similarly to the investigation of new postverbal referents in Turkish just discussed, we selected 19 frequent referents such as *man*, *dog*, *car1*, and *car2*, and coded them for their information status as “new” (the first mention of an entity without any identifying information) or “given” (all the subsequent mentions) (Riester and Baumann, 2017). This yielded 4,961 new and 10,881 given referents.

We identified all new and given referents that were part of unexpected non-canonical structures, that is, “*the* + new” and “*a* + given” referents. All other structures in which referents appeared, including expected canonical structures (“*a* + new” and “*the* + given”), were marked as “other.” Contrary to what would be expected from the literature, we found non-canonical patterns not only in bilinguals’ productions [see (13) and (15)], but also in those of monolinguals [see (14) and (16)], and we found this for both the “*the* + new” pattern [(13) and (14)] and the “*a* + given” pattern [(15) and (16)]:

- (13) Oh my god, I just saw a car get rear-ended, there was this really cute dog and he ran out in front of it because this guy dropped his ball and the dog chased it onto the street. And then they had to pick up the groceries [*new*]. [USbi50FD_isE]
- (14) I’m calling about incident number F16, I was there at the time it happened. There was a blue car and a white car both coming down a path, they both made a right. As they made a right, a man had a ball, and it went into the street into the pathway. The dog [*new*] ran after the ball. [USmo07FE_fsE]
- (15) There was this couple with a ball on one side and a lady packing groceries on another. The dude from the couple was holding a ball [*given*]. [USbi03MG_iwE]
- (16) Then a dog starts barking and he runs in the street. So this guy is driving down the street and he sees a dog [*given*]. [USmo10ME_isE]

Hence, we did not find any qualitative differences between monolinguals and bilinguals in this domain. In order to check for a possible impact of bilingualism on quantitative distributions

while taking into account inter-individual speaker variation, we ran two binomial generalized linear mixed effects models, one for new and one for given referents. The dependent variable was Determiner (“*the* + new”/“*a* + given” vs. “other”) and the independent variables were Bilingualism (bilingual vs. monolingual), Setting (formal vs. informal) and Mode (spoken vs. written) (see “**Supplementary Appendix 2**” for full model specifications and summaries).

The results indicate neither a main effect of bilingualism, nor its interactions with other variables, meaning that we have no evidence of differences between HSs and monolinguals. Both groups produced similar numbers of non-canonical “*the* + new” referents, ranging from 6.6% of all given referents by heritage speakers and 5.6% by monolinguals in the formal written situation, to 9.1% by heritage speakers and 11.1% by monolinguals in the informal spoken situation. The two groups did not differ in the production of “*a* + given” referents either: for this pattern, the percentages of non-canonical referents were much smaller and ranged from 0.47% of all new referents by heritage speakers and a complete absence by monolinguals in the informal written situation, to 0.75% in the informal spoken situation by heritage speakers and 1.5% by monolinguals. Overall, our data shows that a pattern that has mostly been attributed to bilingualism actually manifests itself in the productions of English monolinguals in the same way as it does in the speech of bilinguals.

Variation: Not Just in Heritage Speakers, but Also in Monolinguals

In the literature, HSs and bilingual speakers in general are often presumed to exhibit higher degrees of variation than “regular” native speakers. Seton and Schmid (2016:341) even claim that “[t]he most striking characteristic that sets bilinguals apart from monolinguals is a larger amount of variability in performance.” Observations from our data suggest a more complex picture and point to intricate dynamics between groups and individual effects and different degrees of dynamicity in linguistic subsystems. The evidence provided to illustrate this point here comes from our corpus data on maj-German, spoken by monolingual and bilingual speakers in Germany (see **Table 1** above).

Variation can be measured in different quantities, or in different qualities, such as a wider range of structures. For example, a bilingual speaker may use different structures from a monolingual, and those may be canonical or non-canonical (some canonical structures may be dispreferred by some speakers and/or in some registers). In our data, we find that all syntactic dependencies, such as different types of objects, particles, modifiers, etc., are used by mono- and bilinguals alike (see **Figure 3**). Hence, we do not find qualitative differences here: bilinguals (as a group) neither avoid certain dependencies used by monolinguals, nor do they exhibit a wider range of dependencies than monolinguals¹⁰.

⁹See figures for maj-English in **Table 1** above. Eight HSs (three with h-German, two with h-Greek, and three with h-Turkish) were excluded from the analysis due to technical reasons.

¹⁰Even though **Figure 3** contains some fine-grained prepositional subcategories (PNA and PND marking accusative/dative complements to prepositions) that go beyond the underlying dependency grammar (Foth, 2006), and dependency parses have been manually corrected, it is possible that more diversity on this level

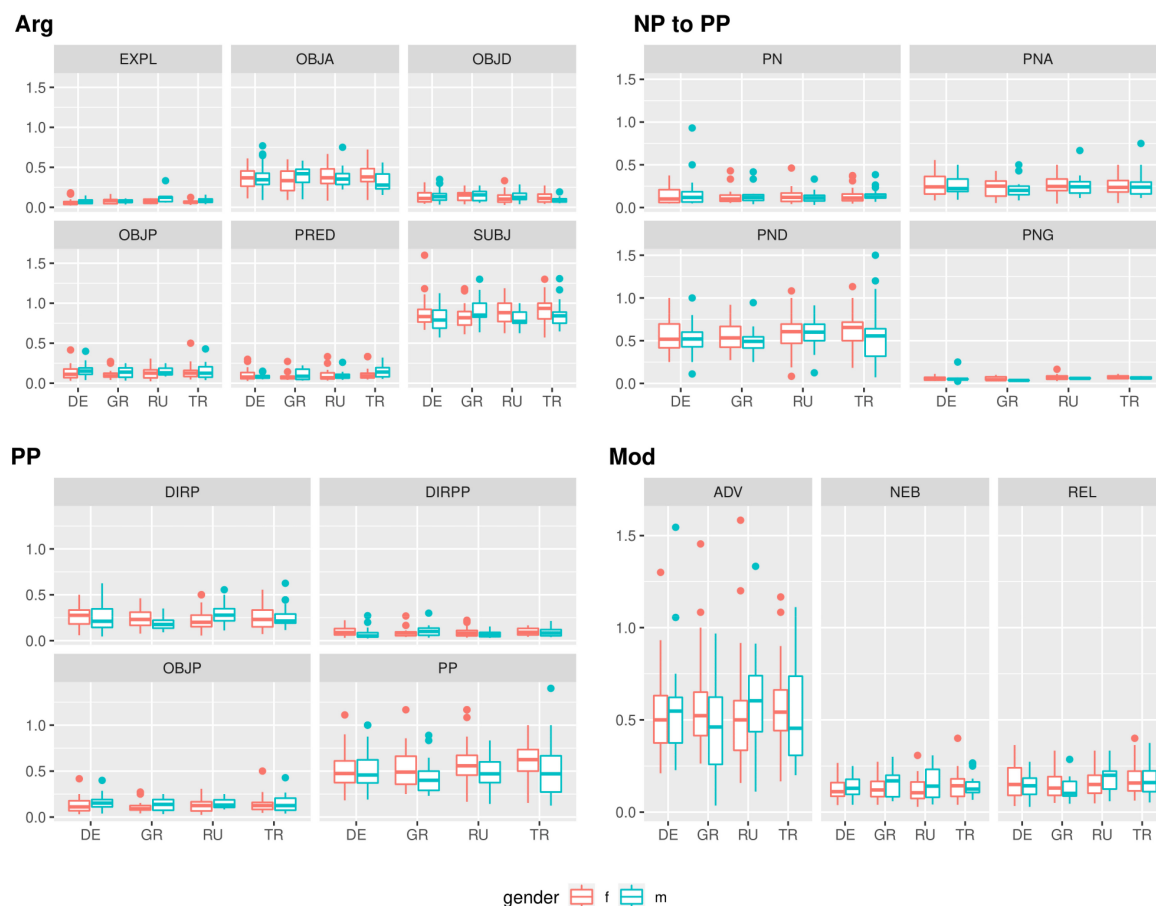


FIGURE 3 | Dependencies in maj-German (formal-written) RUEG texts, normalized by lexical verbs.

Are there quantitative differences in variability, i.e., differences in variance, then? The picture here is more mixed. First of all, both the mono- and the bilingual speakers show large degrees of quantitative variance for some, but not all structures. This is in part an effect of the general frequency of a structure: if a dependency type is overall rare, its frequency of occurrence will exhibit floor effects. However, it also appears that some structures, whether rare or more frequent, are subject to more free choice in the frequency of their realization. This can be due to simple causes, such as the possibility to name all vs. just some of the agents or circumstances in the respective narratives, or recurrence and repetition of previously named entities vs. ellipsis. At the same time, a more or less frequent realization of elements may also be indicative of differences in communicative style (explicitness, emphasis) or formality (assumption of interlocutor expectations; see also Ahern et al., 2019: 487, 488). This can, for example, affect the frequency of modifying structures such as adverbs or relative clauses, which can be used to specify or comment on another dependency. Explicitness, referring here to the tendency to explicitly name more aspects of the environment

or circumstances, may also play a role for different amounts of PP realizations.

However, the realization of PPs can also be due to interactions with the lexicon, for example in the case of prepositional objects required by certain verbs (abbreviated as OBJP – a full list of abbreviations is provided in “**Supplementary Appendix 3**”). Other PPs can be added attributively without lexical constraints. Differences in the frequency of realization of nominal complements to prepositions (PND for dative objects to prepositions, PNA for accusative objects complementing prepositions) can thus be due to lexical diversity (different verbs requiring different complements), but might also point toward case dynamics (identical verbs occurring with different complement cases). Differences in variance can thus represent very different types of phenomena.

In our data, we see variable degrees of variance between structures, but there is barely any evidence for a higher degree of variability in bilinguals. Only the h-Turkish speakers appear to show a higher degree of variance in the realization of (free) prepositional phrases (PP) and dative complements to prepositional phrases (PND), and the effect is rather small. Some bilingual groups appear to show a slightly higher variance in modifying structures (male h-Russian and male h-Greek

only becomes visible under consideration of more fine-grained lexicosyntactic phenomena such as constructions.

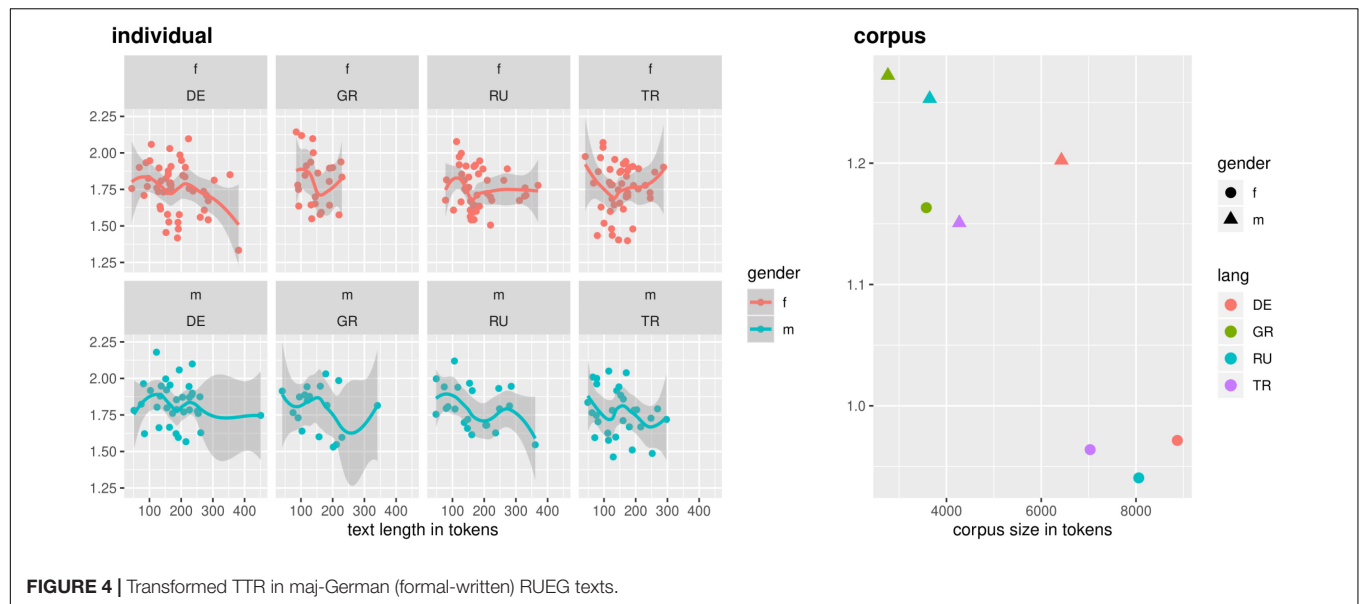


FIGURE 4 | Transformed TTR in maj-German (formal-written) RUEG texts.

for subordinate clauses, marked NEB, male h-Greek and male h-Turkish for adverbs, ADV). However, some chance results are to be expected due to the high number of between-group comparisons.

Overall, there is no clear trend toward higher variability in bilinguals: mono-German speakers are generally within the range of variance found for bilingual speakers, and even have a tendency to be on the upper end of it. In fact, differences between gender groups are generally higher than between speaker groups of different language background but the same gender.

This also appears to be the case for lexical richness, approximated here with a transformed type-token-ratio (TTR, see **Figure 4**)¹¹. We find no strong differences in the distribution of individual speakers in each language group. However, the female mono-German and female h-Turkish bilinguals show the highest degrees of variance in TTR. This may be an artifact of different group sizes (23 female, 10 male) – especially since, in general, we find higher variance in the male groups. This becomes particularly obvious in the distribution by subcorpora: we find much higher TTRs for the male speaker groups compared to the female ones *across language groups*. Male and female speakers use roughly the same amount of lexemes per speaker, but those lexemes converge less between male than female speakers. This suggests divergent effects of idiomaticity or coselectional constraints¹² and might be attributed to communicative style or a higher degree of adjustment to assumptions of interlocutor expectations.

Overall, we find different degrees of variance in our data, but no particular effect for monolinguals vs. bilinguals. In fact,

the monolingual speakers in the RUEG corpus exhibit a degree of variance that reaches, and sometimes surpasses, that of the bilingual speakers. We do not find evidence that variability in the realization of syntactic structures or in lexical richness could be used as a criterion for the definition of a “real” native speaker that would favor monolingual over bilingual speakers.

The Role of Registers

Approaching HSs as native speakers of both their languages, we targeted formal and informal registers in bilinguals and monolinguals alike. Our findings indicate an important role of registers for non-canonical patterns, and this can play out differently in heritage and majority language use. Below we discuss evidence from English and German as majority languages (“Noncanonical Phenomena in Informal or Spoken Registers of Majority Languages”), and German, Turkish, and Greek as heritage languages (“Register Leveling in Heritage Languages”).

Non-canonical Phenomena in Informal or Spoken Registers of Majority Languages

In majority language use, non-canonical patterns are dominantly found in informal and/or spoken registers, and this holds across bilingual and monolingual speaker groups.

Association With Spoken Registers for Non-canonical Patterns in English

In section “Pragmatics” we discussed two non-canonical patterns involving new and given referents in maj-English, namely “*a* + given” and “*the* + new” referents, which appear in bilingual and monolingual English speakers alike. Our further results reveal the influence of spoken vs. written mode on non-canonical article choice. In the two linear mixed effects models reported in section “Pragmatics,” we observed a main effect of mode across speaker groups and formal/informal settings: the spoken mode showed

¹¹The TTR is transformed by the formula $\text{types/tokens} \times 4\sqrt{\text{tokens}}$, or $\text{TTR} \times 4\sqrt{\text{text length}}$ (or corpus size, respectively) and serves only to adjust for text-length dependency. No claims are made to the general usefulness of approximating lexical richness through TTRs.

¹²See Shadrova (2020) for a discussion of measuring coselectional constraints in corpora.

more non-canonical structures than the written mode (for “*the* + new”: $\text{mean}_{\text{formalspoken}} = 7.68\%$, $\text{mean}_{\text{formalwritten}} = 6.12\%$, $\text{mean}_{\text{informalspoken}} = 10.1\%$, $\text{mean}_{\text{informalwritten}} = 7.3\%$; for “*a* + given”: $\text{mean}_{\text{formalspoken}} = 0.69\%$, $\text{mean}_{\text{formalwritten}} = 0.57\%$, $\text{mean}_{\text{informalspoken}} = 1.11\%$, $\text{mean}_{\text{informalwritten}} = 0.23\%$; see “**Supplementary Appendix 2**” for *p*-values and model summaries).

The higher variability in article choice in the spoken mode could be associated with a higher cognitive load of online (spoken) productions compared to offline (written) ones. Spontaneous spoken production often exerts performance pressure since it allows little time for planning and no possibility for changing what has been said (Pullum and Huddleston, 2002:12). This factor is important in L2 research: for instance, Ionin et al. (2021) argue for testing article knowledge of L2 speakers in comprehension rather than production in order to avoid the performance pressure and to evaluate speakers’ implicit sensitivity to (in)definiteness. In our study, performance pressure in the spoken mode might have led the speakers to only consider their own perspective (familiarity with the referent) and, consequently, use the definite article, while ignoring the addressee’s perspective (unfamiliarity with a new referent), which would require the indefinite article.

In addition, we might be witnessing a new development in English: possibly, the definiteness distinction is becoming less strict in spoken spontaneous productions. So far, it is unclear if this is a systematic pattern of English internal dynamics (since over 90% of *the* uses are still canonical), and it needs to be confirmed in future research.

Association With Informal Registers for Non-canonical Patterns in German

In maj-German, non-canonical V3 (as described in “Morphology and Syntax” above) occurs mainly in the informal productions in our corpus data: of 59 V3 cases in maj-German altogether, 53, that is, roughly 90%, are from informal communicative situations,

and this pattern is also evident for normalized frequencies, with 1.24 of 1.37 occurrences per 100 CUs, that is, roughly 91%, from informal communicative situations.

Within the informal settings, we find V3 across spoken (17) and written (18) modes. This suggests that unlike non-canonical article choice in maj-English, V3 in maj-German is primarily associated with informality, independently of mode.

- (17) und dann er lässt sein ball einfach fallen
and then he lets his ball simply fall
[DEbi51MT_isD]

“And then, he just lets his ball fall down.”

- (18) danach er lässt den ball aus versehen fallen
afterward he lets the ball from accident fall
[DEbi58MT_iwD]

“Afterward, he accidentally lets the ball fall down.”

Note, though, that we do find a few occurrences of V3 in formal data, and this also includes one production (formal-written) from a monolingual speaker:

- (19) am helligten tage ein paar mit kinderwagen
at.the bright.light day a couple with stroller
sowie einem fußball ging auf dem fußweg
as.well.as a foot.ball went on the foot.path
[DEmo31MD_fwD]

“In bright daylight, a couple with a stroller went along the sidewalk.”

This, again, integrates bilingual and monolingual speakers alike into the native speaker group: we find the non-canonical V3 pattern across groups, in both groups primarily associated with informal registers, with some exceptions which are also evident in both groups.

For non-canonical bare NPs in maj-German, we similarly found a dominance in informal data, and again this held for bilinguals as well as monolinguals. We also found this pattern in

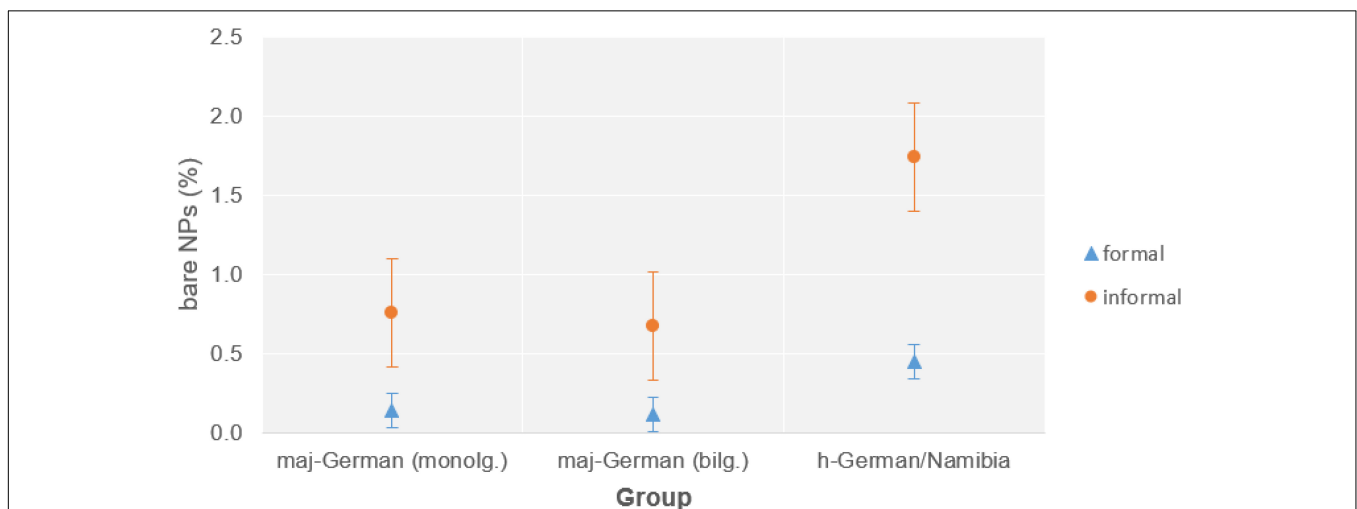


FIGURE 5 | Relative frequencies of non-canonical bare NPs (percentage of all NPs) in German in informal vs. formal settings by three different speaker groups.

h-German in Namibia, where it was even more pronounced, cf. **Figure 5**,¹³ suggesting that such register differentiation can also hold in heritage grammars (but see “Register Leveling in Heritage Languages” below for h-German in the United States).

Register Leveling in Heritage Languages

Some patterns characteristic for informal registers in monolinguals can be generalized to formal registers in bilinguals’ heritage language use, leading to register leveling.

Non-canonical Bare NPs and Word Order in h-German in the United States

One example for this is h-German in the United States. In the United States, speakers do not distinguish between formal and informal settings in their non-canonical word order and NP patterns, unlike (monolingual and bilingual) speakers in Germany.

For h-German V3, we find 1.89 occurrences per 100 CUs in formal settings and 1.57 in informal settings (30 and 25 occurrences, respectively). Mode, however, plays a role here, as 71% of all V3 clauses appear in the spoken mode, pointing to a more general phenomenon that we also saw in non-canonical article choice in maj-English.

H-German speakers in the United States also show a similar distribution of non-canonical bare NPs in formal and informal situations, with 0.15 and 0.17% of all NPs (13 and 10 occurrences), respectively. As the data discussed in “Noncanonical Phenomena in Informal or Spoken Registers of Majority Languages” shows, this is in contrast not only to monolingual and bilingual speakers of maj-German in Germany, but also to h-German speakers in Namibia, where non-canonical NPs were primarily associated with informal registers. This can be related to the different communicative domains for German: in Germany and Namibia, German is used in informal as well as formal settings, including formal schooling, whereas in the United States, it is mostly restricted to informal contexts.

Further evidence for register leveling comes from h-Turkish in Germany. In section “Pragmatics,” we found non-canonical placement of new referents in post-verbal position in both HSs and monolinguals. Our analysis shows that in monolingual data, all the examples of new referents in the post-verbal position occur in informal (spoken and written) settings. HSs, on the other hand, tend to place new referents in the post-verbal position across all communicative situations. Again, this can be related to communicative domains for the heritage language: similarly as for h-German in the United States, h-Turkish speakers in Germany are mostly exposed to informal language at home and with their peers and have less access to formal registers. A phenomenon typical for informal settings in mono-Turkish then spreads into formal settings in h-Turkish.

These findings, then, emphasize that non-canonical phenomena should not always be attributed to heritage languages or bilingualism in general, but they can also be a

feature of informal settings in monolingual varieties. Thus, what might be seen as a consequence of language contact or attrition in a superficial account turns out to be present also in monolingual language use when different communicative situations are taken into account.

Non-canonical Restrictive Relative Clauses in maj- and h-Greek

A further area where we noted register leveling is the distribution of restrictive relative clauses (RCs) in Greek. RCs in Greek function as modifiers of nouns, as in other languages, and always appear in post-nominal position (Chatsiou, 2010). Greek RCs¹⁴ come in two types: they are either introduced by the pronoun *o opios*, literally “the who,” or by the complementizer *pu* “that.” While the pronoun is inflected for gender, case and number and agrees with the nominal head that it modifies in gender and number, *pu* bears no inflection. In the literature on Greek, we find the claim that *pu* appears mostly in colloquial speech, while the pronoun is preferred in formal registers (Holton et al., 1997: 212).

We investigated restrictive RCs for the three groups represented in our corpus: monolingual speakers of maj-Greek in Greece and bilingual speakers of h-Greek in Germany and in the United States. As **Table 3** shows, *pu* RCs are more frequent than *o opios* RCs across the three speaker groups in the different countries, and this holds in both informal and formal settings.

Zooming in on the question of register variation, it appears that monolingual speakers actually prefer *pu* RCs even in formal settings, something that contradicts the claims in grammars of Standard Modern Greek. This points to a process of leveling even in the monolingual group. A similar pattern is exhibited by the h-groups, who also favor *pu* RCs over *o opios* RCs across settings, and a one-way ANOVA test¹⁵ revealed no significant differences between the three groups concerning the production of *pu* RCs in the different communicative settings.

For *o opios* RCs, we find differences in the distribution across registers between groups, as determined by one-way ANOVA tests {formal: [$F(2,171) = 15.99, p < 0.0001$], informal: [$F(2,171) = 8.877, p < 0.0001$]}. Our data indicates that the *o opios* strategy is preferred in the formal register compared to the informal one (although at a lower rate than *pu* RCs). In this case, there is a statistically significant difference between

¹⁴Free RCs are excluded from this account.

¹⁵According to the central limit theorem, the sample distribution will be approximately normally distributed as the sample size is quite large ($N > 50$ items).

TABLE 3 | Quantitative data - distribution of restrictive RCs in different registers per group.

Communicative situation		h-Greek in the U.S.	h-Greek in Germany	Mono-Greek
Informal	<i>Pu</i>	165 (99.4%)	150 (82%)	183 (74.7%)
	<i>o opios</i>	1 (0.6%)	33 (18%)	62 (26.3%)
Formal	<i>Pu</i>	238 (97.9%)	190 (71.7%)	244 (64%)
	<i>o opios</i>	5 (2.1%)	75 (28.3%)	137 (36%)

¹³Note, though, that the corpus data for h-German in Namibia comes only from spoken productions (see “Morphology and Syntax” above).

groups, as determined by a one-way ANOVA [$F(2, 171) = 14.50$, $p < 0.0001$]. Specifically, Bonferroni-adjusted *post hoc* tests enabling pairwise comparisons revealed that the United States group differs from the other two in the production of *o opios* RCs across the different communicative settings ($p < 0.05$). The preference for formal registers holds in particular for the monolingual group, and secondly for HSs in Germany. HSs in the United States rarely use *o opios* RCs regardless of setting [as reported also by Lithoksoou (2019), who investigated part of RUEG corpus].

Taken together, our results indicate register leveling for *pu* RCs across monolingual and bilingual speaker groups, and differences for *o opios* RCs between different heritage speaker communities, with those in Germany aligning with monolingual speakers in Greece.

Such findings underline the relevance of considering register differentiations in bilinguals and monolinguals alike, while integrating HSs into the native speaker continuum.

DISCUSSION

In this manuscript, we argued for a perspective on bilingual heritage speakers as native speakers of both their languages, and demonstrated what can be gained from such an approach. We showed that recognizing heritage speakers as native speakers is justified on conceptual and theoretical grounds, and that this perspective is also a better fit for the empirical data. We presented results from a large-scale, cross-linguistic study that approached bilingual heritage speakers and monolingual speakers on equal terms, rather than using monolinguals as a yardstick for what counts as a competent “native speaker.” In line with this approach, we targeted actual language use that covered broader repertoires than just formal language, and we did so for bilingual and monolingual speakers alike. To this end, we elicited linguistic productions representative of speakers’ natural behavior in formal and informal, written and spoken communicative situations. This provided us with comparable data across registers, languages, contact-linguistic settings, and speaker groups, incorporated in an open-access corpus, the RUEG corpus.

Our findings support the integration of heritage speakers into the native-speaker continuum and show that they can shed light on language variation and change in native grammars. In a number of heritage languages, we find patterns in formal registers that do not appear in the respective majority settings for those languages. However, a closer look shows that these patterns are by no means completely absent from monolingual grammars, but can be associated with informal registers there. Patterns that might otherwise have been attributed to bilingualism could hence inform us on ongoing developments and variation in native grammars.

Our data points to a range of non-canonical patterns in monolinguals’ productions that have so far been considered absent from native grammars. We discussed non-canonical patterns of participle formation and boundary tones in Russian; verb-third and bare NPs in German; new referents

in post-verbal position in Turkish; and non-canonical choices of (in-)definite articles for given vs. new referents in English. Our study showed that these patterns are not restricted to heritage speakers, but they occur systematically in monolinguals as well.

Along similar lines, we found inter-speaker variability not restricted to bilinguals either. In our data for German, monolinguals displayed a degree of lexical and morphosyntactic variation that was sometimes higher than that of bilinguals, further challenging the model of the streamlined native speaker.

In majority language use, non-canonical patterns were dominant in spoken and/or informal registers, and this was true for monolinguals and bilinguals alike. In majority-English, for example, spoken registers featured more non-canonical article choices than written registers in both monolingual and bilingual speakers, and in majority-German, non-canonical word order and bare NPs were associated with informal registers.

In several languages, though, our data points to tendencies of register leveling in heritage contexts. This is presumably due to a lack of formal schooling: heritage speakers are mostly exposed to informal language at home and often have limited access to formal registers. Hence, phenomena typical for informal settings in majority language use can spread to formal settings in heritage languages. This points to different register distributions rather than to distinct grammatical patterns in heritage speakers. Accordingly, where the heritage language is also used in formal schooling, e.g., in the case of heritage German in Namibia, patterns were similar to those found for majority language use (i.e., German in Germany). As our Greek data showed, register leveling can also occur in monolingual speakers: we found that, contrary to claims in the literature, a non-inflected relativizer is preferred in formal as well as informal registers, and we found this for monolinguals and heritage speakers alike.

Some non-canonical patterns in informal settings were more frequent in bilingual speakers. Heritage speakers can hence put a unique spotlight on internal developments. In line with this, several of the non-canonical patterns we found point to extensions of existing, salient variants and/or ongoing language change. This underscores the importance of taking into account a broader range of communicative settings, not just for heritage speakers, but also for monolinguals.

For heritage languages, our findings also indicate that the size and cohesion of speech communities can play a role. In smaller, more widely distributed communities like our populations in the United States, we sometimes found more diverging patterns. This does not imply that those options are outside native grammars, but we might see more variation. Our data point to broader options of non-canonical word order and bare NPs in heritage German in the United States, while heritage German in Namibia patterned with majority German. For heritage Russian, we found a lower frequency of participles in the United States, while heritage Russian in Germany patterned with majority Russian.

Our findings also lead us to reconsider the state-of-the-art on majority languages, given recurring evidence for non-canonical patterns in monolinguals that deviate from what has been

assumed in the literature so far. In the case of the intonational pattern analyzed in Russian, it was in fact the “deviant” pattern, that is, the one that differed from those described in the literature, that we found in monolinguals as well as in one of our heritage language communities, namely in Germany, but not in the other, that is, in heritage Russian in the United States.

Taken together, our findings support current calls to normalize multilingualism. Multilingualism can act as a motor of linguistic developments, and accordingly, multilingual communities can afford us a privileged view into ongoing tendencies of language variation and change. However, this does not make them an exotic, special case. Our findings put multilinguals’ language solidly within native grammars, at levels of structure as well as language use. In order to make full use of the opportunity that multilingual speakers provide for linguistic theory, we need to take into account variation in native grammars, including informal registers, in bilinguals and monolinguals alike.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <http://doi.org/10.5281/zenodo.3236069>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by DGfS Ethics Committee (Deutsche Gesellschaft

für Sprachwissenschaft/German Society for Linguistics). Written informed consent to participate in this study was provided by participants or, in case of minors, their legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

FUNDING

Research for this volume was supported through funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for the Research Unit “Emerging Grammars in Language Contact Situations” (AL 554/13-1, 15-1; AL 1886/2-1, 3-1, 4-1; BU 2690/1-1; GA1424/7-1, 10-1; LU 856/16-1; SCHR 1261/3-1, 4-1; SH 1685/1-1; SZ 263/4-1, 6-1; TR 238/5-1, 6-1, 7-1; WI 2155/10-1, 10-2, 11-1, 12-1, 13-1; and ZE 940/2-1, 4-1).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.717973/full#supplementary-material>

REFERENCES

- Aalberse, S., Backus, A., and Muysken, P. (2019). *Heritage Languages. A Language Contact Approach*. Amsterdam: Benjamins.
- Ackema, P., and Szendrői, K. (2002). Determiner sharing as an instance of dependent ellipsis. *J. Comp. Ger. Linguist.* 5, 3–34.
- Ahern, A., Amenós-Pons, J., and Guijarro-Fuentes, P. (2019). Interpreting mood choice effects in L2 and L1 Spanish: empirical evidence and theoretical implications. *Appl. Linguist. Rev.* 10, 469–490. doi: 10.1515/applrev-2017-0097
- Auer, P. (2013). “Ethnische marker zwischen varietät und stil,” in *Das Deutsch der Migranten*, ed. A. Deppermann (Berlin: de Gruyter), 9–40.
- Bayram, F., Kupisch, T., Pascual y Cabo, D., and Rothman, J. (2019). Terminology matters on theoretical grounds, too! Coherent grammars cannot be incomplete. *Stud. Second Lang. Acquis.* 41, 257–264. doi: 10.1017/S0272263119000287
- Bogdanov, S. I., Evtjuxin, V. B., Knjazev, J. P., et al. (2009). *Morfologija Sovremennogo Russkogo Jazyka: Učebnik Dlja Vysšix Učebnyx Zavedenij Rossijskoj Federacii*. Sankt-Peterburg: Fakul'tet filologii i iskusstv SPbGU.
- Bonfiglio, T. P. (2010). *Mother Tongues and Nations: The Invention of the Native Speaker*. Berlin: de Gruyter.
- Brutt-Griffler, J., and Samimy, K. K. (2001). Transcending the nativeness paradigm. *World Engl.* 20, 99–106. doi: 10.1111/1467-971X.00199
- Bunk, O. (2020). ‘Aber Immer Alle Sagen Das’. *The Status of V3 in German: Use, Processing and Syntactic Representation*. Ph.D. Dissertation. Berlin: Humboldt University Berlin.
- Cejtin, S. N. (2009). *Očerki po Slovoobrazovaniju i Formoobrazovaniju v Detskoj Reči*. Moscow: Znak.
- Chatsiou, A. (2010). *A Lexical Functional Grammar Approach to Modern Greek Relative Clauses*. Ph.D. Dissertation. Colchester: University of Essex.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Cook, V. (2016). “Premises of multi-competence,” in *The Cambridge Handbook of Linguistic Multi-Competence*, eds V. Cook and Li Wei (Cambridge: Cambridge University Press), 1–25.
- Criado-Perez, C. (2019). *Invisible Women: Data Bias in a World Designed for Men*. New York, NY: Vintage Publishing.
- de Saussure, F. (1916). *Cours de Linguistique Générale*, eds C. Bally and A. Sechehaye (Chicago, IL: Open Court).
- Embick, D., White, Y., and Tamminga, M. (2020). “Heritage languages and variation: identifying shared factors” [commentary article on Polinsky, M., and Scontras, G., “understanding heritage languages”]. *Bilingualism* 23, 21–22. doi: 10.1017/S1366728919000476
- Erguvanli, E. E. (1984). *The Function of Word Order in Turkish Grammar*. Berkeley, CA: University of California Press.
- Flores, C., and Rinke, E. (2020). The relevance of language-internal variation in predicting heritage language grammars. *Bilingualism* 23, 25–26. doi: 10.1017/S1366728919000464
- Foth, K. A. (2006). *Eine umfassende Constraint-Dependenz-Grammatik des Deutschen*. Available online at: <https://edoc.sub.uni-hamburg.de/informatik/volltexte/2014/204/> (accessed December 9, 2021).
- Gagarina, N. (2002). “Productivity of the verbal inflection classes (the case of Russian),” in *Future Challenges for Natural Linguistics*, Lincom Studies in Theoretical Linguistics, Vol. 30, eds K. Dziubalska-Kolaczky and J. Weckwerth (München: Lincom), 153–164.
- Gogolin, I. (1994). *Der monolinguale Habitus der multilingualen Schule*. Münster: Waxmann.
- Gogolin, I. (2002). Linguistic and cultural diversity in Europe: a challenge for educational research and practice. *Eur. Educ. Res. J.* 1, 123–138. doi: 10.2304/eeerj.2002.1.1.3
- Golub, I. B. (2001). *Russkij Jazyk i Kul'tura Reči: Učebnoe Posobie*. Moscow: Logos.

- Grice, M., Baumann, S., and Benz Müller, R. (2005). "German intonation in autosegmental-metrical phonology," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, ed. S.-A. Jun (Oxford: Oxford University Press), 55–83. doi: 10.1093/acprof:oso/9780199249633.003.0003
- Grosjean, F. (1982). *Life With Two Languages*. Cambridge, MA: Harvard University Press.
- Grosjean, F. (2010). *Bilingual: Life and Reality*. Cambridge, MA: Harvard University Press.
- Guijarro-Fuentes, P., and Schmitz, K. (2015). The nature and nurture of heritage language acquisition. *Lingua* 164, 239–250. doi: 10.1016/j.lingua.2015.05.008
- Hawkins, J. A. (1984). A note on referent identifiability and co-presence. *J. Pragmat.* 8, 649–659. doi: 10.1016/0378-2166(84)90004-3
- Hawkins, R., Al-Eid, S., Almaboob, I., Athanasopoulos, P., Chaengchenkit, R., Hu, J., et al. (2006). Accounting for English article interpretation by L2 speakers. *EUROSLA Yearb.* 6, 7–25. doi: 10.1075/eurosla.6.04haw
- Hickmann, M., and Hendriks, H. (1999). Cohesion and anaphora in children's narratives: a comparison of English, French, German, and Mandarin Chinese. *J. Child Lang.* 26, 419–452. doi: 10.1017/S0305000999003785
- Holton, D., Mackridge, P., and Philippaki-Warbuton, I. (1997). *Greek: A Comprehensive Grammar of the Modern Language*. London: Routledge.
- Igarashi, Y. (2006). "Intonation patterns in Russian interrogatives – phonetic analyses and phonological interpretations," in *Usage-Based Linguistic Informatics: 3 Prosody and Syntax*, eds Y. Kawaguchi, I. Fonagy, and T. Moriguchi (Amsterdam: John Benjamins), 175–196. doi: 10.1075/ubli.3.11iga
- Ionin, T. (2006). This is definitely specific: specificity and definiteness in article systems. *Nat. Lang. Semant.* 14, 175–234. doi: 10.1007/s11050-005-5255-9
- Ionin, T., and Díez-Bedmar, M. B. (2021). "Article use in Russian and Spanish learner writing at CEFR B1 and B2 Levels: effects of proficiency, native language, and specificity," in *Learner Corpus Research Meets Second Language Acquisition*, eds B. Le Brun and M. Paquot (Cambridge: Cambridge University Press), 10–38. doi: 10.1017/9781108674577.003
- Ionin, T., Choi, S. H., and Liu, Q. (2021). Knowledge of indefinite articles in L2-English: online vs. offline performance. *Second Lang. Res.* 37, 121–160. doi: 10.1177/0267658319857466
- Ionin, T., Zubizarreta, M. L., and Maldonado, S. B. (2008). Sources of linguistic knowledge in the second language acquisition of English articles. *Lingua* 118, 554–576. doi: 10.1016/j.lingua.2006.11.012
- Kornfilt, J. (1997). *Turkish. Descriptive Grammars*. Abingdon-on-Thames: Routledge.
- Krause, T. (2019). *ANNIS: A Graph-Based Query System for Deeply Annotated Text Corpora*. Ph.D. Dissertation. Berlin: Humboldt-Universität zu Berlin. doi: 10.18452/19659
- Kupisch, T., and Rothman, J. (2018). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *Int. J. Biling.* 22, 564–582. doi: 10.1177/1367006916654355
- Lardiere, D., Smith, C. E., Brugos, A., and Micciulla, L. (2004). "Knowledge of definiteness despite variable article omission in second language acquisition," in *Proceedings of the 28th Annual Boston University Conference on Language Development* (Somerville, MA: Cascadia Press), 328–339.
- Leiss, E. (2010). "Koverter Abbau des Artikels im Gegenwartsdeutschen," in *Kodierungstechniken im Wandel. Das Zusammenspiel von Analytik und Synthese im Gegenwartsdeutschen*, eds D. Bittner and L. Gaeta (Berlin: de Gruyter), 137–158.
- Li Wei (2016). "Epilogue: multi-competence and the translanguaging instinct," in *The Cambridge Handbook of Linguistic Multi-Competence*, eds V. Cook and Li Wei (Cambridge: Cambridge University Press), 533–543.
- Lithoksoou, D. (2019). *To sintaktiko Phenomeno Tis Ipotaksis sta Kimena Helinoamerikanon Heritage Omilition (Speakers) Tis Helinikis Glosas (In Greek)*. Athens: Kapodistrian University.
- Loban, W. (1976). *Language Development: Kindergarten Through Grade Twelve. Research Report No 18*. Urbana, IL: National Council of Teachers of English.
- Lohndal, T., Rothman, J., Kupisch, T., and Westergaard, M. (2019). Heritage language acquisition: what it reveals and why it is important for formal linguistic theories. *Lang. Linguist. Compass* 36:165. doi: 10.1111/lnc3.12357
- Lowe, R. J. (2020). *Uncovering Ideology in English Language Teaching*. Berlin: Springer.
- Makarova, V. (2003). "Intonation features in categorization of Russian sentence type," in *Investigations Into Formal Slavic Linguistics IV*, eds P. Kosta and J. Blaszczak (Frankfurt/M., Germany: Peter Lang), 83–95.
- McCawley, J. D. (1993). "Gapping with shared operators," in *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 19, eds J. S. Guenter, B. A. Kaiser, and C. C. Zoll (Berkeley, CA: BLS), 245–253. doi: 10.3765/bls.v19i1.1507
- McGregor, A. (2020). *Sex Matters: How Male-Centric Medicine Endangers Women's Health and What We Can Do about It*. London: Quercus Publishing.
- Meyer, R., and Mleinek, I. (2006). How prosody signals force and focus – a study of pitch accents in Russian yes-no questions. *J. Pragmat.* 38, 1615–1635. doi: 10.1016/j.pragma.2005.05.011
- Montrul, S. (2008). *Incomplete Acquisition in Bilingualism: Re-Examining the Age Factor*. Amsterdam: Benjamins. [Studies in Bilingualism 39]
- Montrul, S., and Polinsky, M. (2011). Why not heritage speakers? *Linguist. Approaches Biling.* 1, 58–62.
- Montrul, S., and Polinsky, M. (2019). "Introduction to heritage language development," in *Oxford Handbook of Language Attrition*, eds M. Schmid and B. Köpcke (Oxford: Oxford University Press), 419–433.
- Myers-Scotton, C. (2006). *Multiple Voices. An Introduction to Bilingualism*. Malden, MA: Blackwell.
- Ortega, L. (2009). *Understanding Second Language Acquisition*. London: Hodder Education.
- Ortega, L. (2016). "Multi-competence in second language acquisition: inroads into the mainstream?," in *The Cambridge Handbook of Linguistic Multi-Competence*, eds V. Cook and Li Wei (Cambridge: Cambridge University Press), 50–76. doi: 10.1017/cbo9781107425965.003
- Payne, J., and Huddleston, R. (2002). "Nouns and noun phrases," in *The Cambridge Grammar of the English Language*, eds R. Huddleston and G. K. Pullum (Cambridge: Cambridge University Press), 323–524. doi: 10.1017/9781316423530.006
- Polinsky, M. (2015). Heritage languages and their speakers: state of the field, challenges, perspectives for future work, and methodologies. *Z. Fremdsprachenforschung* 26, 7–27.
- Polinsky, M. (2018). *Heritage Languages and Their Speakers*. Cambridge: Cambridge University Press.
- Pullum, G. K., and Huddleston, R. (2002). "Preliminaries," in *The Cambridge Grammar of the English Language*, eds R. Huddleston and G. K. Pullum (Cambridge: Cambridge University Press), 1–42. doi: 10.1017/9781316423530.002
- Rathcke, T. (2006b). "Relevance of f0 peak shape and alignment for the perception of a functional contrast in Russian," in *Proceedings of the 3rd conference on speech prosody*, (Dresden), 65–68.
- Rathcke, T. (2006a). "A perceptual study on Russian questions and statements," in *Arbeitsberichte des Instituts für Phonetik und Digitale Sprachverarbeitung (AIPUK 37)*, eds J. Harrington, C. Mooshammer, and F. Kleber (Kiel, Germany: Institut für Phonetik und Digitale Sprachverarbeitung), 51–62.
- Rathcke, T. (2009). *Komparative Phonetik und Phonologie der Intonationssysteme des Deutschen und Russischen. (Comparative Phonetics and Phonology of the Intonation Systems of German and Russian)*. Univ. Dissertation. München: Herbert Utz Verlag.
- Rathcke, T. (2013). On the neutralizing status of truncation in intonation: a perception study of boundary tones in German and Russian. *J. Phon.* 41, 172–185.
- Riester, A., and Baumann, S. (2017). *The RefLex Scheme – Annotation Guidelines*. Stuttgart: SFB. doi: 10.18419/opus-9011
- Roeper, T. (2003). "Multiple grammars, feature-attraction, pied-piping, and the question: is AGR inside TP?" in *(In)vulnerable Domains in Multilingualism*, ed. N. Müller (Amsterdam: Benjamins), 335–360.
- Romaine, S. (1989). *Bilingualism*. Oxford: Basil Blackwell.
- Rothman, J., and Treffers-Daller, J. (2014). A prolegomenon to the construct of the native speaker: heritage speaker bilinguals are natives too! *Appl. Linguist.* 35, 93–98. doi: 10.1093/applin/amt049
- Sarko, G. (2009). "L2 English article production by Arabic and French speakers," in *Second Language Acquisition of Articles: Empirical Findings and Theoretical Implications*, eds M. D. P. García Mayo and R. Hawkins (Amsterdam: Benjamins), 37–66. doi: 10.1075/lald.49.06sar

- Schmid, M. S., and de Leeuw, E. (2019). "Introduction to linguistic factors in language attrition," in *Oxford Handbook of Language Attrition*, eds M. Schmid and B. Köpcke (Oxford: Oxford University Press), 181–190.
- Schroeder, C. (1995). Post-predicate arguments in Turkish. *STUF Lang. Typol. Univ.* 48, 204–218.
- Schroeder, C. (2016). "Clause combining in Turkish as a minority language in Germany," in *Exploring the Turkish Linguistic Landscape. Essays in Honour of Eser Erguvanli-Taylan*, eds M. Güven, D. Akar, B. Öztürk, and M. Keleşir (Amsterdam: Benjamins), 81–101. doi: 10.1075/slcs.175.06sch
- Schroeder, C., Iefremenko, K., and Öncü, M. (to appear). "The post-verbal position in heritage Turkish. A comparative approach with a focus on non-clausal elements," in *Zweischprachigkeit Deutsch-Türkisch. Studien in Deutschland und in den Nachbarländern*, eds Z. Kalkavan-Aydın and Y. Şimşek (Münster: Waxmann).
- Seton, B., and Schmid, M. S. (2016). "Multi-competence and first language attrition," in *Cambridge Handbook of Linguistic Multi-competence*, eds V. Cook and L. Wei (Cambridge: Cambridge University Press), 338–354.
- Shadrova, A. (2020). *Measuring Coselectional Constraint in Learner Corpora: A Graph-Based Approach*. Ph.D. Dissertation. Berlin: Humboldt-Universität zu Berlin. doi: 10.18452/21606
- Snapé, N., García Mayo, M. D. P., and Gürel, A. (2013). L1 transfer in article selection for generic reference by Spanish, Turkish and Japanese L2 learners. *Int. J. Engl. Stud.* 13, 1–28. doi: 10.6018/ijes/2013/1/138701
- te Velde, J. (2017). Temporal adverbs in the Kiezdeutsch left periphery: combining late merge with deaccentuation for V3. *Stud. Linguist.* 71, 301–336.
- Tracy, R. (2002). Growing (clausal) roots: all children start out (and may remain) multilingual. *Linguistics* 40, 653–686.
- Tribushinina, E., van den Bergh, H., Kilani-Schoch, M., Aksu-Koç, A., Dabašinskienė, I., Hrzica, G., et al. (2013). The role of explicit contrast in adjective acquisition: a cross-linguistic longitudinal study of adjective production in spontaneous child speech and parental input. *First Lang.* 33, 594–616.
- Walkden, G. (2017). Language contact and V3 in Germanic varieties new and old. *J. Comp. Ger. Linguist.* 20, 49–81. doi: 10.1007/s10828-017-9084-2
- Wiese, H. (2012). *Kiezdeutsch. Ein neuer Dialekt entsteht*. München: C.H. Beck.
- Wiese, H. (2013). "What can new urban dialects tell us about internal language dynamics? The power of language diversity," in *Linguistische Berichte, Special Issue 19 [Dialektologie in neuem Gewand: Zu Mikro-/Varietätenlinguistik, Sprachenvergleich und Universalgrammatik]*, eds W. Abraham and E. Leiss (Hamburg: Buske), 207–245.
- Wiese, H. (2020). "Language situations: a method for capturing variation within speakers' repertoires," in *Methods in Dialectology XVI*, ed. Y. Asahi (Frankfurt: Peter Lang), 105–117. [Bamberg Studies in English Linguistics].
- Wiese, H., Alexiadou, A., Allen, S., Bunk, O., Gagarina, N., Iefremenko, K., et al. (2019). *RUEG Corpus [Data set]*. Zenodo.
- Wiese, H., and Müller, H. G. (2018). "The hidden life of V3: an overlooked word order variant on verb-second," in *Linguistische Berichte, Special Issue 25 [Non-Canonical Verb Positioning in Main Clauses]*, eds S. Müller and M. Antomo (Hamburg: Buske), 201–224.
- Wiese, H., and Rehbein, I. (2015). "Coherence in new urban dialects: a case study," in *Lingua 172/173: [Special Issue on Coherence, Covariation and Bricolage]*, eds G. Guy and F. Hinskens 45–61. doi: 10.1016/j.lingua.2015.10.016
- Wiese, H., Simon, H., Zimmer, C., and Schumann, K. (2017). "German in Namibia: a vital speech community and its multilingual dynamics," in *Language Contact in the German Colonies: Papua New Guinea and Beyond*, eds P. Maitz and C. A. Volker 221–245.
- Zdorenko, T., and Paradis, J. (2008). The acquisition of articles in child second language English: fluctuation, transfer or both? *Second Lang. Res.* 24, 227–250. doi: 10.1177/0267658307086302
- Zdorenko, T., and Paradis, J. (2012). Articles in child L2 English: when L1 and L2 acquisition meet at the interface. *First Lang.* 32, 38–62. doi: 10.1177/0142723710396797
- Zemskaja, E. A. (1973). *Russkaja Razgovornaja Reč*. Moscow: Nauka.
- Zimmer, C., Wiese, H., Simon, H. J., Zappen-Thomson, M., Bracke, Y., Stuhl, B., et al. (2020). Das Korpus *Deutsch in Namibia (Dnam)*: eine Ressource für die Kontakt-, Variations- und Soziolinguistik. *Dtsch. Sprach.* 48, 210–232. doi: 10.37307/j.1868-775X.2020.03.03
- Zuban, Y., Rathcke, T., and Zerbian, S. (2020). "Intonation of yes-no questions by heritage speakers of Russian," in *Proceedings of 10th Speech Prosody* (Tokyo), 96–100.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wiese, Alexiadou, Allen, Bunk, Gagarina, Iefremenko, Martynova, Pashkova, Rizou, Schroeder, Shadrova, Szucsich, Tracy, Tsehay, Zerbian and Zuban. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.