

ANNA VERSCHIK, HELIN KASK (Tallinn)

ENGLISH DISCOURSE MARKERS IN ESTONIAN-ENGLISH BILINGUAL BLOGS AND VLOGS

Abstract. The paper explores English discourse markers (DMs) in bilingual Estonian-English blogs and vlogs. The research questions are: (1) What are the functions of English DMs that appear in Estonian bilingual speech of vloggers and bloggers? (2) Is there any difference between vlogs and blogs as far as English DMs are concerned? 45 blogs and 8 vlogs were analysed (365,973 and 73,858 tokens respectively). The results show that the prevalent type of DMs were evaluatives, while the share of interactional performatives and markers of discourse structure and force was similar. The attractiveness of expressive meaning explains the preference for evaluative. The tendencies in blogs and vlogs are similar but the number of DMs in vlogs is higher because it is an oral genre.

Keywords: Estonian, English, discourse markers, codeswitching, language contacts.

1. Introduction

The topic of the paper is the usage of English discourse markers (hereafter DMs) in Estonian. DMs are defined as "sequentially dependent elements which bracket units of talks" (Schiffrin 1987 : 31; see discussion in Fuller 2003). DMs are heterogenous linguistic items (words and expressions like *hello, sorry, apparently, no way, oh my god*) that upon removal do not change the meaning of an utterance, and the utterance remains grammatical, although the truth-condition may change in some DMs, for instance, *like* (Fuller 2003: 186). A prototypical DM is illustrated in example (1).¹ The DM does not alter the meaning but is used to direct the discourse.

(1) B: *Anyway, tagasi tulles eilse päeva juurde, siis päev jätkub, jälle töö juures ning eriti rahulikult.*

¹ Hereafter, when an example is from a blog the original spelling is used. The spelling of examples from vlogs are based on automatic transcriptions that were corrected manually if needed. The example number is followed by a letter B or V. This marks if the example is from a blog (B) or a vlog (V).

Received 22 November 2022, accepted 2 March 2023, available online 10 June 2023.

© 2023 the Authors. This is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International Licence CC BY 4.0 (<http://creativecommons.org/licenses/by/4.0>).

'Anyway, coming back to yesterday, the day continues, again at work and especially calmly.'

Some authors, for instance, Matras (2009), Schourup (1999), Wertheim (2003) suggest that conjunctions belong to DMs because they often provide "links between utterances and assumptions" (Schourup 1999). In other words, DMs form a separate class of words that modify an utterance or show connections between and hierarchy of different utterances.

Borrowing of DMs in bilingual settings has become a prominent topic in the literature on contact linguistics since the late 1980s (i.e., Brody 1995; Hill, Hill 1986) and it would be reasonable to look at what happens as far as DMs are concerned in a recent contact situation (English-Estonian). Borrowing of DMs is widely attested in various contact settings (Blankenhorn 2003, Goss and Salmons 2000, Hlavac 2006, Maschler 1998, Peterson 2017, Salmons 1990, Sankoff et al. 1997, Torres and Potowski 2008, Wertheim 2003 to name just some studies). (Maschler 1994), their borrowing should be distinguished from lexical borrowing in general. It has been claimed that borrowing DMs is a diagnostic feature in contact situations (i.e., for so-called contracting languages, see Wertheim 2003: 216 ff.) and, thus, should be included in models of contact-induced language change.

The intensive contacts between Estonian and English are rather recent, starting from the restoration of independence 1991. Over the years English has become more important (see Kask 2021). In 2016, 91% of Estonians in the age of 15–64 claimed they knew at least one foreign language (https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Foreign_language_skills_statistics). English is the primary foreign language students learn at school: in 2019, 99.4% of Estonians in the upper secondary general education studied English (https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Foreign_language_learning_statistics&oldid=542152). Also, the overall proficiency in English among Estonians has risen: in 2008, 32% of Estonians claimed they have active English-language skills, by 2020 it was 47%. The tendency of using English more and more was also noted in a report studying integration of Estonian society: Estonians are using English as a language of work or studies more than Russian for the first time in 2020, and this trend is likely to get stronger (Kivistik 2020 : 33). As of the 2021 census, about 50% of the population speaks at least one language in addition to L1 and 13% two languages. English is the most spoken foreign language in Estonia, and ethnic Estonians are more proficient in English than in Russian, and among the younger generation (the 15–29 age group) 85% are proficient in English (<https://rahvaloendus.ee/en/results/demographic-and-ethno-cultural-characteristics-of-the-population>).

While the relatively recent impact of English on Estonian has received some scholarly attention (Kask 2021; Verschik, Kask 2019a; Verschik 2019) and English DMs in Estonian speech are discussed to some extent, there are no separate studies on the subject except for MA and BA theses (Matvejev 2021; Mägi 2022). There exist studies on DMs in other language pairs where one of the languages is Estonian, either as a dominated (Keevallik 2006a; 2006b) or dominant language (Zabrodskaja 2009).

This article examines the use of DMs in blogs and video blogs, in other words vlogs. Bloggers and vloggers clearly belong to the younger generation

of Estonians that, as already mentioned, is highly proficient in English. While this is only a tiny group and the results cannot be extrapolated to Estonians as a whole, we believe that it would be instructive to look at bilingual language use in this population. According to Matras (2009 : 3) and Weinreich (1968 : 1), any change, including contact-induced, starts in an individual and then may become conventionalized among bilinguals and, later, monolinguals. Therefore, we examine the group who is likely to be in the forefront of innovative language use.

A blog is considered a person's web journal or web home (Crystal 2007 : 15). Blogs are text-based, but vlogs are seen as their natural extension using new technology and providing a more diverse way to communicate with one's audience (Aran, Biel, Gatica-Perez 2014 : 201). Both genres are easily accessible, only a smart device and Internet is needed for posting, this makes blogging and vlogging attractive to many users. Blogs and vlogs are studied because they give a valuable insight to multilingual language use on the Internet. David Crystal (2007 : 15–16) explains that the language used in social media is 'naked'. This means that the language use is more immediate and unedited: bloggers and vloggers can express themselves the way they want, be more fluid, use fresh and novel expressions, seek colorful language use and so, establish their own language policy.

We ask the following research questions:

- (1) What are the functions of English DMs that appear in Estonian-English bilingual language use of vloggers and bloggers?
- (2) Is there any difference between vlogs and blogs as far as English DMs are concerned?

The paper is organised as follows. First, borrowing DMs and reasons thereof are discussed. Attention is paid to both form and functions of DMs. This is followed by the description of our data and methodology, the results and discussion.

2. Discourse markers in bilingual speech

2.1. General remarks

Rather often the relevant literature talks about borrowing DMs. First, it should be noted that we do not think that there are formal differences between code-switching (CS) and borrowings. This is especially true of markers that by definition do not change. Besides, the norms of monolingual and multilingual communities may differ. We adhere to usage-based approaches (Backus 2014) that, instead of formal criteria, suggest a continuum between occasional usage, entrenchment in individual speaker's cognition, and conventionalisation in a community. We will refer to English DMs as borrowings for the sake of simplicity while refraining from postulating conventionalization because we do not have information concerning the latter.

The literature on DM both in monolingual and bilingual settings suggests that the terminology is far from uniform. Some authors use *markers* and *particles* interchangeably (Dajko, Carmichael 2014 : 160), while some distinguish between the two. Fuller (2003) maintains that a particle is a discourse marker if certain conditions are fulfilled. Andersen and Fretheim (2000 : 1) believe that marker is broader than particle. There is a variation in use of the

attributes *discourse* and *pragmatic* (particles/markers): the title of the paper by Fraser (1990) mentions discourse markers and in a later paper (Fraser 1996) pragmatic markers. Klumpp (2022 : 19) talks about different traditions of grammatical description in different languages and uses the term DM as a general neutral term that is not dependent on the tradition of grammatical description of a particular language, and we do the same.

Since such linguistic items do not change the meaning of an utterance but rather modify them, Matras (1998) calls this word class *utterance modifiers*. These are gesture-like words that convey the speaker's attitude and help to structure the information, including conjunctions as well. It appears that in some bilingual situations both markers in the strict sense and conjunctions are borrowed (Wertheim 2003 on the Russian impact on Tatar), while in other situations it is either markers or conjunctions (Muhamedowa 2009 on Russian conjunctions in Kazakh). Since in our data no English conjunctions have been attested, we will not discuss the literature dedicated to the borrowing of conjunctions.

According to Maschler and Schiffrin (2015 : 190), there are three perspectives in the research of DMs: the discourse perspective (Schiffrin 1987; 2006), the pragmatic perspective (Fraser 1990; 2006) and the functional interactional perspective (Maschler 1998; 2000). The approach chosen by Matras (1998; 2009) can be labelled as functional-cognitive and may be considered within the functional interactional linguistic perspective. We follow the ideas of Maschler (1998; 2000) and Matras (1998; 2009) because these scholars have worked on DMs in language contact situations.

Borrowing of various DMs from English into other languages has been considered in the field of Anglicisms research (Mišić Ilić 2017; Peterson 2017; Zenner, van de Mieroop 2017). There are mentions of the pragmatic turn in Anglicisms research (Andersen, Furiassi, Mišić Ilić 2017) and English DMs are labelled as pragmatic Anglicisms (Mišić Ilić 2017). We do believe that borrowing of English DMs into Estonian should be investigated within the context of DMs borrowing in general. Historically speaking, there is nothing new in the fact that DMs are being borrowed; consider Estonian DMs that originate from a proto-Germanic (*jaa, jah* 'yes'; the conjunction *ja* 'and' from the same source or from Swedish) and Russian (*sutsu, suts* 'a bit' < *чуть* 'a bit, barely'; *tutki(t)* 'no way' < *дядьку*).

It is widely attested in the literature that in contact situations pragmatics is susceptible to impact or, in the terms of Matras (1998 : 203), contact vulnerable. The observation that DMs, like other lexical items, are borrowed from a sociolinguistically dominant/majority language may be true but this does not provide an exhaustive explanation why such borrowing occurs. On the contrary, in some situations L1 markers prevail when the speakers attempt to speak L2 (Matras 2009 : 96, example (13)). Also, in language shift DMs from the former L1 may be retained despite the claim that the traces of L1 appear in grammar and phonology rather than in lexicon; this may happen in ethnolects (Verschik 2014 : 50). Borrowed DMs can be emblematic of identity or politeness in outgroup communication when the speakers of language A as L1 do not feel confident in their skills in language B but demonstrate politeness to the speakers of B by the means of inserting some DMs from B like *hello, thank you, sure* and the likes into their L1, while talking to the speakers of A as L1 (this may be the case in Russian-to-Estonian communication when speakers of Russian as L1 do not want to stick to monolingual

Russian when talking to Estonian-speakers, and speak Russian with some Estonian insertions, including DMs see (Verschik 2008 : 153).

These facts require a different explanation than just sociolinguistic dominance and/or the level of proficiency. In the terms of Maschler (1994; 1998), DMs are different from other lexical items because they belong to the level of metalanguage: they comment and modify utterances that belong to the level of propositional language but do not change its meaning. According to Matras (1998), utterance modifiers, as he prefers to refer to DMs, are operators that "are responsible for monitoring and directing the hearer's processing of propositional content". Having one set of DMs helps to ease the cognitive load experienced by a bilingual speaker. Matras (2000 : 521) proposes the notion of pragmatically dominant language that is "the language toward which a speaker directs maximum mental effort at a given instance of linguistic interaction". Pragmatically dominant language is the one easily activated and the default choice and can differ in different communicative situations (Matras 2009 : 98). That is, it is the language that takes less effort to produce.

In our case it means that we can only make conclusions about DMs in the particular genres, blogs and vlogs that use both languages; the same language users may have Estonian-only register that they employ in other situations, but we do not have access to other parts of their repertoire and, therefore, cannot judge about replacement/retention on a large scale.

Not only the reasons for DMs susceptibility to borrowing but also the process and the outcomes should be considered as well. As the empirical data from different contact situations shows, the whole system of DMs may be either entirely (Goss, Salmons 2000) or partially (de Rooij 2000; Hlavac 2006) replaced by the set of borrowed DMs (historically speaking, the system of Estonian DMs is mixed) or, according to Solomon (1995), differentiation of functions (i.e. complementary distribution) is characteristic of the situation of stable bilingualism, that is, language maintenance. Markus (2022) shows that borrowed DMs may coexist with their native equivalents in rather similar functions. As Hlavac (2006 : 1880) notes, not only borrowing (meaning + form) may take place but also copying of functions of borrowed DMs onto their native equivalents: for instance, *da* is being used more frequently in Australian Croatian than in the homeland variety of Croatian due to the adoption of Australian English discourse conventions). In the same spirit, Keevallik (2006a) in her study of Estonian in Sweden shows that some Estonian markers have extended their functions.

To provide a general picture of the "lifecycle" of foreign DMs, Peterson (2017 : 123) suggests a descriptive model that shows a continuum between insertional CS, an intermediate stage that she calls *limited borrowing*, and full borrowing that implies integration into recipient language, adaptation and usage by all social classes. However, one has to distinguish between conventionalization among all speakers of the recipient language and conventionalization among the bilingual speakers, for the need for adaptation may differ in bilingual and monolingual speakers (see Leisiö 2001) because bilinguals differ cognitively (i.e., the multi competence model, Cook 2016; Dewaele 2016). New items and patterns may preserve their original features, compromise forms and bilingual constructions may emerge that are absent from the separate monolingual varieties. In this article, we are not concerned with monolingual speakers or with those bilinguals who use English only occasionally.

Maschler (2000) proposes that the development of borrowed DMs should be viewed in the terms of Auer's (1999) model that, in a nutshell, presents a continuum from CS (other language items are novel, the novelty is visible, CS is meaningful locally) via language mixing (the juxtaposition of languages is meaningful not only locally but globally) towards fused lects (where other language items have lost their novelty, may become obligatory, and the contrast has faded away). At the current stage of English-Estonian language contacts, it is too early to speak about fused lects and conventionalisation; we have only evidence from bilingual speakers who are eager to use English and, therefore, may differ from other speakers of Estonian.

2.2. Classification

The next issue to be considered is the ways of DMs classification. It appears that there is no unity among researchers, partly because of differences in approaches and interests, partly because of the great diversity of empirical data and contact situations. Klumpp (2022 : 30) even suggests that the ways and approaches to DMs have changed and will undoubtedly change in future, and researchers should not be afraid of using their intuition for the description of markers.

There are different classifications, proposed for monolingual varieties (for instance, Hennoste, Klumpp, Metslang 2022 for Estonian; more generally: Fraser 1996; Maschler 1998) as well as for contact situations (such as Wertheim 2003 : 162 ff). There is no agreement among scholars as far as classification of DMs is concerned, and most often the approach is descriptive. Wertheim (2003 : 162) suggests that, based on functions, there are three groups of DMs with several subgroups. While she provides Russian examples, we use English equivalents here to illustrate the classification.

(1) Markers of discourse structure and force (what Fraser (1996) calls discourse markers). These include subordinative, coordinative, and contrastive markers (conjunctions and markers like *as well*), and markers of metacommentary and deixis (metacommentary on text, some of which is deictic reference to preceding and upcoming information: *in brief, for example, what's more*).

(2) Interactional performatives (congruent with Fraser's basic markers); greetings, farewells, words like *thank you, please*, labelled by Wertheim (2003 : 162) as pragmatic idioms.

(3) Evaluatives express the speaker's stance and evaluation of discourse (Wertheim 2003 : 163). It is the largest category and includes: positive evaluation (*cool, great*), negative evaluation (*shame, too bad*); hedges (*only, almost, simply*); epistemics (evaluation of probability / of time and degree: *possibly, almost, at once, totally*).

We are going to use the mentioned classification by Wertheim (2003 : 162 ff) because it is designed for contact situations and because it has been used by various researchers for the cases where one of the languages is Estonian (Keevallik 2006a; 2006b; Verschik 2008 : 154–155; Zabrodskaja 2006 : 744).

3. Data and methodology

Our written data comes from blogs, the corpus consists of 851 blog posts from 45 blogs (365,973 tokens). The corpus of oral data comes from vlogs

and comprises 48 videos from 8 vloggers (10 hours 1 minute 41 seconds of footage, approximately 74,000 tokens). To analyse vlogs, the videos were transcribed using a full-automatic web-based speech recognition programme that is designed for monolingual Estonian (Alumäe, Tilk, Asadullah 2019). English words were automatically replaced with similar words in Estonian, all errors, both in Estonian and English, were corrected manually. Each text was checked twice to avoid possible mistakes. We used the programme Notepad++ where we manually marked contact phenomena, including DMs. Each blog post/vlog episode was saved as a separate file. For quantification, a special programme Corpus Stat² was used.

The studied blogs and vlogs had to meet five criteria: 1) the topic is beauty, lifestyle and/or fashion, so the results reflect the language use in one field; 2) the blogger/vlogger has to live in Estonia and 3) the main language of the blog/vlog is Estonian because our interest is to study the impact of English on the language use of young Estonians living in Estonia;³ 4) the blog/vlog is publicly accessible (at least at the time of posting); 5) the post has to contain at least one lexical element from English. However, proficiency in and previous exposure to English did not play a role when choosing the informants. This is based on modern views on bilingualism that emphasise ability to use rather than measuring proficiency level (for instance, see Cenoz 2013 and references therein).

Finding and classifying DMs may be at times a difficult task because there are phrases and fixed expressions that technically look like autonomous sentences, yet their function is that of DM. Since these phrases have a complex meaning that is not a mechanical sum of separate meanings of each component, they are counted as one token. Consider example (2):

(2) B: *Üks mu sõpradest tahab lausa juba siit linnast endale ka ühe kutsa päriseks varjupaigast võtta, h o w c o o l i s t h a t ?!*

'One of my friends even wants to adopt a dog from a shelter from this city, how cool is that?!'

From the point of view of CS typology, these are instances of code alternation (CA) in Muysken's (2000) terms. CAs are stretches in another language, occur between the clauses, that is, a constituent in language A is followed by a constituent in language B (Muysken 2000 : 7). We were guided by the criteria provided in the numerous definitions of DM (see Section 1 and 2): DMs belong to meta-language and modify the utterance but do not change the meaning; if DMs are removed, the sentence remains grammatical. Apparently, there are borderline cases: some CAs function like DMs and some are prototypical CAs (Demirçay 2017 : 17, 30).

After marking DMs in the texts, we created an Excel table that contained the following information: the particle and the sentence it appeared in, the source (a blog or a vlog and the user) and the function. The function was assigned based on the classification proposed by Wertheim (2003, see section 2.2) and it was done manually, taking into account the context and the

² This program was created by Katrin Tsepelina for the project EKKM 09-85 within the national programme "Estonian Language and Cultural Memory".

³ The fact that these are young Estonians and that they live in Estonia is easily established because (1) they publish pictures/videos of themselves, so it is clear that they are rather younger than older (2) blogs and vlogs contain the relevant information: bloggers and vloggers introduce themselves (20 year old student living in Tallinn).

sentence it was used in. To avoid possible mistakes, the function was checked twice, by both authors.

4. Results

The total number of the overt English lexical items (CSs), i.e., tokens, is 4,600, including 808 DMs (tokens) whose share is 18%. Prototypical CAs (syntactically autonomous stretches in English) are not included here. Firstly, taking into account Wertheim's classification (see Section 2.2) we analyse the function of the DMs in each group and, secondly, we will compare what occurs in blogs and vlogs.

The DMs belonging to the group of markers of discourse structure and force were the least used in our data. This group includes conjunctions but in our data, we did not attest English conjunctions. The DMs from this group were mostly markers of metacommentary and deixis, the most used (i.e. the number of tokens) were *anyway* (30), *by the way* (16), *also* (14) and *actually* (14). These were used to add some side comments or when the blogger/vlogger started to tell a story and got off the topic. Consider example (3) from a blog post where a blogger was describing her trip and how they missed their flight. She adds a side comment that is not related to the main storyline of how they got home. When she returns to the main topic, she starts it with the DM *anyway*.

(3) B: *Nagu kuidas on selline asi üldse võimalik, et mingi lennuk on üle bookitud? A n y w a y, töötajad hakkasid meile kohe uut lendu otsima.*

'Like how is it possible at all that a plane is overbooked? Anyway, the workers started to look for a new flight for us.'

The use of other markers in this group was less frequent. There were few examples of using subordinative markers, e.g., *although* (one time) and *though* (one time), and contrastive markers, e.g., *still* (two times) and *however* (one time). Only one coordinative marker (*so*) was used. Consider example (4) from a blog where a blogger was describing how she likes to write poems and what makes a poem enjoyable. The post is poetic and uses different metaphors but at the end of the post the blogger warns her readers to not send her any poems. She adds an exception and uses a contrastive marker *though*.⁴

(4) B: *Kirjapulki oma Facebooki sõnumitesse ei oota. T h o kui sa tabasid kõik kahemõttelisused ära, siis see on muljetavaldav.*

'I am not expecting any lengthy texts in my Facebook mailbox. Though, if you managed to capture all the ambiguities, then it's impressive.'

The group of interactional performatives contains different DMs that are often used at the beginning or at the end of the post, these are greetings and farewells, e.g., *hello* (95), *cheers* (87), *bye* (7). Some of the informants use English DM for both: for greeting and farewell. It is interesting to note that sometimes other languages are used for this function, too, e.g., *hola*, *au revoir*. Consider example (5) where a blogger starts her post with *hello* and ends with *cheers*. The main text of the post is in Estonian. This is a typical way to start or to end a post.

⁴ In English, *tho(uhg)* may be considered as a conjunction but in the current example it is classified as a contrastive particle it has the meaning 'still, however' (in Estonian, it would be *siiski*) and does not introduce a relative clause. Therefore, not only the form but the meaning and function has to be considered.

- (5) B: *Hello!* [post in Estonian]
Cheers!

There are many examples of using the English DM for apologising, thanking, and asking: *sorry* (35), *thanks* (16) and *thank you* (11), *please* (11). Consider example (6) where a vlogger uses an English DM for thanking and apologising within the same utterance. It is quite common to thank their audience for watching a vlog at the end of the video and often an English DM is used for that. In this case, the vlogger feels that the video was not good, she has expressed in her previous videos that she has been struggling to find ideas for new videos. When she finally publishes a new video, she finds that the video is not good and feels the need to apologise in front of her audience.

- (6) V: *Thanks, et vaatasite, ja sorry selle mõttetu video pärast.*
'Thank you for watching and sorry for this pointless video.'

The largest and the most diverse is the group of evaluatives. Here are subgroups containing DMs to express positive or negative evaluation, hedges, or evaluation of probability. One of the most used DM was *oh my god* or its abbreviation *omg* (both considered as one token), in total 74 tokens. The reason why this specific DM is so popular is not yet known; it would be useful to look into Estonian youth language use in general; so far, Estonian youth language research is in its initial stages, and nothing can yet be said about frequency. This could mark the informant's positive emotions, excitement and sometimes relief. Consider example (7) where the vlogger uses the expression to emphasise her pure joy of going to an event she has been wanting to go to for a very long time.

- (7) V: *Ah, oh my god, kõik mu unistused täituvad!*
'Ah, oh my god, all my dreams do come true!'

However, there are several instances where the DM *oh my god* is used to express disappointment or even anger. Consider example (8) where a blogger talks about the weather and while it seems to her that everyone else around her is enjoying hot weather, she does not like it. She also uses capital letters to stress her viewpoint.

- (8) B: *O M G KUIDAS MA EI SALLI PALAVUST.*
'Oh my god I can't stand the heat.'

Positive evaluation was also expressed with DMs *wow* (20), *love* (marking appreciation etc.) (10), *yay* or *yey* (considered as one token, 10 occurrences in total), *amazing* (3) etc. Consider example (9) where a blogger uses the DM *yey* to express that she is happy over something. Mostly, DMs occur at the beginning of the utterance, but this is not a rule, and they can also be used at the end of the utterance.

- (9) B: *Otsustasin vestlusele minna ja saingi tööle, y e y!*
'I decided to go to the job interview, and I got the job, yey!'

DMs expressing negative evaluation were often swearwords, e.g., *fuck* (16), *damn* (9) and *holy shit* (3). Also, DMs like *awkward* (2) and *eww* (1) were used to describe a weird or an uncomfortable situation. Consider example (10) where a blogger looks at a photo of herself and does not like her outfit.

- (10) B: *Ma kandsin oma pidžaamapükse — a w k w a r d.*
'I was wearing my pajama pants — awkward.'

Among the group of evaluatives the following epistemic markers were often used, e.g., *ever* (7), *obviously* or its abbreviation *obv* (5), *indeed* (4), *probably* (4). Consider example (11) where a vlogger is talking about how she is filming a new video and she wants to stress that she is doing it for her audience. She uses the English DM *obviously* to highlight it.

- (11) V: *Filmin ühte uut lahedat videot. Teile, o b v i o u s l y.*
'I'm filming a new cool video. For you, obviously.'

Epistemics were also used to evaluate time, e.g., *again* (2), *finally* (1), *once and for all* (1). Consider example (12) where a blogger has been waiting for spring so she could wear lighter clothes. Warmer weather is arriving, and the blogger is excited about it. She uses the DM *finally* to mark the time.

- (12) B: *Peagi on käes aeg kevadriiete jaoks, f i n a l l y!*
'Soon it's time to wear spring clothing, finally!'

Hedges were used only a few times, e.g., *somehow* (1), *or something* (1) and *I guess* (1). Consider example (13) where a blogger did her make-up and she felt insecure about it. She uses an English DM *somehow* to mark it.

- (13) B: *Mul pole olnud mahti ega aega väga harjutada ja olin ka seda paletti kätte võttes ebakindel, aga s o m e h o w tuli mul väga kena silma-meik välja.*
'I haven't had the chance or the time to practice and I was insecure when using this palette but somehow the eye make-up turned out really nice.'

Secondly, we analysed by group what DMs are used most frequently (see Figure 1). The largest group in our data is evaluatives (in total 396 tokens, or 49%), e.g., *damn*, *oh my god*, this is followed by interactional performatives (297 tokens, or 37%), e.g. *sorry*, *cheers*, and the smallest group is markers of discourse structure and force (115 tokens, or 14%), e.g. *also*, *last but not least*. However, it is important to note that the figures in the group of interactional performatives were heavily affected by one blogger who started and ended each post with an English greeting and farewell. As she is well represented in our corpus, 160 DMs (tokens) belonging to the group of interactional performatives (54% of DMs in that group) were used by her.

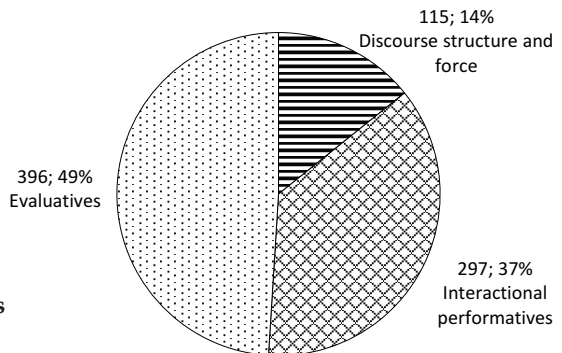


Figure 1. The distribution of DMs (tokens).

When comparing blogs and vlogs, there are noticeable differences concerning the two larger groups (see Figure 2). In blogs, the share of evaluatives and interactional performatives are roughly the same: respectively, 257 and 246 tokens (45% and 43%). In blogs, DMs of discourse structure and force constituted a noticeably smaller group than the two others: 70 tokens (12%). Notably, in vlogs evaluatives are heavily prevailing: 139 tokens (59.1%) while the number and share of interactional performatives and markers of discourse structure and force are rather similar: respectively, 51 and 45 tokens (22% and 19%).

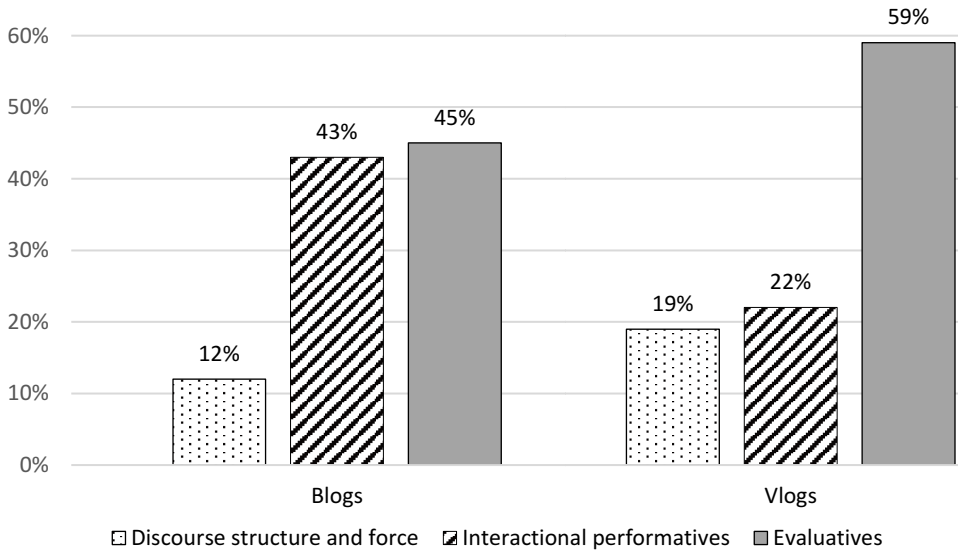


Figure 2. The share of DMs (tokens) in blogs and vlogs.

To get a better overview of the differences in blogs and vlogs, the share of DMs (tokens) were also analysed leaving out the one blogger who is heavily represented in our corpus and is characterised by overuse of interactional performatives (see Figure 3). This blogger used 177 DMs (tokens) in total (31% of DMs in all blogs together): 6 (8.6%) DMs belonged to the group of discourse structure and force, 160 (65%) to the group of interactional performatives and 11 (4.3%) to the evaluatives group. When the data from this particular blogger is left out, the results for blogs and vlogs are similar: evaluatives heavily prevail (246 tokens, 62%) and the share of interactional performatives and markers of discourse structure and force are comparable (respectively, 86 tokens, 22% and 64 tokens, 16%).

As illustrated in Figure 3, after the exclusion of the blog that tends to overuse interactional performatives, the size of the functional groups of DMs in blogs and vlogs is roughly the same: evaluatives heavily prevail, staying around 60%, this is followed by interactional performatives (22%) and the group of discourse structure and force is the smallest (around 17.5%). All DMs of the latter group belong to the subgroup of metacommentary and deixis, and no conjunctions were attested.

To compare how often DMs are used in blogs and in vlogs, the occurrence of DMs (tokens) was calculated per 1,000 words. In blogs, there are 1.6 DMs per 1,000 words and in vlogs it is 3.2 DMs. Thus, in vlogs DMs are used twice as much as in blogs.

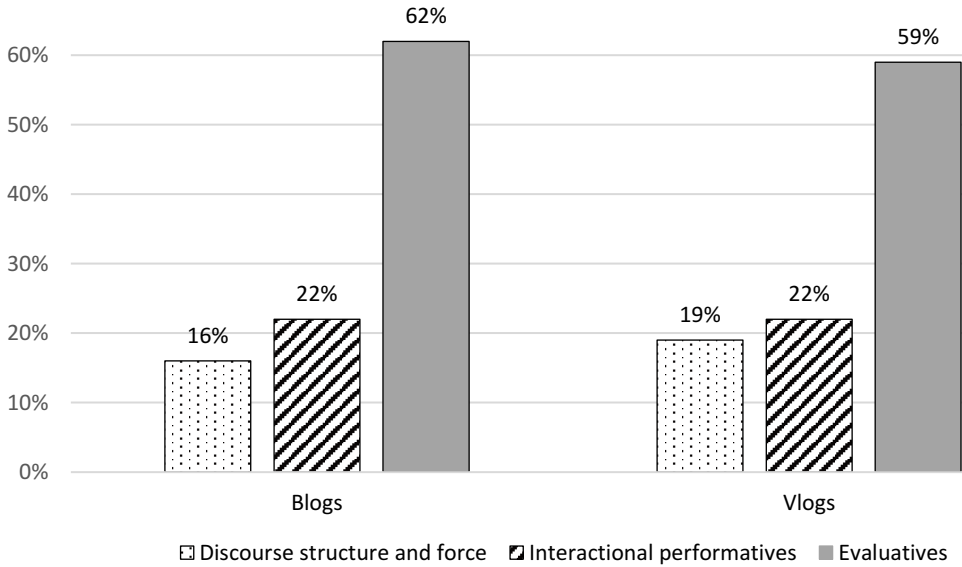


Figure 3. The share of DMs (tokens) in blogs and vlogs leaving out one blog.

Among 808 DMs in blogs and vlogs, there were 174 types of markers. The most diverse is the group of evaluatives where there are 127 types of DMs, the other two categories are roughly the same: 29 different DMs in the category of discourse structure and 22 in the category of interactional performatives. The most used DMs (tokens) were *hello* (95) and *cheers* (87). To specify, these were the DMs that one blogger heavily used (respectively 77 and 82). Taking this into consideration, the most used DM was *oh my god* (74). This was followed by *sorry* (35), *anyway* (30) and *wow* (20). Other DMs were used less than 20 times (see Figure 4).

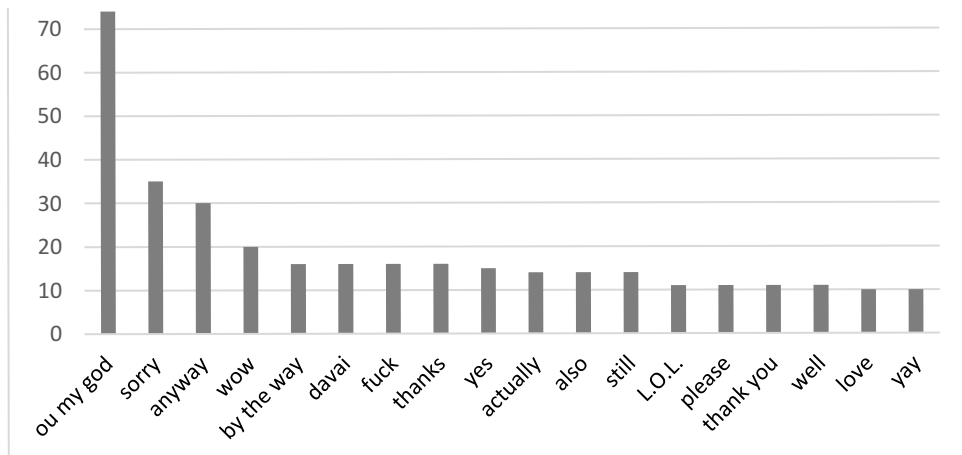


Figure 4. 18 most used DMs.

To sum up, all three functional groups of DMs are represented in our corpus but certain functional sub-groups (e.g., evaluatives) prevail.

5. Discussion and conclusions

Our results are, of course, based on our data and cannot be extrapolated to all English-Estonian bilingual speakers. To make generalizations, the findings need to be checked across other genres and text types. The answer to the research question (1) concerning the functions of DMs, the answer is as follows. DMs belonging to all three functional groups (markers of discourse structure and force, represented by markers of metacommentary and deixis, interactional performatives, and evaluatives) are present both in vlogs and blogs. However, conjunctions (a subgroup of DMs of discourse structure and force) are strikingly absent.

Research question (2) focuses on the similarities and differences in blogs and in vlogs. The share of the three functional groups is rather similar in blogs and in vlogs, with a clear preference for evaluatives. At the same time, in our data, there are twice as many DMs in vlogs per 1,000 words than in blogs (accordingly 3.2 and 1.6 DMs). This is explicable with the difference of medium (oral, spontaneous vs. written, edited). To maintain a normal speech flow, usually elements from pragmatically dominant language are activated first. Matras (1998 : 281) has explained that this is an automatic process to reduce tension. One might say that language use in vlogs is not spontaneous but rather often rehearsed and follows the script prepared in advance. Nevertheless, unlike in blogs, this is an oral genre that differs from written genres, and self-repairs, pauses, searching for the right wording occur there. Apparently, the vloggers speak in a way that they assume to be acceptable to their audience, and this kind of language use is characteristic to Estonian-English vlogging.

The latter is in line with the results in other studies that deal with English DMs in Estonian in vlogs (Matvejev 2021; Mägi 2022). The corpora in questions are small, yet the tendency is clear. The prevalence of evaluatives and the lack of conjunctions (in all genres, i.e., podcasts, blogs and vlogs) calls for further investigation into why this is the case. Still, a provisional explanation may be provided.

Matras (2012 : 42) proposes various borrowability hierarchies for different kinds of linguistic items, including conjunctions and discourse markers. He suggests that items with contrastive meaning are more prone to borrowing than items with other meanings (see also Matras 2005) and first candidates for borrowing are contrastive conjunctions. The suggestion is based on empirical data from various contact situations. Although in his earlier work Matras (1998) uses the umbrella term utterance modifiers that includes both DMs in the classical sense and conjunctions, he provides two separate hierarchies for them. However, a possible general borrowability continuum is needed for all utterance modifiers. We suggest that borrowability is linked to meaning (and this follows from the fact that contrastive meaning is more prone to borrowing than addition, concession etc.).

As stated in the descriptive model that combines cognitive and linguistic factors (Verschik 2019 : 76), certain types of meaning make a linguistic item more cognitively prominent and noticeable and, therefore, a likely candidate for copying (borrowing). Items with specific meaning (typically, nouns but not only) but also with expressive and figurative meaning would be copied earlier than items with a more abstract meaning (see also Verschik 2020). DMs have a more specific meaning (expressing speakers' attitudes) than conjunctions

that show relations between utterances, and DMs that convey the speakers' attitude (that is, evaluatives) would then prevail over other DMs. Here meaning outweighs frequency because it would be expected that conjunctions and grammatical words in general are more frequent (Backus, Verschik 2012). However, this may be different for interactive genres, and it has to be acknowledged that most of the studies on DM borrowing referred here are based on conversational data in situations where people actually talk to each other and where the need for interactional performatives is therefore higher than in the monologic genres described in the current study.

The case of the blogger whose overuse of interactional performatives is distinctive deserves attention for several reasons. First, a small community (and a small corpus) calls for attention to the role of an individual speaker (recall that language contacts are not about systems but about multilingual resources in a multilingual's mind and their employment in a given communicative situation, Matras 2009 : 38). Second, in bilingual English-Estonian blogs certain post templates have emerged (Verschik, Kask 2019b : 314), for instance, Estonian-dominant entries with titles (and introductory/concluding phrases) in English. Here again the specifics of blogs as genre (monologic, not real-time communication, edited, more features of written texts, etc., see Kask 2021) plays a role.

In sum, the vlogs and blogs exhibit similar tendencies as far as the presence of English DMs is concerned. The prevalence of evaluatives among DMs suggests that in further research we need a closer look into various types of meaning as a predictor of chronology of contact-induced language change.

Acknowledgments. The research was supported by the Ministry of Education and Research of Republic of Estonia, grant EKKD33 "Data and corpora of Estonian children and youth multilingual communication".

The publication costs of this article were covered by the Estonian Academy of Sciences and the Estonian Research Council grant PUT 1290.

Addresses

Anna Verschik
Tallinn University
E-mail: annave@tlu.ee

Helin Kask
Tallinn University
E-mail: helin.kask@hotmail.com

REFERENCES

- Alumäe, Tanel, Tilk, Ottokar, Asadullah 2019, Advanced Rich Transcription System for Estonian Speech. — *Human Language Technologies — The Baltic Perspective* 308, Amsterdam, 1–8.
- Andersen, Gisle, Frøtheim, Thorstein 2000, Introduction. — *Pragmatic Markers and Propositional Attitude*, Amsterdam—Philadelphia, 1–16.
- Andersen, Gisle, Furiassi, Cristiano, Mišić Ilić, Biljana 2017, The Pragmatic Turn in Studies of Linguistic Borrowing. — *Journal of Pragmatics* 113, 71–76.
- Aran, Oya, Biel, Joan-Isaac, Gatica-Perez, Daniel 2014, Broadcasting Oneself: Visual Discovery of Vlogging Styles. — *IEEE Transactions on Multimedia* 16 (1), 201–215.
- Auer, Peter 1999, From Codeswitching via Language Mixing to Fused Lects: Toward a Dynamic Typology of Bilingual Speech. — *International Journal of Bilingualism* 3 (4), 309–332.

- B a c k u s, A d 2014, Towards a Usage-Based Account of Language Change: Implications of Contact Linguistics for Linguistic Theory. — Questioning Language Contact, Leiden—Boston, 91—118.
- B a c k u s, A d, V e r s c h i k, A n n a 2012, Copyability of (Bound) Morphology. — Copies versus Cognates in Bound Morphology, Boston—Leiden, 123—149.
- B l a n k e n h o r n, R e n a t e 2003, Pragmatische Spezifika der Kommunikation von Russlanddeutschen in Sibirien. Entlehnung von Diskursmarkern und Modifikatoren sowie Code-Switching, Berlin.
- B r o d y, J i l l 1995, Lending the "Unborrowable": Spanish Discourse Markers in Indigenous American Languages. — Spanish in Four Continents. Studies in Language Contact and Bilingualism, Washington, 132—147.
- C e n o z, J a s o n e 2013, Defining Multilingualism. — Annual Review of Applied Linguistics 33, 3—18.
- C o o k, V i v i a n 2016, Premises of Multi-Competence. — The Cambridge Handbook of Linguistic Multi-Competence, Cambridge, 1—25.
- C r y s t a l, D a v i d 2007, Language and the Internet, Cambridge—New York.
- D a j k o, N a t h a l i e, C a r m i c h a e l, K a t i e 2014, But qui c'est la différence? Discourse Markers in Louisiana French: The Case of *but* vs. *mais*. — Language in Society 43, 159—183.
- D e m i r ç a y, D e r y a 2017, Connected Languages. Effects of Intensifying Contact between Turkish and Dutch, Utrecht.
- D e w a e l e, J e a n - M a r c 2016, Multi-Competence and Personality. — The Cambridge Handbook of Linguistic Multi-competence, Cambridge, 403—419.
- F r a s e r, B r u c e 1990, An Approach to Discourse Markers. — Journal of Pragmatics 14, 383—395.
- — 1996, Pragmatic Markers. — Pragmatics 6 (2), 167—190.
- — 2006, Towards a Theory of Discourse Markers. — Approaches to Discourse Particles, Amsterdam, 189—204.
- F u l l e r, J a n e t 2003, Discourse Marker Use Across Speech Contexts: A Comparison of Native and Non-Native Speaker Performance. — Multilingua 22, 185—208.
- G o s s, E m i l y L., S a l m o n s, J o s e p h 2000, The Evolution of a Bilingual Discourse Marking System: Modal Particles and English Markers in German-American Dialects. — International Journal of Bilingualism 4 (4), 469—484.
- H e n n o s t e, T i i t, K l u m p p, G e r s o n, M e t s l a n g, H e l l e 2022, Diskursuse markerid ja pragmaatika. Sissejuhatuseks. — KK, 3—18.
- H i l l, J a n e K., H i l l, K e n n e t h C. 1986, Speaking Mexicano. Dynamics of Syncretic Language in Central Mexico, Tucson.
- H l a v a c, J i m 2006, Bilingual Discourse Markers: Evidence from Croatian-English Code-Switching. — Journal of Pragmatics 38 (11), 1870—1900.
- K a s k, H e l i n 2021, English-Estonian Code-Copying in Estonian Blogs and Vlogs, Tallinn.
- K e e v a l l i k, L e e l o 2006a, Keelekontakt ja pragmaatika. — Teoreetiline keeleteadus Eestis II, Tartu, 85—96.
- — 2006b, Pragmaatiliste partiklite laenuüübid rootsieesti keeles. — Mitmõkeelisus ja keelevaihtus õdagumeresoomõ maiõ pääl, Võro (Võro Instituudi toimõndusõq 18), 116—150.
- K i v i s t i k, K a t s 2020, Eesti elanike keeleoskus ja -kasutus, inimestevahelised kontaktid, sallivus ning eesti keele väärtustamine. — Eesti integratsiooni monitooring 2020, Tallinn, 28—43.
- K l u m p p, G e r s o n 2022, Väikeste uurali keelte diskursuspartiklite kirjeldamise vajadus ja keerukus. — KK, 19—34.
- L e i s i ö, L a r i s a 2001, Integration and Gender Assignment in Finland Russian. — SKY Journal of Linguistics 14, 87—119.
- M ä g i, M a i k e n 2022, Ingliskeelsed diskursuspartiklid taskuhäälingus "Päevapokaal", Tallinn.
- M a r k u s, E l e n a 2022, Diskursusepartiklid isuri keele Soikkola murdes. — KK, 74—94.
- M a s c h l e r, Y a e l 1994, Meta-Languaging and Discourse Markers in Bilingual Conversation. — Language in Society 23, 325—366.
- — 1998, On the Transition from Codeswitching to a Mixed Code. — Code-Switching in Conversation, London, 125—149.

- — 2000, What Can Bilingual Conversation Tell Us about Discourse Markers? Introduction. — *International Journal of Bilingualism* 4 (4), 437–445.
- M a s c h l e r, Yael, S c h i f f r i n, Deborah 2015, Discourse Markers. Language, Meaning, and Context. — *The Handbook of Discourse Analysis*, Malden—Oxford—Chichester, 189–221.
- M a t r a s, Yaron 1998, Utterance Modifiers and Universals of Grammatical Borrowing. — *Linguistics* 36 (2), 281–331.
- — 2000, Fusion and the Cognitive Basis for Bilingual Discourse Markers. — *International Journal of Bilingualism* 4 (4), 505–528.
- — 2005, The Full Extent of 'Fusion': A Test Case for Connectivity and Language Contact. — *Prozesse des Wandels in historischen Spannungsfeldern Nordostafrikas/Westasiens, Würzburg (Kulturelle und sprachliche Kontakte 3)*, 241–255.
- — 2009, *Language Contact*, Cambridge—New York.
- — 2012, An Activity-Oriented Approach to Contact-Induced Language Change. — *Dynamics of Contact-Induced Language Change*, Berlin, 17–52.
- M a t v e j e v, Maiken 2021, *Inglise diskursuspartiklid eesti vlogides*, Tallinn.
- M i š i ć I l i ć, Biljana 2017, Pragmatic Borrowing from English into Serbian: Linguistic and Sociocultural Aspects. — *Journal of Pragmatics* 113, 103–115.
- M u h a m e d o w a, Raihan 2009, The Use of Russian Conjunctions in the Speech of Bilingual Kazakhs. — *Journal of Bilingualism* 13 (3), 331–356.
- M u y s k e n, Pieter 2000, *Bilingual Speech. A Typology of Code-Mixing*, Cambridge.
- P e t e r s o n, Elizabeth 2017, The Nativization of Pragmatic Borrowings in Remote Language Contact Situations. — *Journal of Pragmatics* 113, 116–126.
- d e R o o i j, Vincent A. 2000, French Discourse Markers in Shaba Swahili Conversations. — *International Journal of Bilingualism* 4 (4), 447–467.
- S a l m o n s, Joseph 1990, Bilingual Discourse Marking: Code Switching, Borrowing, and Convergence in Some German-American Dialects. — *Linguistics* 28 (3), 453–480.
- S a n k o f f, Gillian, T h i b a u l t, Pierrette, N a g y, Naomi, B l o n d e a u, Hélène, F o n o l l o s a, Maria-Odile, G a g n o n, Lucie 1997, Variation in the Use of Discourse Markers in a Language Contact Situation. — *Language Variation and Change* 9 (2), 191–217.
- S c h i f f r i n, Deborah 1987, *Discourse Markers*, Cambridge (*Studies in Interactional Sociolinguistics* 5).
- — 2006, Discourse Marker Research and Theory: Revisiting *and*. — *Approaches to Discourse Particles*, Amsterdam, 315–338.
- S c h o u r u p, Lawrence 1999, Discourse Markers. — *Lingua* 107, 227–265.
- S o l o m o n, Julie 1995, Local and Global Functions of a Borrowed Native Pair of Discourse Markers in a Yucatan Mayan Narrative. — *Proceedings of the 21st Annual Meeting of the Berkeley Linguistic Society*, Berkeley, 287–298.
- T o r r e s, Lourdes, P o t o w s k i, Kim 2008, A Comparative Study of Bilingual Discourse Markers in Chicago Mexican, Puerto Rican, and MexiRican Spanish. — *Journal of Bilingualism* 12 (4), 263–279.
- V e r s c h i k, Anna 2008, *Emerging Bilingual Speech. From Monolingualism to Code-Copying*, London—New York.
- — 2014, Estonian-Russian Code-Copying in Russian-Language Blogs: Language Change and a New Kind of Linguistic Awareness. — *Negotiating Linguistic Identity*, Oxford—Bern—Berlin—Brussels—Frankfurt—New York, 59–87.
- — 2019, English-Estonian Code-Copying in Blogs: Combining a Contact Linguistic and Cognitive Approach. — *Cognitive Contact Linguistics*, Berlin, 51–80.
- — 2020, Explaining the Emergence Order of Contact-Induced Language Change Phenomena in Language Maintenance. — *Philologia Estonica Tallinnensis* 5, Tallinn, 316–350.
- V e r s c h i k, Anna, K a s k, Helin 2019a, English-Estonian Code-Copying: Comparing Blogs and Vlogs. — *Applied Linguistics Review*, Berlin, 1–29.
- — 2019b, Estonian-English Code Alternation in Fashion Blogs: Structure, Norms and Meaning. — *Multilingualism in the Baltic States*, London, 307–335.
- W e i n r e i c h, Uriel 1968, *Languages in Contact. Findings and Problems*. The Hague.

- W e r t h e i m, Suzanne 2003, Discourse Pragmatics as a Means of Contact-Induced Change. — Selected Proceedings of the CLIC-LISO Ninth Annual Conference on Language, Interaction and Culture, 1–18.
- Z a b r o d s k a j a, Anastassia 2006, Eestivene keel(evariant): kas samm segakoodi poole? — KK, 736–750.
- — 2009, Russian-Estonian Language Contacts: Grammatical Aspects of Language Use and Change, Tallinn.
- Z e n n e r, Eline, v a n d e M i e r o o p, Dorien 2017, The Social and Pragmatic Function of English in Weak Contact Situations: Ingroup and Outgroup Marking in the Dutch Reality TV Show Expeditie Robinson. — Journal of Pragmatics 113, 77–88.

АННА ВЕРШИК, ХЕЛИН КАСК (Таллинн)

АНГЛИЙСКИЕ МАРКЕРЫ ДИСКУРСА В ДВУЯЗЫЧНЫХ ЭСТОНСКО-АНГЛИЙСКИХ БЛОГАХ И ВЛОГАХ

В статье рассматриваются английские маркеры дискурса в двуязычных эстонско-английских блогах и влогах. В исследовании поставлены вопросы: (1) Каковы функции английских маркеров дискурса в двуязычной речи и письме влогеров и блогеров? (2) Есть ли различия в употреблении английских маркеров дискурса в блогах и влогах? Проанализировано 45 блогов и 8 влогов (соответственно 365 973 и 73 858 токенов). Выяснилось, что преобладающий тип маркеров — это оценочные слова, в то время как доля интеракционных перформативов и маркеров структуры дискурса была схожей. Привлекательность экспрессивного значения для заимствования объясняет предпочтение оценочных слов. И в блогах, и во влогах похожая тенденция, но количество маркеров дискурса больше во влогах, поскольку это устный жанр.

ANNA VERSCHIK, HELIN KASK (Tallinn)

INGLISE DISKURSUSEMARKERID EESTI-INGLISE KAKSKEELSETES BLOGIDES JA VLOGIDES

Artikkel käsitletakse eesti-inglise kakskeelsetes blogides ja vlogides kasutatud inglise diskursusemarkereid. Uurimisküsimused on: (1) mis on eesti kakskeelsete blogijate ja vlogijate keelekasutuses esinevate inglise diskursusmarkerite funktsioonid? ja (2) kas inglise diskursusmarkerite kasutamisel esineb blogide ja vlogide puhul erinevusi? Kokku uuriti 45 blogi ja 8 vlogi (vastavalt 365 973 ja 73 858 sõnet). Analüüs näitab, et diskursusemarkeritest kasutatakse ennekõike hinnangusõnu ning suhtlusperformatiivide ja diskursuse struktuuri markerite osakaal on sarnane. Enim esineb just hinnangusõnu, kuna need on oma ekspressiivse tähenduse tõttu kasutamiseks atraktiivsed. Üldiselt on inglise diskursusemarkerite esinemise tendentsid blogides ja vlogides ühesugused, kuid vlogides esineb neid rohkem, seda vlogide suulise žanri tõttu.